David H. Romer and Justin Wolfers, Editors

Brookings Papers
ON ECONOMIC ACTIVITY

SPRING 2010

ELSBY, HOBIJN, and ŞAHİN
on the Labor Market in the Great Recession

NALEWAIK
on the Income- and Expenditure-Side Measures of Output

RAMEY and RAMEY
on the Rug Rat Race

GREENSPAN
on the Financial Crisis

BLANCHARD, DAS, and FARUQEE
on the Impact of the Crisis on Emerging Markets

PHILIPSON, SEABURY, LOCKWOOD, GOLDMAN, and LAKDAWALLA
on Geographic Variation in Health Care Spending

Brookings Papers
ON ECONOMIC ACTIVITY

SPRING 2010

DAVID H. ROMER
JUSTIN WOLFERS
Editors

BROOKINGS INSTITUTION PRESS
Washington, D.C.

THE BROOKINGS INSTITUTION
1775 Massachusetts Avenue, N.W., Washington, D.C. 20036

ISSN 0007-2303
ISBN-13: 978-0-8157-0513-0

Brookings Papers
ON ECONOMIC ACTIVITY

SPRING 2010

PURPOSE

The *Brookings Papers on Economic Activity* publishes research in macroeconomics, broadly defined, with an emphasis on analysis that is empirical, focuses on real-world events and institutions, and is relevant to economic policy. Papers are presented and discussed at conferences twice each year, and the papers and discussant remarks are published in the journal several months later. The intended audience includes analysts from universities, research institutions, governments, and business. The subject matter encompasses all fields of economic inquiry relevant to macroeconomics, including business cycles; development and long-term growth; the distribution of income, wealth, and opportunities; financial markets; international capital and foreign exchange markets; fiscal and monetary policy; international trade; labor markets; the public sector and the welfare state; energy; the environment; political economy; regulation and industrial organization; education; health; and demography.

The conference and the journal are based upon the work partially supported by the National Science Foundation under Grant No. 0752779 and the Alfred P. Sloan Foundation. The papers and discussant remarks reflect the views of the authors and not necessarily the views of the funding organizations or the staff members, officers, or trustees of the Brookings Institution.

CALL FOR PAPERS

Although most papers that appear in the *Brookings Papers* are solicited by the editors, the editors welcome submitted proposals. Editorial decisions are generally made about nine months in advance of each conference. Therefore, proposals should be received by December 1 for the following fall conference and by June 1 for the following spring conference. Proposals should be no more than five double-spaced pages and should be sent to brookingspapers@brookings.edu.

ACCESSING THE JOURNAL

For information about previous issues of the journal, participants in this conference, and agendas for upcoming conferences, visit www.brookings.edu/economics/bpea/bpea.aspx. To purchase subscriptions or single copies, visit www.brookings.edu/press, or contact the Brookings Institution Press at 866-698-0010 or P.O. Box 465, Hanover, PA 17331-0465. Brookings periodicals are available online through both the Online Computer Library Center (contact OCLC subscription department at 800-848-5878) and Project Muse (http://muse.jhu.edu). Archived issues of the *Brookings Papers on Economic Activity* are available through JSTOR (www.jstor.org).

EDITORS, PANEL ADVISERS, AND STAFF FOR THE EIGHTY-NINTH CONFERENCE

George A. Akerlof *University of California, Berkeley*
Olivier J. Blanchard *International Monetary Fund and MIT*
William C. Brainard *Yale University*
Mitali Das *International Monetary Fund*
Michael W. L. Elsby *University of Michigan*
Hamid Faruqee *International Monetary Fund*
Dana P. Goldman *University of Southern California*
Alan Greenspan *Greenspan Associates LLC*
Bart Hobijn *Federal Reserve Bank of San Francisco*
Darius N. Lakdawalla *University of Southern California*
Lee M. Lockwood *University of Chicago*
N. Gregory Mankiw *Harvard University*
Jeremy J. Nalewaik *Board of Governors of the Federal Reserve System*
George L. Perry *Brookings Institution*
Tomas J. Philipson *University of Chicago*
Garey Ramey *University of California, San Diego*
Valerie A. Ramey *University of California, San Diego*
David H. Romer *University of California, Berkeley*
Ayşegül Şahin *Federal Reserve Bank of New York*
Seth A. Seabury *RAND Corporation*
Justin Wolfers *University of Pennsylvania*
Michael Woodford *Columbia University*
Jean-Marie Callan *Brookings Institution*
Michael Treadway *Brookings Institution*
Lindsey B. Wilson *Brookings Institution*

GUESTS WHOSE WRITINGS OR COMMENTS APPEAR IN THIS ISSUE

Martin Neil Baily *Brookings Institution*
Alan S. Blinder *Princeton University*
Steven N. Braun *Council of Economic Advisers*
Christopher D. Carroll *Johns Hopkins University*
Susan M. Collins *University of Michigan*
Richard N. Cooper *Harvard University*
David M. Cutler *Harvard University*
Steven J. Davis *University of Chicago*
Francis X. Diebold *University of Pennsylvania*
Kathryn M. Dominguez *University of Michigan*
Kristin J. Forbes *Massachusetts Institute of Technology*
Benjamin A. Friedman *Harvard University*
Robert J. Gordon *Northwestern University*
Robert E. Hall *Stanford University*
Jan Hatzius *Goldman Sachs*
Erik Hurst *University of Chicago*
Lawrence F. Katz *Harvard University*
Alan B. Krueger *U.S. Department of the Treasury*
J. Steven Landefeld *Bureau of Economic Analysis*
Mark B. McClellan *Brookings Institution*
Daniel W. Sacks *University of Pennsylvania*
Matthew D. Shapiro *University of Michigan*
Robert Shimer *University of Chicago*
Christopher A. Sims *Princeton University*
Jeremy C. Stein *Harvard University*
Betsey Stevenson *University of Pennsylvania*
Phillip L. Swagel *Georgetown University*
Linda L. Tesar *University of Michigan*
George von Furstenberg *University of Indiana*

Editors' Summary

THE BROOKINGS PANEL ON ECONOMIC ACTIVITY held its eighty-ninth conference in Washington, D.C., on March 18 and 19, 2010. The recent financial crisis and ensuing recession continue to dominate the minds of leading economists, and this conference was no exception. Three of the papers in this volume assess macroeconomic developments in light of these remarkable events, examining the downturn in the U.S. labor market, the vulnerability of the financial system, and the spread of the crisis to emerging market countries. In each case the authors illustrate how economic institutions mediated the consequences of the macroeconomic shocks. A fourth paper, which addresses how best to measure GDP, is also highly relevant, showing that an alternative to the most commonly used measure would have yielded a clearer early warning of the size and scope of the U.S. downturn. The two remaining papers compile interesting new data that speak to ongoing longer-term debates about the balance between work and family and about health care reform.

IN THE FIRST PAPER in this issue, Michael Elsby, Bart Hobijn, and Ayşegül Şahin provide a heroic real-time analysis of recent labor market outcomes, comparing the recession that began in late 2007 with earlier downturns. All major measures of labor market conditions—including changes in unemployment, employment, participation, and hours—indicate that this most recent recession has been more severe than any since the Great Depression. The impact of the recession has been widespread, as unemployment rates among most major socioeconomic groups have exceeded previous postwar peaks.

Yet this recession also mirrors previous downturns in many respects. As in those recessions, the total decline in labor input is about one-third due to a shorter workweek and two-thirds due to fewer people working. Labor force participation has fallen, muting the impact of this decline on the unemployment rate. And the sharpest impacts of this recession also follow the pattern observed in earlier downturns, with men suffering more than women, the young more than the old, and the less educated and racial minorities bearing disproportionate impacts.

The authors then turn to examining inflows and outflows from unemployment. They find that inflows into unemployment rose sharply, particularly in the early stages of the recession, and that the subsequent rise in unemployment largely reflects a rise in the duration of unemployment spells. Yet the rate at which workers separate from jobs has not risen—a fact that suggests a change in the composition of separations toward fewer quits (which often involve job-to-job flows) and more layoffs. The important role of layoffs early in this recession represents a departure from recent downturns, but it parallels earlier severe recessions. Outflows from unemployment (the flip side of the rise in duration of the typical unemployment spell) have been strikingly similar across demographic groups, and hence demographic differences in the impact of this recession—as in previous downturns—are largely driven by the different rates at which members of each group typically enter unemployment.

Looking forward, Elsby, Hobijn, and Şahin note that the rise in inflows to unemployment has abated, and that the rate at which workers are exiting unemployment has fallen further than in previous recessions. Consequently, the key to subsequent recovery will be further rises in the unemployment exit rate. Indeed, perhaps the most distinctive feature of this recession is the recent record low in the exit rate, which is also reflected in current record rates of long-term unemployment. Unfortunately, recent job vacancy data suggest that the Beveridge curve, which relates unemployment and vacancies, has shifted outward, perhaps because of a decline in the efficiency with which job seekers are being matched with available jobs. In turn, outflows from unemployment are lower than might be expected on the basis of the vacancy-unemployment ratio, which, the authors argue, may be partly (but only partly) due to the temporary extension of unemployment insurance for the long-term unemployed. Because the long-term unemployed tend to exit unemployment only very slowly, outflows from unemployment may remain depressed for some time, dampening the recovery. Even so, the authors note that the emerging long-term unemployment problem in the United States remains small relative to the stagnation that virtually halted the recovery of European labor markets in the 1970s and 1980s.

IN THE SECOND PAPER, Jeremy Nalewaik turns to a critically important issue in economic measurement. GDP, a country's overall economic output, can be measured either as the sum of all final expenditures or as the sum of all incomes earned. Yet despite the conceptual equivalence, the measure based on expenditure—which Nalewaik calls GDP(E)—has often differed substantially over recent decades from the measure based on income, or GDP(I). Currently, the Bureau of Economic Analysis, the official source of data for both measures, emphasizes the expenditure-based measure as its "top line" measure, and the income-based measure (which the bureau calls gross domestic income, or GDI) rarely receives much mention in public discussion.

Nalewaik compellingly demonstrates that this emphasis is misplaced. Real-time estimates of the income-based measure of GDP growth have yielded a much more reliable picture of the contours of the business cycle than the expenditure-based measure. He makes his case in three steps. First, he runs an array of horserace regressions, assessing the relative weight that one should put on real-time GDP(I) and GDP(E) data in predicting each of a wide range of measures of business cycle conditions, including changes in the unemployment rate, employment growth, the slope of the yield curve, growth in stock prices, and periods of recession. In each case he finds that GDP(I) vastly outperforms GDP(E). Likewise, GDP(I) does a better job of predicting the future path of many of these business cycle indicators, as well as the GDP predictions of professional forecasters and next quarter's growth in both GDP(E) and GDP(I) themselves. In fact, the only variable that GDP(E) significantly helps predict is the final revised value of growth in this quarter's GDP(E). And even on this score, the regressions using data since the mid-1990s suggest putting about equal weight on GDP(E) and GDP(I).

Second, Nalewaik turns to evaluating the estimates after they have been thoroughly revised. Since the 1980s, the gap between the revised measures has been highly cyclical, with GDP(E) recording a shallower and less distinct business cycle. Digging into the construction of the estimates, he concludes that GDP(E) is constructed from data sources that appear to miss important parts of the business cycle. And indeed, he shows that the final GDP(I) data are much more highly correlated with numerous other indicators of business conditions than are the final GDP(E) data.

Finally, Nalewaik shows that GDP(I) has identified the beginning of each of the last four recessions more quickly than GDP(E). Indeed, one reason that there was some debate as to whether the economy had entered a recession in late 2007 is that the expenditure-based measure continued to show economic growth throughout 2008.

The paper concludes with a modest proposal: that the Bureau of Economic Analysis emphasize as its top-line estimate of GDP growth a weighted average of GDP(E) and GDP(I), placing at least as much weight on GDP(I) as on GDP(E). We believe that Nalewaik has presented an overwhelming case and hope that the bureau will be responsive; until then, macroeconomists would do well to make themselves more familiar with income-based measures of GDP.

IN THE THIRD PAPER, Garey Ramey and Valerie Ramey bring to light a rather extraordinary recent trend in Americans' use of time: parents—and in particular highly educated parents—have greatly increased the amount of time they spend on childcare activities. In time-use surveys from 1965 to 1995, mothers recorded an average of about 12 hours per week looking after their children, and the gap between college-educated mothers and those with less education was about 1 hour. Yet by 2007 this time commitment had risen to 21 hours per week for college-educated mothers, and to 16 hours per week for non-college-educated mothers. Similar changes were observed among fathers: the rise in their childcare time was smaller in absolute terms, but larger proportionally. These are macroeconomically important shifts, representing around $300 billion in forgone wages, and the change in time use is roughly comparable to the effect of a typical recession on work hours.

The authors present a novel hypothesis for these observations. The child population has grown with the baby-boom "echo," but ever-more-valuable spots in elite colleges have not increased commensurately. In response, parents, and especially college-educated parents, are engaged in a "rug rat race," making ever-increasing investments of their time in activities that they believe will help build a compelling college application for their children. Just as in an arms race, or as in the original "rat race" among urban white-collar workers, this rivalry can lead to overinvestment in some activities relative to the social optimum.

The authors document several facts consistent with their explanation: the rise in time spent with children paralleled the rise in the number of graduating high school seniors; much of this rise reflects time spent caring for older children, and in particular transporting them to extracurricular activities; and the trend toward increasing childcare time is less evident in Canada, where college admissions are less rivalrous. The authors also assess—and reject—a number of competing explanations, including changes in who becomes a parent (the rise in average childcare hours remains even when averaging across all adults); rising incomes (an insufficient explanation given the moderate income elasticity of childcare time); increasing safety concerns (survey data suggest that such concerns actually fell over the relevant period); greater enjoyment of childcare (which predicts, counterfactually, that fertility would also rise); and more flexible work schedules (which cannot explain why the rise is even greater among nonworking mothers). The facts so carefully catalogued by the authors will surely generate further research, and with it, even more hypotheses about just what factors are driving these enormous and important changes in family and work life.

IN THE FOURTH PAPER, Alan Greenspan offers his diagnosis of the recent financial crisis and his proposals for reducing the chances of future crises. The seeds of the crisis, in his view, were sown by a period of historically low real interest rates, unprecedented macroeconomic stability, and low inflation. These developments led to large increases in investors' willingness to take on risk and, partly as a result, to the rapid growth of home prices in the mid-2000s. This price growth in turn fueled (and was reinforced by) an explosion of securitization of mortgage loans into assets whose risk characteristics were often poorly understood and that were often held by highly leveraged institutions. When home prices began to fall in 2006, the result was a cascade of financial failures and contagion. Greenspan assigns some of the blame for the crisis to failures of regulatory oversight, but he finds no evidence that the conduct of monetary policy played a role: economic theory, time-series evidence from the United States, and cross-country evidence all suggest that the central bank's decisions about its interest rate target over a period of a few years are not a major driver of home prices.

Greenspan then turns to the issue of how to reduce the risk of future crises. He argues that policymakers face daunting empirical difficulties in fully understanding risks and in identifying asset bubbles and potential incipient crises in real time. This implies that policies that require regulators to forecast financial instability are unlikely to succeed, especially considering the political and practical difficulties in continually adjusting regulation in response to economic developments.

Instead, he argues, the system needs to be designed so that it is broadly robust to shocks. One key feature of such a system would be increased capital requirements for financial institutions. Based on historical relationships, he estimates that these could be as high as 10 to 15 percent without impairing the functioning of the banking system. Such requirements would need to apply both to existing regulated banks and to the "shadow" banking system and be accompanied by ample collateral and liquidity requirements. Finally, Greenspan argues that it is essential to address the problem of financial institutions that are "too big to fail," either by breaking them up or by putting in place mechanisms that subject their equity holders and creditors to the possibility of large losses without threatening the stability of the financial system.

In the fifth paper, Olivier Blanchard, Mitali Das, and Hamid Faruqee investigate the short-run impact of the global financial crisis on emerging market countries. They begin with a simple reduced-form model to identify possible channels of transmission. Some channels involve trade, through reduced demand for a country's exports when its trading partners enter a crisis. Others involve financial markets, through reduced demand for a country's assets and increases in risk premia. The authors argue that it is crucial to recognize the adverse effects of depreciation of the home currency on real debt burdens, and the possibility that depreciation may reduce net exports in the short run. Once these complications are introduced, even a comparatively barebones model allows for a potentially rich set of effects of the initial shocks and for complex interactions with the policy responses.

Blanchard, Das, and Faruqee then turn to the cross-country data. They find evidence of effects working in the expected directions. In late 2008 and early 2009, countries whose trading partners suffered larger shortfalls in growth relative to precrisis forecasts suffered substantially larger growth shortfalls themselves, suggesting an important impact through trade. And countries that had more debt coming due during the crisis also suffered much larger growth shortfalls, suggesting an important impact through financial markets.

At the same time, no simple story explains the different effects of the crisis across countries. Although both trade and financial variables typically are significant when both are included in the regressions, a substantial portion of the variation in growth remains unexplained. The results also imply that a hypothetical country with no trade or financial exposure to the rest of the world would nonetheless have suffered a significant growth shortfall from the precrisis prediction, suggesting that more was at work than the channels the authors focus on. The authors are unable to detect any large role of reserve holdings, the exchange rate regime, or the fiscal response in determining the short-run impact of the crisis.

The paper concludes by looking at three countries in detail: Latvia, Russia, and Chile. The contrast between Russia and Chile is particularly striking. Much about the two countries before the crisis was similar: both are financially open economies whose exports are dominated by commodities. Yet Russia had one of the largest growth shortfalls, while Chile's shortfall was below the average. The different outcomes are not entirely mysterious, however: Chile's stronger institutions and longer track record of sound policies seem to have prevented a net capital outflow, whereas Russia's attempt to use its reserves to stem what proved to be overwhelming pressure for depreciation led to very large capital outflows.

In the final paper, Tomas Philipson, Seth Seabury, Lee Lockwood, Dana Goldman, and Darius Lakdawalla examine geographic variation in health care utilization and spending. An important line of inquiry—most prominently associated with the Dartmouth Atlas project—has documented large disparities in health care use and spending across regions of the country. These disparities cannot be explained by differences in observed patient demographics or disease prevalence, and regions using more health care do not exhibit substantially better outcomes. But the authors note that these findings are largely based on data from Medicare, which is a public program. By contrast, private payers may have stronger incentives to restrain costs and utilization, and hence greater incentives to eliminate wasteful procedures. On the flip side, government-run insurers have greater bargaining power, which they may use to restrain costs.

In their empirical analysis, the authors compare health care use and spending records of employees and retirees of 35 Fortune 500 firms with patient records from a survey of Medicare beneficiaries. In order to analyze samples with roughly comparable health status, they focus only on patients with a diagnosis of heart disease. The authors find that the variance of health care utilization across 99 metropolitan areas tends to be lower in the private than in the public sector, although this finding is sensitive to controlling appropriately for differences in the demographic and health status of the two samples. The geographic variation in health care spending (as opposed to utilization), on the other hand, is generally lower in the public sector. The authors highlight the need for further research on the determinants and benefits of health care utilization and spending in the private sector.

MICHAEL W. L. ELSBY
University of Michigan

BART HOBIJN
Federal Reserve Bank of San Francisco

AYŞEGÜL ŞAHİN
Federal Reserve Bank of New York

The Labor Market in the Great Recession

ABSTRACT From the perspective of a wide range of labor market outcomes, the recession that began in 2007 represents the deepest downturn in the postwar era. Early on, the nature of labor market adjustment displayed a notable resemblance to that observed in past severe downturns. During the latter half of 2009, however, the path of adjustment exhibited important departures from that seen during and after prior deep recessions. Recent data point to two warning signs going forward. First, the record rise in long-term unemployment may yield a persistent residue of long-term unemployed workers with weak search effectiveness. Second, conventional estimates suggest that the extension of Emergency Unemployment Compensation may have led to a modest increase in unemployment. Despite these forces, we conclude that the problems facing the U.S. labor market are unlikely to be as severe as the European unemployment problem of the 1980s.

Since December 2007, labor market conditions in the United States have deteriorated dramatically. The depth and duration of the decline in economic activity have led many to refer to the downturn as the "Great Recession." In this paper we document the adjustment of the labor market during the recession and place it in the broader context of previous postwar downturns. What emerges is a picture of labor market dynamics with three key recurring themes:

—From the perspective of a wide range of labor market outcomes, the recession that began in 2007 (hereafter "the 2007 recession")[1] represents the deepest downturn in the postwar era.

—Early on, the nature of labor market adjustment in the 2007 recession displayed a notable resemblance to that observed in past severe downturns.

—During the latter half of 2009, however, the path of adjustment exhibited important departures from that seen during and after prior deep recessions.

These broad conclusions arise from a detailed investigation of the behavior of labor market stocks and flows over the course of the downturn.[2] Our point of departure, in section I, is to document patterns over time in key labor market indicators—unemployment, employment, labor force participation, and hours per worker—during the 2007 recession. No matter what indicator of labor market activity we consider, the deterioration of labor market conditions during this recession is the worst on record since the late 1940s. Rates of unemployment among most major subgroups of the labor market reached postwar highs. From the perspective of the labor market, the 2007 recession is truly a Great Recession.

As noted above, we nonetheless observe that many dimensions of these key indicators mirror those seen in past recessions. Labor force participation declined, reflecting the modest procyclicality observed in many postwar recessions; the relative contributions of the intensive and the extensive margins (that is, of changes in hours per worker and in the number of workers employed) to the decline in total labor input typify the conventional one-third hours to two-thirds bodies split observed in the past; and the constellation of demographic groups most affected—younger workers, male workers, less educated workers, and workers from ethnic minorities—is reminiscent of previous downturns.

1. We adopt this terminology because although the recession is widely believed to have ended in 2009, as of this writing the Business Cycle Dating Committee of the National Bureau of Economic Research (NBER) has not yet fixed an end date. In some of our figures, specification of an end date is unavoidable and is not intended as a firm judgment as to when the recession ended.

2. A drawback of the real-time nature of our analysis is that a detailed treatment of the cyclical behavior of wages is infeasible. Although timely aggregate compensation data are available, such data are plagued by countercyclical composition biases, as low-skilled workers are more likely to lose their jobs in time of recession. As emphasized by Solon, Barsky, and Parker (1994), obtaining an accurate sense of real-wage cyclicality requires the use of longitudinal microdata that are available in a less timely manner.

It is well known that changes in aggregate unemployment in the United States mask substantial variation in underlying worker flows, a point emphasized by Olivier Blanchard and Peter Diamond (1990). Reflecting this fact, in section II we investigate the sources of increased unemployment in the 2007 recession by analyzing the behavior of unemployment flows. This analysis reveals that both increased inflows into unemployment and declines in the rate at which workers flow out of the unemployment pool play crucial roles in accounting for the recent upswing in unemployment. As in previous severe recessions, the initial ramp-up in unemployment was accompanied by a sharp rise in inflows. In contrast to the claims of some recent literature on unemployment flows (Hall 2005, Shimer 2007), elevated rates of inflow in time of recession appear not to be a relic of past downturns, but rather a distinctive feature of severe recessions, both old and modern. The behavior of the outflow rate also mirrors that observed in past deep recessions: as the wave of inflows receded in the latter stages of the 2007 recession, the outflow rate continued to fall. Reflecting the distinctive severity of the downturn, recent data have seen the outflow rate reach a postwar low.

Measures of unemployment flows for different labor force groups yield an important message on the sources of the disparate trends in unemployment across those groups: higher levels and greater cyclical sensitivity of joblessness among young, low-skilled, and minority workers, both in this and in previous downturns, are driven predominantly by differences in rates of entry into unemployment between these groups and others. In sharp contrast, a striking feature of unemployment exit rates is a remarkable uniformity in their cyclical behavior across labor force groups—the declines in outflow rates during this and prior recessions are truly an aggregate phenomenon.

In the remainder of section II, we take advantage of a unique opportunity to assess the role of labor turnover in the 2007 recession. This is the first full upswing in unemployment covered by the new Job Openings and Labor Turnover Survey (JOLTS), which reveals some stark findings. In contrast to the behavior of unemployment inflows, rates of separation of workers from employers did not rise in the 2007 recession. This suggests support for a hypothesis offered by Robert Hall (2005): increases in unemployment inflows may have little to do with increased rates of job loss, but merely are a symptom of declining rates of job finding among potential job-to-job movers. Our analysis of the JOLTS data points to a different story: increased inflows into unemployment are driven predominantly by a change in the *composition* of separations toward layoffs, which are likely to result in unemployment, and away from quits, which often represent workers flowing to new jobs upon separation. Job loss played a key role in driving increased unemployment in the 2007 recession.

We close our analysis in section III by assessing the outlook for the recovery of the labor market in the wake of the current downturn. Motivated by the recent subsidence of inflows into unemployment and the historic decline in the outflow rate from unemployment, we emphasize the importance of a rebound in the latter for future reductions in unemployment and highlight a potential cause for concern in recent data. The postwar U.S. labor market has been characterized by two remarkably stable aggregate relationships: the inverse co-movement of unemployment and vacancies—the Beveridge curve—and the positive association between the outflow rate from unemployment and the vacancy-unemployment ratio, a point noted by Robert Shimer (2005). The latter half of 2009 witnessed a break from these relationships, with unemployment rising higher than implied by the historical Beveridge curve, and the outflow rate from unemployment falling significantly below the path implied by the past relationship with the vacancy-unemployment ratio.

These trends resemble those observed in the breakdown in efficiency of matching jobs with workers that accompanied the European unemployment problem of the 1980s, raising the concern of persistent unemployment, or hysteresis, in U.S. unemployment going forward. We consider a range of possible causes of hysteresis, including sectoral mismatch, the extension of the duration of unemployment insurance benefits, the dependence of unemployment outflow rates on the duration of unemployment, and reductions in the rates of worker flows—what Blanchard (2000) has termed "sclerosis." Recent data point to two warning signs. First, the historic decline in unemployment outflow rates has been accompanied by a record rise in long-term unemployment. We show that this is likely to result in a persistent residue of long-term unemployed workers with relatively weak search effectiveness, depressing the strength of the recovery. Second, conventional estimates of the impact of longer unemployment benefit duration on the length of unemployment spells suggest that the extension of Emergency Unemployment Compensation starting in June 2008 is likely to have led to a modest increase in long-term unemployment. Nonetheless, we conclude that, despite these adverse forces, they have not yet reached a magnitude that would augur a European-style hysteresis problem in the U.S. economy in the long run.

I. Basic Facts about the Labor Market in the 2007 Recession

The recession that started in December 2007 has been severe according to many measures, not least in terms of its effect on the labor market. In this section we review the recent behavior of some of the main aggregate measures of labor market outcomes and place the recent deterioration in labor market conditions in the broader historical context of previous postwar recessions.

I.A. Unemployment, Employment, Labor Force Participation, and Hours per Worker

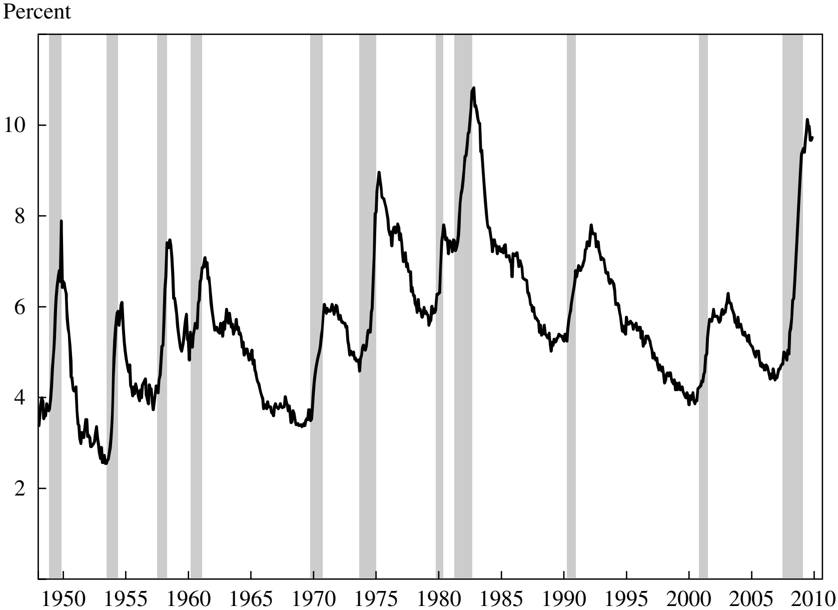
The main labor market indicator on which much of this paper will focus is the unemployment rate. To set the stage, figure 1 displays the published time series for the civilian unemployment rate from Current Population Survey (CPS) data. The 2007 recession figures prominently in this series. Unemployment rose from a prerecession low of 4.4 percent to reach 10.1 percent in October 2009. This increase—5.7 percentage points—is the largest postwar upswing in the unemployment rate. It dwarfs the rise in joblessness in the two previous recessions, in 1990–91 and 2001, when in each case unemployment rose by approximately 2.5 percentage points. It dominates even the severe recession of 1973–75 (4.4 percentage points) as well as the combined effects of the consecutive recessions of the early 1980s (5.2 percentage points). There is little doubt that the present downturn is the deepest since World War II from the perspective of the labor market.[3]

In what follows we will closely examine the rise in unemployment in the present downturn. But it is helpful at this point to place the increase in joblessness in the broader context of other, related labor market indicators. We consider two sets of measures: first, the relationship between the rise in unemployment and the decline in employment during the downturn, and second, the role of the decline in employment relative to the decline in hours per worker in accounting for the contraction in total labor input.

THE DECLINE IN EMPLOYMENT. The unemployment rate at a given point in time u_t can be related to the level of employment E_t and the size of the labor force L_t by the simple identity $u_t = 1 - (E_t / L_t)$. This identity suggests a simple metric for gauging the relative roles of variation in employment

3. Of course, even the current ramp-up in the unemployment rate is overshadowed by that witnessed during the Great Depression. In 1929 the unemployment rate stood at 3.2 percent, rising to 25.2 percent by 1933, a 22-percentage-point rise in 4 years. Indeed, such is the extremity of the Great Depression that adding it to any plot renders the postwar variation in joblessness very difficult to perceive.

Figure 1. Unemployment Rate, 1948–2010[a]



Source: Bureau of Labor Statistics data.
a. Monthly data, seasonally adjusted. Shading indicates recessions.

and labor force participation in accounting for the upswing in unemployment, since

$$\mathrm{d}u_t = (1 - u_t)\left[\mathrm{d}\log(L_t/P_t) - \mathrm{d}\log(E_t/P_t)\right], \tag{1}$$

where P_t denotes the working-age population. The increase in the unemployment rate over the course of a recession can be decomposed into two parts, accounted for by logarithmic variation in the labor force participation rate and in the employment-population ratio.

Figure 2 shows results of such an exercise. It plots the cumulative log deviations from trend of the employment-population ratio and the labor force participation rate, both taken from the CPS, for each of the last six recessions. Figure 2 conveys two related messages. First, the record upswing in the unemployment rate observed in figure 1 is mirrored by a record contraction in employment: employment declined relative to trend by 7 log points from the start to the trough of the 2007 recession, dominating the severe recession of the mid-1970s as well as the joint effects of the consecutive recessions of the early 1980s.

Figure 2. Cumulative Deviations from Trend of the Employment-Population Ratio and of Labor Force Participation in Six Recessions[a]

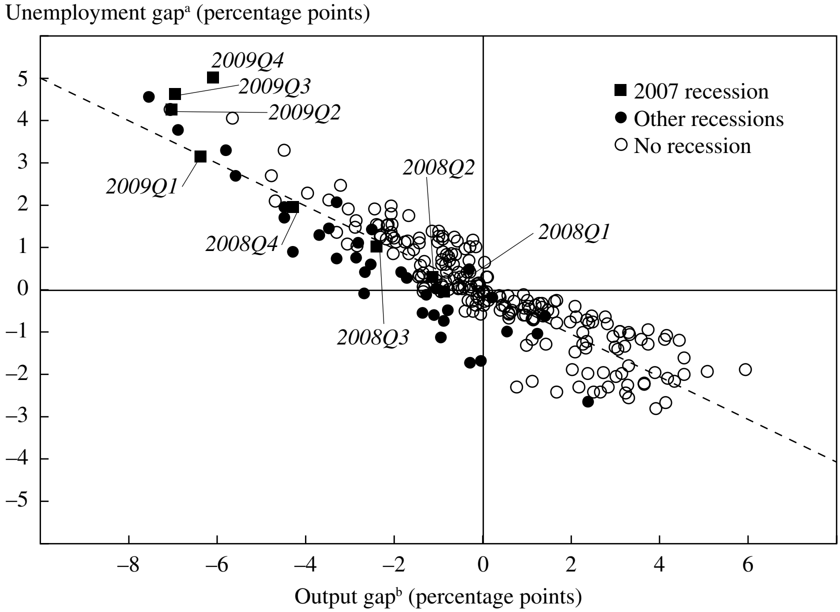
Source: Authors' calculations using BLS data.
a. Trend is estimated using a Hodrick-Prescott filter with a smoothing parameter of 100000.

Second, rather than contributing to the rise in unemployment, a reduction in labor force participation of around 2 log points muted the rise in joblessness in the 2007 recession. Figure 2 also reveals that the 2007 recession is no exception in this respect: almost all of the earlier downturns also exhibit at least a mild procyclicality of labor force participation.

An interesting aspect of the response of labor force participation in the 2007 recession is that it seems to have had two stages. Mary Daly, Hobijn, and Joyce Kwok (2009a) note that during the first part of the recession, the labor force participation rate remained unexpectedly high. From May to December 2009, however, the labor force participation rate fell by 1.2 percentage points, its steepest decline since the 1950s.

UNEMPLOYMENT AND GDP (OKUN'S LAW). One of the most robust aggregate statistical relationships for the U.S. economy is the inverse co-movement between changes in the unemployment rate and growth in GDP—Okun's law (Okun 1962). Figure 3 displays a version of the Okun's law relationship updated to include the 2007 recession. It plots the quarterly deviation

Figure 3. Okun's Law, 1949–2009



Source: Authors' calculations using BEA, BLS, and CBO data.
a. Deviation of the actual unemployment rate from its trend.
b. Deviation of actual GDP from its trend, as estimated by the CBO.

from trend of the unemployment rate against the contemporaneous percentage deviation from trend of GDP, using estimates by the Congressional Budget Office (CBO) of the nonaccelerating-inflation rate of unemployment (NAIRU) and potential output up to January 2010.[4] The regression line is based on the observations from 1949 through 2007, thus excluding the Great Recession. In the absence of large movements in potential output and the NAIRU, Okun's law implies that for every 2 percentage points that output falls below trend, the unemployment rate will increase by about 1 percentage point.

This rule of thumb performs remarkably well in the first part of the 2007 recession, from 2008Q1 through 2009Q1, as indicated in figure 3. Thus, as we have noted of other dimensions of the 2007 downturn, the adjustment of the labor market until the second quarter of 2009 is by no means an outlier relative to past recessions. The last nine months of 2009, however, wit-

4. Detrended unemployment and output data based on Hodrick-Prescott-filtered series yield very similar results.

nessed an important departure from Okun's law: even though overall economic activity, as measured by GDP, rebounded in the second half, the unemployment rate continued to rise. This recent divergence between output and the labor market can be traced to high average labor productivity growth during that period,[5] resulting in an increase in the unemployment rate in 2009 that surprised policymakers and forecasters alike. The exceptionally strong productivity growth during the early recovery also occurred during the jobless recoveries that followed the previous two recessions. We revisit the implications of this pattern for the current outlook in section III.[6]

HOURS VERSUS BODIES. The evidence presented thus far has pertained solely to measures of the number of persons in or out of work, and not to the number of hours worked per employed person. Here we summarize the behavior of each of these measures and identify their relative importance in driving the recent contraction in total labor input. Our point of departure is another simple accounting identity, namely, that total labor input H_t is the product of employment E_t and hours per worker h_t. It follows that the logarithmic decline in total hours worked during the recession may be decomposed into the sum of the respective logarithmic declines in E_t and h_t.

Figure 4 performs this simple accounting exercise using data on employment and weekly hours per worker in the nonfarm business sector from the Labor Productivity and Costs program of the Bureau of Labor Statistics (BLS).[7] It plots the cumulative log declines in employment and hours per worker for each of the last six recessions.[8] The figure shows that although the 2007 recession is unusual in its severity, the adjustment of the labor market in this recession resembles that observed in prior recessions on two important dimensions. First, the reduction in hours per worker is steeper

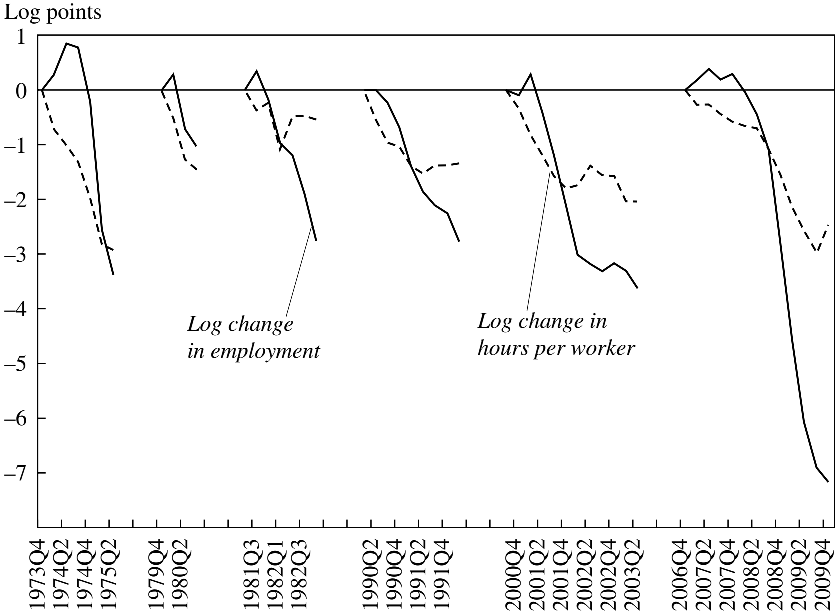
5. Mulligan (2009, 2010) argues that the current downturn has been qualitatively different from previous severe recessions in that productivity growth remained normal while labor supply shifted to the left. He concludes that a reduction in labor supply or an increase in labor market distortions, or both, are major factors in the 2007 recession.

6. Nalewaik (this volume) suggests that the deviations from Okun's law are less severe when one considers gross domestic income, the income-based measure of output, rather than GDP, which is based on the expenditure side of the national accounts. For a detailed analysis of the recent behavior of Okun's law, see Gordon (2010).

7. The BLS series identifiers used for employment and weekly hours per worker are, respectively, PRS85006013 and PRS85006023. In constructing these series, the BLS combines data from the Current Employment Statistics and the CPS. Employment here includes both payroll employees and self-employed and unpaid family workers.

8. The recession dates used to construct figure 4 differ slightly from the official recession dates established by the Business Cycle Dating Committee of the NBER. They correspond to the quarters around the NBER recession dates over which total hours worked are observed to decline.

Figure 4. Cumulative Declines in Employment and Weekly Hours per Worker in Six Recessions



Source: Authors' calculations using BLS data.

than that in employment in the early stage of all six recessions, with the contraction in employment becoming dominant later on. Second, employment in the 2007 recession fell by 7 log points, as figure 2 showed, but hours per worker also contracted, by 3 log points. (Total labor input thus declined by 10 log points, again more than in any other postwar recession.) This 70:30 bodies-hours split is in line with the conventional wisdom since at least Arthur Okun (1962) that the extensive margin (the number employed) accounts for around two-thirds of the cyclical variation in labor input. Reiterating this point, figure 4 also reveals that across the last six recessions, variation in employment accounts for approximately 50 to 80 percent of the decline in total labor input.

I.B. Who Has Been Hit Hardest?

Underlying the acute surge in joblessness documented in figures 1 through 4 is a rich heterogeneity in the structure of unemployment across different groups in the labor force. Here we document this heterogeneity in the experience of unemployment across groups, focusing on four dimensions: sex, age, race, and educational attainment.

To assess the quantitative importance of these differences, table 1 reports the ratio of the rise in each group's unemployment rate to the rise in the overall unemployment rate for the last five downturns, using data from the CPS. If the rise in unemployment were spread uniformly across different subgroups of the labor market, the ratios in table 1 would all equal 1. Instead we find that males, younger workers, and less educated workers, as well as members of ethnic minorities, experience steeper rises in joblessness during all recessions, including the 2007 recession.[9]

One aspect of the results in table 1 is worth highlighting. Although many commentators on the present downturn have emphasized its character as a "mancession," table 1 reveals that all recessions have affected male workers disproportionately; the mancession is not a new phenomenon. Şahin, Joseph Song, and Hobijn (2009) show that this pattern can be traced to the fact that industries in which male workers are concentrated, such as construction and durable goods manufacturing, are particularly sensitive to the business cycle.

I.C. Accounting for the Composition of the Labor Force

Heterogeneity in the experience of unemployment across labor force groups is an important characteristic of joblessness in the 2007 recession. Recent decades have witnessed dramatic changes in this heterogeneity. We focus here on one particular dimension that has a crucial bearing on historical comparisons of unemployment rates: age structure. The labor force has become older since the 1980s as the baby-boom generation has aged—a point emphasized by Shimer (1998, 2001).[10] Accounting for such

9. This finding echoes those of an abundant literature that has documented differences in the cyclical sensitivity of different demographic groups (see Clark and Summers 1981, Gomme and others 2004, Kydland 1984, and Mincer 1991, for example).

10. The online appendix to this paper (available on the *Brookings Papers* webpage at www.brookings.edu/economics/bpea.aspx, under "Conferences and Papers") presents compositional adjustments for the full interaction of age, sex, race, and education, as well as for each dimension individually. Although the changing sex composition of the labor force has had very little impact, composition by race and education plays a role. The influx of immigrants since the 1970s has led to a greater fraction of Hispanic workers in the labor force, who in turn are more likely to experience an unemployment spell. On the other hand, increased educational attainment since the 1980s has shifted the structure of the labor force toward better-educated workers, who face lower unemployment rates on average (see Farber and Western 2010 for more on this topic). Shimer (1998) cautions against adjustments for educational composition, however. Workers with higher unobserved ability are likely to face lower unemployment rates conditional on education. As workers become more educated over time, the innate ability of each education group will decline, leading to an increase in that group's unemployment rate. In addition, if the educational distribution shifts, employers may simply revise the educational requirements of jobs, leading to no real effect on the unemployment rate.

Table 1. Changes in the Unemployment Rates of Selected Demographic Groups Relative to Overall Unemployment in Five Recessions

Group	*Group average unemployment rate, 1976Q1–2009Q4 (percent)*	*Ratio of the change in the indicated group's unemployment rate to the change in the overall employment rate*[b]				
		1979–80 recession	*1981–82 recession*	*1990–91 recession*	*2001 recession*	*2007 recession*
Sex						
Male	6.2	1.42	1.19	1.13	1.14	1.19
Female	6.3	0.42	0.75	0.85	0.84	0.78
Age (years)						
16–24	12.6	1.60	1.27	1.58	1.67	1.62
25–54	5.0	1.02	1.01	0.93	0.98	0.96
55 and over	3.5	0.23	0.65	0.91	0.60	0.64
Educational attainment						
Less than high school	8.8	1.30	1.56	1.89	1.20	1.47
High school diploma	5.4	1.31	1.23	1.01	0.89	1.19
Some college	4.3	0.88	0.72	0.92	1.11	0.99
College degree or higher	2.6	0.18	0.42	0.42	0.73	0.53
Race or ethnicity						
White	5.4	0.95	0.90	0.90	0.86	0.95
Black	12.1	1.28	1.65	1.61	1.72	1.42
Hispanic	8.8	n.a.	n.a.	1.73	0.77	1.36
Asian	4.6	n.a.	n.a.	n.a.	1.19	0.86
Memorandum: change in overall unemployment rate[c] (percentage points)		1.9	3.3	2.3	2.2	5.5

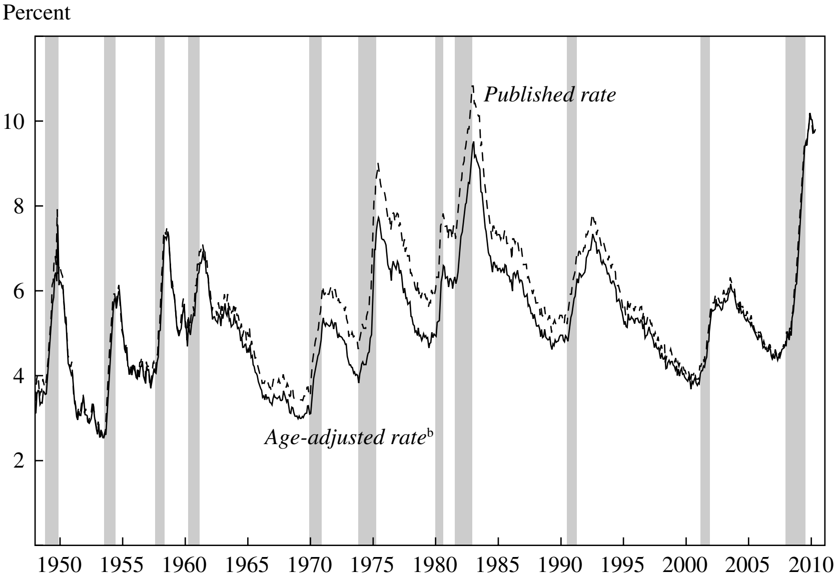
Source: Authors' calculations using BLS data. Note that all calculations are based on quarterly data.

a. Data for Hispanics are available only from 1982Q1 to 2009Q4, and for Asians only from 2000Q1 to 2009Q4. n.a. = Not available.

b. Periods covered are, respectively, 1979Q2–1980Q3, 1981Q2–1982Q4, 1990Q2–1992Q3, 2000Q4–2003Q3, and 2007Q1–2009Q4. These dates may not correspond exactly to those of recessions as dated by the NBER Business Cycle Dating Committee.

c. Measured from trough to peak, except for the 2007 recession, where it is measured from its most recent trough in 2007Q1 to 2009Q4.

Figure 5. Age-Adjusted Unemployment Rate, 1948–2010[a]



Source: Authors' calculations using BLS data.
a. Monthly data. Shading indicates recessions.
b. Rate that would prevail if the age structure of the labor force (the shares of workers aged 16-24, 25-34, 35-44, 45-54, and 55 and over) remained constant at its 2009 level.

compositional changes can paint a different picture of aggregate unemployment trends because each of these different labor force groups is systematically more or less likely than others to experience spells of unemployment.

We implement a simple method for controlling for the impact of changes in the age composition of the labor force on trends in aggregate unemployment: we fix the labor force shares for each age group to their level at some reference date and then trace out the implied composition-adjusted unemployment series. Figure 5 performs this exercise using the most recent labor force shares and reveals an interesting finding: accounting for changes in age composition leads to a substantial downward revision of past unemployment rates, such that the age-adjusted unemployment rate in the 2007 recession reached its highest level in the postwar period.

II. Labor Market Flows in the 2007 Recession

Another defining characteristic of the U.S. labor market is that it is in continual flux. Even when the aggregate economy is tranquil, many workers flow in and out of employment and unemployment. In time of recession these flows come into focus as proximate determinants of increases in joblessness: Does unemployment rise as a result of increased inflows as workers lose their jobs? Or does it rise because unemployed workers increasingly fail to find new jobs? Or is it some combination of the two?

Based on the shallow downturns of 1990–91 and 2001, recent research has argued that the nature of labor market adjustment in time of recession has radically shifted in recent years. Hall (2005a, p. 397) states that "in the modern U.S. economy, recessions do not begin with a burst of layoffs." Echoing this, in his study of unemployment flows, Shimer (2007, abstract) concludes that "fluctuations in the employment exit probability are quantitatively irrelevant during the last two decades."[11] Instead, in this view, increased unemployment duration, or a decline in the rate at which workers flow out of the unemployment pool, drives the entirety of contemporary variation in unemployment.

In contrast, a long line of research on labor market flows before the last two recessions came to the conclusion that cyclical ramp-ups in unemployment are driven by both margins, inflows and outflows.[12] More recent work has revived this conclusion and identified a clear pattern to unemployment flows in recessions: increases in unemployment are preceded by sharp rises in unemployment inflows, followed by more prolonged periods of elevated unemployment duration.[13] That literature pointed toward cyclical ramp-ups in unemployment being driven by both margins, with inflows being relatively more dominant early in recessions.

The 2007 downturn provides an opportunity to assess these conclusions: is a diminished role of job loss a feature of modern recessions, or of shallow recessions? We explore this question using updated estimates of unemployment transitions from a variety of data sources.

11. Shimer (2007) uses the term "employment exit probability" to refer to the probability of entering unemployment. We do not use this terminology because employment exit can be taken to mean a flow from employment to either unemployment or nonparticipation in the labor force, and may even be taken to mean any separation from employment, which would also include job-to-job flows.

12. See, among others, Perry (1972), Marston (1976), Blanchard and Diamond (1990), and Baker (1992).

13. See Braun, De Bock, and DiCecio (2006), Davis (2006), Elsby, Michaels, and Solon (2009), Fujita and Ramey (2009), Kennan (2006), and Yashiv (2008).

II.A. The Ins and Outs of Unemployment in the 2007 Recession

A first glimpse of the dynamics of unemployment flows can be obtained from published time series from the CPS.[14] Shimer (2007) describes a method that uses monthly series on the number of workers employed, the number unemployed, and the number unemployed for less than 5 weeks to infer the rates at which workers enter unemployment and unemployed workers exit unemployment. His point of departure is the following description of the path of the unemployment stock U_t:

$$\mathrm{d}U/\mathrm{d}t = s_t\left(L_t - U_t\right) - f_t U_t, \tag{2}$$

where s_t and f_t are, respectively, the unemployment inflow and outflow rates, L_t is the labor force, and t indexes months. Although some recent literature has referred to s_t and f_t as "separation" and "job-finding" rates, respectively, we instead use the terms "inflow" and "outflow" rates, for two reasons. First, many separations from employers do not result in a flow into unemployment, a point to which we return in section II.C. Second, f_t includes flows from unemployment to nonparticipation as well as to employment. The cyclical properties of the outflow rate in the 2007 and prior recessions are almost identical to those of transitions from unemployment to employment in longitudinally linked microdata.[15] We focus on the outflow rate because it is the proximate driving force for the changes in the unemployment rate, and because it is much more transparent to compute.[16]

The goal of the analysis is to relate variation in the unemployment rate $u_t = U_t/L_t$ to variation in the flow hazards s_t and f_t. To that end, we first need

14. Throughout the remainder of this section we focus on unemployment flows estimated from CPS time series, rather than the longitudinally matched monthly CPS microdata (the so-called gross flows data). This choice is informed by the fact that important measurement issues accompany the use of the gross flows data, including spurious transitions driven by measurement error in reported labor market states in consecutive monthly surveys, nonrandom attrition from the sample, and discrepancies between published changes in aggregate labor market stocks and those implied by the gross flows.

15. It is difficult to make strong statements on the importance of the distinction as one uses increasingly disaggregated data. The reason is that as one disaggregates the CPS data further, cell sizes start getting smaller and sampling variance worsens, yielding noisy estimates. This problem is aggravated when one uses longitudinally linked microdata, as in practice only a fraction of the CPS sample can be matched across months.

16. An implicit assumption underlying equation 2 is that all inflows into unemployment originate from employment, $L_t - U_t$. In fact, as we will show in what follows, in the United States a substantial fraction of inflows originate from nonparticipation. We relax this simplifying assumption in section II.C.

to estimate these flow rates. Following Shimer (2007), we compute the monthly outflow probability,

$$F_t = 1 - \left[\left(U_{t+1} - U_{t+1}^{<1}\right)/U_t\right], \tag{3}$$

where $U_{t+1}^{<1}$ is the stock of workers who report having been unemployed for less than one month.[17] Intuitively, the term inside the brackets is the fraction of the unemployed in month t who remain unemployed the next month, the complement of which is the monthly outflow probability. This can then be mapped into a Poisson outflow hazard rate $f_t = -\log(1 - F_t)$.

Obtaining an estimate of the inflow rate is slightly more involved. Assuming that the flow hazards s_t and f_t and the labor force L_t are constant between surveys, one can solve equation 2 forward one month to obtain

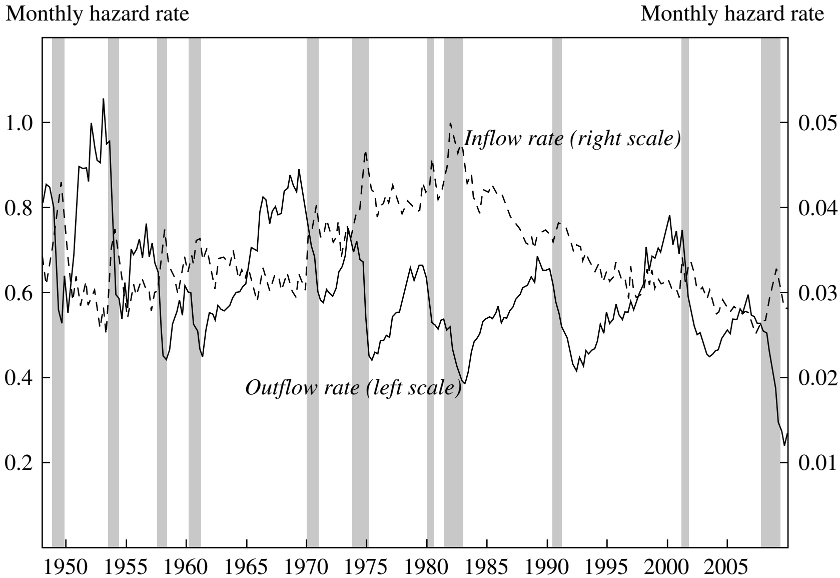
$$U_{t+1} = \lambda_t U_t^* + \left(1 - \lambda_t\right) U_t. \tag{4}$$

Here unemployment is a weighted average of the flow steady-state level of unemployment $U_t^* = s_t L_t/(s_t + f_t)$ and last month's unemployment U_t, with the weight given by the monthly rate of convergence to the steady state, $\lambda_t = 1 - e^{-(s_t+f_t)}$. Since we observe the labor force and unemployment stocks in each month, with an estimate of the outflow rate f_t in hand, equation 4 becomes a nonlinear equation that can be solved for the inflow rate s_t. As emphasized by Shimer (2007), this procedure for estimating s_t implicitly corrects for a time aggregation bias arising from inflows within a given month exiting before the next month's survey.

Figure 6 plots quarterly averages of the estimated monthly time series for the rates of inflow to and outflow from unemployment, using the most recent CPS data up to 2009Q4. The figure highlights a number of interesting properties of the dynamics of unemployment flows in past recessions. First, as emphasized in the entirety of research on unemployment flows, both old and new, the outflow rate from unemployment is markedly procyclical, exhibiting systematic and prolonged downswings in all recessions. Second, the inflow rate into unemployment is countercyclical, exhibiting sharp upswings at the onset of all recessions that tend to subside quickly by the end of the recession. Third, the response of unemployment

17. As noted by Polivka and Miller (1998) and Abraham and Shimer (2001), the published BLS time series on short-term unemployment displays a discontinuous decline following the CPS redesign in 1994, due to a change in the way unemployment duration was recorded. We correct the published postredesign series by rescaling it by a factor of 1.16. See Elsby, Michaels, and Solon (2009) for more details.

Figure 6. Unemployment Inflow and Outflow Rates, 1948–2009[a]



Source: Authors' calculations using BLS data.
a. Quarterly averages of monthly data. Shading indicates recessions.

inflows in the relatively mild recessions that began in 1990 and 2001 appears to be muted in comparison to other episodes, a point that echoes the recent conclusions of Hall (2005b, 2007) and Shimer (2007).

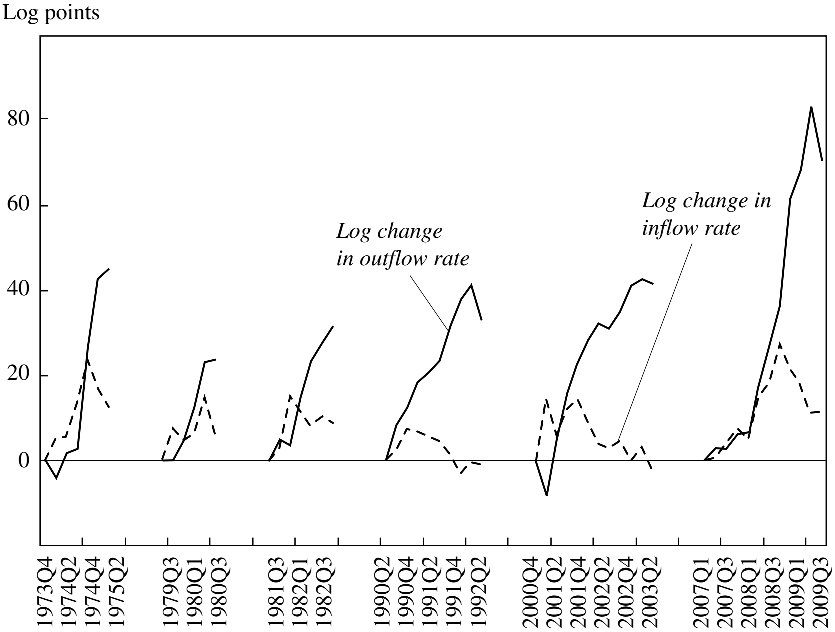
At this point we can return to the question that motivated this part of our analysis: to what extent is the cyclical ramp-up in unemployment accounted for by changes in these flow hazard rates? Elsby, Ryan Michaels, and Gary Solon (2009) provide a simple method for answering this question. Their starting point is the observation, noted by many analysts of U.S. unemployment flows, that the U.S. unemployment rate is very closely approximated by its flow steady-state value, that is:

$$u_t \equiv U_t / L_t \approx u_t^* \equiv s_t / (s_t + f_t). \quad (5)$$[18]

Equation 5 is useful for our purposes because it provides a link between variation in the unemployment stock and variation in the constituent flow

18. To see why this is so, note that the sum of the inflow and outflow rates $s_t + f_t$ typically exceeds 0.5 on a monthly basis in the United States. An implication is that the rate of convergence to flow steady state λ_t in equation 4 tends to be very high in practice.

Figure 7. Cumulative Changes in Unemployment Inflow and Outflow Rates in Six Recessions[a]



Source: Authors' calculations using BLS data.
a. Quarterly averages of monthly data.

hazard rates. Elsby, Michaels, and Solon (2009) show that simple log differentiation of this approximate relationship implies that

$$(6) \qquad \Delta u_t \approx \beta_{t-1}\left[\Delta \log s_t - \Delta \log f_t\right], \text{ where } \beta_{t-1} = u_{t-1}\left(1 - u_{t-1}\right).$$

Equation 6 has a simple message: to compare changes in inflow and outflow rates on an equal footing with respect to changes in unemployment, all one needs to do is compare the logarithmic variation in each of the flow hazards.

Figure 7 depicts the results from applying this decomposition of unemployment variation for each recession since 1973. We identify start and end dates for each recessionary ramp-up in unemployment since 1973 and compute the cumulative logarithmic difference in inflow and outflow rates relative to their values at the start of the recession. In many ways figure 7 confirms the qualitative picture suggested in figure 6. In all recessions, inflows account for a substantial fraction of unemployment variation early on and then subside. In contrast, the contribution of the outflow rate becomes more dominant as each recession progresses.

For our current focus, there are two noteworthy aspects of figures 6 and 7. First, mirroring the conclusions of section I on labor market stocks, the behavior of unemployment flows in the initial stages of the current downturn bears a striking resemblance to the dynamics of unemployment flows in past severe recessions. The early quarters of the current ramp-up in unemployment are characterized by a wave of inflows that has since partly receded. The contribution of the inflow rate is almost identical to that observed in the 1974 downturn. Thus, to return to the question that motivated this analysis, sharp spikes in the rate of inflow into unemployment appear to be a feature of severe recessions, rather than only of older ones.

Figures 6 and 7 also shed light on what is new about the current downturn. Figure 6 reveals that the unemployment outflow rate fell to a historic low of 24 percent in 2009Q3. This is not just a consequence of the secular trend toward declining outflow hazards shown in figure 6: figure 7 shows that the outflow rate fell by over 80 log points in the current downturn, more than in any of its postwar counterparts, echoing the conclusion of section I that this is the deepest postwar downturn in terms of labor market outcomes. We return to this phenomenon in section III, when we discuss its implications for the recovery.

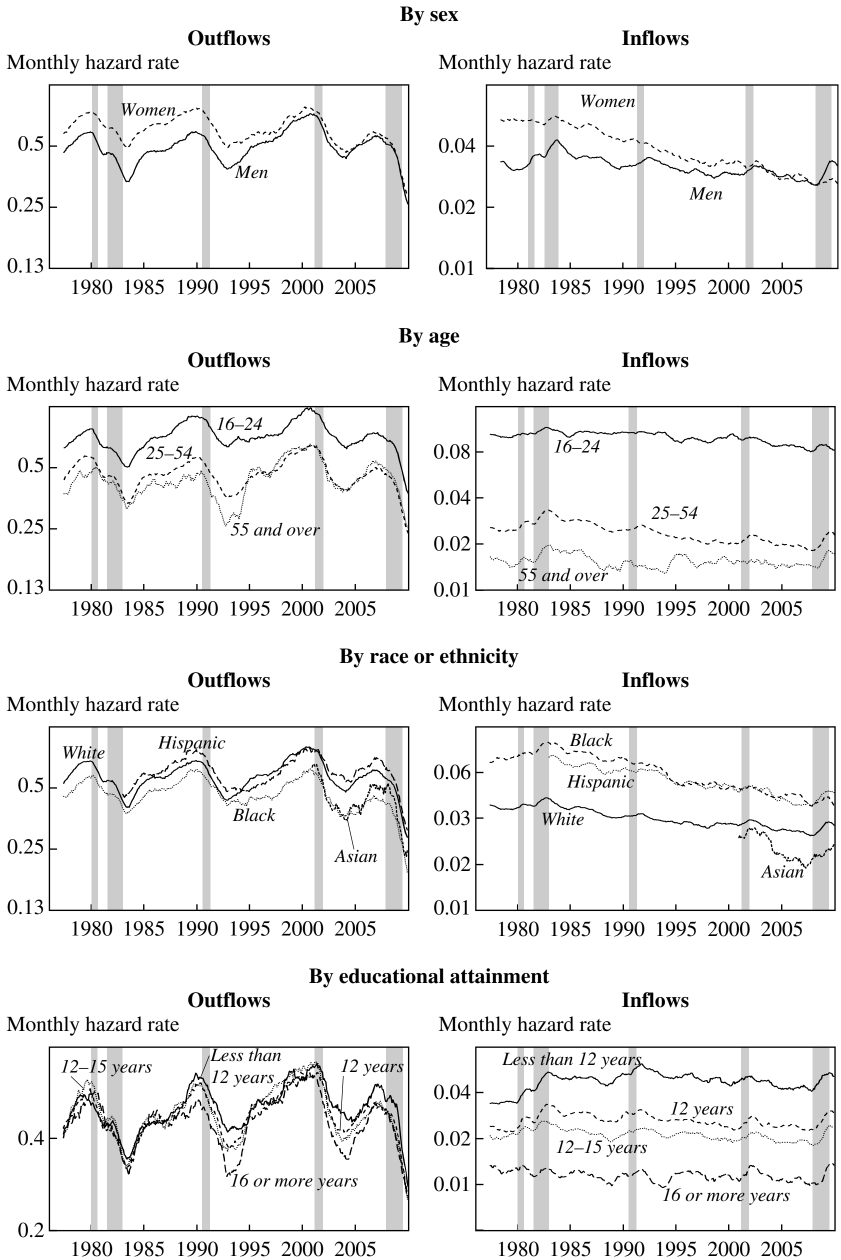
II.B. Unemployment Flows by Labor Force Group

In section I.B we showed that changes in unemployment rates have differed substantially across demographic groups during the 2007 recession, with some groups hit harder than others. We now look into the sources of this heterogeneity by examining unemployment flows across groups.

We focus on the same four dimensions of heterogeneity as in section I.B. Estimation of the flow hazards for each labor force group mirrors the aggregate analysis above.[19] Figure 8 displays the series for the inflow and outflow hazards for each group. They are calculated as 12-month moving averages to smooth out noise induced by the greater sampling variance that

19. The BLS publishes seasonally unadjusted estimates of unemployment by duration starting from the mid-1970s by sex, age, and race. As in section I.B, for education groups we use the CPS monthly microdata files from January 1976 onward to construct measures of the number unemployed less than 5 weeks, the total number unemployed, and the total number employed, by group. We then seasonally adjust the raw data using the Census' X12 procedure and compute the monthly outflow and inflow rates using the analogues to equations 3 and 4 that hold for each group. As before, we also correct for discontinuities in the series for short-term unemployment by group induced by the redesign of the CPS in 1994.

Figure 8. Unemployment Flows by Demographic or Educational Group, 1976–2009[a]



Source: Authors' calculations using BLS data.
a. Twelve-month moving averages of monthly data. Shading indicates recessions. Logarithmic scale.

accompanies these more disaggregated series. In accordance with the message of equation 6, the flow hazards are plotted on log scales.

Figure 8 has a rich set of implications for the structure of joblessness across groups. Perhaps its most prominent feature is the remarkable uniformity in both the levels and the cyclical behavior of outflow rates across groups within each dimension (left-hand panels). Most striking are the series by education group, for which the exit rates are virtually indistinguishable since 1976 (echoing the findings of Mincer 1991). In the 2007 recession the log decline in outflow hazards is almost identical across groups in all dimensions. Reductions in the outflow rate that accompany recessions, from both a qualitative and a quantitative perspective, are truly an *aggregate* phenomenon.

In stark contrast, there are large differences in rates of inflow into unemployment across groups (right-hand panels of figure 8). Comparison of these with the heterogeneity of unemployment across groups in table 1 reveals a close link: the same groups that face high unemployment rates—young workers, less educated workers, and workers from ethnic minorities—also face markedly high rates of entry into unemployment. This comparison indicates that the bulk of the large differences in unemployment across groups observed in table 1 is driven by differences in each group's propensity to enter unemployment, rather than differences in the duration of their spells.

In addition to revealing large differences in the levels of unemployment across groups, table 1 demonstrated that some groups face greater increases in unemployment in time of recession. What can account for this? Recalling equation 6, we can write the change in group j's unemployment rate as

$$(7) \quad \Delta u_{j,t} \approx \beta_{j,t-1}\left[\Delta \log s_{j,t} - \Delta \log f_{j,t}\right], \text{ where } \beta_{j,t-1} = u_{j,t-1}\left(1 - u_{j,t-1}\right).$$

One possibility, then, is that these groups simply faced larger logarithmic changes in their constituent flow hazards. Figure 8 reveals that this is precisely what accounts for the surge in unemployment of men relative to women in the current recession: male and female outflow rates have been essentially identical, but men have faced a much larger increase in inflows—a point emphasized by Şahin and others (2009).[20]

20. These authors explore this phenomenon using longitudinally linked monthly CPS microdata to estimate labor market flows among unemployment, employment, and nonparticipation. Consistent with figure 8, they find that for men the employment-to-unemployment transition rate increased more than it did for women, whereas the unemployment-to-employment transition rate declined proportionally across the two groups.

But this is not the whole story. For age, race, and education groups, there is little difference in the cyclicality of unemployment flows, and what differences exist tend to predict the opposite of the pattern depicted in table 1. For example, in the 2007 recession, outflow rates among young workers aged 16 to 24 fell just as much as for older workers, and their inflow rates have hardly risen. Yet in table 1 the unemployment rate among 16- to 24-year-olds rose substantially more than aggregate unemployment.

The answer to this puzzle lies in equation 7: for values of the group-specific unemployment rates $u_{j,t}$ observed in table 1 (as for all values lying below one-half), $\beta_{j,t-1}$ is increasing in $u_{j,t-1}$. Thus the higher the unemployment rate faced by an individual group, the greater the responsiveness of the group's unemployment rate to changes in its constituent flow hazards. Intuitively, equation 7 implies that changes in the flow hazards have a logarithmic influence on unemployment: a doubling of, for example, the inflow hazard leads to an almost doubling of the unemployment rate. The higher a group's unemployment rate, then, the more cyclically sensitive that rate is.

Figure 8 reveals that this observation can account entirely for the greater cyclical sensitivity of unemployment among youth, ethnic minorities, and the less educated in the 2007 recession, and indeed in all recessions over the sample period. Combining this with our earlier observation that the bulk of the differences in unemployment levels, and thereby of β_j, across groups can be attributed to differences in rates of entry into unemployment yields an interesting implication: the majority of the variation in both the levels and the cyclical sensitivity of group unemployment rates can be accounted for by differences in the *level* of inflow rates across groups.

II.C. The Role of Job Loss in the 2007 Recession

The previous sections have shown that unemployment inflows are a proximate driving force of the increase in unemployment in the 2007 recession, and that they play an important role in accounting for cross-sectional differences in the level and cyclicality of unemployment across groups. It is tempting to conclude that this constitutes evidence that *job loss* has played a key role in the 2007 recession. In this section we delve into this observation to uncover the mechanisms that can account for these elevated inflow rates.

We address two important conceptual distinctions. First, as mentioned above, estimates of the unemployment inflow rate, s_t based on equation 4, are based on the implicit assumption that all inflows into the unemployment pool originate from employment rather than nonparticipation. In fact, new entrants and reentrants to the labor force account for around 40 percent of the unemployment stock. Consequently, estimates of s_t conflate two economically distinct driving forces for entry into unemployment: flows from nonparticipation brought about by the process of labor force entry, and flows from employment to unemployment that are associated with elevated rates of job loss.

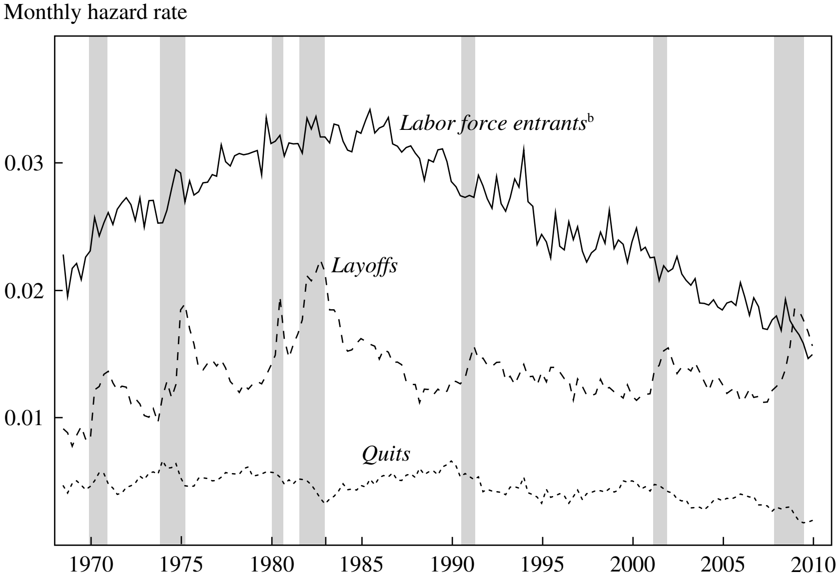
Second, job loss is often taken to mean a separation from an employer rather than an inflow into the unemployment pool. But workers leaving an employer can, and frequently do, line up new jobs without an intervening unemployment spell, a point that has been made since Peter Mattila (1974) and more recently by Bruce Fallick and Charles Fleischman (2004) and Éva Nagypál (2008). In what follows, we bring to bear a range of additional data that speak to these distinctions.

UNEMPLOYMENT INFLOWS BY REASON FOR UNEMPLOYMENT. It is possible to distinguish among different sources of unemployment flows using publicly available monthly time series from the CPS on the total number unemployed and the number unemployed for less than 5 weeks by reason for unemployment. We focus on three main reasons for unemployment: job loss (layoffs), job leaving (quits), and labor force entry.[21] An important benefit of this breakdown is that the first two categories originate from employment whereas the third originates from nonparticipation, allowing us to distinguish flows from employment to unemployment associated with job loss from the flows from nonparticipation to unemployment that accompany labor force entry.[22]

21. One can further decompose job losers into those on temporary versus those on permanent layoff, and labor force entrants into new entrants and reentrants. We do not distinguish among these, principally because the redesign of the CPS in 1994 led to substantial changes in the definitions of these subgroups and associated discontinuities in their time series. See Polivka and Miller (1998).

22. A potential concern when distinguishing between job leavers and job losers in the CPS data is that the distinction, much like the unemployment-nonparticipation distinction, can be blurred. Poterba and Summers (1984) find that although few job losers alter their reported reason for unemployment from month to month, around 25 percent of job leavers in May 1976 reported in the next month's survey that they lost their job. We are less concerned about this for two reasons. First, as shown in figure 9, job leavers make up such a small fraction of unemployment inflows that such response error is unlikely to distort the job loser inflow rate, our primary focus in this section. Second, we will show in figure 11 that the cyclical properties of the job loser inflow rate implied by household responses in the CPS are strikingly similar to those of the layoff separation rate implied by establishment responses in the JOLTS data.

Figure 9. Unemployment Inflows by Reason for Unemployment, 1968–2009[a]



Source: Authors' calculations based on the method of Elsby, Michaels, and Solon (2009) using BLS data.
a. Quarterly averages of monthly data. Shading indicates recessions.
b. Includes both reentrants and new entrants.

Elsby, Michaels, and Solon (2009) describe how these data can be used to infer estimates of unemployment flows by reason for unemployment.[23] Figure 9 shows that, as these authors emphasize, all of the observed countercyclicality in the aggregate inflow rate noted above is driven by a markedly countercyclical layoff inflow rate. The quit inflow rate is comparatively very low and mildly *pro*cyclical, thereby dampening the observed countercyclicality of aggregate inflows. In addition, inflows due to labor force entry are essentially acyclical, further moderating the rise in the aggregate inflow rate in time of recession.

The impression given by figure 9, and one that is a unifying theme of the present paper, is that the behavior of unemployment inflows by reason

23. There is a slight difference between the method used by Elsby, Michaels, and Solon (2009) to compute inflow rates by reason for unemployment and that used by Shimer (2007) to compute the aggregate inflow rate. Elsby and coauthors use a discrete time correction for time aggregation bias, whereas Shimer uses a continuous time correction. The results reported in Elsby, Michaels, and Solon (2009) suggest that this difference is not quantitatively important.

in the current downturn is again very reminiscent of past recessions. The behavior of the layoff inflow rate in particular suggests a simple two-way classification of recessionary episodes: deep recessions, such as that starting in 1974, the Volcker disinflation period of the early 1980s, and the present downturn, are characterized by markedly elevated layoff inflow rates; milder recessions, such as those starting in 1990 and 2001, are typified by a more modest increase in inflows due to layoffs. Again, the message of the 2007 recession is that severe modern recessions share many of the characteristics of deep recessions in the past.

EVIDENCE FROM LABOR TURNOVER. The fact that unemployment inflows rose markedly in the 2007 recession, and that layoff inflows dominated that trend, is suggestive of job loss playing a key role in driving cyclical rises in unemployment. But it is not necessarily conclusive. As noted by George Perry (1972) and recently reemphasized by Hall (2005), elevated rates of inflow into unemployment need not be the outcome of elevated rates of separation from employers: increased inflows in time of recession can occur if workers are increasingly unable to line up new jobs immediately upon separation. Under this alternative hypothesis, countercyclical inflows are a symptom of declining rates of job finding among potential job-to-job movers, rather than of elevated rates of job loss.

The 2007 recession provides a unique opportunity to assess these competing hypotheses: it is the first recession covered from its onset by the new Job Openings and Labor Turnover Survey (JOLTS).[24] This is crucial for our present purpose because it provides a representative measure of the rate at which employed workers separate from their employers in the United States. More formally, denote the separation rate from employers by σ_t, and the employment-to-unemployment inflow rate by s_t^{eu}. Note that a measure of the latter is given by the sum of the layoff and quit inflow rates presented above, $s_t^{eu} = s_{l,t} + s_{q,t}$. It follows that we can relate σ_t and s_t^{eu} simply according to

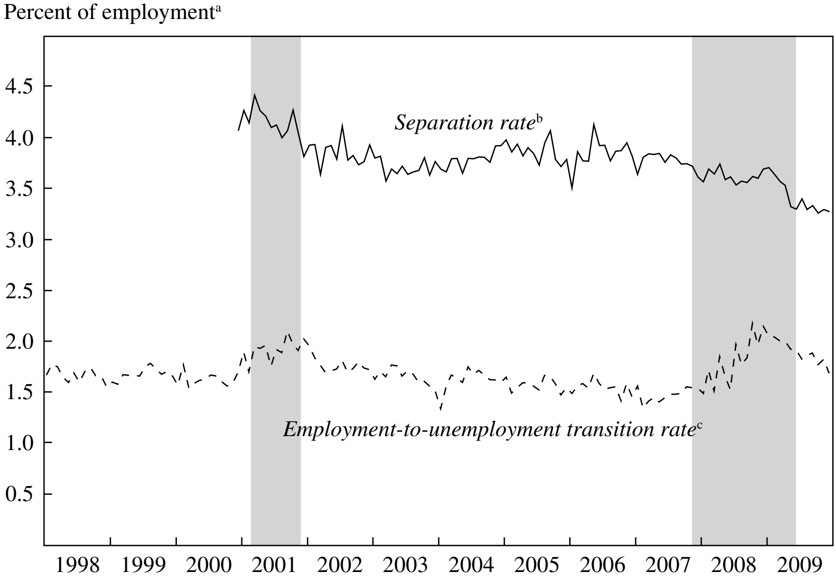
$$s_t^{eu} = p_t \sigma_t, \tag{8}$$

where p_t denotes the probability that a worker who separates from her employer in month t subsequently flows into unemployment.

Figure 10 plots the published JOLTS time series for the separation rate σ_t and the employment-to-unemployment transition rate s_t^{eu} implied by the CPS data. These series reveal a stark set of facts. First, the two rates differ

24. JOLTS data are available only back to December 2000 and therefore miss part of the ramp-up in unemployment in the 2001 recession.

Figure 10. Separation Rate and Employment-to-Unemployment Transition Rate, 1998–2009



Source: Authors' calculations using BLS data.

a. Monthly data and rates. Shading indicates recessions.

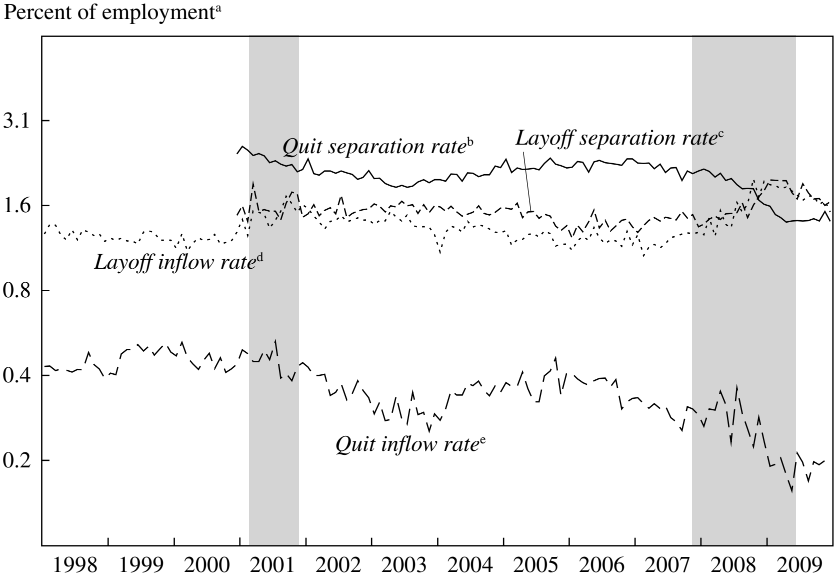
b. Includes all separations of workers from employers (job losers and job leavers), whether leading to unemployment or not.

c. Includes all transitions from employment to unemployment.

substantially at all points in time, a fact that is suggestive of the abundance of job-to-job transitions in the U.S. economy, as emphasized by Fallick and Fleischman (2004) and Nagypál (2008). Second, whereas the employment-to-unemployment inflow rate has increased in the current downturn, the separation rate has, if anything, *fallen* slightly. At first blush, then, it would seem that the elevated rate of inflow into unemployment during the 2007 recession is driven wholly by reductions in the rate at which workers line up new jobs.

The results presented in figure 10 would seem to provide ample support for Hall's (2005) hypothesis that in today's economy, job loss has little to do with increased unemployment in time of recession. We argue that such a conclusion would be premature. It has long been recognized that the relatively modest cyclical behavior of separations masks substantial cyclicality in its constituent elements: quits and layoffs. These tend to display markedly opposite cyclical patterns: the quit rate moves *pro*cyclically, whereas the

Figure 11. Separations and Unemployment Inflows from Quits and Layoffs, 1998–2009



Source: Authors' calculations using BLS data.
a. Monthly data and rates. Shading indicates recessions. Logarithmic scale.
b. Includes all voluntary separations, whether transitioning to another job or to unemployment or to nonparticipation.
c. Includes all involuntary separations, whether transitioning to another job or to unemployment or to nonparticipation.
d. Includes only involuntary separations, leading to unemployment.
e. Includes only voluntary separations, leading to unemployment.

layoff rate moves *counter*cyclically.[25] Figure 11 plots economy-wide layoff and quit rates from the JOLTS data for the current downturn and reveals that, as with unemployment flows, the behavior of labor turnover is again remarkably consistent with historical trends in these series.

Accounting for the difference between quits and layoffs allows a more revealing investigation of the relationship between separations and unemployment inflows than in equation 6. The employment-to-unemployment transition rate can be decomposed as follows:

(9) $$s_t^{eu} = p_{l,t}\sigma_{l,t} + p_{q,t}\sigma_{q,t} = \underbrace{\left[\omega_t p_{l,t} + \left(1-\omega_t\right)p_{q,t}\right]}_{p_t}\sigma_t,$$

25. See, for example, Slichter (1919), Woytinsky (1942), Akerlof, Rose, and Yellen (1988), and Anderson and Meyer (1994).

where subscripts l and q, respectively, denote layoffs and quits, $\sigma = \sigma_l + \sigma_q$ is the aggregate separation rate, and $\omega = \sigma_l/\sigma$ is the share of layoffs in aggregate separations. Equation 9 therefore highlights an additional channel by which employment-to-unemployment transitions may increase, namely, through changes in the *composition* of separations (layoffs versus quits) that occur during recessions ω.[26]

Figure 11 clarifies this point. It depicts the quit separation rate σ_q from the JOLTS data along with the quit inflow hazard into unemployment s_q derived from the CPS data using the method described in the previous section. At all points in time, workers who quit their previous job face a very low probability of subsequently entering unemployment: p_q averages just 16 percent over the sample period. Job-to-job flows drive an important wedge between separations and unemployment inflows due to quits. It is for this reason that quits account for only a small fraction of unemployment inflows. In addition, the implied series for p_q displays no cyclical pattern: it fell steadily from approximately 20 percent in 2001 to 14 percent in 2009. These two observations—that p_q is small, and that it has not risen in the current downturn—account for why the contribution of quits to increased unemployment inflows is not significant in the current downturn.

A quite different story holds for layoffs. Figure 11 shows that, at all points in time, laid-off workers face a very high probability of entering unemployment: p_l averages 91 percent since 2001. Job-to-job flows do not appear to be prevalent among laid-off workers. Moreover, although the gap between the separation and the inflow rates for layoffs closed in the early part of the current downturn, the rise in p_l accounts for only a small fraction of the overall rise in unemployment inflows, and for perhaps one-quarter of the overall rise in the layoff inflow rate.

Figure 11 therefore provides a unique perspective on the rise in unemployment inflows during the 2007 recession. As suggested by Hall (2005), elevated rates of entry into unemployment are not driven by increases in the overall rate at which workers separate from employers. But in contrast to the claims of recent literature, job loss nonetheless plays a crucial role in accounting for recessionary unemployment: increased inflows into unemployment can be traced to a shift in separations during

26. As with so much of the analysis of unemployment flows, this compositional point was first noted by Perry (1972), who refers to workers flowing into potential unemployment as possessing "lottery tickets" for avoiding entry into unemployment. In his words, "Those who enter the flow because they quit voluntarily have better lottery tickets than those who enter it because they are laid off. Since quits fall and layoffs rise when unemployment rises, the quality of the average lottery ticket of workers in the pool . . . will deteriorate. . . ." (p. 267).

the recession toward layoffs, and these laid-off workers are very likely to flow into unemployment. An increase in the layoff rate therefore played a central role in accounting for the increased rate of entry into unemployment in the 2007 recession.

III. Outlook for Recovery in the Labor Market

Until now we have concentrated on analyzing the behavior of labor market stocks and flows associated with the rise in unemployment in the 2007 recession. In this section we turn to the prospects for the labor market going forward.

Two features of figure 6 provide a first glimpse of the central features that will guide the recovery. First, since the spike in the unemployment inflow rate has largely subsided, the key to any future decline in unemployment is a recovery of the outflow rate. Second, the decline in the outflow rate that has accompanied the 2007 recession has been much more severe than in past recessions, making its recovery all the more salient.

One can think of the relative strength of the rebound in the outflow rate as determined by two things. First, how many new job openings will be created? Second, for a given increase in the number of vacancies, how quickly will the pool of unemployed find new jobs?

III.A. Vacancy Creation

Job creation reflects the overall health of the economy, and it is expected that as aggregate activity recovers, vacancy creation will also start to increase. However, many factors affect the timing and the level of vacancy creation during recoveries.

One positive factor for the recovery from the 2007 recession is the additional strength in vacancy creation due to the alleviation of the credit constraints that resulted from the financial crisis. Moreover, since the resolution of the financial crisis is likely to cause a substantial decline in aggregate and individual uncertainty, firms' willingness to hire could increase significantly. In particular, the passing of the crisis implies a drastic reduction in the probability of a detrimental aggregate economic outcome. As Ben Bernanke (1983) points out, such a reduction in the probability of "bad news" will increase the likelihood that firms will make the decisions to invest and hire, which are costly to reverse.

There are also reasons to imagine that the factors that explain the jobless recoveries of the 1990–91 and 2001 recessions are likely to be absent during the current episode. Tim Willems and Sweder van Wijnbergen

(2009) argue that labor hoarding can explain the jobless recoveries following the two earlier recessions. Labor hoarding is more likely during shallow recessions but much less likely during a deep recession like that of 2007, which exhibited sharp rises in rates of job loss. Similarly, Thijs van Rens (2004) and Kathryn Koenders and Richard Rogerson (2005) have argued that firms used the previous two recessions as an opportunity to improve their organizational efficiency and productivity. Since the 2001–07 expansion was neither exceptionally long nor very strong, it seems that the forces that might have limited hiring after the 1990–91 and 2001 recessions are much less likely to have a large and persistent effect during this recovery. However, the strength in productivity growth in the second half of 2009 that led to the deviation from Okun's law depicted in figure 3 may suggest that these forces are still present.

On the downside, some firms have considerable unused labor capacity in the form of part-time workers. As of December 2009, part-time workers who would prefer to work full-time made up 6.7 percent of total employment. Daly, Hobijn, and Kwok (2009b), among others, have argued that the pace of hiring relative to output growth during the recovery could be slowed by firms first increasing the hours of those already employed.

Finally, there are reasons to suspect that labor market changes over the last two decades will render any sharp reversal in employment less likely. For example, firms' use of temporary layoffs has declined, and with it the possibility of increasing employment at low cost.[27] In addition, the sharp recovery following the 1980s recession may have been aided by the reversal of the disinflationary monetary policy that instigated the recession in the first place, a feature the 2007 recession does not share.

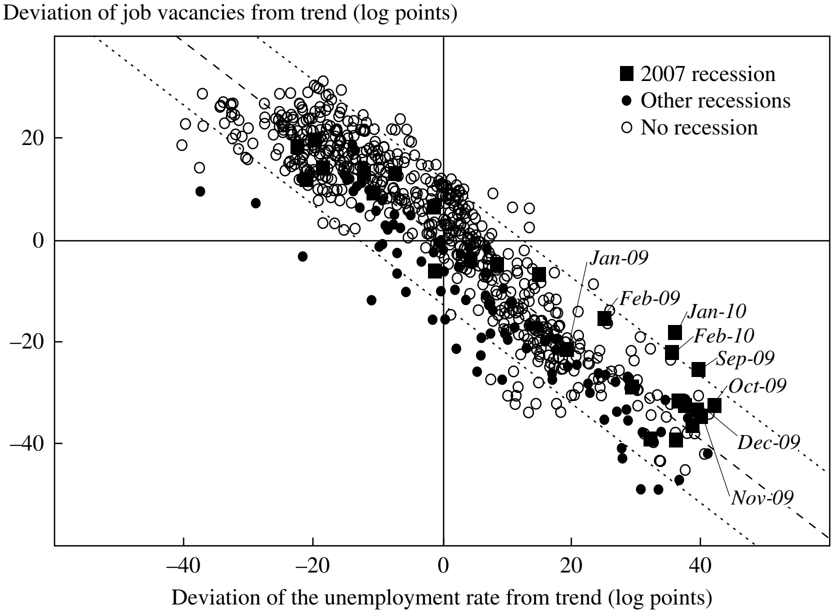
III.B. Match Efficiency and the Beveridge Curve

An important concern for the strength of the recovery is that even if firms create new jobs, it may be harder than in the past to match workers with appropriate job openings. Figures 12 and 13 reveal the main reason for this concern.[28] Figure 12 illustrates, for the period 1951–2009, the rela-

27. See Groshen and Potter (2003) for a detailed discussion.

28. Figures 12 and 13 are updated versions of figures 4 and 6 in Shimer (2005). For expositional purposes we plot monthly rather than quarterly data. To account for this change in frequency, we use a value of 2700000 for the smoothing parameter of the Hodrick-Prescott (HP) filter, which is used to filter the trend in log levels of all variables. This corresponds to the value that Shimer (2005) uses, corrected for the change in frequency using the factor for stock variables derived by Ravn and Uhlig (2002). The vacancy series is based on Barnichon (2010), who builds a vacancy posting index for the years 1951–2009 by combining information from the total print and online help-wanted advertising indexes with the

Figure 12. The Beveridge Curve, 1951–2010[a]



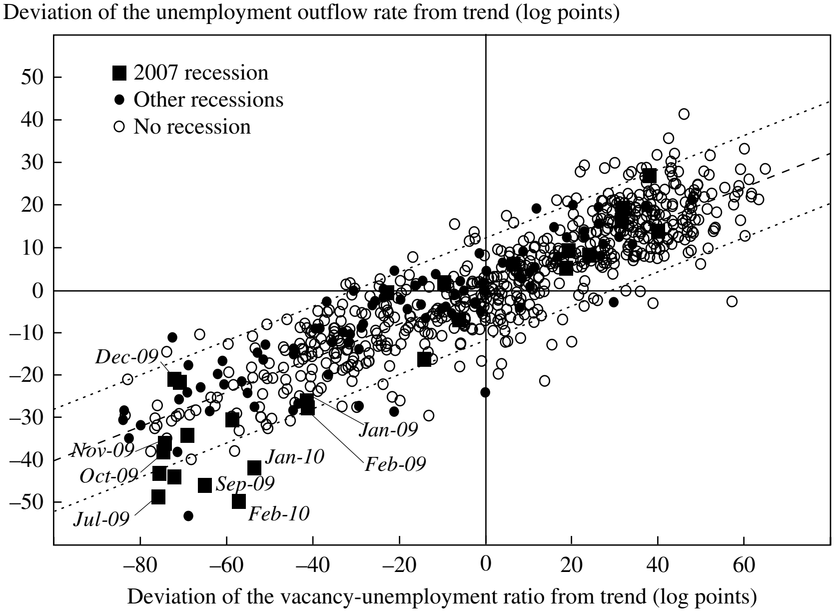
Source: Authors' calculations using data from BLS and Barnichon (2010).
a. Monthly data. Dotted lines are 90 percent confidence intervals around the fitted regression line.

tionship between logarithmic deviations from Hodrick-Prescott-filtered trends of vacancies and of the unemployment rate—the Beveridge curve. The fitted regression line is based on all observations before 2008, and 90 percent confidence intervals are shown. As noted by Shimer (2005), historically there has been a remarkably stable negative association between job openings and the unemployment rate. As the figure shows, during the fall of 2009 the unemployment rate was higher than would be implied by the historical Beveridge curve.

Figure 13 investigates the sources of this deviation from past trends. It plots the logarithmic deviations from Hodrick-Prescott-filtered trends of

JOLTS data. As discussed in Shimer (2005), the growth of Internet vacancy postings since the mid-1990s, together with newspaper consolidation and the equal opportunities legislation of the 1960s, makes it hard to compare the level of vacancies over time. Shimer uses a low-frequency HP filter to remove these trends. In addition, the series we use from Barnichon (2010) are robust to a range of possible higher-frequency paths for the diffusion of Internet vacancy postings. The cyclical component of the vacancy series that we use moves consistently with economic activity over the business cycle.

Figure 13. The Matching Function, 1951–2010[a]



Source: Authors' calculations using data from BLS and Barnichon (2010).
a. Monthly data. Dotted lines are 90 percent confidence intervals around the fitted regression line.

the outflow rate from unemployment f_t, and of the ratio of the number of vacancies to the number of unemployed, a measure of labor market tightness. Shimer (2005) refers to the remarkably stable positive relationship between these measures as the "matching function." The figure reveals that the recent divergence from the Beveridge curve can be traced to the outflow rate being substantially lower than would be suggested by the matching function relationship observed over much of the postwar period. The substantial decline in the outflow rate witnessed in the latter part of 2009 (figure 6) therefore represents a significant outlier in the context of the historical matching function.

The recent breakdown of the Beveridge curve and matching function relationships shown in figures 12 and 13 is evocative of the similar breakdown in match efficiency during the period of high European unemployment in the 1980s and 1990s (see, for example, figure 11 in Layard, Nickell, and Jackman 1991). This raises the concern that the U.S. economy may become plagued by the same persistently high unemployment rates that Europe experienced—the so-called hysteresis effect. In practice, hysteresis can arise through a number of channels. We highlight a few of these possibilities here and attempt to gauge their relevance in the current downturn.

MISMATCH BETWEEN WORKERS AND JOBS. One potential reason for a persistent reduction in match efficiency is a mismatch between the skills of workers and the skill requirements of job openings. For example, Erica Groshen and Simon Potter (2003) have argued that the jobless recoveries after the 1990–91 and 2001 recessions were in large part due to structural reallocation of workers across sectors.[29] They claim that this reallocation led to a mismatch in the skill mix that resulted in a slower adjustment of the labor market than in previous recessions. More recently, Edmund Phelps (2008) has reiterated this concern with respect to construction and finance workers in the 2007 recession.

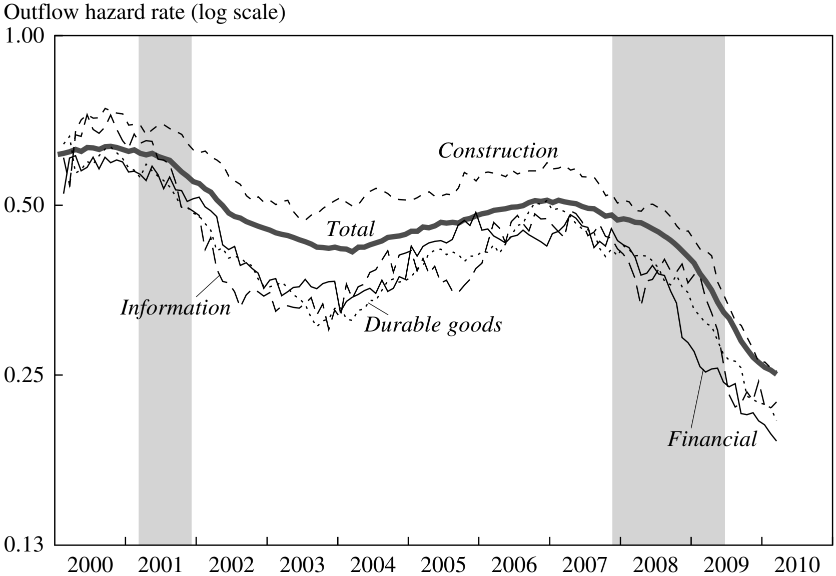
This reallocation argument suggests that workers formerly employed in sectors in structural decline will have a harder time finding new jobs than other workers. That is, it implies a divergence in outflow rates from unemployment between these two groups of workers. Figure 14 addresses this question by showing unemployment outflow hazard rates conditional on the industry in which a worker was employed at the start of the unemployment spell. If anything, these outflow rates have converged rather than diverged as the structural reallocation argument implies.[30]

Besides a mismatch in skills, an additional concern is the potential emergence of geographical disparities in the location of workers and of job openings. This issue came into focus in the 2007 recession amid concerns that, given the decline in home prices that accompanied the recession, job applicants are more reluctant to apply for and accept jobs that are not within commuting distance and would require them to sell their home. Fernando Ferreira, Joseph Gyourko, and Joseph Tracy (2008), using data from the American Housing Survey for 1985–2005, find that homeowners with

29. Related to this argument, Aaronson, Rissman, and Sullivan (2004) point out that the need to reallocate labor across sectors in the 1990–91 and 2001 recessions, which were accompanied by jobless recoveries, seemed no greater than in earlier ones. Valletta and Cleary (2009) reach the same conclusion for the 2007 recession.

30. Although this finding is suggestive, it need not imply that skill mismatch is not an issue in the 2007 recession. For example, it may be the case that skill mismatch exists but occurs *within* industry classifications. In that case, disaggregation by industry would be too broad to detect an increase in skill mismatch. However, estimation of further disaggregated unemployment flows is limited by the increased sampling variance that would accompany additional splitting of the CPS sample.

Figure 14. Unemployment Outflow Rates in Selected Industries



Source: Authors' calculations using BLS data.
a. Twelve-month moving averages of non-seasonally adjusted data. Shading indicates recessions.

negative equity are less likely than other homeowners to move.[31] Their results cannot be easily extrapolated to the 2007 recession but still point to a potentially important negative effect of housing-related problems on labor market recovery, since geographic mobility is an important part of adjustment to shocks in the U.S. labor market, as emphasized by Blanchard and Lawrence Katz (1992).[32]

SCLEROSIS AND DURATION DEPENDENCE. Associated with the record rise in the unemployment rate in the 2007 recession has been a surge in long-term unemployment. The fraction of the labor force unemployed for more than 6 months has increased by a staggering 3.5 percentage points to a postwar high of 4 percent, 1.5 percentage points higher than the previous

31. Some commentators on the 2007 recession have pointed to recent data showing that the rate of domestic migration in the United States has reached a postwar low. However, it is difficult to discern how much of this decline is associated with the recession; rates of internal migration have been falling as a secular phenomenon since the mid-1980s (see, for example, Frey 2009).

32. This implication of Blanchard and Katz (1992) has been the source of some dispute, however. See, for example, Rowthorn and Glyn (2006).

high in 1983. Likewise, average unemployment duration has risen to a historic high of more than 30 weeks—the mirror image of the historic low in the unemployment outflow rate noted in section II. Here we explore the effects of these depressed unemployment flows on the likely path of the recovery, what Samuel Bentolila and Giuseppe Bertola (1990) and Blanchard (2000) have referred to as "sclerosis" in the European context.[33]

A first potential source of sclerosis relates to the effect of reductions in unemployment outflow rates on the speed of adjustment of the unemployment rate. This point can be clarified in terms of equation 4: reductions in the pace of worker reallocation, $s_t + f_t$, lead to reductions in the responsiveness of unemployment to changes in flow steady-state unemployment, $u_t^* = s_t/(s_t + f_t)$. This matters for the recovery of unemployment in the wake of the 2007 recession: a by-product of the historically low outflow rate reached during this recession is that the rate of convergence of unemployment to its flow steady state, λ_t in equation 4, has also arrived at a postwar low. Thus, even if firms start to hire again, the outflow rate rebounds, and flow steady-state unemployment recovers, the actual unemployment rate may exhibit a delayed reaction.

Quantitatively, however, we find that these effects are likely to be small. Although the recent trough in the monthly outflow rate of 24 percent is a record low by historical U.S. standards, it remains very high in comparison with rates in Europe during the 1980s, which fell below 8 percent in many European economies.[34] To put this in perspective, the half-life of a deviation of unemployment from flow steady state, which stood at a little over 1 month before the current downturn in the United States, has risen to just under 3 months in recent U.S. data but is not even close to the 9 months to a year experienced in Europe in the 1980s and early 1990s.[35]

A second source of sclerosis is the persistence in the decline of the outflow rate itself. Previous literature has identified the duration composition of unemployment as a key potential driving force for such persistence (Blanchard 2000). Specifically, a pervasive feature of U.S. unemployment flows

33. We use the term "sclerosis" in the sense of Blanchard (2000, p. 2): "Flows decrease, individual unemployment duration increases, and so does the proportion of long-term unemployed."

34. Hobijn and Şahin (2009, table 1) report average duration distributions of unemployment spells, and Elsby, Hobijn, and Şahin (2009) document the behavior of inflow and outflow rates over time for a broad number of industrialized countries. Even the unemployment-to-employment transition rate for the United States (currently around 20 percent on a monthly basis) substantially exceeds the outflow rate (the sum of the unemployment-to-employment and unemployment-to-nonparticipation flow rates) in many European countries.

35. These figures are computed from estimates in Elsby, Hobijn, and Şahin (2009, figure 3).

is that average rates of outflow from unemployment decline as the duration of unemployment spells rises—so-called negative duration dependence—a point noted first by Hyman Kaitz (1970) and more recently by Shimer (2008).[36] Several explanations have been proposed for such an outcome, including depreciation of the skills of the unemployed (Pissarides 1992, Ljungqvist and Sargent 1998), employers' ranking of job applicants by the duration of their unemployment spell (Blanchard and Diamond 1994), and statistical discrimination by employers against the long-term unemployed (Lockwood 1991).

Here we highlight some potential reasons why such duration dependence can matter for labor market conditions over the cycle. Noting that the aggregate outflow probability F_t can be expressed as a share-weighted sum of the outflow probabilities faced by each duration group d, $F_t = \Sigma_d \omega_{dt} F_{dt}$, it follows that changes in the aggregate outflow probability over time can be decomposed according to

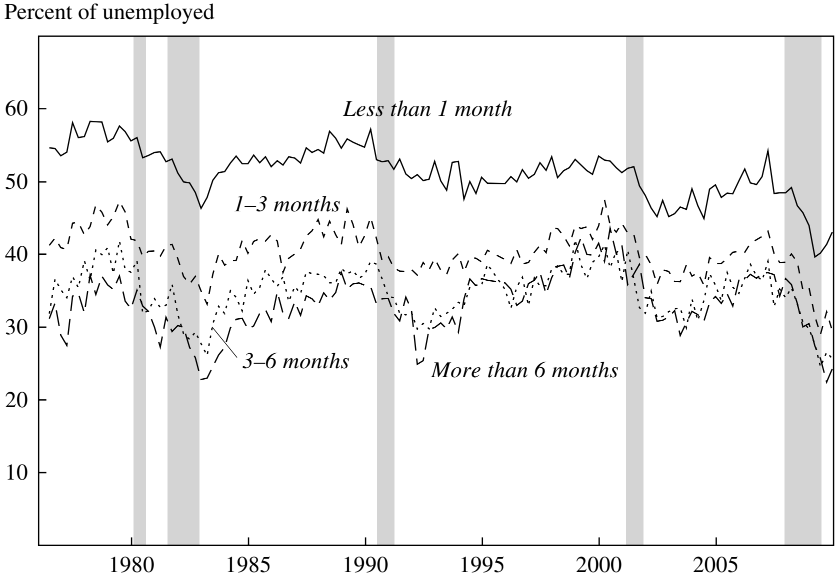
$$\Delta F_t = \sum\nolimits_d \left(\omega_{dt} \Delta F_{dt} + \Delta \omega_{dt} F_{dt-1} \right). \tag{10}$$

Equation 10 summarizes two potential concerns about the role of duration dependence in the 2007 recession. First, given the surge in long-term unemployment, it is tempting to hypothesize that workers with longer unemployment spells have increasingly become disenfranchised from the labor market, leading to a disproportionate decline in their outflow rates. Such an effect would be captured by the first term in parentheses in equation 10.

Figure 15 addresses this question by presenting time series for a range of outflow rates for workers with different unemployment durations. Specifically, we use longitudinally linked monthly CPS microdata from 1976 onward to compute the probability that a worker unemployed for a given duration exits unemployment within a month. Figure 15 plots the associated hazards for durations of less than 1, 1 to 3, 3 to 6, and 6 or more months. Consistent with the literature on negative duration dependence in unemployment exit rates, the hazard for exiting unemployment declines as duration rises. More important for the hypothesis under discussion, however, there is no evidence that exit rates have fallen disproportionately among the high-duration unemployed in the last five recessions. Rather,

36. As noted by Kaitz (1970), this phenomenon may take the form of "spurious" duration dependence that arises from dynamic selection (Salant 1977), or of "true" duration dependence whereby the accumulation of unemployment duration has a causal effect on outflow rates.

Figure 15. Unemployment Outflow Probabilities by Duration of Unemployment, 1976–2009[a]



Source: Authors' calculations using CPS data.
a. Quarterly averages of monthly data. Shading indicates recessions.

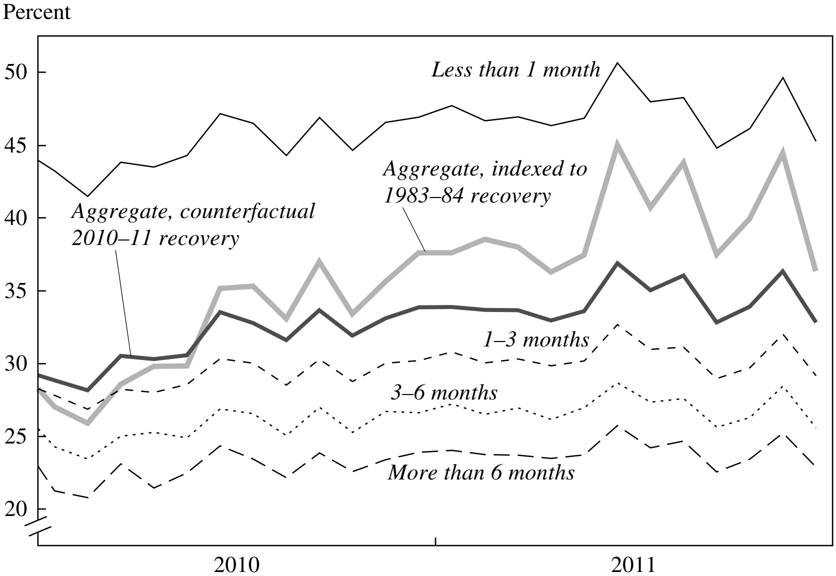
just as we showed in section II.B on unemployment flows by group, the cyclicality of outflow rates displays an extraordinary regularity across duration groups. In sum, there appears to be little evidence that elevated rates of joblessness are a symptom of diminished employment opportunities for the long-term unemployed in the 2007 or any other recession.[37]

However, equation 10 also reveals that duration dependence can affect the cyclicality of the aggregate outflow rate through changes in the duration structure of unemployment, $\Delta\omega_{dt}$. Formally, a simple description of the stock of unemployed workers of duration d over time t is

$$u_{dt+1} = \left(1 - F_{d-1t}\right)u_{d-1t}, \tag{11}$$

37. Interestingly, this conclusion mirrors the results of Machin and Manning's (1999, p. 3086) detailed analysis of the long-term unemployment problem in Europe: "While the longterm unemployed do leave unemployment at a slower rate than the shortterm unemployed, this has always been the case and their relative outflow rate has not fallen over time."

Figure 16. Simulated Unemployment Outflow Rates, 2010–11[a]



Source: Authors' calculations.

a. Monthly data. The simulation assumes that outflow rates for each duration group recover at the same rate as in the 1983–84 recovery. For purposes of comparison, the rebound in the aggregate outflow rate in that period is plotted.

with initial condition u_{0t} given by the unemployment inflow derived in section II. It follows that the unemployment share of duration group d is given by

$$\omega_{dt+1} = (1 - F_{d-1t})(u_t / u_{t+1})\omega_{d-1t}. \tag{12}$$

Equation 12 has significant implications for the path of the outflow rate during the recovery. It reveals that the unemployment shares of the high-duration unemployed are persistent, and in particular that they depend on the outflow rates faced by the low-duration unemployed that prevailed in the past. Intuitively, even if outflow rates have moved uniformly across duration groups during the 2007 recession, the historic decline in outflow rates as a whole can result in a persistent residue of long-term unemployed workers who exit unemployment slowly, depressing aggregate outflow rates in the future.

To illustrate the potential importance of this mechanism, figure 16 simulates the future path of the aggregate outflow rate in the wake of the 2007 recession, assuming that outflow rates for each duration group, as well as the aggregate inflow rate, rebound in proportion to what was witnessed in the last recovery from a deep recession, that of 1983–84. For comparison, figure 16 also plots an alternative path for the aggregate outflow rate, indexed to the actual recovery observed in 1983–84.

Figure 16 suggests that the accumulation of long-term-unemployed workers in the 2007 recession can indeed have quantitatively important effects on the rebound in the outflow rate during the recovery. Whereas the aggregate outflow rate rebounded by around 30 percent in the 1983–84 recovery, the simulated path for the upcoming recovery augurs a more lackluster 15 percent.

The difference between these two paths is largely due to the low outflow rates prevailing at the end of 2009. Hence, even if these rates were to rebound at the same growth rate as in 1983, they would remain at a lower level than in 1983, leading to a higher average duration of unemployment, even in the long run. Although this is definitely a cause for concern, it is unlikely that this mechanism will lead to the degree of persistence in the outflow rate that marked the hysteresis seen in European unemployment in the 1980s and 1990s. The simple reason is that the long-term unemployed in the United States flow out of unemployment at a rate that is *four times higher* than the aggregate outflow rates in continental Europe reported in Elsby, Hobijn, and Şahin (2009).

THE ROLE OF EMERGENCY UNEMPLOYMENT COMPENSATION. One particularly salient reason for a temporary decline in match efficiency relates to the temporary extension of federal Emergency Unemployment Compensation (EUC) that began in June 2008. In addition to the regular 26 weeks of unemployment insurance (UI), workers may be eligible for 53 additional weeks of EUC as long as Congress continues to extend it.[38] Conventional economic theory suggests that this lengthening of the expected duration of unemployment benefits will place downward pressure on the unemployment outflow rate seen in figure 13, as those searching for a job become more selective about which job offers they accept.

Existing research on the effects of UI benefits suggests a strong positive relationship between their maximum duration and the average unemployment

38. EUC is divided into four tiers (20 weeks, then 14, then 13, and finally another 6 weeks); recipients must reapply when each tier expires. In addition to these 53 extra weeks, most states offer extended benefits of up to 20 weeks. The number claiming these benefits has been relatively small.

spell. Estimates suggest that a 1-week increase in potential benefit duration is associated with an increase in the average duration of the unemployment spells of UI recipients of around 0.08 to 0.20 week (see Moffitt 1985, Katz and Meyer 1990, Meyer 1990, Card and Levine 2000, Krueger and Meyer 2002). According to these estimates, then, a 53-week extension in potential benefit duration would be associated with an average 4.2- to 10.6-week increase in unemployment duration among UI recipients.[39] Since the fraction of unemployed workers claiming some form of UI benefits has averaged 50 percent in the 2007 recession, this suggests something like a 2.1- to 5.3-week increase in overall unemployment duration. Over the course of the 2007 recession, average unemployment duration surged from 16.5 weeks to 30.2 weeks, a 13.7-week increase. This back-of-the-envelope calculation therefore suggests that EUC can account for as much as 15 to 40 percent of the rise in aggregate unemployment duration. This is a potentially substantial effect, which corresponds to between 0.7 and 1.8 percentage points of the 5.5-percentage-point rise in the unemployment rate.

There are reasons to believe, however, that the true effect of extended UI benefits on unemployment duration is likely to be at the lower end of these estimates. Many of the larger estimates of the effect are based on data from the 1970s and 1980s, when temporarily laid-off workers, who are more responsive to the generosity of UI, made up a larger fraction of unemployment. In addition, many of the larger estimates in the literature are based on empirical strategies that identify the effect of UI by exploiting differences in benefit schedules across states and time. As Card and Levine (2000) point out, however, many states extend UI benefits as a response to poor job-finding prospects in recessions, so that this approach may overstate the true disincentive effect of UI. Indeed, Card and Levine's estimates based on an exogenous policy change lie at the low end of the range of effects, suggesting a more modest impact of EUC.

NOT ALL VACANCIES ARE ASSOCIATED WITH JOB CREATION. A final reason for the observed decline in match efficiency could be that the measured stock of vacancies overstates the true number of job openings in the economy. Evidence from microdata on vacancies presented by Steven Davis, Jason Faberman, and John Haltiwanger (2009, figure 5) suggests that establishments whose employment is not growing nevertheless post vacancies. They estimate that these firms have a vacancy rate of about 2 percent of

39. This calculation assumes that, upon entering unemployment, all unemployed workers anticipate that benefit duration will be extended by 53 weeks. In that sense it is an upper bound on the response.

employment. Interestingly, this is about equal to the aggregate vacancy rate observed during the second half of 2009. This suggests that a substantial part of the vacancies reported in the latter half of 2009 may be associated not with job creation, but rather with a minimum level of vacancy postings that exists regardless of the level of net job growth.

Taken together, our analysis of the decline in match efficiency observed in the latter stages of the 2007 recession points to two potentially important driving forces: the existence of a substantial residue of long-term unemployed workers with relatively weak search effectiveness, and the extension of EUC. Taking these separately, one might imagine that the temporary nature of EUC implies that the labor market will recover as these benefits are withdrawn, whereas the structural nature of the long-term unemployment problem will cause it to endure well into the recovery. However, there are likely to be important interactions between the two factors. A major impetus for the introduction of the EUC program was in fact the rise in long-term unemployment that accompanied the recession. Thus an enduring long-term unemployment problem could mean that the political will to withdraw EUC may take some time to materialize.

IV. Conclusion

Our detailed analysis of the adjustment of the labor market in the current downturn reveals it to be the deepest deterioration in labor market outcomes on record in the postwar era. Every indicator of labor market activity suggests that the recession has been unique in both its depth and its duration. Rates of joblessness among all groups in the labor market have reached historic postwar highs. There is little doubt that it is a Great Recession.

Nonetheless, our analysis suggests that many of the features of labor market dynamics in the Great Recession through the latter half of 2009 are strikingly similar to those seen in earlier recessions. This is true of the behavior of employment and the labor force participation rate, the use of the intensive versus the extensive margin in the adjustment of labor input, and the differential impact on demographic groups, with young workers, male workers, less educated workers, and workers from ethnic minorities hit harder than others.

In terms of the underlying flows, just as in earlier deep recessions, increased joblessness in the current downturn can be traced to both increased rates of inflow into unemployment and increased duration of unemployment spells, with higher inflows relatively more important early on in the downturn. This suggests that the more modest response of unemployment inflows in the 1990–91 and 2001 recessions is a feature of mild recessions rather than of modern ones.

Further analysis of worker turnover data from the new Job Openings and Labor Turnover Survey provides a unique perspective on the driving forces of job loss in the 2007 recession. Recent literature has emphasized the relatively acyclical behavior of the rate at which workers separate from employers, suggesting that job loss plays only a limited role in driving recessionary unemployment. Combining data from JOLTS and the CPS reveals that increased inflows into unemployment have been driven predominantly by a change in the composition of separations toward layoffs, which are very likely to lead to unemployment, and away from quits, which are very likely to lead to a new job upon separation. Thus, contrary to recent claims, increases in layoffs have played a key role in driving increased unemployment in the recession.

Although the labor market response in the early stages of the 2007 recession has resembled that in prior downturns, more recent evidence suggests an important divergence from past trends. Most prominently, rates of exit of unemployed workers from joblessness have slowed to record low levels, drawing into focus the importance of a rebound in outflow rates for the recovery. Recent data point to two key factors. First, the record rise in long-term unemployment associated with the recession is likely to yield a persistent overhang of workers facing long unemployment spells, slowing the recovery. Second, the extension of EUC starting in June 2008 is likely to have led to a modest increase in long-term unemployment in the recession.

Despite these unfavorable forces, recent data suggest that the problems facing the U.S. labor market going forward are unlikely to be as severe as the European hysteresis problem of the 1980s. Although the jobless in the United States are exiting unemployment at a historically slow rate, they nonetheless leave unemployment as much as four times faster than their counterparts in continental Europe in the 1980s. Looking ahead, then, a tentative expectation is for a lackluster recovery, but one not nearly as dismal as seen in Europe in the past.[40]

40. Even after the unemployment rate recovers, labor market disturbances associated with the recession are likely to have important and potentially long-lasting effects on workers. Since Ruhm (1991), and Jacobson, LaLonde, and Sullivan (1993), research has emphasized that the negative effects of displacement go beyond a temporary unemployment spell, as displaced workers often suffer substantial wage losses even after reemployment. Sullivan and von Wachter (2009) argue that job displacement might also have an effect on mortality, with annual death hazards 10 to 15 percent higher for high-seniority displaced male workers

ACKNOWLEDGMENTS We are especially grateful to Bruce Fallick, Lawrence Katz, Ryan Michaels, Robert Shimer, Gary Solon, and the editors for particularly detailed comments and suggestions. Thanks also to Mary Daly, Steven Davis, Henry Farber, Robert Hall, John Haltiwanger, Marianna Kudlyak, Simon Potter, and Jonathan Willis for constructive comments. We would like to thank Regis Barnichon for providing us with the Composite Help-Wanted Index data, and Joyce Kwok, Joseph Song, and Theodore Wiles for their outstanding research assistance.

20 years after displacement. The recession might also have negative effects on the careers of new labor market entrants. Oreopoulos, von Wachter, and Heisz (2006) find that students graduating in a recession start work at lower-paying employers, with permanent effects on low-skilled graduates.

References

Aaronson, Daniel, Ellen R. Rissman, and Daniel G. Sullivan. 2004. “Can Sectoral Reallocation Explain the Jobless Recovery?” Federal Reserve Bank of Chicago *Economic Perspectives* (Second Quarter): 36–49.

Abraham, Katharine G., and Robert Shimer. 2001. “Changes in Unemployment Duration and Labor Force Attachment.” In *The Roaring Nineties: Can Full Employment Be Sustained?* edited by Alan B. Krueger and Robert Solow, pp. 367–420. New York: Russell Sage Foundation and Century Foundation Press.

Akerlof, George A., Andrew K. Rose, and Janet L. Yellen. 1988. “Job Switching and Job Satisfaction in the U.S. Labor Market.” *BPEA,* no. 2: 495–582.

Anderson, Patricia M., and Bruce D. Meyer. 1994. “The Extent and Consequences of Job Turnover.” *BPEA: Microeconomics* 177–236.

Baker, Michael. 1992. “Unemployment Duration: Compositional Effects and Cyclical Variability.” *American Economic Review* 82, no. 1: 313–21.

Barnichon, Regis. 2010. “Building a Composite Help-Wanted Index.” Finance and Economics Discussion Series 2009-35. Washington: Board of Governors of the Federal Reserve System.

Bentolila, Samuel, and Giuseppe Bertola. 1990. “Firing Costs and Labour Demand: How Bad Is Eurosclerosis?” *Review of Economic Studies* 57, no. 3: 381–402.

Bernanke, Ben S. 1983. “Irreversibility, Uncertainty, and Cyclical Investment.” *Quarterly Journal of Economics* 98, no. 1: 85–106.

Blanchard, Olivier. 2000. “The Economics of Unemployment: Shocks, Institutions, and Interactions.” Lionel Robbins Lectures. London School of Economics.

Blanchard, Olivier Jean, and Peter Diamond. 1990. “The Cyclical Behavior of the Gross Flows of U.S. Workers.” *BPEA,* no. 2: 85–143.

———. 1994. “Ranking, Unemployment Duration, and Wages.” *Review of Economic Studies* 61(3): 417–34.

Blanchard, Olivier Jean, and Lawrence F. Katz. 1992. “Regional Evolutions.” *BPEA,* no. 1: 1–75.

Braun, Helge, Reinout De Bock, and Riccardo DiCecio. 2006. “Aggregate Shocks and Labor Market Fluctuations.” Working Paper no. 2006-004A. Federal Reserve Bank of St. Louis.

Card, David, and Phillip B. Levine. 2000. “Extended Benefits and the Duration of UI Spells: Evidence from the New Jersey Extended Benefit Program.” *Journal of Public Economics* 78, no. 1–2: 107–38.

Clark, Kim B., and Lawrence H. Summers. 1981. “Demographic Differences in Cyclical Employment Variation.” *Journal of Human Resources* 16, no. 1: 61–79.

Daly, Mary, Bart Hobijn, and Joyce Kwok. 2009a. “Labor Supply Responses to Changes in Wealth and Credit.” FRBSF Economic Letter 2009-05. Federal Reserve Bank of San Francisco.

———. 2009b. “Jobless Recovery Redux?” FRBSF Economic Letter 2009-18. Federal Reserve Bank of San Francisco.

Davis, Steven J. 2006. "Job Loss, Job Finding, and Unemployment in the U.S. Economy over the Past Fifty Years: Comment." In *NBER Macroeconomics Annual 2005,* edited by Mark Gertler and Kenneth Rogoff, pp. 139–57. MIT Press.

Davis, Steven J., R. Jason Faberman, and John C. Haltiwanger. 2009. "The Establishment-Level Behavior of Vacancies and Hiring." Working Paper no. WP-09-14. Federal Reserve Bank of Philadelphia Research Department.

Elsby, Michael, Bart Hobijn, and Ayşegül Şahin. 2009. "Unemployment Dynamics in the OECD." Working Paper no. 14617. Cambridge, Mass.: National Bureau of Economic Research.

Elsby, Michael W. L., Ryan Michaels, and Gary Solon. 2009. "The Ins and Outs of Cyclical Unemployment." *American Economic Journal: Macroeconomics* 1, no. 1: 84–110.

Fallick, Bruce, and Charles A. Fleischman. 2004. "Employer-to-Employer Flows in the U.S. Labor Market: The Complete Picture of Gross Worker Flows." Finance and Economics Discussion Series 2004-34. Washington: Board of Governors of the Federal Reserve System.

Farber, Henry S., and Bruce Western. 2010. "The Distribution of Unemployment, Employment, and Hours in the Current Recession: A Preliminary Look." Princeton University.

Ferreira, Fernando, Joseph Gyourko, and Joseph Tracy. 2008. "Housing Busts and Household Mobility." Working Paper no. 14310. Cambridge, Mass.: National Bureau of Economic Research.

Frey, William H. 2009. "The Great American Migration Slowdown: Regional and Metropolitan Dimensions." Brookings.

Fujita, Shigeru, and Garey Ramey. 2009. "The Cyclicality of Separation and Job Finding Rates." *International Economic Review* 50, no. 2: 415–30.

Gomme, Paul, Richard Rogerson, Peter Rupert, and Randall Wright. 2004. "The Business Cycle and the Life Cycle." In *NBER Macroeconomics Annual 2004,* edited by Mark Gertler and Kenneth Rogoff, pp. 415–461. MIT Press.

Gordon, Robert J. 2010. "Okun's Law, Productivity Innovations, and Conundrums in Business Cycle Dating." Northwestern University.

Groshen, Erica L., and Simon Potter. 2003. "Has Structural Change Contributed to a Jobless Recovery?" *Current Issues in Economics and Finance* 9, no. 8. Federal Reserve Bank of New York.

Hall, Robert E. 2005a. "Employment Efficiency and Sticky Wages: Evidence from Flows in the Labor Market." *Review of Economics and Statistics* 87, no. 3: 397–407.

———. 2005b. "Job Loss, Job Finding, and Unemployment in the U.S. Economy over the Past Fifty Years." In *NBER Macroeconomics Annual 2005,* edited by Mark Gertler and Kenneth Rogoff, pp. 101–37. MIT Press.

———. 2007. "How Much Do We Understand about the Modern Recession?" *BPEA,* no. 2: 13–28.

Hobijn, Bart, and Ayşegül Şahin. 2009. "Job-Finding and Separation Rates in the OECD." *Economics Letters* 104: 107–11.

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan. 1993. "Earnings Losses of Displaced Workers." *American Economic Review* 83, no. 4: 685–709.

Kaitz, Hyman B. 1970. "Analyzing the Length of Spells of Unemployment." *Monthly Labor Review* 93, no. 11: 11–20.

Katz, Lawrence F., and Bruce D. Meyer. 1990. "The Impact of the Potential Duration of Unemployment Benefits on the Duration of Unemployment." *Journal of Public Economics* 41: 45–72.

Kennan, John. 2006. "Job Loss, Job Finding, and Unemployment in the U.S. Economy over the Past Fifty Years: Comment." In *NBER Macroeconomics Annual 2005,* edited by Mark Gertler and Kenneth Rogoff, pp. 159–64. MIT Press.

Koenders, Kathryn, and Richard Rogerson. 2005. "Organizational Dynamics over the Business Cycle: A View on Jobless Recoveries." *Federal Reserve Bank of St. Louis Review* 87, no. 4: 555–79.

Krueger, Alan B., and Bruce D. Meyer. 2002. "Labor Supply Effects of Social Insurance." In *Handbook of Public Economics,* vol. 4, edited by Alan J. Auerbach and Martin Feldstein. Amsterdam: North Holland.

Kydland, Finn E. 1984. "Labor-Force Heterogeneity and the Business Cycle." *Carnegie-Rochester Conference Series on Public Policy* 21: 173–208.

Layard, Richard, Stephen Nickell, and Richard Jackman. 1991. *Unemployment: Macroeconomic Performance and the Labour Market.* Oxford University Press.

Ljungqvist, Lars, and Thomas J. Sargent. 1998. "The European Unemployment Dilemma." *Journal of Political Economy* 106, no. 3: 514–50.

Lockwood, Benjamin. 1991. "Information Externalities in the Labour Market and the Duration of Unemployment." *Review of Economic Studies* 58, no. 4: 733–53.

Machin, Stephen, and Alan Manning. "The Causes and Consequences of Long-term Unemployment in Europe." In *Handbook of Labor Economics,* vol. 3C, edited by Orley C. Ashenfelter and David Card, pp. 3085–3139. Amsterdam: Elsevier.

Marston, Stephen T. 1976. "Employment Instability and High Unemployment Rates." *BPEA,* no. 1: 169–203.

Mattila, J. Peter. 1974. "Job Quitting and Frictional Unemployment." *American Economic Review* 64, no. 1: 235–39.

Meyer, Bruce D. 1990. "Unemployment Insurance and Unemployment Spells." *Econometrica* 58, no. 4: 757–82.

Mincer, Jacob. 1991. "Education and Unemployment." Working Paper no. 3838. Cambridge, Mass.: National Bureau of Economic Research.

Moffitt, Robert. 1985. "Unemployment Insurance and the Distribution of Unemployment Spells." *Journal of Econometrics* 28, no. 1: 85–101.

Mulligan, Casey. 2009. "What Caused the Recession of 2008? Hints from Labor Productivity." Working Paper no. 14729. Cambridge, Mass.: National Bureau of Economic Research.

———. 2010. "Aggregate Implications of Labor Market Distortions: The Recession of 2008–9 and Beyond." Working Paper no. 15681. Cambridge, Mass.: National Bureau of Economic Research.

Nagypál, Éva. 2008. "Worker Reallocation over the Business Cycle: The Importance of Employer-to-Employer Transitions." Northwestern University.

Okun, Arthur M. 1962. "Potential GNP: Its Measurement and Significance." In *Proceedings of the Business and Economics Statistics Section of the American Statistical Association,* pp. 98–104. Alexandria, Va.

Oreopoulos, Philip, Till von Wachter, and Andrew Heisz. 2006. "The Short- and Long-Term Career Effects of Graduating in a Recession: Hysteresis and Heterogeneity in the Market for College Graduates." Working Paper no. 12159. Cambridge, Mass.: National Bureau of Economic Research.

Perry, George L. 1972. "Unemployment Flows in the U.S. Labor Market." *BPEA,* no, 2: 245–78.

Phelps, Edmund S. 2008. "U.S. Monetary Policy and the Prospective Structural Slump." Speech at the Seventh Annual BIS Conference on Monetary Policy, Lucerne, Switzerland, June 26–27.

Pissarides, Christopher A. 1992. "Loss of Skill during Unemployment and the Persistence of Employment Shocks." *Quarterly Journal of Economics* 107, no. 4: 1371–91.

Polivka, Anne E., and Stephen M. Miller. 1998. "The CPS after the Redesign: Refocusing the Economic Lens." In *Labor Statistics Measurement Issues,* edited by John Haltiwanger, Marilyn E. Manser, and Robert Topel, pp. 249–86. University of Chicago Press.

Poterba, James M., and Lawrence H. Summers. 1984. "Response Variation in the CPS: Caveats for the Unemployment Analyst." *Monthly Labor Review* (March): 37–43.

Ravn, Morten O., and Harald Uhlig. 2002. "On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations." *Review of Economics and Statistics* 84, no. 2: 371–76.

Rowthorn, Robert, and Andrew J. Glyn. 2006. "Convergence and Stability in U.S. Employment Rates." *Contributions to Macroeconomics* 6, no. 1: Article 4.

Ruhm, Christopher J. 1991. "Are Workers Permanently Scarred by Job Displacements?" *American Economic Review* 81, no. 1: 319–24.

Şahin, Ayşegül, Joseph Song, and Bart Hobijn. 2010. "The Unemployment Gender Gap during the 2007 Recession." *Current Issues in Economics and Finance* 16, no. 2. Federal Reserve Bank of New York.

Salant, Stephen W. 1977. "Search Theory and Duration Data: A Theory of Sorts." *Quarterly Journal of Economics* 91, no. 1: 39–57.

Shimer, Robert. 1998. "Why Is the U.S. Unemployment Rate So Much Lower?" In *NBER Macroeconomics Annual,* vol. 13, edited by Ben S. Bernanke and Julio J. Rotemberg, pp. 11–61. MIT Press.

———. 2001. "The Impact of Young Workers on the Aggregate Labor Market." *Quarterly Journal of Economics* 116, no. 3: 969–1007.

———. 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *American Economic Review* 95, no. 1: 25–49.

———. 2007 "Reassessing the Ins and Outs of Unemployment." Working Paper no. 13421. Cambridge, Mass.: National Bureau of Economic Research.

———. 2008. "The Probability of Finding a Job." *American Economic Review* 98, no. 2: 268–73.

Slichter, Sumner H. 1919. *The Turnover of Factory Labor.* New York: Appleton.

Solon, Gary, Robert Barsky, and Jonathan A. Parker. 1994. "Measuring the Cyclicality of Real Wages: How Important is Composition Bias?" *Quarterly Journal of Economics* 109, no. 1: 1–25.

Sullivan, Daniel, and Till von Wachter. 2009. "Job Displacement and Mortality: An Analysis Using Administrative Data." *Quarterly Journal of Economics* 124, no. 3: 1265–1306.

Valletta, Rob, and Aisling Cleary. 2008. "Sectoral Reallocation and Unemployment." FRBSF Economic Letter 2008-32. Federal Reserve Bank of San Francisco.

Van Rens, Thijs. 2004. "Organizational Capital and Employment Fluctuations." Economics Working Paper 944. Universitat Pompeu Fabra, Barcelona.

Willems, Tim, and Sweder van Wijnbergen. 2010. "Imperfect Information, Lagged Labor Adjustment and the Great Moderation." Discussion Paper 09-063/2. Tinbergen Institute.

Woytinsky, W. S. 1942. *Three Aspects of Labor Dynamics.* Washington: Social Science Research Council.

Yashiv, Eran. 2008. "U.S. Labor Market Dynamics Revisited." *Scandinavian Journal of Economics* 109, no. 4: 779–806.

Comments and Discussion

COMMENT BY

LAWRENCE F. KATZ Michael Elsby, Bart Hobijn, and Ayşegül Şahin have produced a superb descriptive empirical analysis of U.S. labor market stocks (unemployment, employment, and hours of work) and flows (into and out of unemployment) during the 2007 recession. This recession has generated particularly trying times for U.S. workers and their families and clearly merits being denoted as the Great Recession. The authors carefully document that the 2007 recession has been the most severe labor market downturn since World War II as measured by the increase in the unemployment rate, the peak age-adjusted level of the unemployment rate, the rise and level of unemployment durations, and the decline in employment and the employment-population ratio. And they convincingly show that labor market adjustment in this recession has been qualitatively similar to that in past deep downturns along three key dimensions: the demographic composition of the rise in unemployment, with larger increases for males, the less educated, and the young; the split of the contraction in labor input into declines in bodies employed (70 percent) and in hours per worker (30 percent); and the combination of an initial increase in unemployment driven by a sharp rise in unemployment *inflows* from layoffs followed by continuing increases in unemployment from a large decline in unemployment *outflow* rates.

The authors also find that the deterioration in labor market conditions from late 2007 to early 2009 followed the historical cyclical negative relationship between job openings (vacancies) and unemployment (that is, moving down the Beveridge curve with rising unemployment and falling vacancies). In other words, rising unemployment through the first quarter of 2009 looked like a very severe but normal cyclical phenomenon. But they show that the unemployment rate continued rising in 2009 after the job openings rate stabilized, so that the unemployment rate in late 2009 and early 2010 was much higher than would be implied by the historical Beveridge curve. The authors complement this finding with evidence of a downward shift in the job matching function and of deviations from Okun's law (higher unemployment than justified by the GDP gap) starting in the second quarter of 2009 and continuing into early 2010. These patterns suggest the emergence of structural unemployment problems: mismatches between unemployed workers and potential new jobs and/or the exacerbation of the longer-term structural problems associated with trends of rising wage inequality and declining employment opportunities in traditional middle-class jobs (Goldin and Katz 2007).

I find little to disagree with in the paper's excellent empirical analysis and discussion. I do have some concerns about drawing conclusions about the nature of labor market adjustments using data on labor market quantities alone and not using any information on labor market prices (wages). Robert Shimer flags this issue in his comment, and Jordi Galí (2010) examines the U.S. wage Phillips curve and aggregate wage behavior through 2009Q3.

In the remainder of my comment I will briefly discuss two issues. The first is the possible sources of the potential emerging structural unemployment problems suggested by the outward shift in the Beveridge curve since early 2009; the second is the likely longer-term human costs of the sharp rise of unemployment in the Great Recession.

Elsby, Hobijn, and Şahin argue that the substantial extensions of the potential duration of unemployment benefits (up to 99 weeks) in the current downturn could be contributing to the increased duration of unemployment and the outward shift of the Beveridge curve. They emphasize the traditional disincentive effects on job search effort from unemployment insurance (UI). Shimer in his comment posits that an increase in the duration and availability of UI benefits reduces the pressure on real wages from the unemployed and thereby slows labor market adjustment in a downturn. But the most compelling micro research using discrete policy changes or sharp regression discontinuity strategies suggests only modest impacts of UI extensions on search effort and on the duration of unemployment of UI recipients (Card and Levine 2000; Schmieder, von Wachter, and Bender 2009). Furthermore, previous estimates of larger impacts on unemployment duration for the United States (Katz and Meyer 1990) are based on data from the 1970s and early 1980s, in which much of the responsiveness comes from firms and industries using temporary layoffs and from the sensitivity of recall dates to UI benefits. This layoff-recall process is much less important today than it was in the downturns of that era.

UI extensions also have important consumption smoothing benefits for the unemployed (Gruber 1997), and much of the impact on job search effort comes from reducing liquidity (credit constraint) problems rather than traditional job search disincentives (Chetty 2008). Traditional microeconomic estimates of the impact of UI on the unemployment durations of UI recipients further tend to overstate the aggregate impact by ignoring the spillover effects of shorter unemployment spells for unemployed workers *not* receiving UI benefits (Levine 1993). They also ignore the macroeconomic stimulus arising from increased consumption expenditure by UI recipients, which raises both aggregate demand and demand for labor during a deep recession. UI extensions may also improve longer-run employment levels by keeping more of the long-term unemployed attached to the labor market rather than moving onto disability programs. Thus emergency UI extensions are likely to raise contemporaneous measured unemployment by more than they actually reduce employment, since those receiving benefits are more likely than other jobless workers to indicate in labor force surveys that they are searching for work, leading to a shift in the classification of workers from out of the labor force to unemployed.

Regional labor market problems and geographic disparities in the location of job seekers and potential job openings may be an underlying source of structural unemployment problems. Relative to workers in other nations, U.S. workers have always been highly mobile, and their moves in pursuit of new opportunities have enhanced U.S. economic dynamism. High rates of geographic labor mobility have allowed the United States to recover more rapidly from adverse economic shocks and to have smaller regional unemployment differences than European nations with less mobile workforces (Blanchard and Katz 1992).

But the geographic mobility of U.S. workers has declined over the last two decades and has fallen sharply in the Great Recession since 2007 (Frey 2009). Three factors may account for this change. First, the housing market crisis and large home price declines in many regions may have generated a geographic lock-in effect: if homeowners with negative equity are hesitant to sell their home at a loss, mobility from distressed areas will be reduced (Ferreira, Gyourko, and Tracy 2009). Second, the subprime crisis has created economic distress in precisely those fast-growing areas, such as California, Florida, and Nevada, that have absorbed workers from declining regions in the past, thus further slowing the movement of labor from declining to expanding regions that ordinarily helps drive U.S. job recoveries. Third, lingering credit market problems, especially for potential new start-ups, hinder job creation even in economically vibrant locales, reducing labor mobility to these areas.

The sharp cyclical downturn of the Great Recession comes on the heels of a three-decade increase in U.S. wage inequality and educational wage differentials. The former has been linked to rapid skill-biased technological change associated with computerization and to a slowdown in the growth of average educational attainment (Goldin and Katz 2008). The finance boom of the 1990s to 2007, some aspects of globalization and offshoring, and weakening U.S. labor market institutions have exacerbated these wage inequality trends. Technological changes and increased offshoring opportunities over the last 20 years have contributed to a polarization of the U.S. labor market, with strong growth in high-end, high-skill jobs and in traditionally lower-wage jobs in the in-person service sector, but particularly weak demand for traditional middle-class jobs such as manufacturing production jobs and middle management positions (Autor, Katz, and Kearney 2006, Autor 2010). The typical high-wage jobs of non-college-educated men, as well as many middle-class jobs for those with college training, have been hard hit. The housing market boom and bubble of 2002–06 obscured some of these trends by buoying demand for non-college-educated men in construction. The Great Recession has reinforced the longer-term jobs polarization and wage inequality trends, with huge declines in construction, manufacturing, and middle management employment.

These long-term structural labor market problems suggest that substantial mismatches between the skills and aspirations of job losers (especially the long-term unemployed) and the skill requirements and compensation packages of new job openings are likely to emerge as the economy recovers from the Great Recession. Many job losers from sectors such as construction and manufacturing may face difficulties in making the necessary psychological and financial adjustments, as well as in obtaining the training and education required for the new jobs available in the growing (primarily service) sectors.

Elsby, Hobijn, and Şahin conclude from similar large declines in unemployment outflow rates across aggregate industries since late 2007 (their figure 14) that increased sectoral shifts and mismatch are unlikely to be a driving force behind the apparent outward shift in the Beveridge curve. But skills mismatch is difficult to measure using such broad industry classifications. And Jinzu Chen, Prakash Loungani, and Bharat Trehan (2010) document a huge shock to the dispersion of stock market returns across industries at the start of the Great Recession and find that this stock market–based measure of sectoral shocks is a strong predictor of the path of long-duration unemployment rates.

Two particularly worrisome signs suggestive of longer-term structural labor market problems and persistent costs of unemployment from this recession are the concentration of the rise in unemployment among permanent job losers and the huge increase in long-term unemployment. Much research demonstrates that permanently displaced workers and the long-term unemployed face particularly difficult labor market adjustments (Jacobson, LaLonde, and Sullivan 2003, Couch and Placzek 2010).

Workers displaced from long-term jobs in the early-1980s recession faced large earnings declines upon reemployment and still had 20 percent earnings losses 15 to 20 years after displacement (von Wachter, Song, and Manchester 2009). The health consequences of permanent loss of a long-term job are also severe, with a 50 to 100 percent increase in mortality the year following displacement, 10 to 15 percent increases in mortality rates 20 years after displacement, and an implied loss of life expectancy for a worker aged 40 at displacement of 1 to 1.5 years (Sullivan and von Wachter 2009). The health problems and mortality increases from job loss are strongly positively associated with larger permanent earnings losses. A substantial number of permanent job losers also end up on the disability insurance rolls as they become discouraged in their search for new jobs, and many have multiple health problems (Autor and Duggan 2003). Parental job loss also appears to have adverse impacts on children, including poorer schooling outcomes and worse labor market outcomes as adults (Oreopoulos, Page, and Stevens 2008, Stevens and Schaller 2009).

Policies designed to help displaced workers make the transition to new jobs, gain valuable new skills, and reduce their earnings losses may be necessary to try to combat the potential for large and persistent adverse impacts on well-being arising from today's high level of long-term unemployment. Permanent job losers often are reluctant to accept new job offers below their pre-separation wage, and they often spend a long time searching for a job like their previous one, even when prospects are much brighter in other sectors and for other types of jobs. This leads to a form of long-term "retrospective wait unemployment," particularly for long-tenure workers displaced from declining sectors. A potential policy to address these issues and supplement unemployment benefits for likely permanent job losers is wage-loss insurance (also called wage insurance), which (at least temporarily) subsidizes earnings upon reemployment when the wage on the new job is less than that on the old job (Babcock and others 2009). Also, although the economic returns to further education and training at community colleges that lead to degrees and certificates are high for dislocated workers (Jacobson, LaLonde, and Sullivan 2005), the existing employment service programs and job training systems created under the Workforce Investment Act are fragmented and difficult for many workers to navigate. Improvements in reemployment services and access to training and education for permanently dislocated workers could reduce some of the long-term costs of the current downturn.

Sector-focused training programs (also known as sectoral employment programs) have emerged over the last 15 years as a particularly promising approach to workforce development. Sectoral employment programs work closely with local employers to create industry-specific programs that prepare and connect unemployed and underskilled workers to employers seeking to fill skilled vacancies, for example in allied health professions, information technology, and skilled manufacturing jobs. These sectoral employment programs, originally initiated by nonprofit, community-based organizations, have developed strong connections to employers and to the broader community. Early evaluations suggest that well-run versions of these programs can be quite successful in placing workers in high-quality jobs and in improving hourly and annual earnings (Maguire and others 2009).

Finally, the sharp decline in employment opportunities for teenagers and young adults in the Great Recession raises further longer-run worries. Young workers entering the labor market during a deep recession are likely to see reduced earnings for 10 to 15 years thereafter, relative to those graduating from high school or college in more normal times (Oreopoulos, von Wachter, and Heisz 2008, Kahn 2010). The returns to high school and postsecondary training are quite high in the current labor market, suggesting the need for policies to make it easier for young people to stay in school during a severe downturn.

There are some hopeful signs and some worrisome signs in the labor market and enrollment data for youth and young adults in the Great Recession. The employment-population ratio for 16- to 24-year-olds declined sharply (by about 8 percentage points) from the fall of 2007 to the fall of 2009. The decline in employment for this age group has resulted in bifurcated responses, with a rise in enrollment rates and a rise in the share both out of school and out of work (the idleness rate). The school enrollment rate among 20- to 24-year-olds increased by 3 percentage points overall and by 5 percentage points for blacks from the fall of 2007 to the fall of 2009. And the college enrollment rate of new high school graduates reached a record level of 70.1 percent in the fall of 2009 (Bureau of Labor Statistics 2010). The idleness rate among teens and young adults has increased most for males (especially black males). A major open question involves the current activities of and longer-run prospects for these idle males. One worry is that criminal involvement may rise in response to poor legitimate labor market opportunities, leading to longer-run scarring effects in the labor market for those who end up with serious criminal records. But at least through the first half of 2009, property and violent crime rates sharply declined during the Great Recession, suggesting that reduced demand for criminal output has outweighed potential increased criminal activity among more idle youth. Trends in the labor market, educational, and criminal activities of young people will be important to monitor going forward, to contribute to our understanding of the ultimate social consequences of the Great Recession.

REFERENCES FOR THE KATZ COMMENT

Autor, David. 2010. "The Polarization of Job Opportunities in the U.S. Labor Market: Implications for Employment and Earnings." Washington: Center for American Progress and the Hamilton Project (April).

Autor, David H., and Mark G. Duggan. 2003. "The Rise in the Disability Rolls and the Decline in Unemployment." *Quarterly Journal of Economics* 118, no. 1 (February): 157–205.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2006. "The Polarization of the U.S. Labor Market." *American Economic Review* 96, no. 2 (May): 189–94.

Babcock, Linda, William Congdon, Lawrence F. Katz, and Sendhil Mullainathan. 2009. "Notes on Behavioral Economics and Labor Market Policy." Harvard University (December).

Blanchard, Olivier Jean, and Lawrence F. Katz. 1992. "Regional Evolutions." *BPEA,* no. 1: 1–75.

Bureau of Labor Statistics. 2010. "College Enrollment Up among 2009 High School Grads." *TED: The Editor's Desk* (April 28). www.bls.gov/opub/ted/2010/ted_20100428.htm.

Card, David, and Phillip B. Levine. 2000. "Extended Benefits and the Duration of UI Spells: Evidence from the New Jersey Extended Benefit Program." *Journal of Public Economics* 78 (October): 107–38.

Chen, Jinzhu, Prakash Loungani, and Bharat Trehan. 2010. "Explaining the Duration of Unemployment: Sectoral vs. Aggregate Shocks." Washington: International Monetary Fund (March).

Chetty, Raj. 2008. "Moral Hazard versus Liquidity and Optimal Unemployment Insurance." *Journal of Political Economy* 116, no. 2: 173–234.

Couch, Kenneth A., and Dana W. Placzek. 2010. "Earnings Losses of Displaced Workers Revisited." *American Economic Review* 100, no. 1 (March): 572–89.

Ferreira, Fernando, Joseph Gyourko, and Joseph Tracy. 2010. "Housing Busts and Household Mobility." *Journal of Urban Economics* (forthcoming).

Frey, William H. 2009. "The Great American Migration Slowdown: Regional and Metropolitan Dimensions." Washington: Metropolitan Policy Program, Brookings Institution (December).

Galí, Jordi. 2010. "The Return of the Wage Phillips Curve." Working Paper no. 15758. Cambridge, Mass.: National Bureau of Economic Research (February).

Goldin, Claudia, and Lawrence F. Katz. 2007. "Long-Run Changes in the Wage Structure: Narrowing, Widening, Polarizing." *BPEA,* no. 2: 135–65.

———. 2008. *The Race between Education and Technology.* Harvard University Press.

Gruber, Jonathan. 1997. "The Consumption Smoothing Benefits of Unemployment Insurance." *American Economic Review* 87, no. 1 (March): 192–205.

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan. 1993. "Earnings Losses of Displaced Workers." *American Economic Review* 83, no. 4: 685–709.

———. 2005. "Estimating the Returns to Community College Schooling for Displaced Workers." *Journal of Econometrics* 125, nos. 1–2: 271–304.

Kahn, Lisa B. 2010. "The Long-Term Labour Market Consequences of Graduating from College in a Bad Economy." *Labour Economics* 17, no. 2: 303–16.

Katz, Lawrence F., and Bruce D. Meyer. 1990. "The Impact of the Potential Duration of Unemployment Benefits on the Duration of Unemployment." *Journal of Public Economics* 41: 45–72.

Levine, Phillip B. 1993. "Spillover Effects between the Insured and Uninsured Unemployed." *Industrial and Labor Relations Review* 47, no. 1 (October): 73–86.

Maguire, Sheila, Joshua Freely, Carol Clymer, and Maureen Conway. 2009. "Job Training That Works: Findings from the Sectoral Employment Impact Study." P/PV *In Brief* no. 7. Philadelphia, Pa.: Public/Private Ventures (May).

Oreopoulos, Philip, Marianne Page, and Ann Huff Stevens. 2008. "The Intergenerational Effects of Worker Displacement." *Journal of Labor Economics* 26, no. 3 (July): 455–83.

Oreopoulos, Philip, Till von Wachter, and Andrew Heisz. 2008. "The Short- and Long-Term Career Effects of Graduating in a Recession: Hysteresis and Heterogeneity in the Market for College Graduates." IZA Discussion Paper no. 3578. Bonn, Germany: Institute for the Study of Labor.

Schmieder, Johannes, F., Till von Wachter, and Stefan Bender. 2009. "The Effects of Unemployment Insurance on Labor Supply and Search Outcomes: Regression Discontinuity Estimates from Germany." Columbia University (November).

Stevens, Ann Huff, and Jessamyn Schaller. 2009. "Short-Run Effects of Parental Job Loss on Children's Academic Achievement." Working Paper no. 15480. Cambridge, Mass.: National Bureau of Economic Research (November).

Sullivan, Daniel, and Till von Wachter. 2009. "Job Displacement and Mortality: An Analysis Using Administrative Data." *Quarterly Journal of Economics* 124 (August): 1265–1306.

Von Wachter, Till, Jae Song, and Joyce Manchester. 2009. "Long-Term Earnings Losses Due to Mass Layoffs in the 1982 Recession: An Analysis Using U.S. Administrative Data from 1974 to 2004." Columbia University (April).

COMMENT BY

ROBERT SHIMER[1] Michael Elsby, Bart Hobijn, and Ayşegül Şahin offer a thorough and convincing description of some key labor market outcomes during the Great Recession. They focus first on the behavior of traditional stock measures, including employment, unemployment, and hours, and show that by any measure, this has been the deepest labor market contraction during the postwar period. They then turn to the flows in and out of employment, which they show are qualitatively similar to those in previous recessions. In particular, the onset of the recession was accompanied by a sharp spike in layoffs, which pushed many workers into unemployment. This process peaked by the end of 2008, so that by the end of 2009, almost all of the increase in unemployment was accounted for by an unprecedented decline in the unemployment outflow rate. Except for some small anomalies in the relationships among unemployment, vacancies, and the unemployment outflow rate in the latter half of 2009, they conclude that the labor market during the Great Recession behaved just as it did during every previous postwar recession, except for the size of the contraction.

I think this story is basically right as far as it goes, and so I do not want to spend too much time on the details of their analysis.[2] But it also sidesteps one important question: why did the shock that instigated the recession—the financial crisis, the construction contraction, the loss in housing and financial wealth, or whatever it might have been—result in such a big decline in employment and increase in unemployment duration? To answer that, it seems useful to go back to a basic model of labor supply and demand. As I will show, the model fails spectacularly during this recession, but its failure is instructive.

The model I will write down should not be controversial. Most modern theories of employment and hours worked, including the real business

1. I am grateful to Christopher Nekarda of the Federal Reserve for providing me with data in spreadsheet form for figure 3 on the number of workers collecting extended benefits and emergency unemployment compensation.

2. One minor comment: Elsby, Hobijn, and Şahin argue that "the measured stock of vacancies overstates the true number of job openings in the economy." This is based on the observation that even during normal times, many firms list vacancies but do not hire. I think this reflects a misunderstanding of what a vacancy is. A firm has a vacancy if it would like to hire but has not yet done so. In fields where good labor is scarce, vacancies may stay unfilled for months. But there is no reason to think that the desire to hire in these fields should have remained constant during the Great Recession. That is, there is no reason to think that there is a floor on the aggregate vacancy rate.

cycle model (Kydland and Prescott 1982) and the "three equation" New Keynesian model that forms the foundation of Michael Woodford's (2003) analysis, assume that hours worked are determined by the intersection of individuals' labor supply curves with firms' labor demand curves. Even job search models effectively assume that wages and hours are determined by labor supply and demand, but that fluctuations in demand are dampened because search frictions act like an adjustment cost (Shimer 2005, 2010; Rogerson and Shimer 2010).

I focus here on the simplest specification of preferences and technology. A representative individual has period-t utility defined over consumption c_t and hours h_t. Suppose in particular that utility is

$$\log c_t - \frac{\gamma\varepsilon}{1+\varepsilon} h_t^{\frac{1+\varepsilon}{\varepsilon}}$$

where $\gamma > 0$ measures the disutility of work and $\varepsilon > 0$ is the Frisch labor supply elasticity. The most important piece of this parametric assumption is that income and substitution effects cancel, so there is no long-run trend in hours worked.[3] The individual faces a period budget constraint,

$$b_t = a_t + (1 - \tau_t) w_t h_t - c_t.$$

She enters a period with some initial financial wealth a_t, earns a pretax wage w_t per hour of work h_t, pays a proportional labor tax τ_t, and consumes c_t, leaving her with financial wealth b_t, which is then invested in any available assets. I am deliberately vague about the set of available assets; in particular, markets may be complete or incomplete. Combining the first-order conditions for consumption and hours gives

$$\gamma c_t h_t^{\frac{1}{\varepsilon}} = w_t (1 - \tau_t), \tag{1}$$

which equates the marginal rate of substitution between consumption and hours to the after-tax wage. Note that the key assumption is that a worker is free to increase or decrease both her consumption and her labor supply at a fixed wage. I will return to this assumption later.

3. I have discussed elsewhere the importance of this specification of utility for the general results that I present here; see, for example, Shimer (2010) for more details.

Similarly, a representative firm has access to a Cobb-Douglas production technology that uses capital k and labor h to produce output. The firm chooses its inputs to maximize its per-period profits,

$$A_t k_t^{\alpha} h_t^{1-\alpha} - r_t k_t - w_t h_t,$$

where A_t is total factor productivity, α is the capital share of income, and r_t is the rental rate on capital. Letting $y_t = A_t k_t^{\alpha} h_t^{1-\alpha}$ denote total output, the first-order condition for the choice of labor is

$$(1 - \alpha)\frac{y_t}{h_t} = w_t, \tag{2}$$

which equates the marginal product of labor to the wage. This holds as long as the firm is free to vary its labor at a fixed wage. Introducing adjustment costs on capital, for example, also does not affect this conclusion.

Eliminate the wage between equations 1 and 2. Note that c_t and h_t in equation 1 denote consumption and hours per capita. Letting y_t and h_t in equation 2 similarly denote output and hours per capita, I obtain

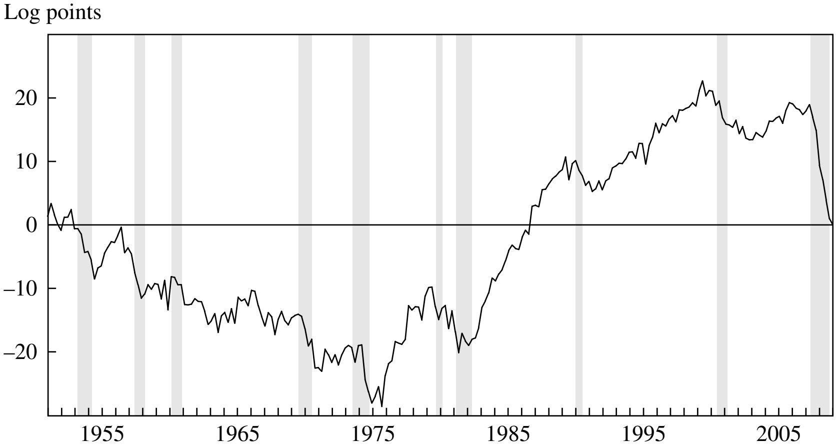
$$(1 - \tau_t)\frac{1 - \alpha}{\gamma} = \left(\frac{c_t}{y_t}\right) h_t^{\frac{1+\varepsilon}{\varepsilon}}. \tag{3}$$

The left-hand side of equation 3 is the proportion of labor income left after taxes, multiplied by the labor share $1 - \alpha$ and divided by the disutility of work γ. The right-hand side is the product of the consumption-output ratio c_t/y_t and hours worked h_t raised to an exponent $(1 + \varepsilon)/\varepsilon \geq 1$. The labor market clearing model predicts some co-movement between the consumption-output ratio and hours per capita in response to a shock to any variable not in this equation, such as a financial crisis, a collapse in construction, or a loss of housing and financial wealth.

To explore whether this relationship is a good description of the data, I use empirical measures of the consumption-output ratio and hours worked in the United States. I measure c as nominal expenditure on nondurable goods and services, and y as nominal GDP.[4] Following Simona

4. Chari, Kehoe, and McGrattan (2007) construct a measure of consumption that includes the flow of services from durables. This does not much affect the results.

Figure 1. Log Labor Wedge, 1951–2009[a]



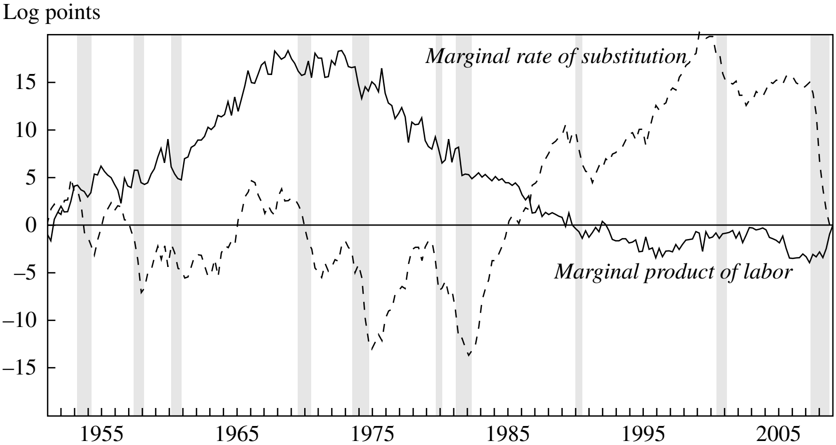
Source: Author's calculations.
a. The labor wedge is $(1 - \tau_t)$, implicitly defined by equation 3. Shaded bands indicate recessions.

Cociuba, Edward Prescott, and Alexander Ueberfeldt (2009), I measure hours per capita as the number of people at work times average hours per person at work divided by the population over age 16 (all from the Current Population Survey).[5] For simplicity's sake, I fix the labor supply elasticity at $\varepsilon = 1$ but stress that all comparisons between the latest recession and previous ones are unaffected by this choice of elasticity. Finally, I normalize $\gamma = (1 - \alpha)(1 - \tau_t)$ at $t = 2009Q4$, the end of the sample.

I plot in figure 1 the logarithm of the right-hand side of equation 3. A few patterns stand out. First, labor market clearing implies significant time variation in labor taxes at low frequencies. Arguably this is consistent with the data; for example, Robert Barro and Charles Redlick (2009) find that the average marginal tax rate increased from 1950 until 1981 and then fell modestly through 2006. More pertinent, the theory implies that taxes must have increased during almost every recession, and in particular shows that $1 - \tau_t$ fell by 19.0 log points during 2008 and 2009. Like many authors before me (see, for example, Parkin 1988; Rotemberg and Woodford 1991, 1999; Hall 1997; Mulligan 2002; and Chari, Kehoe, and McGrattan 2007), I view this prediction of the theory as implausible and instead call these

5. The labor market variables are available from the BLS since June 1976. Data for earlier years can be downloaded from Cociuba's website (sites.google.com/site/simonacociuba/research). I seasonally adjust these monthly data using the Census X11 algorithm and then take quarterly averages.

Figure 2. Marginal Rate of Substitution between Consumption and Hours Worked and Marginal Product of Labor, 1951–2009[a]



Source: Author's calculations.
a. Each series has been reduced by a constant 0.47 percent quarterly growth. Shaded bands indicate recessions.

cyclical movements in $1 - \tau_t$ the "labor wedge," that is, the part of labor market fluctuations that a labor market clearing model cannot explain.

It is worth emphasizing the magnitude of the current increase in the labor wedge. The peak-to-trough decline in $1 - \tau_t$ exceeded 10 log points in only three previous postwar recessions: it was 10.9 log points from 1952Q4 to 1954Q3, 11.2 from 1956Q3 to 1958Q1, and 10.3 from 1979Q4 to 1981Q4. The decline from 2008Q1 to 2009Q4 is nearly twice as large, 19.0 log points. If macroeconomists were already concerned before the Great Recession with the empirical validity of imposing labor market clearing, the assumption should be untenable today.

By breaking equation 3 back down into its components, the marginal rate of substitution (MRS) in equation 1 and the marginal product of labor (MPL) in equation 2, one can get a better understanding of why this theory failed. I measure the real MPL as the ratio of real GDP to total hours and then infer the real MRS as $1 - \tau_t$ times the MPL.[6] Since both the MRS and the MPL have trended upward over time with general growth in the economy, I remove a constant 0.0047 quarterly growth from both lines for visual convenience. The results, presented in figure 2, are stark. The MRS

6. The Bureau of Economic Analysis does not maintain a measure of real consumption of nondurables or of services before 1995, and so I cannot construct this series directly. Using real total consumption gives similar results.

accounts for virtually all of the cyclical movement in the labor wedge.[7] Although there are low-frequency movements in the MPL, business cycle fluctuations are very small. If anything, recessions appear to be periods where the MPL falls. The only exception to this pattern occurs in 2009, when the MPL increased by 4.8 log points (3.4 log points after detrending), an outcome that I will return to shortly.

These findings, or at least the pattern before 2009, are consistent with an environment where the real wage is fairly rigid and firms are always on their labor demand curve. That is, the path of the MPL simply reflects movements in the real wage. A recession, then, is a time when labor demand falls without an offsetting decline in the real wage, lowering the equilibrium level of hours worked. On the other hand, the real wage typically exceeds the MRS, so workers are not on their labor supply curve. The decline in firms' demand for labor during recessions makes this problem particularly acute. Recent work on rigid wages in search models, starting with Robert Hall (2005), offers a theoretical framework in which this possibility can be considered.[8] The patterns that Elsby, Hobijn, and Şahin highlight, including the spike in layoffs early in a recession and the persistent increase in unemployment duration later on, can be understood through the lens of these models.

In closing, I want to consider the unprecedented increase in the MPL, reflecting an unprecedented increase in the real wage, from the first to the fourth quarter of 2009.[9] One possible explanation is that the economy has been shedding its least productive, lowest-wage workers. Concern about this type of compositional effect is the justification that Elsby, Hobijn, and Şahin give for not discussing real wages in their paper (see their footnote 2). But there are some problems with this story. Why don't these compositional effects show up in earlier recessions, when the MPL typically fell or at least remained constant? Why don't they show up earlier in the Great Recession, when total hours were declining at a faster rate? Indeed, it is not even clear

7. See Galí, Gertler, and López-Salido (2007) for a similar conclusion with a different interpretation.

8. Rogerson and Shimer (2010) evaluate the role that search frictions play in macroeconomic models and conclude that the possibility that search may lead to rigid wages is one of its most important roles.

9. This may simply reflect measurement error in GDP, which is subject to numerous revisions. The gap between GDP and gross domestic income might give some support to that hypothesis (Nalewaik, this volume). But even if GDP in the fourth quarter of 2009 is subsequently revised down by a couple of percent, the MPL will still have increased sharply during the year.

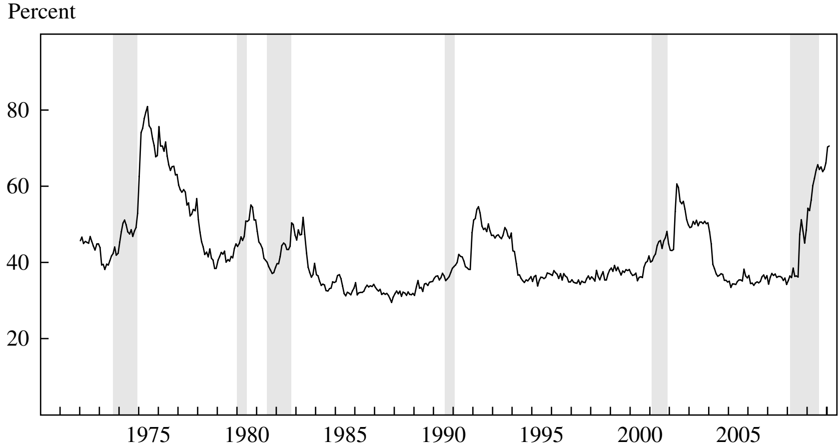
how compositional effects should work at business cycle frequencies. To the extent that wage rigidities force workers out of high-wage jobs into positions for which they are overqualified, the changing composition of jobs artificially inflates the procyclicality of wages and productivity.

This suggests a second possible explanation for the behavior of the MPL. In a low-wage environment with *nominal* wage rigidities, firms may be unable to reduce wages in response to adverse shocks. This leads to layoffs and raises the real wage and the productivity of surviving workers. But Elsby, Hobijn, and Şahin show that the pace of layoffs slowed during 2009. An explanation is still needed for why firms hired relatively few workers during 2009, and the link between that finding and nominal wage rigidities is more tenuous.

A third possibility is that firms are not hiring because of credit market frictions. But if the assumption is that credit market frictions reduced capital investment, the MPL should have fallen, not increased. So instead the model must be one in which credit market frictions reduce firms' ability to hire, either because of difficulties in financing payrolls or because hiring entails upfront recruiting and training costs, with deferred benefits. Since capital purchases may be more easily collateralized than payroll expansions, such a model may be empirically plausible. This hypothesis merits more serious exploration when appropriate data are available.

A fourth possibility is that various well-intentioned government interventions have kept unemployed workers from putting downward pressure on wages. One example is extensions in the potential duration of unemployment benefits, which Elsby, Hobijn, and Şahin discuss. Current law allows workers to collect unemployment benefits for up to 99 weeks in most states, a duration never before experienced in the United States. As I write this in April 2010, extending benefits by another 13 weeks is being debated. As a result of both the policy change and the depth of the recession, 11.5 million workers were collecting benefits in March 2010, 73 percent of all the unemployed. Only once before, in 1975, was the insured unemployment rate higher, as I show in figure 3. This unprecedented extension of benefits dramatically changes the composition of the unemployed population. For example, whereas the uninsured unemployment rate peaked at 5.8 percent in 1983, it never exceeded 3.8 percent during the current recession and was only 2.7 percent in March 2010. It seems unlikely that real wages will fall without more pressure from the unemployed. Viewed through the lens of the MRS and the MPL, the prognosis for a strong labor market recovery without a large preemptive change in labor market policy is poor.

Figure 3. Share of Unemployed Workers Receiving Unemployment Benefits, 1970–2010[a]



Source: U.S. Department of Labor, Unemployment Insurance Weekly Claims Report.
a. Shaded bands indicate recessions.

REFERENCES FOR THE SHIMER COMMENT

Barro, Robert J., and Charles J. Redlick. 2009. "Macroeconomic Effects from Government Purchases and Taxes." Working Paper no. 15369. Cambridge, Mass.: National Bureau of Economic Research.

Chari, V. V., Patrick J. Kehoe, and Ellen R. McGrattan. 2007. "Business Cycle Accounting." *Econometrica* 75, no. 3: 781–836.

Cociuba, Simona E., Edward C. Prescott, and Alexander Ueberfeldt. 2009. "U.S. Hours and Productivity Behavior Using CPS Hours Worked Data: 1947-III to 2009-II." Federal Reserve Bank of Dallas.

Galí, Jordi, Mark Gertler, and J. David López-Salido. 2007. "Markups, Gaps, and the Welfare Costs of Business Fluctuations." *Review of Economics and Statistics* 89, no.1: 44–59.

Hall, Robert E. 1997. "Macroeconomic Fluctuations and the Allocation of Time." *Journal of Labor Economics* part 2, 15, no. 1: S223–50.

———. 2005. "Employment Fluctuations with Equilibrium Wage Stickiness." *American Economic Review* 95, no. 1: 50–65.

Kydland, Finn E., and Edward C. Prescott. 1982. "Time to Build and Aggregate Fluctuations." *Econometrica* 50, no.6: 1345–70.

Mulligan, Casey B. 2002. "A Century of Labor-Leisure Distortions." Working Paper no. 8774. Cambridge, Mass.: National Bureau of Economic Research.

Parkin, Michael. 1988. "A Method for Determining whether Parameters in Aggregative Models Are Structural." *Carnegie-Rochester Conference Series on Public Policy* 29: 215–52.

Rogerson, Richard, and Robert Shimer. Forthcoming. "Search in Macroeconomic Models of the Labor Market." *Handbook of Labor Economics,* vol. 4, edited by Orley Ashenfelter and David Card. Amsterdam: North-Holland.

Rotemberg, Julio J., and Michael Woodford. 1991. "Markups and the Business Cycle." *NBER Macroeconomics Annual* 6: 63–129.

———. 1999. "The Cyclical Behavior of Prices and Costs." In *Handbook of Macroeconomics,* vol. 1B, edited by John B. Taylor and Michael Woodford. Amsterdam: Elsevier.

Shimer, Robert. 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *American Economic Review* 95, no. 1: 25–49.

———. 2010. *Labor Markets and Business Cycles.* Princeton University Press.

Woodford, Michael. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton University Press.

GENERAL DISCUSSION Robert Hall suggested dropping the framework for analyzing unemployment based on inflows and outflows in favor of a richer environment where separations and hires are as important as entry and exit from unemployment. The "ins-and-outs" model is limited because a first-order feature of the labor market escapes it completely, namely, cyclical variation in the fraction of separations that are followed by unemployment rather than by new jobs or exit from the labor force. In a strong labor market, the biggest outflow is into other jobs, with little or no intervening time out of the labor force or in unemployment. But in a recession, that fraction declines dramatically, creating a wide gap between separations and unemployment that Arthur Okun called the "Perry pothole," after George Perry described it in a 1972 Brookings Paper. Hall further noted that the present paper emphasizes a distinction between quits and layoffs, whereas models using bargaining theory do not make that distinction. Whether the worker takes the initiative to leave and the separation is called a quit, or the employer takes the initiative and it is called a layoff, is a detail of governance, not an allocational issue, in those models. It remains a possibility that layoffs differ from quits in a meaningful way, but labor market theory has yet to resolve that question.

Hall observed that vacancies in the most recent recession had been high relative to the high level of unemployment. The result is what looks like a shift of the Beveridge curve, but it could also be that the dynamics matter. Job creation as measured by vacancies has accelerated recently, but unemployment has not yet declined by the corresponding amount. When the job finding rate was high, unemployment dynamics were largely disregarded, but now that the job finding rate has dropped to half its normal level, the dynamics matter. Work by Hal Varian using Google Insight shows a dramatic decline recently in the number of people using the search term "unemployment insurance," which is a good sign. It was possible that vacancies would remain high and that unemployment would decline relatively rapidly, putting the economy back on the historical Beveridge curve.

Hall concluded that the present paper strongly supported the idea that there is no important labor market story for where the recent recession came from. There has been a very large, but directionally normal, response in the labor market, and there is little to suggest that a rapid expansion of demand would not quickly restore full employment.

Responding to Hall, Robert Gordon pointed out that the share of the total decline in hours that has taken the form of involuntary part-time work in this recession is unprecedented. Hence, any increase in the rate of output growth will likely be met disproportionately by firms moving part-time workers back to full-time work. Because the slack in the economy is thus concentrated in hours rather than bodies, it will prevent growth in real GDP from appreciably decreasing the unemployment rate.

Gordon was reminded of his own forthcoming paper on the demise of Okun's law, which documented a systematic structural shift. Comparing data from different periods roughly before and after the mid-1980s, that paper showed that the labor market has become much more responsive to changes in the output gap than in Okun's original formulation, in which two-thirds of any such change was matched by changes in aggregate hours and the remaining one-third by changes in productivity. In the last 25 years the productivity response has all but vanished, so that virtually the entire response now comes from hours. What might explain this shift is the sense that workers have become more disposable. A March 18, 2010, article in the *Economist,* titled "Slash and Earn," highlights how the cyclical behavior of the European and the U.S. economies differs on this score. Europe's typical response to a recession involves more traditional labor hoarding, and thus a collapse in productivity but a smaller decline in jobs for a given change in the output gap. This idea of the disposable American worker might be related to the increase in inequality and the decreased bargaining power of workers in the U.S. economy.

Olivier Blanchard suggested that although both quits and layoffs reflect the realization that it is no longer efficient for a worker and an employer to stay together, they differ in the source of the shock that leads to separation. Quits come from a shock to the worker, whereas layoffs occur when something happens to the profitability of the firm. Blanchard also proposed that differences in time spent searching for work might explain the outliers in the matching function that the authors had observed, or the fact that there is more unemployment than expected given the current level of vacancies. In depressed markets, some job losers are still counted as unemployed but in fact are no longer searching for work. If the necessary data are available, it might be worth correcting the matching function for this.

Steven Davis regarded the evidence presented by Lawrence Katz on the long-term consequences of job loss and displacement as strongly indicating that many workers are in for a difficult time for many years to come. This is an aspect common to severe cyclical downturns to which macroeconomists have paid too little attention. He noted that when one combines the JOLTS data on hires and vacancies with the CPS measure of unemployment, looking through the lens of the simplest Cobb-Douglas matching function with an exponent of around 0.4 on the unemployment rate, things line up beautifully from 2001 to 2007. Hires per vacancy and the transformed unemployment-to-vacancies ratio follow each other closely, consistent with a standard Beveridge curve relationship. The two measures diverged sharply, however, beginning in early 2008; since then there have been far too few hires per vacancy given the unemployment-to-vacancies ratio. Looking at the labor market in this way reveals a more pronounced and earlier departure from the normal pattern, again suggesting that something unusual has happened in the labor market during the Great Recession.

Justin Wolfers responded to two points in Katz's comment. The first was his observation that divorce and job loss are highly correlated. Although this is true in the cross section—people who lose their jobs are indeed more likely to get divorced—the behavior of the divorce rate is in fact completely acyclical. Since the start of the Great Recession, the divorce rate has continued to fall right up until the most recent data. Katz's second point was that over the past year, inflation does not seem to have been trending either upward or downward. That normally suggests that unemployment is near its natural rate, and if that is the case, there is much less reason for optimism about the near-term path of the unemployment rate. Finally, Wolfers noted that Jeremy Nalewaik's paper made the case that the measure of GDP based on income rather than expenditure gives a much more accurate reading in real time. Since December 2006 the expenditure-based measure has risen by 3¾ points more than the income-based measure. Thus, by the income-based measure, productivity might very well be falling, not growing, and unemployment is about where one would expect from Okun's law.

Robert Shimer believed that labor productivity growth would decrease in the near term, but he was not convinced that it would turn negative. He was reminded of a comment he had presented on another paper by Elsby, in which he (Shimer) discussed quits and layoffs as quite distinct things. Before 1994, the CPS used to ask unemployed workers every month why they were not working but did not require that the answers be consistent. In fact, among respondents who were unemployed in two consecutive months, about 30 percent of those who reported in the first month that they were job leavers switched the next month and said they had been fired. In contrast, the switch in the other direction was about 5 percent on average. Moreover, the switches were countercyclical. One would expect to see a boost during a recession, especially a deep recession, in the fraction of people who report themselves as having lost their job rather than quit. So, although the distinction between quits and layoffs is meaningful at some level, there is much spurious measurement of it, and a lot of murky ground in between the two concepts.

Steven Davis echoed Elsby and Blanchard on the importance of the quits-versus-layoffs distinction. He granted that the distinction is blurry, but hardly more so than that between being unemployed and being out of the labor force, yet that has not led economists to abandon the study of unemployment. Davis agreed with Hall that economic theory lacks a satisfactory micro foundation, other than a labeling story, for separations that are not jointly wealth maximizing. But by the same token, there was once a time when economics lacked a satisfactory theory of frictional unemployment, yet that did not prevent its recognition as an important phenomenon. In some of his own work with Jason Faberman and John Haltiwanger, Davis had found strong relationships among layoffs, quits, and hires. In the cross section, they are closely related to job creation and destruction. In good times many workers quit establishments that are shrinking moderately, perhaps because they anticipate bad times and layoffs coming. Much less of this kind of preemptive quitting occurs in weak labor markets. A related phenomenon is the well-known and quite pronounced cyclical variation in the ratio of quits to layoffs. This suggests not only that firms are suffering different kinds of shocks in booms than in busts, but also that workers perceive that their opportunities elsewhere are more limited during weak labor markets, leading them to wait until the ship sinks before abandoning it.

Richard Cooper argued that there is an intrinsic ambiguity, at least in some parts of the labor market, about the distinction between quits and layoffs. For example, some separations that are technically quits occur under circumstances in which the employer has made it clear that the worker's performance was not satisfactory and that the worker's long-term prospects at the firm were poor. Gary Burtless pointed out a simple distinction between quits and layoffs that has real economic significance in the United States: workers who quit are not entitled to unemployment benefits, whereas workers who are laid off are. This has clear implications for how individuals prefer to be labeled when they flow into unemployment.

Valerie Ramey noted a dramatic increase, particularly in the West, and particularly in the construction sector, in the fraction of the labor force who are recent immigrants. Much anecdotal evidence suggests that when the recession hit, many of these immigrants returned to their home countries. This may indicate a more elastic labor supply response among some of these marginal workers, which could have an impact on the overall statistics for unemployment and labor force participation.

JEREMY J. NALEWAIK
Board of Governors of the Federal Reserve System

The Income- and Expenditure-Side Estimates of U.S. Output Growth

ABSTRACT The two official measures of U.S. economic output, gross domestic product (GDP) and gross domestic income (GDI), have shown markedly different business cycle fluctuations over the past 25 years, with GDI showing a more pronounced cycle than GDP. This paper reports a broad range of results that indicate that GDI better reflects the business cycle fluctuations in true output growth. Results on revisions to the estimates, and correlations with numerous other cyclically sensitive variables, are particularly favorable to GDI. The most recent GDI data show the 2007–09 downturn to have been considerably worse than is reflected in GDP.

The Bureau of Economic Analysis (BEA) produces two conceptually identical official measures of U.S. economic output, currently called gross domestic product (GDP) and gross domestic income (GDI). These two measures have shown markedly different business cycle fluctuations over the past 25 years, with GDI showing a more pronounced cycle than GDP. These differences have become particularly glaring over the latest cyclical downturn, which appears considerably worse along several dimensions when measured by GDI. The aim of this paper is to determine which measure better represents the actual business cycle fluctuations in output growth. A wide variety of results suggest the answer is GDI.

Confusion about the information content of the two sets of estimates often starts with the nomenclature. "GDP" can mean either the true output variable of interest or an estimate of that output variable based on the expenditure approach—two very different things. Furthermore, since GDI has a different name than GDP, it may not be initially clear that GDI measures the same concept as GDP using the equally valid income approach. To keep things straight, this paper refers to the true variable of interest as true output, to the expenditure-side estimate of true output as GDP(E), and to the income-side estimate of true output as GDP(I).

The paper presents results for both the initial quarterly output growth estimates (those available three months after the end of the quarter) and the later estimates that have passed through more revisions. After presenting in section I some basic facts about the estimates, I discuss in section II the initial growth rate estimates and present numerous results favoring GDP(I) as the more accurate measure of output growth. First, there is some evidence that initial GDP(I) growth predicts revisions to GDP(E) growth, and no evidence of any tendency for GDP(E) growth to predict revisions to GDP(I) growth. Second, initial GDP(I) growth is the better predictor of a wide variety of business cycle indicators that should be correlated with true output growth. These include all measures of output growth in subsequent periods, the change in the unemployment rate in the current and subsequent periods, employment growth (measured using a household survey) in the current and subsequent periods, the Institute for Supply Management's Purchasing Managers' Index for manufacturing in the current and subsequent periods, changes in stock prices over previous periods, the slope of the Treasury yield curve in previous periods, and forecasts of GDP(E) growth itself from previous periods. Each of these results suggests that GDP(E) growth either is the noisier measure of true output growth or misses fluctuations in true output growth that appear in both GDP(I) growth and the other business cycle indicators. Third, initial GDP(I) growth has identified the onset of each of the last few cyclical downturns more quickly than initial GDP(E).

Section III discusses the latest revised growth rate estimates. I first establish some basic facts about the discrepancies between the fully revised estimates. On average, GDP(I) tends to grow faster than GDP(E) when the economy is expanding robustly, and to lag behind GDP(E) in recessions and in periods when the economy is sluggish. Because of this tendency, the statistical discrepancy between the two output measures is highly negatively correlated with the business cycle. Why is this the case? A thorough analysis of the nature of the source data suggests that GDP(E) misses part of the business cycle and that GDP(I) captures the business cycle better. Statistical analyses reach the same conclusion. First, the nature of the revisions suggests that they add cyclical variation to GDP(I) that is not added to GDP(E), implying that GDP(E) misses some cyclical variation. And second, the latest GDP(I) growth estimates are more highly correlated with a wide range of business cycle indicators, including changes in unemployment, the growth rate of employment, purchasing manager surveys (both manufacturing and nonmanufacturing), changes in stock prices over previous periods, the slope of the Treasury yield curve in previous periods, the spread between high-yield bonds and Treasury securities from previous periods, and indicator variables for officially identified recessions.

Section IV discusses the behavior of the estimates over the most recent cyclical downturn. When measured by GDP(I), output decelerated sooner, fell at a faster rate at the height of the downturn, and recovered less quickly. Drawing on the results from the previous sections and the online appendices,[1] this section discusses how GDP(E) may have missed the severity of the downturn. Section V concludes with some thoughts about the implications of the results for both data users and the BEA.

I. Basic Facts about the Estimates

The BEA's first GDP(E) estimate for the most recent quarter, called the "advance" estimate, is released about a month after the quarter closes. Estimates of most components of GDP(I) for that quarter are included in the advance release, but not all of them; corporate profits and net income from the rest of the world are not released at that time. Those components and GDP(I) are first reported in the second release, about two months after the quarter ends, except for estimates for the fourth quarters, when GDP(I) first appears with the third release, about three months after the quarter ends. To work with a complete time series of the initial growth rates, I focus on these third-release estimates. However, in an online appendix, I repeat the regression results in section II using the second-release estimates for quarters where an estimate of GDP(I) is available and, alternatively, using my own advance GDP(I) estimates constructed using the available income-side components and forecasts of corporate profits and net income from the rest of the world.

Estimates of GDP(E) and GDP(I) growth continue to be revised after the third release. Table 1 shows the variances of the initial (third-release) estimates and the latest estimates, which have passed through more revisions, as well as the correlations between the two estimates; here and throughout the paper ΔGDP(E) and ΔGDP(I) stand for the annualized quarterly growth rates of output implied by the estimates. I focus on two samples here. The first starts in 1978Q3 and is dictated by the start date of the time series of third-release growth rates employed in the paper, which is based on a real-time dataset constructed by the BEA starting in 1978. When analyzing the latest revised estimates, I focus on a shorter sample starting in the

1. Online appendices for all papers in this issue may be found on the *Brookings Papers* webpage (www.brookings.edu/economics/bpea), under "Conferences and Papers."

Table 1. Correlations between and Variances of Initial and Latest Available Estimates of Growth in GDP(E) and GDP(I)[a]

Measure	*Initial ΔGDP(E)*	*Initial ΔGDP(I)*	*Latest ΔGDP(E)*	*Latest ΔGDP(I)*
Correlations, 1978Q3–2009Q3				
Initial ΔGDP(E)	1.00			
Initial ΔGDP(I)	0.95	1.00		
Latest ΔGDP(E)	0.85	0.81	1.00	
Latest ΔGDP(I)	0.77	0.82	0.79	1.00
Correlations, 1984Q3–2006Q4				
Initial ΔGDP(E)	1.00			
Initial ΔGDP(I)	0.90	1.00		
Latest ΔGDP(E)	0.68	0.61	1.00	
Latest ΔGDP(I)	0.63	0.66	0.60	1.00
Variances	*1978Q3–2009Q3*		*1984Q3–2006Q4*	
Initial estimates	8.53	8.90	3.88	3.89
Latest estimates	9.44	10.29	4.23	4.96
Revisions (difference between latest and initial estimates)	2.78	3.60	2.57	3.05

Source: Author's calculations using BEA data.

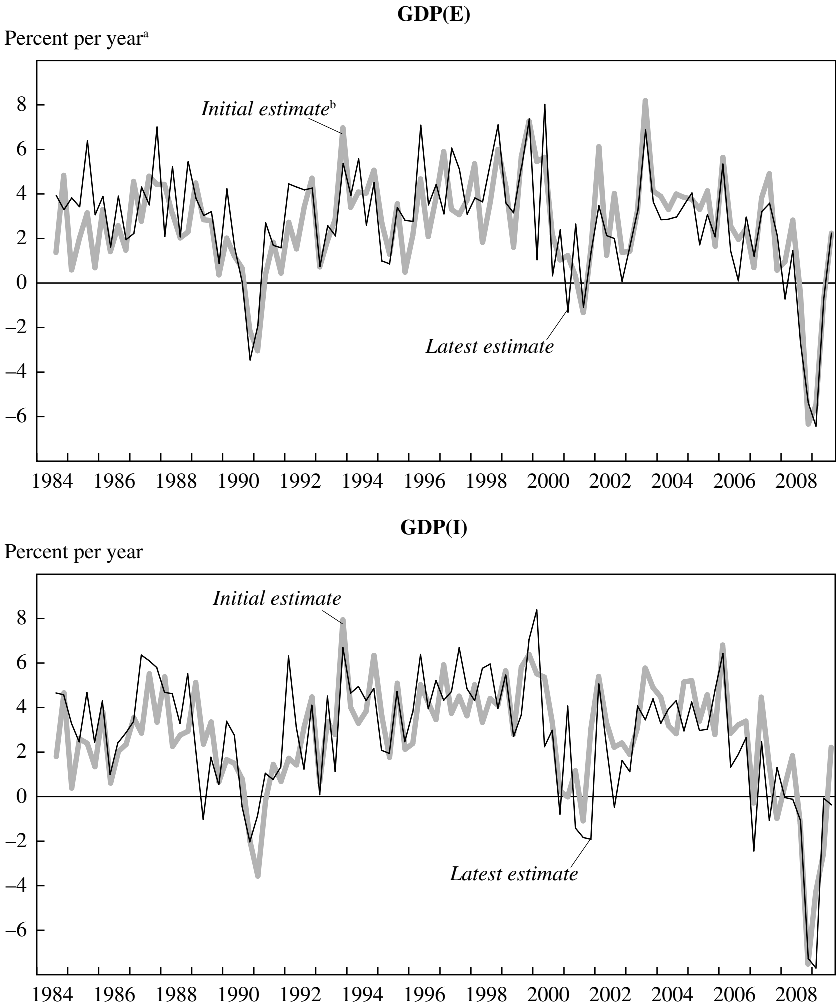
a. "Initial" estimates are those in the third BEA release for each quarter.

mid-1980s, because the divergences between the estimates are particularly stark and highly cyclical over this period.[2] This second sample ends in 2006Q4 to ensure that the latest estimates have been revised to fully incorporate all their major annual source data. Figure 1 plots both ΔGDP(E) and ΔGDP(I) from the mid-1980s to the present.

The top two panels of table 1 show that the correlations of the initial estimates of ΔGDP(E) and ΔGDP(I) with the latest estimates of the same measures are fairly high: 0.85 and 0.82, respectively, for the first sample, and 0.68 and 0.66 for the second. Nonetheless, the revisions do change the estimates in important ways. First, the bottom panel shows that the variance of the revisions is somewhat larger in both samples for ΔGDP(I) than for ΔGDP(E). Moreover, the revisions tend to increase the variance of ΔGDP(I) more than the variance of ΔGDP(E). This suggests that the revisions add information to the latest ΔGDP(I) that is not added to the latest ΔGDP(E).

2. The start date chosen here is the econometric breakpoint marking the beginning of the once widely accepted phenomenon known as the Great Moderation. The precise start date is not particularly important, however; any start date around the mid-1980s gives similar results for the latest estimates.

Figure 1. Initial and Latest Available Estimates of Growth in Real GDP(E) and GDP(I), 1984Q3–2009Q3



Source: Author's calculations using BEA data.
a. Quarterly data, annualized.
b. Estimate from the third release for the indicated quarter.

Finally, and perhaps counterintuitively, the revisions tend to make the two measures less similar, reducing their correlation from 0.90 to 0.60 in the shorter sample.

Given the important differences between the latest estimates of GDP(E) and GDP(I), this paper investigates two questions. First, what is the relative information content of the initial estimates of ΔGDP(E) and ΔGDP(I)? Put differently, how much weight should one place on each of these initial estimates? Second, what is the relative information content of the estimated GDP(E) and GDP(I) growth rates after they have passed through all their revisions? In other words, how much weight should one place on each of these latest, revised growth rates?

Online appendix A provides more background information about GDP(E) and GDP(I). Appendix B discusses the source data used to construct the initial growth rates, and appendix C describes the source data incorporated at the annual and the benchmark revisions.

II. The Information Content of the Initially Estimated Growth Rates

A detailed examination of the source data used to compute the initial (third-release) growth rates of GDP(E) and GDP(I) shows that both estimates suffer from similar types of measurement error problems (see online appendix B). These problems include missing data for a substantial portion of each estimate, sampling errors, and nonsampling errors such as incomplete coverage, survey nonresponse, and incomplete corrections for firm births and deaths. A compelling case for the superiority of either estimate cannot be made from such a detailed examination of source data, so this section proceeds immediately to the more informative statistical tests.[3]

3. In his comment on this paper, Steven Landefeld suggests that for the third-release estimates, a much greater fraction of GDP(I) than of GDP(E) is based on judgmental trends instead of early source data. Almost all of the source data used to compute the third-release estimates are flawed and unrepresentative in some way, and breaking down the data using such a binary classification scheme is a highly subjective exercise. The detailed discussion of the source data in online appendix B suggests that the evidence is less favorable to GDP(E) than this classification scheme suggests. Moreover, if a much greater fraction of GDP(I) were based on trends, one should expect third-release ΔGDP(I) to be much less variable than third-release ΔGDP(E), because trends should have less variance than the actual source data. The summary statistics in table 1 show that this is not the case: third-release ΔGDP(I) is actually slightly more variable than third-release ΔGDP(E).

Table 2 reports the main regression results examining the information content of the initial growth rate estimates. A good place to start is by examining the power of the initial estimates to predict the latest estimates, which incorporate superior source data. Over the full sample, the initial ΔGDP(E) estimates predict well the latest estimates of ΔGDP(E), with initial ΔGDP(I) adding little after conditioning on initial ΔGDP(E). Similarly, initial ΔGDP(I) well predicts latest ΔGDP(I), with initial ΔGDP(E) adding little information after conditioning on initial ΔGDP(I). However, the final two sets of regressions in table 2 show results for a sample starting in 1994Q1; I stop this subsample in 2006Q4 to ensure that the latest estimates have passed through all their annual revisions, but extending the subsample to 2008Q4 produces similar results. The first specification constrains the coefficients on initial ΔGDP(E) and initial ΔGDP(I) to sum to 1, whereas the second does not; the results show that over this sample period, when initial ΔGDP(I) is 1 percentage point above initial ΔGDP(E), initial ΔGDP(E) has subsequently been revised upward about a third (0.28) to two-fifths (0.42) of a percentage point, on average. Dennis Fixler and Bruce Grimm (2006), using a broader set of conditioning variables, also find some tendency for initial ΔGDP(E) to be revised toward initial ΔGDP(I). The last set of results shows that, over this subsample, there remains no significant tendency for initial ΔGDP(E) to predict latest ΔGDP(I).

Initial ΔGDP(I) may have predicted revisions to ΔGDP(E) in this sample because the initial ΔGDP(I) estimates are less noisy than the initial ΔGDP(E) estimates, or because they contain information about true output growth that is missed by the initial ΔGDP(E) estimates but is incorporated into latest ΔGDP(E) through revisions. Each of these explanations is likely part of the story. Averaging the data into year-over-year growth rates eliminates much of the noise in the quarterly data and shows the plausibility of the second explanation. The top panel of figure 2 plots the revisions in the year-over-year (fourth quarter to fourth quarter) growth rate of GDP(E) against the gap between GDP(I) and GDP(E) in the initial estimated year-over-year growth rates.[4] Broadly speaking, two periods drive the positive relationship that emerges.[5]

4. This presentation was suggested to me by William Wascher.

5. The line plots the predicted values from regressing the 13 Q4-over-Q4 growth rates of real GDP(E) on a constant and the gap between the initial estimates of Q4-over-Q4 GDP(I) and GDP(E) growth. The coefficient on the gap is 0.98, with a standard error of 0.44 and an adjusted R^2 of 0.25. I also experimented with corrections that removed the effects of major methodological changes from the revisions; this modification increased the R^2.

Table 2. Regressions Testing the Predictive Content of Initial Estimates of Growth in GDP(E) and GDP(I), 1978Q3–2009Q3 Sample[a]

	Initial ΔGDP(E)			*Initial ΔGDP(I)*				
Dependent variable	*t*	*t − 1*	*t −2*	*t*	*t − 1*	*t −2*	*Constant*	*Adjusted R^2*
ΔGDP(E)$_t$,	0.77			0.12			0.40	0.71
latest estimate	(0.13)			(0.12)			(0.24)	
	0.76	–0.17		0.07	0.27		0.28	0.72
	(0.12)	(0.15)		(0.11)	(0.15)		(0.21)	
		–0.46			0.93		1.42	0.25
		(0.28)			(0.29)		(0.32)	
ΔGDP(E)$_t$,		–0.37			0.80		1.38	0.22
initial estimate		(0.26)			(0.25)		(0.28)	
			–0.33			0.54	1.97	0.05
			(0.22)			(0.26)	(0.33)	
ΔGDP(I)$_t$,	0.02			0.86			0.34	0.66
latest estimate	(0.18)			(0.18)			(0.25)	
	0.03	–0.02		0.81	0.10		0.24	0.66
	(0.19)	(0.18)		(0.20)	(0.17)		(0.29)	
		–0.17			0.66		1.34	0.22
		(0.28)			(0.28)		(0.45)	
ΔGDP(I)$_t$,		–0.17			0.67		1.31	0.26
initial estimate		(0.21)			(0.21)		(0.32)	
			–0.43			0.69	1.91	0.09
			(0.21)			(0.24)	(0.40)	
$(UR_t - UR_{t-1}) \times 4$	0.03			–0.36			1.01	0.57
	(0.08)			(0.08)			(0.18)	
		0.02			–0.30		0.87	0.40
		(0.07)			(0.08)		(0.22)	
			0.05			–0.26	0.69	0.22
			(0.10)			(0.12)	(0.24)	

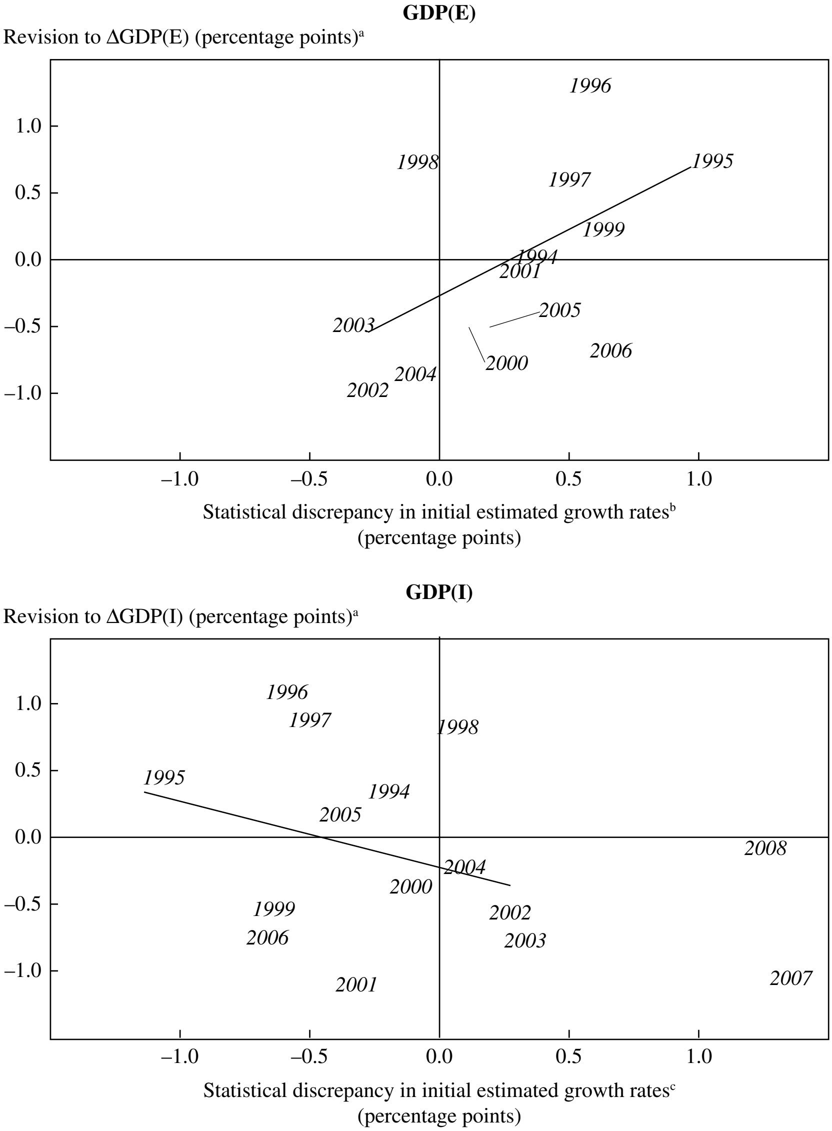
$(E_t^{household}/E_{t-1}^{household})^4$	0.08			0.35			0.06	0.47
	(0.12)			(0.11)			(0.27)	
		−0.03			0.41		0.17	0.38
		(0.10)			(0.13)		(0.33)	
			−0.19			0.43	0.55	0.16
			(0.13)			(0.14)	(0.39)	
$ISM_t^{manuf.}$	0.15			1.33			47.23	0.53
	(0.41)			(0.41)			(0.68)	
		0.41			0.93		47.59	0.43
		(0.38)			(0.37)		(0.66)	
			−0.03			0.83	49.00	0.15
			(0.46)			(0.44)	(0.78)	
Log($S\&P500_t$ −	−0.09			0.73			0.08	0.19
$S\&P500_{t-4}$)/4	(0.30)			(0.33)			(1.08)	
$r_{t-8}^{Treas(10)} - r_{t-8}^{Treas(2)}$	−0.14			0.21			0.51	0.08
	(0.07)			(0.08)			(0.15)	
SPF forecast,	0.06			0.43			0.74	0.51
$\Delta GDP(E)_{t,t}$	(0.11)			(0.12)			(0.28)	
SPF forecast,	0.03			0.22			1.68	0.19
$\Delta GDP(E)_{t,t-1}$	(0.07)			(0.09)			(0.30)	
SPF forecast,	0.08			0.08			2.20	0.11
$\Delta GDP(E)_{t,t-2}$	(0.07)			(0.08)			(0.33)	
$\Delta GDP(E)_t$, latest,	0.58[b]			0.42			−0.29	
1994Q1–2006Q4	(0.16)			(0.16)			(0.20)	
	0.56			0.28			0.31	0.47
	(0.16)			(0.16)			(0.35)	
$\Delta GDP(I)_t$, latest,	0.15[b]			0.85			−0.16	
1994Q1–2006Q4	(0.24)			(0.24)			(0.26)	
	0.13			0.74			0.31	0.41
	(0.23)			(0.22)			(0.59)	

Source: Author's regressions.

a. Each pair of rows reports results of a single regression. "Initial" estimates are those in the third BEA release for each quarter. Regressions are run on the full sample period unless stated otherwise. Numbers in parentheses are Newey-West standard errors using eight lags.

b. Regression coefficients are constrained to sum to 1.

Figure 2. Revisions to Output Growth Measures and the Statistical Discrepancy



Source: Author's calculations using BEA data.

a. Difference between initial and latest available, fourth quarter to fourth quarter.

b. Difference between initial ΔGDP(I) and initial ΔGDP(E), fourth quarter to fourth quarter.

c. Difference between initial ΔGDP(E) and initial ΔGDP(I), fourth quarter to fourth quarter.

First, during the mid- to late 1990s, the gap between initial ΔGDP(I) and initial ΔGDP(E) was consistently positive, as the initial estimates showed GDP(I) growing faster than GDP(E). This phenomenon was discussed in real time by the Council of Economic Advisers (*Economic Report of the President 1997,* pp. 72–74), the Federal Reserve Board (Greenspan 2004), and the BEA itself (Moulton 2000), with conclusions generally favorable to GDP(I). Those conclusions were later vindicated, since ultimately ΔGDP(E) was revised upward toward initial ΔGDP(I). Initial ΔGDP(I) accurately captured information about the brisk pace of economic growth that the initial ΔGDP(E) estimates had missed and that was incorporated only later through revisions (and then probably only partially; see section III). Second, the initial estimates of ΔGDP(I) in 2002 and 2003, after the 2001 recession, showed a more sluggish recovery than the initial estimates of ΔGDP(E), so that the gap between the initial estimates was negative.[6] Ultimately, ΔGDP(E) was revised toward initial ΔGDP(I) again: the recovery was indeed quite sluggish, and this information was reflected in ΔGDP(I) before it appeared in ΔGDP(E). The bottom panel of figure 2, which plots revisions in the fourth quarter–to–fourth quarter growth rate of GDP(I) against the gap between GDP(I) and GDP(E) in the initial estimated year-over-year growth rates, shows no tendency for ΔGDP(I) to be revised toward ΔGDP(E); if anything, ΔGDP(I) tends to be revised in the opposite direction from the initial gap over this period.

This particular set of revision results uses a short sample and should therefore be taken with a grain of salt. As a robustness check, I used past issues of the *Survey of Current Business* to extend the sample back in time to 1966Q4; the results, reported in online appendix D, show a marginally statistically significant tendency for initial ΔGDP(E) to be revised toward initial ΔGDP(I) over this long sample. After the data have passed through their first annual revision, a statistically significant tendency for initial ΔGDP(E) to be revised toward initial ΔGDP(I) in subsequent revisions appears yet again, in regressions using the 1978Q3–2009Q3 sample. So it would probably be unwise to ignore these revision results entirely. However, if for some reason an analyst did decide to ignore the predictability of the revisions, then the weight that analyst should place on the initial

6. As online appendix C outlines, since 2002 the BEA has incorporated information from the Quarterly Census of Employment and Wages (QCEW) into its wage and salary estimates a couple of months after its third release. These QCEW revisions provided much of the information contained in ΔGDP(I) on the relative sluggishness of the recovery; see the discussion in Nalewaik (2007a). The year-over-year growth rates for fourth quarters available in real time reflect these QCEW revisions.

estimates is entirely determined by the weight he or she places on the latest, fully revised estimates. An analyst who believes that latest ΔGDP(E) is more accurate than latest ΔGDP(I) should believe that initial ΔGDP(E) is more accurate than initial ΔGDP(I), and vice versa. So the results outlined in the next section, addressing the paper's second question, are also critical to answering the first.

However, one can make considerable further progress on the paper's first question directly, by examining the predictive power of the initial estimates for other important cyclical indicators. Broadly speaking, these regressions help establish which estimate is more informative about the business cycle, but they also help answer the narrower question of which is the better estimate of true output growth. The inferior estimate of true output growth, containing relatively more noise or classic measurement error, should have a lower signal-to-noise ratio and be the inferior predictor of cyclical indicators correlated with true output growth, all else equal. (This assumes that the noise in the output growth estimates is uncorrelated with the measurement error in the other cyclical indicators, and I have chosen the other cyclical indicators carefully to avoid this problem.) An estimate may be inferior not only because it is noisier, but also because it misses more fluctuations in true output growth (that is, contains less news or signal about true output growth) than the other estimate. But, again, such an inferior estimate should have a lower signal-to-noise ratio and be the inferior predictor of cyclical indicators that reflect those missing fluctuations in true output growth.

Returning to table 2, the top panel shows that as a cyclical indicator of where output growth is headed, initial ΔGDP(I) is superior to initial ΔGDP(E). The initial estimates of ΔGDP(I) are positively related to output growth in the next quarter, whether the latter is measured by ΔGDP(E) or ΔGDP(I), initial or latest. Conditional on initial ΔGDP(I), initial ΔGDP(E) contains no information about output growth next quarter and may actually be negatively related to it. This result holds two quarters ahead as well, using either initial estimate of output growth. Following the logic outlined above, these results may obtain because initial ΔGDP(E) is noisier than initial ΔGDP(I), so that its signal about true output growth in subsequent periods is obscured, or because initial ΔGDP(E), but not initial ΔGDP(I), misses some of the shocks that produce serially correlated fluctuations in true output growth.

I examine next the relationship of the initial estimates to other cyclical variables that should be correlated with true output growth. To avoid correlated measurement errors and spurious correlation, I examine only variables that are not used in the construction of either GDP(E) or GDP(I). As outlined in the appendices, the GDP(E) and GDP(I) estimates make little use of the Current Population Survey (CPS), a monthly household survey used to produce estimates of the unemployment rate.[7] As one of the most important indicators of the business cycle, the unemployment rate is a good variable to use as a starting point for this analysis.

Table 2 shows that initial ΔGDP(I) has a strong negative relationship with the contemporaneous change in the unemployment rate and negatively predicts changes to the unemployment rate one and two quarters ahead, whereas the coefficients on initial ΔGDP(E) are insignificant and have the wrong sign when conditioning on initial ΔGDP(I). Again, this may be because initial ΔGDP(E) is noisier than initial ΔGDP(I), or because initial ΔGDP(E) misses fluctuations in true output that both appear in ΔGDP(I) and are reflected in the differenced unemployment rate.

The next set of regressions in table 2 shows results using quarterly annualized employment growth computed from the household survey data, adjusted for breaks introduced by Census updates to the population. Initial ΔGDP(I) is positively related to employment growth this quarter, as well as one and two quarters ahead, but initial ΔGDP(E) contains little additional information about employment growth beyond that contained in initial ΔGDP(I).

Broadening the results beyond labor market variables, the next set of regressions uses the PMI (formerly called the Purchasing Managers' Index) from the Institute for Supply Management (ISM) manufacturing survey. The ISM measure is computed quite differently from GDP(E) and GDP(I). It is an aggregation of several diffusion indexes, so that even though the companies participating in the ISM survey also participate in the surveys used to estimate GDP(E) and GDP(I), the measurement errors likely behave quite differently. Initial ΔGDP(I) explains the contemporaneous, one-quarter-ahead, and two-quarters-ahead movements in the ISM measure better than initial ΔGDP(E), which in fact provides no statistically significant information conditional on initial ΔGDP(I).

7. At first blush, some analysts might suspect that GDP(I) must be more correlated with the unemployment rate than GDP(E), because "income" is in the name and the unemployment rate is a labor market concept. However, this reasoning is incorrect. Of the various components of the two output measures, one may expect based on a priori considerations that compensation will have a higher-than-average correlation with unemployment, but the other components of GDP(I) should then have a lower-than-average correlation, since all the components of GDP(I) add up to the same conceptual measure of output as GDP(E). For example, stories in the press recently have suggested that some of the recent rebound in corporate profits was facilitated by weakness in the labor market, allowing firms to cut compensation costs.

Business cycle analysts use a host of other variables to predict ΔGDP(E), most notably different asset prices, and since these asset prices are not used in constructing the output growth estimates, they are prime candidate variables for testing the information content of the initial estimates. However, asset prices typically predict output growth in subsequent quarters, rather than being predicted by output growth, so to get the timing correct, I regress lagged values of these asset prices on the two initial output growth measures. This is a somewhat odd specification but still quite instructive. The results essentially reveal which initial estimate is more consistent with market expectations of the business cycle from earlier periods.

The first asset-price specification regresses the logarithm of the change in the S&P 500 stock price index from the end of quarter $t-4$ to the end of quarter t on the two initial output growth measures in quarter t. Initial ΔGDP(I) is strongly positively related to this measure of stock price changes, whereas the coefficient on initial ΔGDP(E) is insignificant and negative. The next specification examines the slope of the yield curve, measured as the difference in yields between 10- and 2-year Treasury notes. This variable is most closely related to the output growth measures about 2 years hence; a regression of this measure from quarter $t-8$ on the two initial output growth measures in quarter t yields a coefficient on initial ΔGDP(I) that is significant and has the correct (positive) sign, and a coefficient on initial ΔGDP(E) that is significant but has the wrong sign.

The final set of testing variables I employ consists of median forecasts of output growth from the Survey of Professional Forecasters (SPF). These forecasters are trying to predict initial ΔGDP(E), presumably inclusive of any measurement errors. However, if the forecasters do not yet have access to the source data used to compute ΔGDP(E) for the quarter they are trying to predict, their forecasts will likely reflect general information about the state of the economy, which may be better related to initial ΔGDP(I) than to initial ΔGDP(E). This may be the case even for the current-quarter forecasts, because the survey occurs relatively early in the quarter, before the analysts have much GDP(E) source data. The results show that those current-quarter forecasts are well explained by initial ΔGDP(I), with initial ΔGDP(E) providing no incremental explanatory power. The SPF forecasts for quarter t, made in the first half of quarter $t-1$, are also better explained by initial ΔGDP(I) in period t than by initial ΔGDP(E) in period t. Forecasters' expectations for how the economy will move in the current quarter and the next appear to play out more fully in the initial ΔGDP(I) estimates than in initial ΔGDP(E).

Given the tighter relationship of initial ΔGDP(I) to all these business cycle indicators, placing full weight on the initial ΔGDP(E) estimates and no weight on the initial ΔGDP(I) estimates necessarily implies several things: first, that one cares about true output growth in the current quarter only (and not about the business cycle more broadly or even where true output growth is headed next quarter); second, that one believes that the latest ΔGDP(E) estimates reflect all available information about true output growth, so that neither latest ΔGDP(I) nor any other variable provides any additional marginal information about true output growth; third, that one believes that the superior explanatory power of initial ΔGDP(I) for various other cyclical indicators reveals nothing about the relative accuracy of initial ΔGDP(I) and initial ΔGDP(E) as estimates of true output growth; and fourth, that one discounts entirely the evidence contained in revisions. Regarding the first point, this may be a reasonable position for the BEA to take, but for analysts it is less clear: true output growth may be the only variable of interest for some purposes, but not for others. The dismissal of initial ΔGDP(I)'s ability to predict other indicators could be justified only if initial ΔGDP(I) contains variation uncorrelated to true output growth but correlated with all the other dependent variables employed in table 2, including actual forecasts of output growth. This is clearly an extreme position. A much more plausible explanation is that initial ΔGDP(I) is more highly correlated with true output growth than initial ΔGDP(E), and that true output growth is correlated with all these other cyclical indicators. Similarly, the view that latest ΔGDP(E) reflects all available information about true output growth is quite an extreme position in favor of the accuracy of latest ΔGDP(E), and the results in the next section suggest that latest ΔGDP(I) does contain a considerable amount of information about true output growth that latest ΔGDP(E) misses.

The regression results in table 2 are broadly consistent with those in Nalewaik (2007a), where I use Markov switching models to show that ΔGDP(I) identifies cyclical turning points more quickly than ΔGDP(E) in real time. Specifically, at the start of the 1980, 1981–82, 1990–91, and 2001 recessions (as defined by the Business Cycle Dating Committee of the National Bureau of Economic Research, NBER), real-time estimates of a Markov switching model using ΔGDP(E) alone put the probability that the economy was in a low-growth state at 52 percent, 40 percent, 45 percent, and 23 percent, respectively. Substituting ΔGDP(I) in the model produced much more accurate probabilities: 78 percent, 44 percent, 72 percent, and 70 percent. Most of the research in Nalewaik (2007a) was carried out in 2005, and the subsequent cyclical downturn was the first out-of-sample test of the main hypotheses of the paper. The model using ΔGDP(I) again performed much better around the start of the downturn in real time; it also performed better than some popular models using monthly indicators (see section IV).

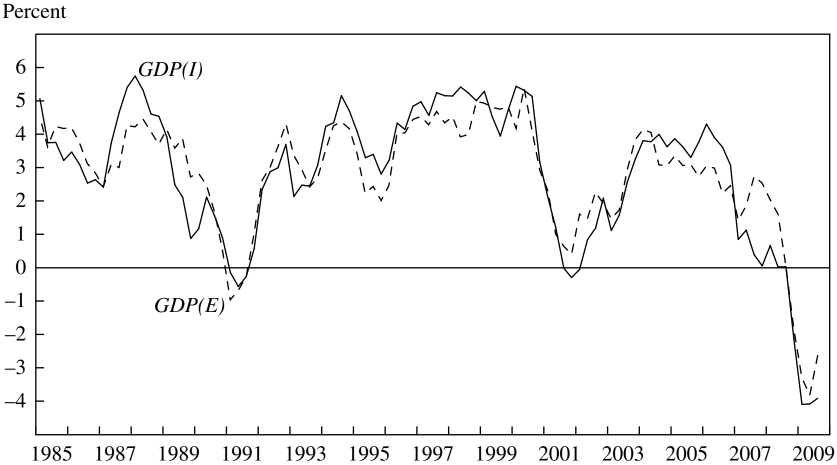
Although this section has focused on growth rates derived from the BEA's third releases, the information content of growth rates calculated from the two preceding releases is of critical importance for analysis in real time. Online appendix D reports results for these vintages, as well as results for the estimates once they have passed through their first annual revision. Briefly, when an official second-release estimate of ΔGDP(I) is available, the results using second-release growth rates are very similar to those reported in this section using the third-release growth rates. And as discussed above, the results using growth rates based on the first annual revision are even more favorable to ΔGDP(I) than the results using third-release growth rates, showing a statistically significant tendency for ΔGDP(E) to be revised toward ΔGDP(I) over the full sample.

For the advance estimates, when an official ΔGDP(I) estimate is not available, the situation is quite different. I use profits data from previous quarters and NIPA components that are reported in the advance release to generate the profits forecast used in constructing the "advance ΔGDP(I)" examined in the online appendix. However, some companies have already reported their quarterly profits for the latest complete quarter at the time of the advance release, and incorporating this information may produce a much-improved "advance ΔGDP(I)" estimate. That said, these rather limited advance-release ΔGDP(I) estimates perform poorly compared with the official advance-release ΔGDP(E) estimates, which better predict most of the business cycle variables used in this section. In addition, when predicting latest ΔGDP(I), about a two-thirds weight should be placed on the advance-release ΔGDP(E), and only about a one-third weight on the constructed advance-release ΔGDP(I) estimates. This suggests that the initial estimates of corporate profits produced by the BEA are highly informative and cannot be easily predicted based on lagged profits or other available NIPA variables. For fourth-quarter second-release estimates, when official profits numbers remain unavailable, this is presumably the case as well.

III. Information Content of the Latest Growth Rate Estimates

This section begins by showing that the latest estimates of GDP(I) and GDP(E) exhibit markedly different cyclical properties since the mid-1980s. I then discuss the evidence from revisions suggesting the superiority of the

Figure 3. Year-over-Year Growth Rates of GDP(E) and GDP(I), 1985Q1–2009Q3[a]



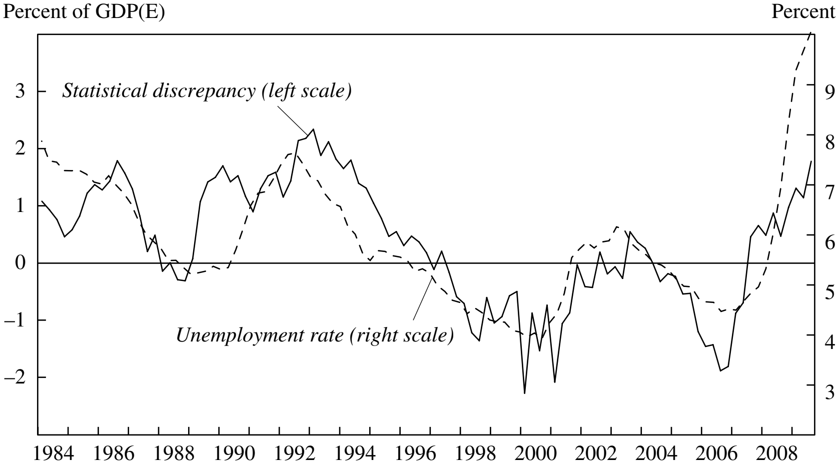
Source: Author's calculations using BEA data.
a. Calculations use latest available estimates of GDP(E) and GDP(I).

cyclical information in latest GDP(I), and I examine the relationship of the latest estimates to other cyclically sensitive variables.

III.A. The Cyclicality of the Latest Estimates

Table 1 showed that the correlation between ΔGDP(E) and ΔGDP(I) dropped sometime around the mid-1980s, and the divergences between the estimates became highly cyclical around that time. Figure 3 shows this using year-over-year growth rates: GDP(I) rose faster than GDP(E) through most of the 1990s boom and the comparatively short boom period from 2004 to 2006; in contrast, GDP(I) growth fell below GDP(E) growth in the 2001 recession and in the latest cyclical downturn. (Of course, these data are subject to further annual and benchmark revisions.) Figure 4 plots the statistical discrepancy between GDP(E) minus GDP(I), as a percent of GDP(E), and the unemployment rate; work by Charles Fleischman first examined this relationship, to my knowledge. Fleischman and John Roberts (2010) have studied the relationships among GDP(E), GDP(I), the unemployment rate, and other variables in the context of a state space model of the business cycle. Their work points to the unemployment rate as an excellent measure of the state of the business cycle; it also suggests that GDP(E) is measured with more error than GDP(I). Figure 4 shows that the measurement errors in either GDP(I) or GDP(E) are clearly systematically

Figure 4. Statistical Discrepancy between GDP(E) and GDP(I) and the Unemployment Rate, 1984Q1–2009Q3[a]



Source: Author's calculations using BEA and Bureau of Labor Statistics data.
a. Estimates use latest available data as of February 2010.

related to the business cycle, and the statistical discrepancy is not noise, as is commonly assumed.

To better understand the relationship depicted in figure 4, consider a very simple model. In Nalewaik (2008) I showed why the type of model outlined below is an incomplete characterization of the growth rates of GDP(E) or GDP(I), and I proposed models that fit the evidence better. However, the model is useful for the limited purpose of framing the subsequent discussion. Let true output be Y_t^*, and assume that it can be decomposed into a trend τ_t and a cycle ψ_t, so that $Y_t^* = \tau_t + \psi_t$. The unemployment rate U_t is governed by an Okun's law relationship:

$$U_t - U_t^n = \gamma\left(Y_t^* - \tau_t\right) = \gamma\psi_t, \gamma < 0.$$

Now assume that GDP(I) and GDP(E) are systematically either too cyclical or not cyclical enough, so that:

$$GDP\left(E\right)_t = \tau_t + \alpha_E\psi_t$$
$$GDP\left(I\right)_t = \tau_t + \alpha_I\psi_t.$$

Then the statistical discrepancy SD_t = GDP(E)$_t$ − GDP(I)$_t$ = $(\alpha_E - \alpha_I)\psi_t$, and assuming that the systematic mismeasurement is not identical for the two estimates, one should observe a relationship between the discrepancy and the unemployment rate:

$$U_t - U_t^n = \frac{\gamma}{\alpha_E - \alpha_I}(SD_t), \gamma < 0.$$

The strong positive relationship shown in figure 4 then implies that $\alpha_E < \alpha_I$, that is, that the magnitude of the cycle is smaller in GDP(E) than in GDP(I). Table 3 reports regressions investigating this relationship. The first three regressions show that the unemployment rate captures more than 60 percent of the variability of the statistical discrepancy from 1984Q3 through 2006Q4, and although the discrepancy is highly autocorrelated, the unemployment rate remains significant when an AR1 term is added. The last three columns of table 3 show specifications in first differences, to isolate the higher-frequency variation in the data. The first difference exhibits some negative autocorrelation, but the coefficient on the differenced unemployment rate remains positive, and the relationship is highly significant when the differenced unemployment rate is lagged one quarter. These regression results confirm that the statistical discrepancy is not noise, even in differences.

Having established that $\alpha_E < \alpha_I$ within the context of this very stylized model, three possibilities can be considered:

—Both GDP(I) and GDP(E) are more cyclical than true output, so that $\alpha_I > \alpha_E > 1$. In this case GDP(E) represents the cycle in true output better than GDP(I).

—Both GDP(I) and GDP(E) are less cyclical than true output, so that $\alpha_E < \alpha_I < 1$. In this case GDP(I) represents the cycle in true output better than GDP(E).

—GDP(E) is less cyclical and GDP(I) is more cyclical than true output, so that $\alpha_E < 1 < \alpha_I$. In this case GDP(I) represents the cycle in true output better than GDP(E) if $\alpha_I - 1 < 1 - \alpha_E$.

These possibilities frame the detailed discussion of the source data incorporated into the latest estimates in online appendix C. Plenty of evidence suggests that GDP(E) misses part of the business cycle, implying that the first possibility is unlikely. Some of the construction components of GDP(E) are smoothed; in particular, the additions and alterations ("adds and alts") component of residential structures is smoothed using a 3-year moving average. This is problematic, because smoothed estimates inherently understate the magnitude of business cycle accelerations and decelerations. Although "adds and alts" is a small component of GDP(E), it may have taken on outsized importance in the late-2000s downturn and

Table 3. Regressions Explaining the Statistical Discrepancy between GDP(E) and GDP(I)[a]

	Levels specifications[b]								
	1984Q3–2006Q4			*1948Q1–1984Q2*			*Differences specifications*[c]		
Independent variable	*3-1*	*3-2*	*3-3*	*3-4*	*3-5*	*3-6*	*3-7*	*3-8*	*3-9*
Statistical discrepancy (SD) lagged one quarter SD_{t-1}	0.93		0.75	0.74		0.72			
	(0.05)		(0.12)	(0.06)		(0.06)			
Change in SD lagged one quarter ΔSD_{t-1}							−0.33	−0.35	−0.36
							(0.17)	(0.18)	(0.16)
Unemployment rate U		0.88	0.25		0.10	0.03			
		(0.11)	(0.11)		(0.05)	(0.02)			
Current change in unemployment rate ΔU_t								0.28	
								(0.23)	
Change in unemployment rate lagged one quarter ΔU_{t-1}									0.84
									(0.25)
Constant α	−0.01	−4.69	−1.36	0.15	0.07	−0.02	−0.04	−0.03	−0.01
	(0.05)	(0.64)	(0.62)	(0.05)	(0.28)	(0.12)	(0.05)	(0.04)	(0.04)
Adjusted R^2	0.83	0.63	0.84	0.55	0.06	0.55	0.10	0.10	0.22

Source: Author's regressions.

a. Each column represents a different regression specification. In the levels specifications, the dependent variable is the statistical discrepancy, the difference between GDP(E) and GDP(I), in the current quarter; in the differences specifications, it is the change in the statistical discrepancy from the previous quarter. Newey-West standard errors with eight lags are in parentheses.

b. The regression equation is $SD_t = \alpha + \beta SD_{t-1} + \theta U_t + \varepsilon$.

c. The regression equation is $\Delta SD_t = \alpha + \beta \Delta SD_{t-1} + \theta_0 \Delta U_t + \theta_1 \Delta U_{t-1} + \varepsilon$.

may have contributed to some of the fluctuations in the discrepancy around the 1990–91 recession.

Probably more important, over most of this sample the type of annual surveys used to compute the goods-producing sector of GDP(E) simply did not exist for most of the (enormous) services-producing sector. As a consequence the BEA was forced to cobble together estimates based on trade-source, administrative, and regulatory data, which may have missed part of the business cycle. For example, these data sources may miss fluctuations in the output of sole proprietors and some small businesses, both of which are highly cyclical. And the activities of many types of financial services companies or entities may have been missed by the regulatory data used by the BEA to compute personal consumption expenditures (PCE) for financial services. The magnitude of the booms and busts in financial services, then, may not be fully reflected in the PCE component of GDP(E) or exports of services. However, many of these firms and entities likely did file tax forms, so their activities would have been represented in the tax data used to compute GDP(I). This could explain part of the increase in the statistical discrepancy in 1989, 2001, and the latest episode.

Online appendix C also discusses potential reasons why GDP(I) might be too cyclical. It is possible that some capital gains, which should be excluded from the BEA's definition of output, were misreported to the Internal Revenue Service as ordinary income and thus included in the tax data used to compute GDP(I). Capital gains are likely to be highly procyclical, so failure to exclude them could have made GDP(I) more cyclical than true output. Although the evidence on this is thin, the third possibility above might be slightly more likely than the second. However, the evidence in favor of GDP(E) understating the cycle is stronger than the evidence in favor of GDP(I) overstating the cycle, so if the third possibility holds, it is probably the case that $\alpha_I - 1 < 1 - \alpha_E$.

Columns 3-4 through 3-6 of table 3 show that the statistical discrepancy is much less cyclical before the mid-1980s. Why might that be the case? Although PCE for services has always been a relatively large share of GDP(E), averaging 30 percent from 1947 to 1984, its share shot up to an average of 43 percent from 1985 to 2009 and reached 48 percent in 2009. As the share of services PCE has increased, the measurement problems in GDP(E) may have become more severe and more plainly visible. In addition, booms and busts in financial services may have accounted for a much larger share of the *variability* of the business cycle since the mid-1980s, a period that includes the junk bond boom and bust (as well as the savings and loan boom and bust) from the mid- to late 1980s, the day-trading boom in the mid- to late 1990s and subsequent stock market crash from 2000 to 2002, and the mortgage securitization boom and bust from 2002 to 2008. GDP(E) may have missed much of this variation. But whatever the reason, since the really interesting divergences between the latest estimates occur in the post-1984Q3 period, the remainder of this section focuses on this sample.

III.B. Information in the Revisions about the Latest Estimates

Consider the following hypothetical example. Two time series estimate the same unobserved variable of interest. The two time series happen to be identical, but they are known to be subject to considerable measurement error and may deviate widely from the true variable of interest. Suppose new information becomes available that leads to large revisions to one of the estimates, bringing it closer to the truth, while the other estimate remains unrevised. Which estimate is now better? Obviously, the estimate that was revised is better: it is now clear that it was far off initially and that the revisions corrected some or all of that measurement error, and that the estimate that was not revised remains far off. More generally, if two estimates start out identical, or reasonably close, and the revisions improve both estimates, then the estimate that is revised more will, on average, tend to be better than the estimate that is revised less. This is the underlying logic of Fixler and Nalewaik (2007). Table 1 shows that the initial estimates of ΔGDP(I) and ΔGDP(E) do start out with a very high correlation, but ΔGDP(I) is revised more. Although the evidence in section II suggests that ΔGDP(I) starts out as the better estimate, if one makes the relatively uncontroversial assumption that the revisions improve the estimates, then the larger revisions imply that ΔGDP(I) has expanded its lead. Fixler and Nalewaik (2007) use this revisions evidence to establish bounds on the optimal weights to be placed on ΔGDP(I) and ΔGDP(E), and the bounds are favorable to ΔGDP(I).

The revisions increase the variance of ΔGDP(I) more than the variance of ΔGDP(E), implying that they add some news, or actual variation in true output growth, to ΔGDP(I) that is not added to ΔGDP(E) (see Mankiw, Runkle, and Shapiro 1984, Mankiw and Shapiro 1986, and Fixler and Nalewaik 2007). This variation in true output growth missed by latest ΔGDP(E) and captured by latest ΔGDP(I) is closely related to the business cycle. In particular, Nalewaik (2007b) shows that the revisions tend to reduce ΔGDP(I) by more than they reduce ΔGDP(E) in low-growth states, so the extent of the weakness of true output growth in low-growth states appears to be part of the information missing from ΔGDP(E) but appearing in ΔGDP(I) through its more informative revisions. Since this weakness in low-growth states appears in neither initial estimate and remains missing in latest ΔGDP(E), if latest ΔGDP(E) is correct, the revisions showing this relative weakness in latest ΔGDP(I) must damage the estimates. More broadly, any suggestion that latest ΔGDP(E) is better than latest ΔGDP(I) would seem to imply that the variability added to ΔGDP(I) through the revisions moves it further away from true output growth. This seems hard to believe and, if carried to its logical conclusion, implies that the BEA should stop revising ΔGDP(I) and allocate its resources elsewhere. I do not think anyone at the BEA would seriously advocate taking that step. In contrast, the standard interpretation of the revisions is less problematic for the BEA: the revisions improve both ΔGDP(E) and ΔGDP(I), but the source data incorporated into ΔGDP(E) are simply not as informative as the source data incorporated in ΔGDP(I). But in that case, latest ΔGDP(I) is likely the better estimate.

III.C. Relationship to Other Business Cycle Variables

The logic behind the tests reported in table 4 is similar to the logic behind the regression results in table 2, but the table switches the regression order and reports results from pairs of regressions, one regressing latest ΔGDP(I) and one regressing latest ΔGDP(E) on each cyclical indicator. The R^2s in the second and third columns show that latest ΔGDP(I) is more highly correlated with every one of these cyclical indicators. Online appendix D repeats this exercise using annual instead of quarterly data, and the results are quite similar.[8] ΔGDP(I) is more highly correlated with lagged stock price changes, the lagged slope of the yield curve, and the lagged spread between high-yield corporate bonds and Treasury bonds (using a somewhat shortened sample).[9] It is more highly correlated with short and long differences of

8. Much of the source data incorporated at annual revisions are at an annual frequency, with no information on quarterly patterns, so the quarterly numbers are likely less reliable than the annual. For example, BEA analysts are confident that employee gains from exercising nonqualified stock options net out of the annual GDP(I) estimates (since profits fall by the same amount as the increase in compensation), but they are concerned that the quarterly pattern within years may be distorted (see Moylan 2008).

9. In his comment, Steven Landefeld suggests that stock market fluctuations may be more highly correlated with ΔGDP(I), because capital gains may be "leaking" into ΔGDP(I). If this is the case, the correlation between changes in stock prices and ΔGDP(I) should be contemporaneous, especially at an annual frequency, since a rising stock market translates immediately into a capital gain. Online appendix D shows that the evidence does not support this: using the annual output growth measures, ΔGDP(E) is slightly more highly correlated with the contemporaneous change in the stock market, whereas ΔGDP(I) is more highly correlated with the stock market change from one year earlier. The evidence is more suggestive of either the stock market anticipating changes in true output, or changes in the stock market affecting true output with a lag, with true output better represented by ΔGDP(I). See also Nalewaik (2008).

Table 4. Regressions of Growth in GDP(I) and GDP(E) on Selected Business Cycle Indicators[a]

	Adjusted R^2		Regression coefficients (β)		
Independent variable	$GDP(I)_t$	$GDP(E)_t$	$GDP(I)_t$	$GDP(E)_t$	*Probability p that the βs are equal*
$\text{Log}(\text{S\&P500}_t/\text{S\&P500}_{t-7})/7$	0.14	0.08	0.29	0.20	0.03
			(0.06)	(0.08)	
$r_{t-1}^{\text{HYcorporate}} - r_{t-1}^{\text{Treasury (7yr)}}$ [b]	0.28	0.19	−0.67	−0.51	0.06
			(0.10)	(0.14)	
$r_{t-2}^{\text{HYcorporate}} - r_{t-2}^{\text{Treasury (7yr)}}$ [b]	0.20	0.11	−0.57	−0.41	0.02
			(0.13)	(0.13)	
$r_{t-3}^{\text{HYcorporate}} - r_{t-3}^{\text{Treasury (7yr)}}$ [b]	0.18	0.06	−0.54	−0.30	0.00
			(0.16)	(0.15)	
$r_{t-8}^{\text{Treasury (10yr)}} - r_{t-8}^{\text{Treasury (2yr)}}$	0.05	0.01	0.70	0.36	0.02
			(0.37)	(0.38)	
$(\text{UR}_t - \text{UR}_{t-1}) \times 4$	0.26	0.24	−1.47	−1.32	0.40
			(0.27)	(0.30)	
$\text{UR}_t - \text{UR}_{t-4}$	0.25	0.10	−1.74	−1.04	0.00
			(0.29)	(0.32)	
$\text{UR}_{t+2} - \text{UR}_{t-2}$	0.35	0.21	−2.19	−1.59	0.02
			(0.28)	(0.34)	
$\text{UR}_{t+4} - \text{UR}_t$	0.24	0.18	−1.81	−1.46	0.08
			(0.31)	(0.31)	
$(E_t^{\text{household}}/E_{t-1}^{\text{household}})^4$	0.30	0.20	1.00	0.76	0.11
			(0.17)	(0.21)	
$E_t^{\text{household}}/E_{t-4}^{\text{household}}$	0.20	0.12	1.01	0.74	0.00
			(0.19)	(0.21)	
$E_{t+2}^{\text{household}}/E_{t-2}^{\text{household}}$	0.34	0.24	1.36	1.06	0.04
			(0.19)	(0.22)	
$E_{t+4}^{\text{household}}/E_t^{\text{household}}$	0.23	0.18	1.14	0.93	0.13
			(0.23)	(0.20)	
$\text{ISM}_t^{\text{manuf.}}$	0.33	0.19	0.29	0.20	0.01
			(0.05)	(0.06)	
$\text{ISM}_t^{\text{nonmanuf.}}$ [c]	0.29	0.18	0.33	0.24	0.20
			(0.08)	(0.05)	
Recession dummies	0.29	0.24	−5.05	−4.28	0.46
			(0.43)	(0.77)	

Source: Author's regressions using BEA data.

a. The sample period is 1984Q3–2006Q4 except where noted otherwise. Newey-West standard errors with eight lags are in parentheses.

b. The sample period is 1988Q3–2006Q4.

c. The sample period is 1997Q3–2006Q4.

the unemployment rate, both contemporaneously and at leads and lags; the same holds true for the household survey measure of employment growth. Recall that there is no reason to suspect these measures to be spuriously correlated with ΔGDP(I). ΔGDP(I) is more highly correlated with the manufacturing ISM index and, when a shorter sample is used, with the nonmanufacturing ISM index as well. It is also more highly correlated with dummies for NBER recessions (see also Nalewaik 2007a).[10]

As in table 2, latest ΔGDP(I) may be more highly correlated with all these variables because latest ΔGDP(E) is contaminated with more noise. But the interpretation of the revisions offered in the previous subsection suggests a second type of measurement error, namely, that latest ΔGDP(E) misses variation in true output growth that appears in all these cyclical indicators and is picked up by latest ΔGDP(I). No matter which type of measurement error drives the results, latest ΔGDP(I) is the better estimate of true output growth. For the more econometrically oriented reader, however, the table provides formal tests derived in Nalewaik (2008) that differentiate between the two types of measurement error. The tests reject the hypothesis that the second type of error, called lack of signal error (LoSE) in Nalewaik (2008), does not contaminate latest ΔGDP(E). Regressions are run of each estimate on a testing variable, under the maintained assumption that the testing variable captures some of the variation missing from one estimate but included in the other. Nalewaik (2008) shows that the LoSE biases the regression coefficient on the testing variable toward zero, so the regression using the estimate that contains more LoSE yields a coefficient closer to zero. Note that it is measurement error of the LoSE form in the *dependent* variable that causes this attenuation bias, precisely the opposite of the conventional wisdom about classic measurement error, namely, that it is measurement error in the *explanatory* variable that causes attenuation bias. Testing the equality of the coefficients on the testing variable across the two regressions, Nalewaik

10. Note that these (often substantially higher) correlations are evidence against the crude model outlined in section III.A. In that model ΔGDP(E) and ΔGDP(I) contain rescaled versions of the same business cycle fluctuations, in which case the R^2 must be equal across the two regressions. That is clearly not the case: ΔGDP(I) contains different business cycle fluctuations, which also show up in these other business cycle variables. Nalewaik (2008) uses essentially this same argument to reject a crude rescaling model in favor of the LoSE model; see below. Nevertheless, both the LoSE model and the rescaling model say the same thing, broadly speaking: GDP(E) growth misses some of the business cycle fluctuations in true output growth, which show up in GDP(I) growth as well as in other variables.

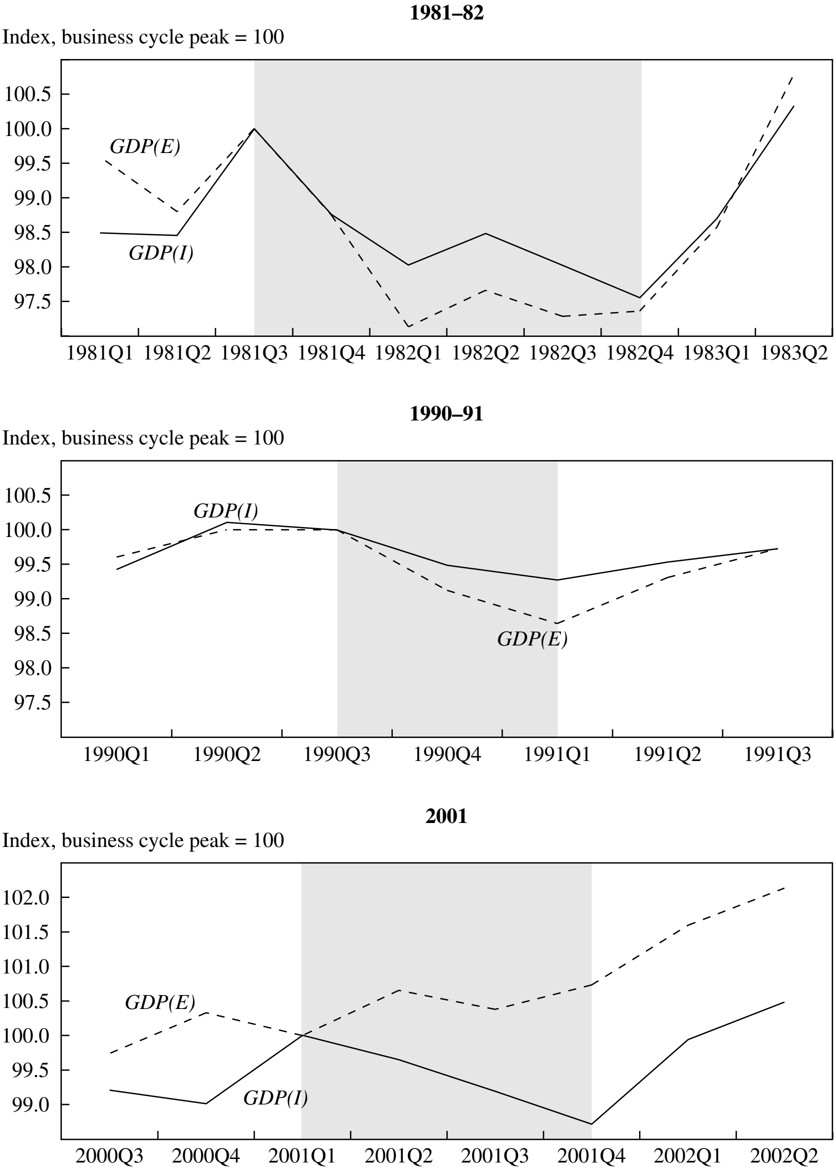
(2008) rejects using the asset price variables employed in the first five specifications of table 4, using a slightly different sample. Under the maintained assumption, ΔGDP(I) contains more signal about true output growth than ΔGDP(E)—signal that is reflected in stock and bond prices. Table 4 shows that this missing signal also appears in the differenced unemployment rate, household survey employment growth, and the ISM measures. The coefficients in table 4 are all larger, in absolute value, when ΔGDP(I) is the dependent variable; a relatively large amount of noise in ΔGDP(E) cannot explain these results, but a relatively large amount of LoSE can.

Less formal comparisons of GDP(E) and GDP(I) with other sources of information about the business cycle are also informative. In particular, one can compare the peaks and troughs in GDP(E) and GDP(I) with the NBER peak and trough dates. Grimm (2005) does this; figure 5 shows the results graphically for the three recessions before the most recent one. The one case where GDP(I) suggests a different dating than the NBER's is the 1990–91 recession: GDP(I) starts declining during the NBER peak quarter whereas GDP(E) is flat, but since the monthly peak was July 1990, the 1990Q3 GDP(I) decline seems consistent with the NBER dating. In the 1981–82 recession, the NBER called the trough in 1982Q4, the same quarter as the trough in GDP(I), whereas GDP(E) calls the trough three quarters earlier. In the 2001 recession it is difficult to discern any real cyclical downturn in GDP(E), whereas the NBER peak and trough dates line up perfectly with those of GDP(I). These peak and trough dates summarize the information in several other reliable indicators, and the fact that they line up better with GDP(I) is again suggestive that GDP(I) is the better estimate.

IV. The Estimates over the 2007–09 Cyclical Downturn

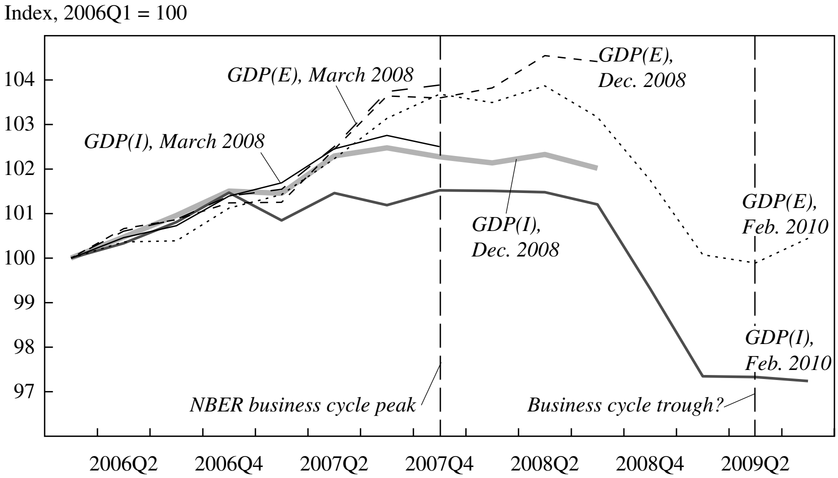
The recent downturn looks considerably worse when output is measured using GDP(I) instead of GDP(E). Several differences can be cited. First, the effect on output appears sooner in GDP(I), which shows a sharp deceleration even before the NBER peak in late 2007. This deceleration was somewhat evident in the real-time estimates of GDP(I), but more important, the recession itself was much more evident in the real-time estimates of GDP(I) than in the real-time estimates of GDP(E). Second, the steepness of the plunge in output in late 2008 and early 2009 appears worse in the GDP(I) estimates. And third, with the BEA's February 2010 data release, the decline in output now appears more prolonged, extending into the summer of 2009.

Figure 5. GDP(E) and GDP(I) in Three Recessions[a]



Source: Author's calculations using BEA data.
a. Shading indicates NBER recessions

Figure 6. GDP(E) and GDP(I) Estimates of Different Vintages, 2006Q1–2009Q3



Source: Author's calculations using BEA data.

Figure 6 shows GDP(E) and GDP(I) as they were measured at different dates, indexing them to 2006Q1 because the initial levels of the two series are different and have changed with revisions. The estimates released in March 2008, shortly after the start of the recession, both trend upward at a similar pace through 2006 and the first half of 2007 but then diverge considerably. GDP(I) shows the economy in a much more vulnerable state in late 2007, with output essentially flat over the second half of the year. GDP(E) shows little of this vulnerability in the second half; although growth was weak in 2007Q4, that weakness came on the heels of estimated annualized growth of almost 5 percent in 2007Q3.[11]

11. At the end of March 2008, the bivariate Markov switching model using ΔGDP(I) in Nalewaik (2007a) estimated a probability of around 90 percent that the economy had downshifted to a low-growth state by 2007Q4, and this probability remained well above 50 percent throughout 2008. At the same time, a Markov switching model using ΔGDP(E) alone estimated a probability of less than 20 percent that the economy had downshifted to a low-growth state, a probability that remained low through most of 2008 (for example, it was 27 percent at the end of September), reaching 50 percent only after the BEA's advance 2008Q3 estimates released at the end of October. Models based on monthly indicators did no better: an implementation of the Diebold-Rudebusch (1996) monthly indicators model, based on Kim and Nelson (2000), did not rise above 50 percent until early November 2008, with the BEA's release of its initial personal income numbers for September 2008. The behavior of these models shows that real-time assessments of the state of the business cycle can be meaningfully improved by looking at GDP(I).

The GDP(E) and GDP(I) estimates released at the end of December 2008 followed the NBER's identification of December 2007 as a business cycle peak. All four of the monthly indicators that the NBER uses to date business cycles had peaked in late 2007 and early 2008, and GDP(I) was trending down slightly through the first three quarters of 2008 as well. GDP(E) was the only anomaly, showing continued growth at an annual rate of almost 2 percent in the first half of 2008.

The latest BEA estimates, those of February 2010, also depicted in figure 6, represent a downward revision of the initial ΔGDP(E) estimates for 2008 toward the initial ΔGDP(I) estimates, a continuation of the recent pattern in revisions discussed in section II. Revisions have also reduced GDP(I), but mainly in the first half of 2007. The latest estimates show that GDP(I) was essentially flat over the four quarters of 2007, declining in 2007Q1 and 2007Q3. These latest estimates suggest that the recent cyclical downturn caused a measurable deceleration in aggregate output much earlier than is commonly believed. Meanwhile, the latest estimate of GDP(E) shows no such early deceleration; instead GDP(E) grew 2.5 percent over the four quarters of 2007, about the same as in 2006. These differences over 2007 produce the bulk of the enormous swing in the statistical discrepancy observed in figure 4, from around −1.9 percent of GDP(E) in late 2006 to +1.8 percent in 2009Q3.

The current estimates show ΔGDP(E) actually slightly weaker than ΔGDP(I) in the first three quarters of 2008, but the current estimates of ΔGDP(I) then show a steeper downturn over the worst part of the recession. The current annualized ΔGDP(I) estimates for 2008Q4 and 2009Q1 are −7.3 and −7.7 percent, respectively, worse than the ΔGDP(E) estimates of −5.4 and −6.4 percent.

Finally, the latest ΔGDP(I) estimates for 2009Q3, released in late February 2010 and incorporating numbers from the Quarterly Census of Employment and Wages (see online appendix C), have called into question the timing of the trough of the recession. Before these numbers were released, a conventional wisdom was emerging that the recession had likely ended late in the second quarter of 2009, perhaps in June, with the economy resuming growth in 2009Q3. Figure 6 shows a modest rebound in GDP(E) in 2009Q3, but no rebound in GDP(I). Personal income less transfer payments and employment—two of the four indicators most emphasized by the NBER Business Cycle Dating Committee—continued to decline in 2009Q3.

What is one to make of these important differences between GDP(E) and GDP(I) over this cycle? All these estimates remain subject to considerable future revision, but the source data are most concrete for 2007, which happens to be the period of greatest widening of the statistical discrepancy. Currently, two of the components of GDP(I), corporate profits and proprietors' income, incorporate IRS tax returns data through 2007, and declines in these two income categories account for the bulk of the deceleration in GDP(I) that year. Nonfarm proprietors' income (without inventory and capital consumption adjustments) increased by about \$68 billion (nominal) in 2006 and fell \$54 billion in 2007,[12] a deceleration of about \$122 billion. The biggest declines in 2007 were in real estate, construction, and finance and insurance, as well as (less explicably) mining (see table 6.12D of the National Income and Product Accounts). As noted in online appendix C, some of the decline in proprietors' income may have represented a decline in capital gains from house flipping, which should not be included in the relevant concept of output. Real estate proprietors' income fell \$24 billion in 2007, but it also fell \$14 billion in 2006, suggesting that this type of mismeasurement cannot explain much of the widening of the statistical discrepancy in 2007. Construction proprietors' income decelerated from a \$6 billion increase in 2006 to a \$14 billion decline in 2007, with the current estimates showing a \$46 billion decline in 2008. Part of this decline in proprietors' income should probably have shown up in lower spending on residential improvements, but as discussed earlier and in online appendix C, the BEA's averaging of its raw source data will tend to miss such a large deceleration. Currently, the raw estimates of improvements spending from the Census show declines of 4 and 14 percent in 2007 and 2008, respectively, steeper than the current BEA estimates of 1 and 4 percent. If the Census numbers are correct, GDP(E) should be \$5 billion lower in 2007 and about \$22 billion lower in 2008, so this also explains only a small portion of the widening of the statistical discrepancy.[13]

12. The 2007 decline in the raw IRS tax numbers was larger, about \$66 billion (see table 7.14 of the National Income and Product Accounts), but the BEA reduced this figure with various adjustments.

13. Some other data sources suggest much larger declines in spending on residential improvements. For example, Greenspan and Kennedy (2005, 2007) use Flow of Funds data, and Mian and Sufi (2009) use data from credit rating agencies, to show that households extracted a very large amount of home equity in the mid-2000s, before banks cut credit lines in 2007 and 2008 and equity extraction dropped dramatically. Using survey evidence that households spend about a third of extracted home equity on home improvements—see Brady, Canner, and Maki (2000), Canner, Dynan and Passmore (2002), and Greenspan and Kennedy (2007) and the references therein—updated Greenspan-Kennedy estimates give declines in spending on home improvements of \$66 billion in 2007 and \$80 billion in 2008. Of course, this does not necessarily imply causality from equity extraction to spending, because households may have found other financing options when home equity lines of credit dried up.

Corporate profits increased by about \$183 billion in 2006 and fell \$48 billion in 2007—a deceleration of about \$230 billion—and the current estimates for 2008 show a decline of \$181 billion. The biggest decline in profits in 2007 was in the finance and insurance industry, where a \$54 billion decline in 2007 followed an increase of \$4 billion in 2006 and a massive increase of about \$180 billion from 2000 to 2005. Looking more broadly, the sum of corporate profits, proprietors' income, and wages and salaries for the finance and insurance industry fell by close to 4 percent in 2007, while PCE for financial services increased by more than 12 percent. Although the categories are not strictly comparable, these numbers are difficult to reconcile unless there is severe measurement error in either the income measures or PCE.[14] In November 2007, in a technical note to its preliminary third-quarter GDP release, the BEA raised the issue of its ability to strip out capital losses (bad debt expenses and asset write-downs) from its initial estimates of financial companies' profits, but the availability of the tax data for 2007 likely made these subtractions much easier. The problems now appear more concentrated in the measurement of financial services PCE and services more generally on the expenditure side, as discussed in the previous section and in online appendix C. Given the advent of the financial crisis and the disappearance of many securitization markets in the second half of 2007, a 12 percent growth rate for financial services PCE seems implausibly high. To get a sense of the magnitudes involved, consider that a decline in financial services PCE of 4 percent would have lowered GDP(E) in 2007 by \$76 billion from its current level, and by more if PCE missed the boom in financial services output over prior years. More recently, profits in the finance and insurance industry fell an additional \$91 billion in 2008 (with proprietors' income and wage and salary income also falling), while financial services PCE increased once again. Since the tax data have not yet been incorporated for 2008, some risk remains that the income declines were too steep, but again it seems implausible that financial services PCE continued its uninterrupted growth.

Overall, this evidence suggests that although there may be problems on both sides of the accounts, they are likely more severe on the expenditure side. Given that, the latest downturn was likely substantially worse than the current GDP(E) estimates show. Output likely decelerated sooner, fell at a

14. The output of financial services could also have shown up in exports, or as an intermediate input into the production of other industries.

faster pace at the height of the downturn, and recovered less quickly than is reflected in GDP(E)—and in conventional wisdom.

V. Concluding Thoughts

Considerable evidence suggests that the growth rate of GDP(I) better captures the business cycle fluctuations in true output growth than does the growth rate of GDP(E). For the initial growth rate estimates, the revisions evidence over the past 15 years, the correlations with other business cycle indicators, and the recent behavior of the estimates around cyclical turning points all point to this conclusion. For the latest estimates that have passed through their cycle of revisions, careful consideration of the nature of the source data, statistical analysis of the information added by the revisions, and statistical tests, as well as informal comparisons with other business cycle indicators, again all suggest that GDP(I) growth is better than GDP(E) growth at tracking fluctuations in true output growth.

These results strongly suggest that economists and statisticians interested in business cycle fluctuations in U.S. output should pay attention to the income-side estimates and consider using some sort of weighted average of the income- and expenditure-side estimates in their analyses. The evidence in this paper clearly suggests that the weights should be skewed toward GDP(I), but even a 50–50 average would be a marked improvement over an estimate that places all its weight on GDP(E). It would also follow the lead of the Council of Economic Advisers, who, after concluding in the 1997 *Economic Report of the President* that GDP(I) might be better than GDP(E), have subsequently given some weight to the income-side estimates in their productivity analyses: see *Economic Report of the President 2008,* p. 39, and *Economic Report of the President 2009,* pp. 47–48.

The results here also have implications for the BEA. When a new quarterly estimate of GDP(I) growth becomes available, the evidence here shows that it is likely to be a better estimate of output growth than the corresponding GDP(E) estimate. However, the first GDP(E) estimate for any given quarter, the advance estimate, is typically released about a month before the first GDP(I) estimate, and GDP(I) is delayed an additional month when the BEA is producing estimates for fourth quarters. As noted above, these delays occur because the BEA has incomplete information on corporate profits and is not comfortable releasing earlier estimates of profits. In general, the profits information released by the BEA appears tremendously useful, and the BEA does have some information on profits at these earlier release dates. An advance estimate of GDP(I) based on the available profits information might be quite helpful for real-time assessment of the speed of economic growth. Earlier release of the fourth-quarter GDP(I) estimates, so that an estimate is available at least as early as the BEA's second release, might be similarly helpful; the BEA has still not released an estimate of GDP(I) growth for the fourth quarter of 2009 as of this writing in mid-March 2010. What the BEA decides will depend on how much information on profits is really available at these earlier dates, and a thorough assessment of this issue seems in order.

The BEA, the Census Bureau, and the Bureau of Labor Statistics doubtless will continue making improvements in their estimates where feasible, and the good news is that there have been substantial improvements over the past few years. The data on services have progressed by leaps and bounds, with the advent of the Census' Quarterly Services Survey in 2003, the recent expansions in the coverage of this survey, and the expansions in the coverage of the Census' Service Annual Survey. Further improvements are in train: in December 2010 the estimates from the Service Annual Survey will roughly double in coverage, expanding to mimic the sector coverage of the Economic Census. These data should improve the estimates of PCE and GDP(E).

Despite these improvements, however, problems with the output growth estimates will inevitably remain, and lack of coverage of services is only one of several important limitations of GDP(E). All the results in this paper suggest that the current reporting practice of the BEA, which puts nearly exclusive emphasis on GDP(E) over GDP(I), is suboptimal statistically. The BEA creates tremendous value by producing an income-based estimate of output growth, but current BEA reporting practice downplays that estimate so much that many analysts may not even be aware of its existence. The BEA's typical press release rarely discusses GDP(I), and it is reported only toward the back of the release, and then as a nominal level, requiring analysts to deflate and compute annualized quarterly growth rates themselves to arrive at a number comparable to headline real GDP(E) growth.

If the BEA finds the results here persuasive, there are several incremental steps it could take to increasing the prominence of GDP(I). Most obviously, the BEA could report real annualized growth rates of GDP(I) in its press releases, preferably in table 1 of the release so they can easily be compared with the annualized growth rates of real GDP(E). Second, it could give those annualized growth rates more prominence in the text of the press releases, discussing them at a level of detail similar to its current discussion of GDP(E). The BEA's discussion of the corporate profits estimates could be rolled into a more general discussion of GDP(I). Third, the BEA could bring more balance to its statements about the reliability of GDP(E) and GDP(I). Steven Landefeld, Eugene Seskin, and Barbara Fraumeni (2008, p. 211) take a small step in this direction by stating, "what these studies remind users is that it is useful to look at growth in both GDP and gross domestic income in assessing the current state of the economy."

Featuring two measures of output growth in the same press release would raise communication challenges, and the BEA might fear that such a practice might prove too confusing for casual analysts.[15] Here the example of other countries is relevant: the United Kingdom and Australia, for example, report an average of the two sides of the accounts as their featured output growth measure. The BEA has considered taking this step in the past (see, for example, Moulton 2000), and it could report such an average of GDP(E) and GDP(I) as "GDP(A)."[16] The BEA could employ optimal weights guided by statistical analysis, as in Fixler and Nalewaik (2007), but the results here suggest that featuring even a simple 50–50 average would be a marked improvement over the current practice of featuring GDP(E) alone.

15. The BLS does report two estimates of the monthly change in employment in its employment report, one from the establishment survey and one from the household survey, but there are clear statistical reasons for favoring the establishment survey number at the monthly frequency. For the case of GDP(E) and GDP(I), making the case in favor of one measure over the other is more complicated.

16. Of course, the components of GDP(E) will not sum to the top-line GDP(A), nor will the components of GDP(I), and this may be confusing for some analysts. But if the evidence in this paper is convincing, the components of GDP(E) already do not sum to true output or even to the best estimate of true output; in fact, the sum of the components of GDP(E) misses important, systematic variation in true output. Reporting an average would simply make these facts explicit. Over the long run, allocating parts of the discrepancy to different components of GDP(E) and GDP(I) may be the right thing to do, but this would be an extremely complicated task, and much research would need to be done beforehand. But if the BEA does attempt to go down this path at some point, it should do so in a transparent and easily replicable fashion. The BEA is to be commended for its transparency in reporting the statistical discrepancy and should do nothing to compromise this transparency.

References

Abel, Andrew B., and Ben S. Bernanke. 2001. *Macroeconomics,* 4th ed. Boston: Addison Wesley.

Brady, Peter J., Glenn B. Canner, and Dean M. Maki. 2000. "The Effects of Recent Mortgage Refinancing." *Federal Reserve Bulletin* (July), pp. 441–50.

Bureau of Economic Analysis. 2002. "Corporate Profits: Profits before Tax, Profits Tax Liability, and Dividends." Methodology Paper. Washington (September).

———. 2008a. "Updated Summary of NIPA Methodologies." *Survey of Current Business* (November), pp. 8–25.

———. 2008b. "Concepts and Methods of the U.S. National Income and Product Accounts (Introductory Chapters 1–4)." Washington (July).

Canner, Glenn, Karen Dynan, and Wayne Passmore. 2002. "Mortgage Refinancing in 2001 and Early 2002." *Federal Reserve Bulletin* (December): 469–81.

Diebold, Francis X., and Glenn D. Rudebusch. 1996. "Measuring Business Cycles: A Modern Perspective." *Review of Economics and Statistics* 78, no. 1: 67–77.

Fixler, Dennis, and Bruce Grimm. 2006. "GDP Estimates: Rationality Tests and Turning Point Performance." *Journal of Productivity Analysis* 25: 213–29.

Fixler, Dennis J., and Jeremy J. Nalewaik. 2007. "News, Noise, and Estimates of the 'True' Unobserved State of the Economy." FEDS Working Paper 2007-34. Washington: Board of Governors of the Federal Reserve System,

Fleischman, Charles, and John Roberts. 2010. "A Multivariate Estimate of Trends and Cycles." Washington: Board of Governors of the Federal Reserve.

Greenspan, Alan. 2004. "Risk and Uncertainty in Monetary Policy." *American Economic Review* 94, no. 2: 33–40.

Greenspan, Alan, and James Kennedy. 2005. "Estimates of Home Mortgage Originations, Repayments, and Debt on One-to-Four-Family Residences." FEDS Working Paper 2005-41. Washington: Board of Governors of the Federal Reserve System.

———. 2007. "Sources and Uses of Equity Extracted from Homes." FEDS Working Paper 2007-20. Washington: Board of Governors of the Federal Reserve System.

Grimm, Bruce T. 2005. "Alternative Measures of U.S. Economic Activity in Business Cycles and Business Cycle Dating." BEA Working Paper 2005-05. Washington: Bureau of Economic Analysis.

Landefeld, J. Steven, Eugene P. Seskin, and Barbara M. Fraumeni. 2008. "Taking the Pulse of the Economy: Measuring GDP." *Journal of Economic Perspectives* 22 (Spring): 193–216.

Mankiw, N. Gregory, and Matthew D. Shapiro. 1986. "News or Noise: An Analysis of GNP Revisions." *Survey of Current Business* (May): 20–25.

Mankiw, N. Gregory, David E. Runkle, and Matthew D. Shapiro. 1984. "Are Preliminary Announcements of the Money Stock Rational Forecasts?" *Journal of Monetary Economics* 14, no. 1: 15–27.

Mian, Atif R., and Amir Sufi. 2009. "House Prices, Home Equity-Based Borrowing, and the U.S. Household Leverage Crisis." Working Paper no. 15283. Cambridge, Mass.: National Bureau of Economic Research.

Moulton, Brent R. 2000. "Getting the 21st-Century GDP Right: What's Underway?" *American Economic Review* 90, no. 2: 253–58.

Moylan, Carol E. 2008. "Employee Stock Options and the National Economic Accounts." *Survey of Current Business* (February): 7–13.

Nalewaik, Jeremy J. 2007a. "Estimating Probabilities of Recession in Real Time Using GDP and GDI." FEDS Working Paper 2007-07. Washington: Board of Governors of the Federal Reserve System.

———. 2007b. "Incorporating Vintage Differences and Forecasts into Markov Switching Models." FEDS Working Paper 2007-23. Washington: Board of Governors of the Federal Reserve System.

———. 2008. "Lack of Signal Error (LoSE) and Implications for OLS Regression: Measurement Error for Macro Data." FEDS Working Paper 2008-15. Washington: Board of Governors of the Federal Reserve System.

Comments and Discussion

COMMENT BY

FRANCIS X. DIEBOLD[1] The topic of Jeremy Nalewaik's paper, the measurement of aggregate output, is of central importance. His case for preferring GDP(I) to GDP(E) is well argued and in certain key respects persuasive. Henceforth it will be impossible for macroeconomic analyses to proceed comfortably simply using GDP(E), as if the choice between GDP(E) and GDP(I) were inconsequential. Exclusive focus on GDP(E) will require justification and may have to be abandoned.

In my view, however, universal prescriptions (which Nalewaik does not offer, but others might) are unlikely to emerge. Rather, the comparative merits of GDP(E) and GDP(I) depend on the context. That is, use of one measure or the other will likely produce different answers for some questions and effectively indistinguishable answers for others. I will substantiate this claim in two contexts: aggregate output measurement and business cycle measurement. I will emphasize, moreover, that the important issue is not which of the two is "better," but rather how best to combine them, and what is ultimately added by GDP(I). I will argue that GDP(I) has much to add for aggregate output measurement, and little to add for business cycle measurement.

Consider first the choice of measure for aggregate output. This is the context in which Nalewaik primarily works, and in which, in my view, his assertions are most persuasive. He argues from a variety of perspectives that GDP(I) may be superior to GDP(E). That is initially surprising—

1. For helpful comments I thank the participants at the Brookings Panel conference, especially Robert Hall, Christopher Sims, and Justin Wolfers. For research support I thank the National Science Foundation and the Real-Time Data Research Center at the Federal Reserve Bank of Philadelphia.

indeed, shocking—given the near-universal neglect of GDP(I). But one must not overinterpret the result. Even if one grants that several arguments favor GDP(I) over GDP(E), one must also recognize that there is no need to choose one or the other. Instead, there may be gains from combining the two.

Consider forming a combined GDP measure, GDP(C), by taking a convex combination of GDP(E) and GDP(I):

$$\text{GDP}(\text{C}) = \lambda\text{GDP}(\text{E}) + (1 - \lambda)\text{GDP}(\text{I}).$$

This is just a "portfolio" of the two measures. Under conditions from the forecast combination literature (see, for example, Diebold 2007), the optimal portfolio weight λ^* is

$$\lambda^* = \frac{1 - \varphi\rho}{1 + \varphi^2 - 2\varphi\rho},$$

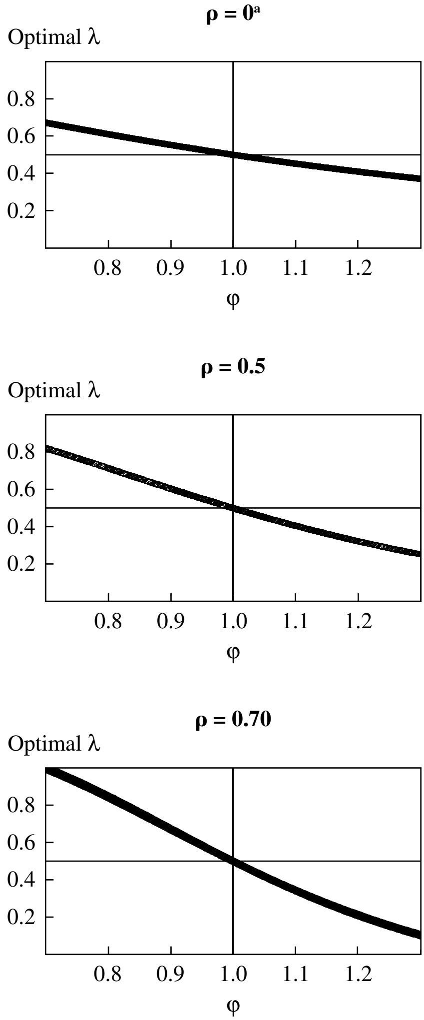
where $\varphi = \text{var}(e_{\text{GDP(E)}})/\text{var}(e_{\text{GDP(I)}})$, $\rho = \text{corr}(e_{\text{GDP(E)}}, e_{\text{GDP(I)}})$, $e_{\text{GDP(E)}} = \text{GDP} - \text{GDP(E)}$, and $e_{\text{GDP(I)}} = \text{GDP} - \text{GDP(I)}$. It is natural and desirable that λ^* depend on the variance ratio $\varphi = \text{var}(e_{\text{GDP(E)}})/\text{var}(e_{\text{GDP(I)}})$. In particular, as $\text{var}(e_{\text{GDP(E)}})$ increases relative to $\text{var}(e_{\text{GDP(I)}})$, the optimal weight on GDP(E) drops, other things equal. It is similarly natural that λ^* depend on ρ, which determines the benefits of portfolio diversification.

I illustrate the situation in figure 1, which plots λ^* as a function of φ, for various values of ρ. For $\varphi = 1$, the optimal weight on GDP(E) is always ½, and the optimal weight drops toward zero as φ increases. The speed with which it drops, moreover, increases as ρ increases.[2]

The key observation is that, except for extreme values of φ or ρ, or both, both GDP(E) and GDP(I) should receive significant weight in an informed assessment of aggregate output. Suppose, for example, that $\varphi = 1.1$, that is, that $\text{var}(e_{\text{GDP(E)}})$ is 10 percent greater than var $(e_{\text{GDP(E)}})$, and that $\rho = 0.5$, that is, that $e_{\text{GDP(E)}}$ and $e_{\text{GDP(I)}}$ are positively correlated, but not overwhelmingly so). Then the middle panel of the figure indicates an optimal GDP(E) weight of $\lambda^* = 0.4$. Weights near 0 or 1 would require extreme variance ratios, or extreme correlations, or both. Optimal weights may, however, be time

2. As ρ increases, the gains from diversification decrease, and so one diversifies less, other things equal.

Figure 1. Optimal Portfolio Weights of GDP(E) and GDP(I) Given the Error Variance Ratio for Various Correlations of GDP(E) and GDP(I)



Source: Author's calculations.

a. $\rho = \text{corr}(e_{GDP(E)}, e_{GDP(I)})$; $\lambda^* = \dfrac{1 - \varphi\rho}{1 + \varphi^2 - 2\varphi\rho}$, $\varphi = \text{var}(e_{GDP(E)})/\text{var}(e_{GDP(I)})$, where $e_{GDP(E)}$ = GDP – GDP(E) and $e_{GDP(I)}$ = GDP – GDP(I).

varying, reflecting changes in measurement error variances and covariances (over the business cycle, for example).

Now consider measuring the business cycle, another task of central importance, as also emphasized in Nalewaik's paper. A key insight, emphasized by Arthur Burns and Wesley Mitchell (1946) and Robert Lucas (1977), and clearly reflected, for example, in the National Bureau of Economic Research's business cycle dating methodology, is that the business cycle is not about any single variable (including GDP). That is, many indicators of business conditions (including GDP) are *related* to the business cycle, but no single indicator *is* the business cycle.

The so-called dynamic factor model embodies the Burns-Mitchell-Lucas insight and has become a standard tool for empirical characterization of the business cycle (see, for example, Sargent and Sims 1977, Stock and Watson 1989, Diebold and Rudebusch 1996, and Aruoba and Diebold 2010). In the dynamic factor framework, one treats the state of the business cycle as latent, with observed indicators of business conditions providing noisy signals, and uses the Kalman filter to produce optimal estimates of the business cycle from the noisy signals.

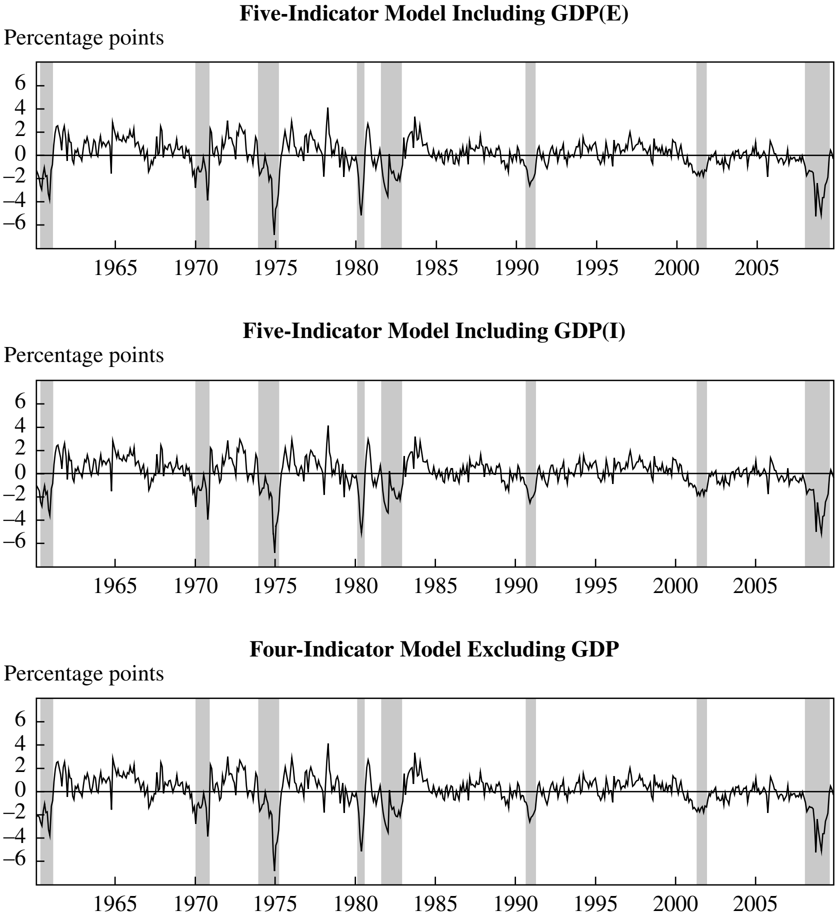
Does the choice of GDP(E) or GDP(I) matter for business cycle measurement, which, as I have emphasized, involves monitoring not only GDP but also a variety of other business conditions indicators? I will address this question using a five-variable dynamic factor model nearly identical to that of the Federal Reserve Bank of Philadelphia based on payroll employment, industrial production, personal income less transfers, manufacturing and trade sales, and GDP (see Aruoba and Diebold 2010 for details).[3]

Figure 2 shows the business cycle factor extracted using several versions of the five-indicator dynamic factor model. The top panel uses GDP(E), and the middle panel uses GDP(I). The difference is negligible. Evidently, given the information in the other four indicators, it makes no difference which estimate of GDP is included as a fifth. Indeed, the bottom panel, based on a four-variable model that simply excludes GDP, produces a nearly identical business cycle factor.

To conclude, Nalewaik's insightful and eye-opening paper deserves significant attention. As I have emphasized, however, the relevant question is not likely to be, "Which of GDP(E) and GDP(I) is better?" or "Which of GDP(E) and GDP(I) should economists use?" Rather, it is how best to

3. The Philadelphia Fed model is described at www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index/.

Figure 2. Extracted Business Cycle Real Activity Factor Using Alternative Output Measures, 1960–2010



Source: Author's calculations.

blend GDP(I) with GDP(E). GDP(I) has much to contribute in some contexts, and little in others.

REFERENCES FOR THE DIEBOLD COMMENT

Aruoba, S. Borağan, and Francis X. Diebold. 2010. "Real-Time Macroeconomic Monitoring: Real Activity, Inflation, and Interactions." *American Economic Review* 100, no. 2: 20–4.

Burns, Arthur F., and Wesley C. Mitchell. 1946. *Measuring Business Cycles.* New York: National Bureau of Economic Research.

Diebold, Francis X. 2007. *Elements of Forecasting,* 4th ed. Mason, O.: Thomson South-Western Publishing.

Diebold, Francis X., and Glenn D. Rudebusch. 1996. "Measuring Business Cycles: A Modern Perspective." *Review of Economics and Statistics* 78, no. 1: 67–77.

Lucas, Robert E. 1977. "Understanding Business Cycles." *Carnegie-Rochester Conference Series on Public Policy* 5: 7–29.

Sargent, Thomas J., and Christopher A. Sims. 1977. "Business Cycle Modeling without Pretending to Have too Much a Priori Economic Theory." In *New Methods in Business Cycle Research: Proceedings from a Conference.* Federal Reserve Bank of Minneapolis.

Stock, James H., and Mark W. Watson. 1989. "New Indexes of Coincident and Leading Economic Indicators." *NBER Macroeconomics Annual* 4: 351–94.

COMMENT BY

J. STEVEN LANDEFELD[1] Jeremy Nalewaik's paper is an excellent piece of research. BEA appreciates work such as this on measurement issues related to its economic accounts, as well as the opportunity to discuss it in more detail.[1] External research, complemented by research at BEA, has long been the source of a wide range of statistical improvements, ranging from chain indexes to hedonic indexes. Although I have a number of questions about the conclusions outlined in this paper, it will certainly serve as the basis for several future research endeavors: first, for further research on the sources of apparent cyclical patterns in the statistical discrepancy between GDP and GDI—what Nalewaik refers to as GDP(E) and GDP(I), respectively; second, for reconciliation with related BEA work on revisions; third, for further work on the balancing of income, production, and expenditure now done in the industry accounts, with particular attention to their use in balancing annual GDP and GDI estimates; fourth, for exploration of the means by which BEA can better present data on GDI and the range of revisions in GDP and GDI estimates without unduly confusing the general community of users; and, finally, and most important, for continued work on improving the early source data for both GDP and GDI to address the measurement issues raised in this paper.[2]

1. These comments reflect the very helpful ideas and calculations of my colleagues, Brent Moulton, Dennis Fixler, Bruce Grimm, and Shaunda Villones.

2. For more information see Fixler and Nalewaik (2009) and Fixler and Grimm (2002, 2006).

In general, the conclusion that the gross domestic income measure of aggregate output is deserving of attention is noncontroversial. The national accounts have double-entry accounts for purposes not only of providing multiple estimates of the breakdown of GDP by expenditure and income, but also of providing a check on the consistency of the two sets of estimates and identifying and correcting sources of discrepancies. As far as I can determine, BEA has never suggested that GDP is the "true" estimate of output, or that GDI is not a meaningful and useful measure of economic activity.

However, the main conclusion of this paper is that GDI is a better indicator than GDP of economic activity over the business cycle. My own conclusions are as follows: First, the evidence suggests that GDP and GDI provide roughly the same picture of economic activity over the business cycle and that a review of the source data and performance of the two measures favors GDP rather than GDI, but both have their strengths and weaknesses. Second, any gain in accuracy from averaging the GDI and GDP estimates is likely to be small. And third, some of the measurement concerns raised in this paper about the ability of GDP and GDI to fully capture changes in the economy over the business cycle are in the process of being resolved, thanks to new quarterly source data on services from the Census Bureau and more comprehensive monthly data on wages and salaries from the Bureau of Labor Statistics. Other concerns, especially those related to the cyclicality of corporate profits and other variables on the income side, are less tractable and will require further research.

QUALITY OF SOURCE DATA FOR GDP AND GDI. In contrast to the paper's assessment, I would describe the source data for the early GDI estimates as considerably less complete, consistent, and timely than the source data for the early GDP estimates. As a result, a significantly smaller share of the early estimates for GDP is based on trend extrapolators rather than directly on source data. Moreover, BEA views the GDP source data as generally superior to the GDI source data, because they are collected for statistical purposes and based on a consistent set of survey definitions designed to be used with the national accounts. They are collected by the Census Bureau as part of a consistent set of business surveys using the same universe and samples to collect monthly, quarterly, annual, and comprehensive (once every five years) data. In contrast, the source data for the GDI estimates are mainly taken from financial statements or collected by a variety of regulatory and tax agencies for nonstatistical purposes. These "administrative" data utilize a wide range of concepts and definitions, many of which differ significantly from those used in the national accounts. They also differ in scope and coverage. As a result, the income-side source data—especially for profits, proprietors' income, rental income, and interest income—differ significantly over time because of changes in business accounting and tax rules, changes in business practices, and changes in business conditions. The estimates from these sources also vary for the same time period, raising concerns about the consistency of estimates compiled from a combination of these sources. (For example, BEA's initial corporate profits estimates are based on companies' financial reports and financial accounting rules, whereas the latest estimates are based on companies' tax reports and IRS accounting rules.) Finally, significant tax incentives and corporate reporting requirements can bias information based on business, financial, and tax records. BEA takes great pains to adjust these administrative data to provide information consistent with the national accounts in terms of definition, scope, and timing, but such adjustments are challenging.

Whereas 86 percent of the early GDP estimates is based on some form of direct monthly or quarterly source data, only 37 percent of the early GDI estimates is based on such data; the rest is based on ratio adjustments, judgment, or trend estimators. The largest extrapolations are for the following: nonwage compensation, or supplements, which account for 18 percent of compensation and 10 percent of GDI; wages and salaries for nonproduction and supervisory workers, which include irregular payments and account for 45 percent of compensation and 21 percent of GDI; interest expense and rental income, which account for 8 percent of GDI; and proprietors' income, which includes large adjustments for misreporting and accounts for 8 percent of GDI (table 1).[3]

For the major components of GDI, revisions to later vintages of the estimates are sometimes significant. For example, the initial estimates of total wages and salaries have been subject to significant revision when the quarterly administrative (payroll tax) data become available in the Quarterly Census of Employment and Wages (QCEW). These revisions reflect the fact that although production and nonsupervisory workers account for roughly two-thirds of employment, they account for only a little more than

3. The 13 percent of the third estimates of GDP that is trend based is mainly in service components of personal consumption expenditures, including "other" services, "other" transportation, medical services, recreation, personal care, other personal business services, education and research, and religious and welfare services—as well as "other" state and local expenditures.

Table 1. Source Data for Gross Domestic Income, Estimates for 2007[a]
Percent of GDI

Component of GDI	*Judgmental trend*	*Early source data*	*Description of early source data or estimation method*
Compensation of employees			
Wages and salaries			
Nonsupervisory and production workers		17.1	BLS Current Employment Statistics: payroll survey
Supervisory/nonproduction workers	20.6		Judgmental extrapolation based on payroll survey
Government		7.7	Payroll survey employment and ECI
Supplements	10.3		Judgmental trend extrapolation
Taxes on production and imports, less subsidies			
Property taxes	2.8		Judgmental trend extrapolation
Other		4.1	Federal Monthly Treasury data; Census data for sales taxes
Net interest and miscellaneous payments	6.8		FDIC data for commercial banks; judgmental trend extrapolation based on interest rates for most of the remainder
Business current transfer payments	0.7		Judgmental trend extrapolation
Proprietors' income	7.8		Judgmental trend extrapolation based on BLS payroll survey, Census data, and other indicators
Rental income of persons	1.0		Mixture of actual source data and judgmental extrapolation
Corporate profits		8.5	Census Quarterly Financial Report, FDIC, Compustat data
Current surplus of government enterprises	0.0		Judgmental trend extrapolation
Consumption of fixed capital	12.5		Judgmental extrapolation based on BEA capital stocks
Total	62.6	37.4	
Percent based on early source data that are conceptually consistent with annual or benchmark data		11.8	

Source: Bureau of Economic Analysis.
a. BLS = Bureau of Labor Statistics; ECI = Employment Cost Index; FDIC = Federal Deposit Insurance Corporation.

half of wages and salaries, and the fact that the payroll survey does not capture stock options, bonuses, and other irregular payments.[4]

Further, although the QCEW data, which are available 4 months after the advance GDP report, cover virtually all workers, they are quite volatile and have proved to be extremely difficult to measure on a seasonally adjusted basis. Moreover, once annual QCEW data are received, there can be significant revisions in the quarterly data.

Corporate profits are even more difficult to measure, and early estimates based on corporate financial statements can differ significantly from both the economic accounting measure from BEA and the tax-based measure from the IRS. According to BEA's revision studies, corporate profits have the largest mean absolute revision of any component of GDP or GDI, except for farm proprietors' income.

The large revisions to profits reflect a number of factors, including the large differences between financial and tax accounting rules and BEA's economic accounting conventions; the use of financial corporate data for public companies to extrapolate profits for private or S corporations; and the possible effects of capital gains and losses or "unusual" losses—which should be excluded from GDI—in the source data for profits. And the final profits numbers differ widely depending on the source of the data. For 2005, profits as reported in the IRS Statistics of Income (SOI) increased by 43 percent, S&P operating profits by 9 percent, and Census Bureau Quarterly Financial Report profits by 15 percent. The mean absolute difference between the highest and the lowest estimate of growth in profits from 1999 to 2007 was 23 percentage points, with the largest differences recorded in 2001. Although many of these differences are relatively easily resolved, many others, such as those surrounding major changes in the economy, changes in accounting rules and practices, or changes in tax law, can be quite difficult.

The GDP estimates are, of course, not without their own limitations. As Nalewaik points out, one of the most important has been the absence of a timely, comprehensive data source for services in the early GDP estimates. Extrapolators for services may well have contributed to the tendency of the early GDP estimates to understate the decline in GDP during

4. Beginning with the first quarter of 2010, BEA estimates of wages and salaries reflect newly available monthly tabulations of hours and earnings for *all employees* on private nonfarm payrolls from the BLS's expanded current employment statistics program. However, the new BLS monthly data do not include certain types of irregular pay, such as bonuses and stock options, which are included in the QCEW data.

contractions and the increase during the early stages of expansions (Fixler and Grimm 2002).[5]

One of the important advantages of the GDP estimates is that source data (mainly Census data) are quite timely: only about 25 percent of GDP is estimated using trend-based data for the first (or, as BEA calls it, the advance) estimate of GDP. That estimate is available approximately 1 month (25 days) after the end of the quarter, whereas sufficient source data for the first GDI estimate are not available until 2 months after the end of the quarter (3 months in the case of the fourth quarter). Also, the share of trend-based data for the first estimate of GDI is 63 percent, significantly higher than even the third GDI estimate.

However, the most important advantage of the GDP source data is the ability to develop an integrated benchmark for the GDP estimates once every 5 years based on detailed, high-quality data from the Economic Census. Equally important, the monthly and quarterly Census Bureau data are conceptually consistent with the definitions used by the Census Bureau for their every-5-year benchmark and annual data. Although this consistency does not provide clear evidence that GDP is closer to "true" production, for many users the conceptual consistency of the monthly, quarterly, annual, and every-5-year Economic Census data is a major source of comfort.

THE CYCLICALITY OF THE LATEST ESTIMATES. Nalewaik's conclusions on the superiority of GDI rest mainly on his reading of the source data. As suggested above, a careful and detailed analysis suggests that the source data for the early GDI estimates are significantly weaker than those for the early GDP estimates. Nalewaik argues that the benchmarking procedures and the extrapolation of services make the GDP estimates too smooth, but he discounts the likelihood that the failure to fully remove capital gains and losses makes GDI too cyclical. Yet firms do seem to have the ability to time their receipts, expenses, and recognition of unusual losses in ways that would overstate the cyclicality of recorded profits relative to underlying economic activity. Firms may tend, for example, to recognize unusual losses when the overall economy and competing firms' sales and profits

5. As a result of a multiyear Census Bureau initiative to expand its services surveys, through new quarterly and expanded services, BEA is making substantial progress in improving the GDP source data for services. The Census Bureau's plans call for completing its program to provide complete coverage in the quarterly and annual services surveys by 2012.

are down and the losses are likely to have a smaller effect on investor perceptions and stock prices.

It is also known that quarterly wage data have included capital gains in the form of stock options, which were not taken out until the annual revisions, when profits data that excluded them were available. It also seems plausible that the misreporting adjustment that BEA applies to IRS data is countercyclical, yet BEA's misreporting adjustment is proportional and varies little over time, implying that measured income would be too cyclical.

Nonetheless, there is probably something to both sides of the argument over the cyclicality of GDP and GDI. Indeed, to address the issue, BEA is taking steps, outlined below, to improve the accuracy of services estimates through the incorporation of new quarterly services data in GDP, and to improve the early wage and salary estimates in GDI. Corporate profits will remain an issue, but this work by Nalewaik on the cyclicality of the statistical discrepancy suggests directions for future research.

THE RELATIONSHIP OF GDI AND GDP TO COINCIDENT MEASURES OF ECONOMIC ACTIVITY. Nalewaik shows that the cyclical measures that he examines move more closely in tandem with GDI than with GDP. Part of that closer correspondence relates to his use of income-type variables that either are used to measure GDI or can be seen as proxies for income-side measures of GDP. An examination of a broader set of cyclical measures, such as retail sales and manufacturing sales, shows that GDP has a closer correlation over 1984–2009 with these variables than does GDI. This partly reflects the fact that these variables either are used to measure GDP or can be seen as proxies for product-side measures of GDP. Moreover, an examination of nine business cycle measures—nonfarm employment, private services payroll, manufacturing employment, nonmanufacturing employment, personal income less transfers, industrial production, manufacturing sales, retail sales, and the nonmanufacturing ISM index—shows that their correlations with GDP and GDI are very similar, with only a slightly larger correlation for one or the other.[6] The average correlation of these variables

6. The correlations are as follows: nonfarm employment with GDP, 0.73, with GDI, 0.77; private services payroll with GDP, 0.69, with GDI, 0.73; manufacturing employment with GDP, 0.68, with GDI, 0.74; nonmanufacturing employment with GDP, 0.68, with GDI, 0.67; personal income less transfers with GDP, 0.65, with GDI, 0.76; industrial production with GDP, 0.50, with GDI, 0.60; manufacturing sales with GDP, 0.75, with GDI, 0.69; retail sales with GDP, 0.65, with GDI, 0.57; nonmanufacturing ISI index with GDP, 0.57, with GDI, 0.66.

with GDP was 0.66, compared with 0.69 for GDI, even though five of the nine variables are income-type variables.

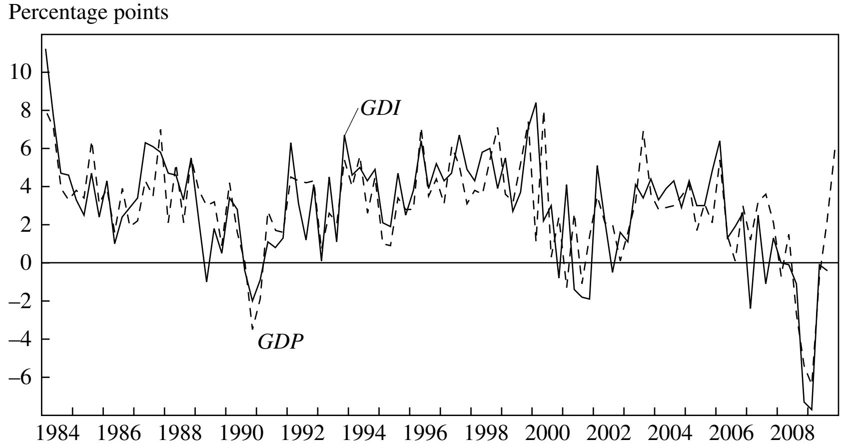
I have not examined the leading index indicators discussed by Nalewaik because they do not seem to be a meaningful measure of the accuracy of either GDP or GDI. Stock prices, the yield curve, and high-yield bond spreads are not, as Nalewaik notes, measures of economic activity but rather leading indicators that are used to try to predict economic activity. As the former custodian of the leading indicators—which have been described as measurement without theory—I can report that they have a less-than-stellar history of predicting GDP *and* the business cycle, especially when examined in real time. The S&P 500 index, for example, may be a good financial indicator, but it has a checkered history as a leading economic indicator.

ACCURACY OF THE GDP AND GDI ESTIMATES IN 1984–2006. My review of most of the evidence marshaled by Nalewaik and a review of the current (rather than the revised, or latest) estimates from BEA suggest similar trend growth and cyclical patterns for the GDP and GDI estimates in 1984–2006. Both GDP and GDI provide very similar estimates of trend growth. Looking at revisions to the GDP and GDI estimates at the time of the comprehensive benchmarks—which are based on the Economic Censuses of 1982, 1987, 1992, 1997, and 2002—one can see that nominal GDP and GDI were revised by an average of 1 to 2 percent, and the growth rates for those 5-year periods were revised by about one-quarter of 1 percentage point.

A number of revision studies have shown that GDP and GDI estimates are both reliable indicators of general economic activity, as defined by whether growth is fast or slow relative to trend, whether growth is accelerating or decelerating, which of the major components are contributing to growth, and trends in saving and other major components of GDP. Dennis Fixler and Nalewaik (2009) have found that the revisions are larger around turning points, and given the degree of extrapolation in both the GDP and the GDI estimates, this makes sense. However, as I show in figures 1 and 2, the general patterns exhibited by the early estimates of both GDP and GDI are quite similar. In the last three business cycles, both early estimates show roughly the same peak, slowing pattern, trough, and recovery pattern. Most of Nalewaik's figures look at the differences in the revised GDP and GDI data, and except for 2007, the revised, or latest, data also show the same general cyclical patterns for GDP and GDI (figure 3).

ACCURACY OF THE GDP AND GDI ESTIMATES OVER 2007–09. The estimates of GDI and GDP for 2007–09 show a much larger cumulative drop in GDI than in GDP: GDI declined 4.1 percent from the fourth quarter of 2006 to

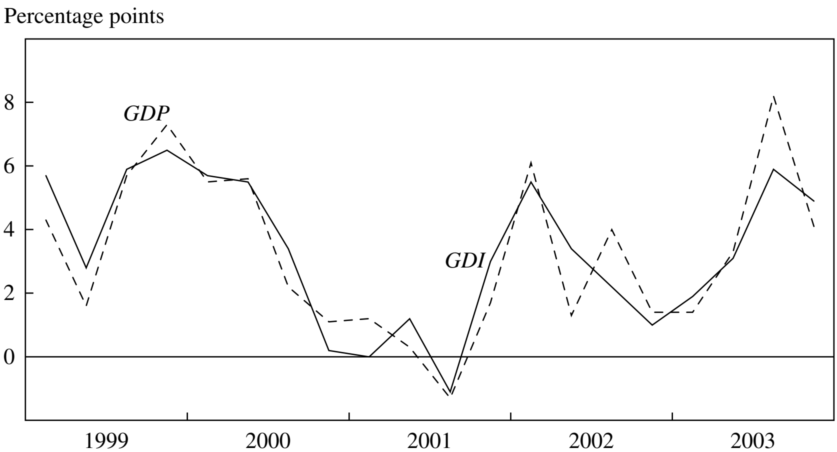
Figure 1. Changes in Real GDI and GDP, Latest-Release Estimates, 1984–2009[a]



Source: Bureau of Economic Analysis.
a. Quarterly data, annualized.

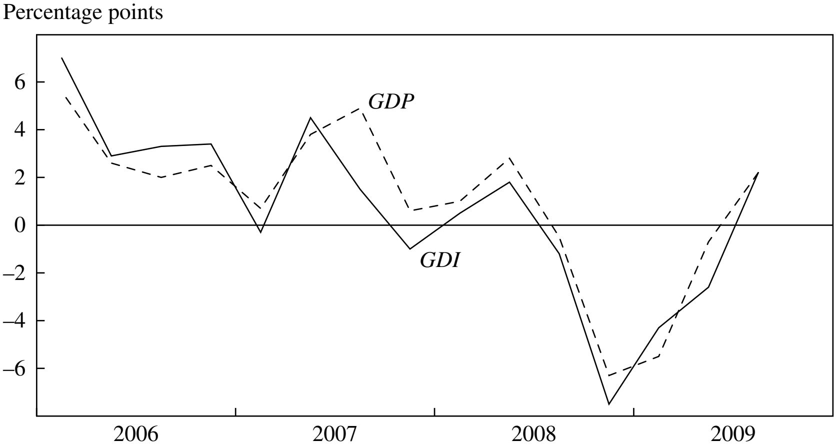
the second quarter of 2009 (the trough in both GDP and GDI), while GDP declined 1.2 percent. Almost all of the cumulative difference occurs in 2007; between the relative peak in the fourth quarter of 2007 and the second quarter of 2009, the declines in GDP and GDI were much closer, with GDP declining 3.7 percent and GDI 4.1 percent.

Figure 2. Changes in Real GDI and GDP, Third-Release Estimates, 1999–2003[a]



Source: Bureau of Economic Analysis.
a. Quarterly data, annualized.

Figure 3. Changes in Real GDI and GDP, Third-Release Estimates, 2006–09[a]



Source: Bureau of Economic Analysis.
a. Quarterly data, annualized.

Although both GDP and GDI growth began to show signs of weakness in 2007, with quarters of sharply lower and even negative growth, the latest estimates show GDI increasing only 0.1 percent over the four quarters of 2007, while GDP increased 2.5 percent. The main source of the slowdown in GDI comes from profits and proprietors' income, both of which declined in 2007. During 2007, compensation, including wages and salaries, continued to grow. This growth in compensation in GDI seems consistent with the 2.5 percent growth in GDP, which in turn seems consistent with the 0.8 percent growth in employment. Also, the residual growth in productivity (as measured by GDP per employee) of about 1.7 percent does not seem so high as to suggest that GDP growth was overestimated relative to employment (average productivity growth since 1995 averaged 2.5 percent).

The difference between the GDP and GDI estimates for 2007 (and early 2008) seems to turn on the accuracy of the profits and proprietors' income data. As discussed above, converting IRS data and financial report data to a national income and product accounts basis is extremely difficult, especially during periods of rapid change in markets. Firms have the ability to adjust the timing of their expenses and receipts and when they recognize unusual gains and losses. Tax law changes and changes in the economy can also affect the consistency of profits over time. The year 2007 marked the beginning of the financial crisis, and although BEA does its best to exclude unusual gains and losses, profit estimates for banks and other financial institutions were particularly challenging during and after the crisis. Estimating proprietors' income is challenging as well. IRS studies suggest that for each dollar reported to the IRS, another dollar is not reported. BEA therefore roughly doubles the annual estimate reported by the IRS. Unfortunately, only two comprehensive IRS estimates of underreporting have been published in the last 25 years: the 1988 IRS Taxpayer Compliance Measurement Program and the 2001 IRS National Research Program. So BEA's doubling may capture the long-run trend in compliance, but it may not be appropriate at times of significant change in the economy, and as noted above, it may cause measured income to be too cyclical.

It is also worth noting that the National Bureau of Economic Research, using data on employment, sales, and a number of other cyclical indicators including GDP, placed the cyclical peak in December 2007. This cyclical dating then counts most of 2007 as a period of expansion, which is consistent with growth in GDP, but not with the flat-to-declining pattern of GDI.

Overall, my reading of the behavior of the GDP and GDI data for 2007 and early 2008 suggests that GDP looks more consistent with the behavior of employment and unemployment than GDI. However, both sets of estimates will see further revisions, so interested readers should stay tuned.

SHOULD WE AVERAGE GDP AND GDI? Averaging GDP and GDI, using the two-to-one weighting recommended by Nalewaik, should produce an estimate of output that changes less in subsequent annual revisions than GDP has in recent years, but the statistical gain would not be large, on average, relative to the average revision. Moreover, the value of averaging must be weighed against two disadvantages: first, that of having larger revisions between the advance and the second estimate (when the income data are introduced), and second, the risk of having anomalous revisions that could reduce confidence in the overall accuracy of the national accounts.

Although the use of GDI and other real-time data may be able to reduce revisions to the early output estimates, the relative gain is likely to be small. Based on the estimates in the paper, the use of GDI over 1994–2006 would have reduced the mean absolute revision in the early GDP estimate by 0.2 percentage point, but the mean absolute revision to GDP over this period was 1.25 percentage points.[7] That is not insignificant, but the relative size, along with the fact that the early GDI and GDP estimates present a similar picture of the business cycle, needs to be considered in any proposal to produce a weighted average. Also, users of economic data may

7. Based on a replication of Nalewaik's equation, the mean absolute revision in 1994–2006 to GDP would be reduced from 1.25 to 1.18 percentage points.

perceive problems with an average GDP growth rate that is not consistent with the growth in the subcomponents for either GDP or GDI.

In the past, BEA has presented the idea of averaging the estimates to its key users. BEA's Advisory Committee, the Federal Reserve Board, and other users of the national accounts have consistently told us that if we want to balance GDP and GDI, we should continue to publish separate estimates of both, along with the statistical discrepancy, and then produce a balanced set of accounts that allocate the discrepancy using a replicable, statistically based method. BEA has been working on a methodology for balancing the input-output and industry accounts, but we do not think it is feasible to develop balanced quarterly GDP and GDI accounts. However, BEA will explore means of better presenting and highlighting the GDP and GDI estimates in ways that meet the differing needs of the various users.

BEA will also continue to work with the BLS, the Census Bureau, and the IRS to improve the source data for both GDI and GDP. Incorporation of the next steps in the expansion of the quarterly services survey should continue to bring significant improvement in the source data for GDP. Recent efforts by the BLS to collect data on all types of income—including bonuses, stock options, and other irregular payments—were unsuccessful, but the recent incorporation of the new BLS data on wages and salaries for all workers should significantly improve the estimates of wages in GDI. Finally, legislation now in prospect that would allow BEA, the BLS, and the Census Bureau more consistent access to tax data could be very helpful in reconciling the large differences between financial and tax accounting data.

REFERENCES FOR THE LANDEFELD COMMENT

Fixler, Dennis J., and Bruce T. Grimm. 2002. "Reliability of GDP and Related NIPA Estimates." *Survey of Current Business* 82, no. 1 (January): 9–27.

———. 2006. "GDP Estimates: Rationality Tests and Turning Point Performance." *Journal of Productivity Analysis* 25, no. 3: 213–29.

Fixler, Dennis J., and Jeremy J. Nalewaik. 2009. "News, Noise, and Estimates of the 'True' Unobserved State of the Economy." Working paper. Washington: Bureau of Economic Analysis.

GENERAL DISCUSSION Christopher Sims asserted that the relevant issue is whether GDP(I) or GDP(E) is better by itself as an indicator, and not whether putting GDP(I) or GDP(E) into a factor model makes a difference in the outcome, as Francis Diebold had sought to demonstrate. If GDP(E) were simply GDP(I) plus noise, then the result would be the same, because Diebold is extracting the nonnoise part of GDP(I) and GDP(E). Sims claimed that what is of real interest is which measure, GDP(I) or GDP(E), does most of the work in the factor model. Traditional aggregates, which are in a sense informal factors, can come close to extracting the main business cycle factor. It would be interesting to test a bivariate factor model to see whether an underlying factor in both GDP(I) and GDP(E) is closer to one or the other, and to ask how close GDP(I) or GDP(E) comes by itself. Francis Diebold responded that one could indeed treat GDP(E) and GDP(I) as indicators and extract a factor from them, but that that is a different exercise, in no way superior to or more appropriate than the results he had reported, which answer different but equally important questions.

Robert Hall noted that the press release from the NBER Business Cycle Dating Committee announcing its determination of the December 2007 business cycle peak said specifically that the committee consulted real gross domestic income. Clearly its usefulness is not news to the committee. The committee also rejected an approach to defining real activity that mixes a number of indicators together, because the mix always overweights manufacturing. As the manufacturing share of GDP has declined, it is important to avoid what would become substantial double counting. Hall further reported that the notional set of indicators that the committee listed does not reveal the weights the committee applies. The committee is focused on finding two things: the best possible measures of output and aggregate employment. The committee looks at the modern economy mainly with economy-wide, not sectoral, measures.

Hall found the paper persuasive on the point that the best early estimate of output ought to use a lot of highly relevant variables. The quality of the early estimates of both GDP(E) and GDP(I) could be improved by giving weight to private forecasts as well as the early data available to the BEA. Because the government might be squeamish about releasing an output estimate that relies mainly on forecasts and correlations, the BEA ought to consider leaving the close-to-real-time estimates—the nowcasting—to others. At a minimum, users of the BEA's early estimates should be consulting private nowcasts as well.

Phillip Swagel thought the paper raised a basic question about the very nature of a recession. Consider the debate over unemployment versus output measures: if real GDP growth were 1 percent for a considerable period, there would surely be net job losses. Would that be a recession, or not? The question is, What does one hinge the start and end dates on: measures of output or measures of the labor market?

Matthew Shapiro seconded Sims's comment that it would be useful to have the bivariate factor model calculate the optimal portfolio weights of

GDP(E) and GDP(I). He felt Diebold's illustrative model was misleading because it assumes that all the variance is error. If most of the variance were signal, the results might be quite different. The right weights will depend on the relative amounts of signal and noise in the two data series. Philip Howrey had done something similar 20 years ago, attempting to assign weights to the Bureau of Labor Statistics' household and establishment surveys. His analysis put about 80 percent of the weight on the establishment survey, which has now become the conventional wisdom.

Jan Hatzius noted that before the BEA reported estimates of GDP(E) and GDP(I) for the fourth quarter of 2009, the Federal Reserve's Flow of Funds tables already included an estimate of the fourth-quarter statistical discrepancy between the two, which shows an increase of about $130 billion over the previous quarter. That implies an estimate of real GDP(I) growth for the fourth quarter of just over 2 percent annualized. Hatzius was curious about how much, if any, weight should be put on that. He also observed that many analysts are concerned at the moment about the deviation between the performance of large firms and that of small firms, and what that might mean for preliminary estimates of GDP. Which of the two GDP measures is more vulnerable to that deviation? Finally, Hatzius wondered why other countries put more weight on income-based measures. Is it because they have different data sources, with different strengths and weaknesses than their U.S. counterparts, or do they simply reach a different conclusion about how important it is not to confuse the public, for example by taking an average of different measures?

Robert Gordon did not accept the characterization of the debate over output versus employment in the business cycle dating context as a tug-of-war between proponents of one or the other. Rather, he saw it as an econometric problem, one that involved studying the breakdown of changes in output and the output gap into their components, starting from the simple identity that output is equal to aggregate hours times output per aggregate hour, that is, aggregate productivity.

Gordon also acknowledged that the Okun's law relationship is quite different today from what it was in the mid-1980s, with unemployment becoming much more responsive to output than in the original formulation. Whereas in Okun's day aggregate hours responded by two-thirds of any movement in the output gap, today hours respond more than one for one. Further, the Great Recession witnessed departures from this relationship, with productivity growing faster and hours falling further than even the post-1986 equation would have predicted.

Gordon went on to point out that Nalewaik's figure 6, which shows the behavior of real GDP(I) and GDP(E) estimates over the most recent recession, reduces the residuals. In fact, the erroneous division of the equation between productivity and hours is almost eliminated. But this happened before the recession started; it is mainly a story about 2007. Gordon found it reassuring to have at least a partial explanation of why productivity had looked so good in the last 2 years.

Benjamin Friedman encouraged making a sharper distinction between two conceptual questions. The first is whether output or employment is the more meaningful concept for judging turns in business cycles. The second is which of the two statistics, GDP(E) or GDP(I), does a better job of measuring what we understand by output. The second question arises only because the statistical agencies use double-entry bookkeeping, which in a world of imperfect measurement necessarily leads to discrepancies. Even in a world of perfect measurement, where GDP(I) and GDP(E) are always identical, the first issue would still be a question, but the issue addressed by the paper would go away.

Friedman noted that the paper showed a very strong historical correlation between the statistical discrepancy between the two output measures and unemployment, which suggests that the question of what is going either unmeasured or mismeasured that gives rise to the difference is not just about, for example, the superiority of one or another source of data, but rather involves substantive questions of economics. He encouraged further analysis of what these measurement problems are. One might think the discrepancy is just noise, but the correlation he cited shows that it is not pure noise. Steven Davis added that the correlation suggests that the discrepancy is cyclically varying, which implies that determining the optimal weights for an average is more complicated.

Davis also remarked on Steven Landefeld's discussion of the administrative record inputs to the GDP(I) side, each of which is somewhat different in scope. Recognizing that making adjustments for these differences is challenging, he hoped that the BEA would drill down deeper and investigate the extent to which these administrative data sources line up when the discrepancies are more fully taken into account. There are potentially tremendous advantages to relying on administrative records when possible; their comprehensive nature creates much greater opportunities for disaggregation by type of activity, location, and other dimensions.

Steven Braun complimented the BEA for showing its dirty laundry, in the sense of making its best estimates of both GDP(E) and GDP(I) available so that economists could analyze the statistical discrepancy. Oral tradition among analysts of BEA data has it that before 1980, the BEA managed the statistical discrepancy and did not allow it to change very much. Braun also said that what he would most like to see changed in the federal statistical system was not anything that the BEA does, but rather the way the BLS publishes its productivity data, preferring that the numerator for the productivity calculation be a weighted average of the two output measures.

Justin Wolfers expressed concern about Diebold's approach, on the grounds that the usefulness of any data is in what they reveal about the underlying real factor. When looking for the causes of the current recession, it matters for the diagnosis whether it appears that labor productivity was growing or falling. As Braun had argued, it is a question of getting the numerator right, and it has real economic significance.

Wolfers went on to make a plea for Landefeld to give the paper another chance. He thought the concern over users being confused by an average of GDP(E) and GDP(I) was overblown: sophisticated users are already taking an average. Moreover, when the choice is framed as one measure or the other, it is all too easy to regard whichever choice the administration's economists make as political. In any case, if the BEA is determined to rely on only one measure, all the metrics in the paper except one say it should be the income-based measure.

Alan Blinder observed that the paper shows that the GDP(E) revision, which previously had been thought not to be forecastable, is in fact forecastable by GDP(I). This alone, in his view, was a sufficient reason for the BEA to use it.

Steven Landefeld, responding to some of the concerns raised, noted that Europe uses GDP(I), probably for lack of the kind of sources available in the United States for expenditure-side data. Europe is now in the process of developing a set of monthly indicators for sales and output. He also commented that the gain in accuracy from "nowcasting" is relatively small. It does not dramatically change things. With respect to averaging, what some users, including many members of the BEA's advisory committee, have requested is, rather than a simple average, a statistical methodology that shares it out to components, creating a consistent picture. The BEA is working on something similar in its industry accounts.

Landefeld added that from his experience as the custodian of the index of leading indicators and previous research, he did not expect to find the use of cyclical indicators very helpful in improving the accuracy of GDP(E) or GDP(I), but that the BEA would continue to research and attempt to address the sources of the statistical discrepancy over the business cycle. He also noted that the BEA intended to better highlight GDP(I) in its reports.

GAREY RAMEY
University of California, San Diego

VALERIE A. RAMEY
University of California, San Diego

The Rug Rat Race

ABSTRACT After three decades of decline, the amount of time spent by parents on childcare in the United States began to rise dramatically in the mid-1990s. This increase was particularly pronounced among college-educated parents. Less educated mothers increased their childcare time by over 4 hours per week, and college-educated mothers increased theirs by over 9 hours per week. Fathers showed the same patterns, but with smaller magnitudes. Why would highly educated parents increase the time they allocate to childcare at the same time that their returns from paid employment have skyrocketed? Finding no empirical support for standard explanations, such as selection or income effects, we argue instead that increased competition for college admissions may be an important factor. We provide empirical support for our explanation with a comparison of trends between the United States and Canada, across ethnic groups in the United States, and across U.S. states.

As time in paid work has increased over the last four decades, time spent on most home production activities has trended downward (see, for example, Robinson and Godbey 1999, Bianchi, Robinson, and Milkie 2006, Aguiar and Hurst 2007). One notable exception, however, is time spent on childcare. Suzanne Bianchi (2000) and Liana Sayer, Bianchi, and John Robinson (2004) show that despite shrinking families, parents in the late 1990s reported spending as much or more time on childcare than parents in earlier decades.

In this paper we show that there has in fact been a dramatic increase in time spent on childcare. Linking 13 time-use surveys between 1965 and 2008, we show that after declining for several decades, time spent per week on childcare started increasing in the mid-1990s. The trends follow a pronounced S-shaped pattern, rising markedly from the mid-1990s to the early 2000s and then flattening out. Moreover, the increase in childcare time has been twice as great for college-educated as for less educated parents. This differential trend is particularly puzzling in view of the sharp increase in the average wages of college-educated individuals over much the same period. We also show that an important component of the increase in childcare time was time spent on *older* children, and in particular on coordinating and transporting them to their activities.

Our estimates imply increases in average weekly hours of childcare time ranging from 3 hours per week for less educated fathers to more than 9 hours per week for mothers with a 4-year college degree. The implications for the allocation of time are large by any metric. According to our estimates, the time spent on childcare by the entire adult population in 2008 is equal to almost 20 percent of the time spent on paid work. The increase in average weekly time spent on childcare during a 10-year period from the early 1990s to the early 2000s was equal to 70 percent of the absolute decline in work hours during the "Great Recession" that began in late 2007. If those hours are valued at the market wage, the cost of the increase in childcare time amounts to over $300 billion per year.

The literature has offered several explanations both for why childcare time has increased and for why it is greater among more educated parents (see, for example, Bianchi and others 2006, p. 87; Aguiar and Hurst 2007; Guryan, Hurst, and Kearney 2008). These explanations include selection effects, income effects, safety concerns, greater enjoyment of childcare, and more flexible work schedules. We test each of these and find that they are not consistent with the data.

The inability of existing explanations to account for the evidence leads us to offer a new explanation for the upward trends in childcare time. We argue that much of the increase, particularly among college-educated parents, may be a response to an increase in the perceived return to attending a good college, coupled with an increase in competition for college admissions. The size of college-bound cohorts rose dramatically beginning in the mid-1990s, coincident with the increase in childcare time. John Bound and Sarah Turner (2007) provide evidence that these larger cohorts are associated with increasingly severe cohort crowding at quality schools. The increased scarcity of college slots appears to have heightened rivalry among parents, which takes the form of more hours spent on college preparatory activities. In other words, the rise in childcare time resulted from a "rug rat race" for admission to good colleges.

To clarify the mechanics of this explanation, we develop a simple theoretical model in which college admission depends on parents' choice of time spent preparing their children for college. College-educated parents are assumed to have a comparative advantage in preparation time. When slots at good colleges are relatively plentiful, the marginal slots are filled by children of less educated parents. Competition among these parents then determines the preparation required for admission. When good slots become relatively scarce, rivalry for the marginal slots shifts to the college-educated parents, who are better able to compete. A rug rat race emerges among these parents, driving up both admissions requirements and the time spent on childcare.

We provide support for this explanation using three comparisons. First, we compare childcare trends in the United States with those in Canada. The two countries are similar along many dimensions but differ in one respect that is key to our explanation: the Canadian higher education system lacks a steep prestige hierarchy, so that Canadians do not experience the same intense rivalry to gain admission into higher-rated colleges. Thus our theory predicts that time spent in childcare by more educated Canadians should not have increased by as much as it has among their U.S. counterparts. Employing time-use data from Canada's General Social Survey, we show that time spent in childcare by more educated Canadian parents changed very little over this period, corroborating our theory. Second, we show that black and Hispanic parents in the United States spend less time in childcare than white parents. Since affirmative action policies may attenuate the rivalry for scarce slots for underrepresented minorities, there may be less pressure for them to spend time on childcare. Third, we use Bound, Brad Hershbein, and Bridget Long's (2009) measure of competition for college admission to demonstrate a positive correlation between the degree of competition and childcare time across U.S. states.

The paper proceeds as follows. Section I documents trends in childcare over the 1965–2008 period. The standard explanations are evaluated in section II. Section III documents that competition for college increased over this period, develops our new explanation, and reports the empirical evidence in its favor. Section IV concludes.

I. Trends in Time Spent in Childcare

Long-term trends in time spent in care of children have been the subject of many studies in sociology (for example, Bryant and Zick 1996; Robinson and Godbey 1999; Sayer and others 2004). It has long been noted that college-educated mothers devote more time to childrearing than less educated mothers (see, for example, U.S. Department of Agriculture 1944, Leibowitz 1974, Bianchi and others 2006, Guryan and others 2008). Here we document that since the mid-1990s there has been a substantial increase in childcare time as well as a widening of the gap between college-educated and less educated parents.

I.A. Data Description

To document these trends, we use information from 13 nationally representative time-use surveys from 1965, 1975, 1985, 1992–94, 1995, 1998, 2000, and annually from 2003 through 2008. All of the surveys are based on time diary information, which is considered to be the most reliable measure of how individuals spend their time. Table A1 of the online data appendix provides details about the surveys.[1] We use the American Heritage Time Use Study (AHTUS) versions of the 1965, 1975, 1985, and 1992–94 surveys (Fisher and others 2006) and the original versions of the other surveys (Robinson, Bianchi, and Presser 2001, Bianchi and Robinson 2005, Bureau of Labor Statistics 2010).

The key measurement issue is the extent to which the surveys give consistent measures over time. The potentially problematic surveys are the 1992–94 survey and the Bureau of Labor Statistics (BLS) surveys starting in 2003. Many childcare researchers believe that the 1992–94 survey undercounts primary childcare activities (Robinson and Godbey 1999, Bianchi and others 2004, Bianchi and others 2006). Using results from other time-use studies that are not part of the AHTUS but are considered comparable to the earlier surveys, Allard and others (2007, footnote 19) argue that the 1992–94 survey is not comparable. That survey suggests that time spent in childcare was 1 hour per week lower in the early 1990s than in 1985, whereas the 1995 survey suggests that it was 1 hour per week *higher.* Thus any drops in childcare time between 1985 and 1992–94 may be due to problems with the 1992–94 survey. Another important drawback of that survey is its lack of information on key controls, such as marital status.

Concerns have also been expressed about the comparability of the 2003–08 BLS surveys with the earlier surveys. However, Allard and others (2007) compare the 2003 BLS survey with the 2000 Survey Research Center national survey of parents and find very similar estimates of primary time spent in childcare (but not of secondary time, that is, time when the parent is engaged in another, primary activity while also engaged in childcare). The 2000 survey was designed to be comparable to the earlier

1. Online appendices for all papers in this issue may be found on the *Brookings Papers* webpage (www.brookings.edu/economics/bpea), under "Conferences and Papers."

surveys, so it appears that the increase in time spent on childcare in the BLS surveys relative to earlier surveys is real rather than due to methodological differences.

Fortunately, the 1965, 1975, 1985, 1995, 1998, and 2000 surveys all involved John Robinson as a principal investigator. As a result, the coding of activities is very similar across these surveys. Because these surveys span the period in which childcare began trending upward, we feel confident that the trends we find in time spent in childcare reflect actual trends rather than changes in activity classification.

We use a comprehensive measure of childcare that includes care of infants, care of older children, medical care of children, playing with children, helping with homework, reading to and talking with children, dealing with childcare providers, and travel related to childcare. The online data appendix gives details of the activity codes used.

I.B. Trends in Total Childcare

To study changes in childcare over time, we regress individual-level time spent on childcare on various sets of controls. Most of our results are based on the following simple descriptive model:

$$CH_{it} = X_{it}\beta + \varepsilon_{it},$$

where CH_{it} is the number of hours per week spent on childcare by person i in year t, X_{it} is a set of controls, and ε_{it} captures other, omitted factors affecting childcare time. Our sample consists of parents aged 18–64 who are not students, where "parent" is defined as anyone having a child under 18 years in the same household.[2] We use the recommended weights from the various studies, normalized so that a representative individual in 1965 has the same weight as a representative individual in 2008. In addition to the year of the survey, X_{it} may include dummy variables for the age group of the parent (ages 18–24, 25–34, 35–44, 45–54, and 55–64), whether or not the parent has a 4-year college degree, the interaction of the college degree dummy with the survey year, a dummy for the parent's marital status, the number of children in the household, the number of children squared, dummy variables for the age of the youngest child (1 or less, 2, 3–5, 6–9, 10–13, and 14–17), and the number of children under age 5.

2. We use this definition because most earlier time-use studies did not specifically identify parents. In 2003–08, mothers spent only 27 minutes a week more on childcare on average than the average for all women who lived in a household with children. One of the reasons we omitted students from all of our samples was to avoid misrecording a college student living at home with younger siblings as a parent.

We first consider time spent in childcare by mothers. In our benchmark specification we do not condition on any choice variables that may be correlated with education level; thus the only control variables used are the five age categories of the women, as defined above. The omitted dummy variables are survey year 1975, less than college education, and ages 25–34.[3]

Column 1-1 of table 1 shows the results of this estimation. The levels effects for the survey year dummies show that the average amount of time spent by mothers on childcare decreased from 1965 to 1975, and again in 1985 and 1992–94. Recall, however, that many analysts believe that the 1992–94 survey undercounted childcare time, so this estimate may not indicate an actual decrease. Mothers' childcare time in 1995 was 1.74 hours more per week than in 1975, and by 2000 it had risen to nearly 4 hours more. From 2003 through 2008, less educated mothers spent about 4 hours more per week in childcare activities than they did in 1975.

Of additional interest are the coefficients on the interactions between survey year and college education. After a trough in the mid-1990s, these coefficients begin to grow in the late 1990s. These estimates, combined with the coefficients on survey year and college education, show that from 1965 to 1995, college-educated mothers spent between 0.03 and 2 more hours per week on childcare than did non-college-educated mothers. Beginning in 1998, however, this differential underwent a dramatic increase: college-educated mothers spent over 3 hours more per week in 1998, roughly 5 hours more in 2000 and 2003, and over 6 hours more in 2004 and 2005. Between 1998 and 2008, the college differential in every year was at least double the highest differential observed between 1965 and 1995.

The top panel of figure 1 depicts these trends graphically for mothers in the 25–34 age group; the trends for other age groups follow the same time pattern, differing only in the time-invariant constant term. For both education groups, time spent in childcare rose beginning in 1995, but after 1998 the upward movement was much sharper among college-educated mothers, following a pronounced S-shaped pattern. As discussed earlier, problems with the 1992–94 study make it likely that the true value for that period was somewhat higher.

3. We use 1975 as the omitted year because we will later be comparing the 2000 and later surveys with the 1975 survey using the more complete set of controls that are available for those surveys, but not for other years.

Table 1. Baseline Regressions Identifying Trends in Childcare Time[a]

	Mothers		*Fathers*	
Independent variable[b]	*Age controls only 1–1*[c]	*Full set of controls 1–2*[d]	*Age controls only 1–3*[c]	*Full set of controls 1–4*[d]
1965	1.636		0.290	
	(0.613)**		(0.512)	
1985	–0.369		0.005	
	(0.689)		(0.583)	
1992–94	–1.013		0.210	
	(0.552)		(0.516)	
1995	1.744		1.232	
	(0.883)*		(0.821)	
1998	1.842		3.102	
	(0.933)*		(0.856)**	
2000	3.928	3.936	4.522	4.472
	(0.640)**	(0.603)**	(0.579)**	(0.587)**
2003	4.676	4.527	3.184	3.446
	(0.398)**	(0.380)**	(0.340)**	(0.352)**
2004	4.071	4.065	3.444	3.449
	(0.435)**	(0.413)**	(0.367)**	(0.378)**
2005	3.992	3.628	3.327	3.520
	(0.436)**	(0.415)**	(0.372)**	(0.383)**
2006	4.324	4.122	3.104	3.286
	(0.443)**	(0.421)**	(0.375)**	(0.387)**
2007	4.227	3.898	3.277	3.395
	(0.452)**	(0.430)**	(0.377)**	(0.389)**
2008	4.288	3.983	4.44	4.324
	(0.450)**	(0.429)**	(0.382)**	(0.393)**
College	0.026	–0.633	0.854	0.873
	(0.900)	(0.851)	(0.579)	(0.585)
1965 × college	2.048		0.093	
	(2.174)		(1.229)	
1985 × college	1.873		–0.321	
	(1.671)		(1.142)	
1992–94 × college	1.373		–0.422	
	(1.264)		(0.972)	
1995 × college	0.741		1.799	
	(2.153)		(1.642)	
1998 × college	3.117		2.134	
	(2.052)		(1.666)	
2000 × college	4.868	4.149	–0.406	–0.496
	(1.479)**	(1.389)**	(1.098)	(1.109)
2003 × college	4.999	3.948	1.913	1.132
	(1.015)**	(0.958)**	(0.688)**	(0.696)
2004 × college	6.344	5.478	1.425	0.903
	(1.062)**	(1.002)**	(0.733)	(0.741)
2005 × college	6.038	5.293	2.514	2.043
	(1.073)**	(1.012)**	(0.746)**	(0.754)**

(continued)

Table 1. Baseline Regressions Identifying Trends in Childcare Time[a] (*Continued*)

Independent variable[b]	*Mothers* Age controls only 1–1[c]	*Mothers* Full set of controls 1–2[d]	*Fathers* Age controls only 1–3[c]	*Fathers* Full set of controls 1–4[d]
2006 × college	4.109	3.015	2.296	1.692
	(1.072)**	(1.011)**	(0.741)**	(0.749)*
2007 × college	5.291	4.456	2.138	1.511
	(1.073)**	(1.012)**	(0.755)**	(0.764)
2008 × college	4.659	3.872	0.900	0.682
	(1.074)**	(1.013)**	(0.745)	(0.753)
Constant	11.656	−3.251	3.565	−2.655
	(0.342)**	(0.542)**	(0.300)**	(0.496)**
No. of observations	24,342	21,659	17,806	15,829
Adjusted R^2	0.09	0.21	0.06	0.10

Source: Authors' regressions using data from various time-use surveys.

a. The dependent variable is hours per week spent by parents (all persons aged 18–64, except students, caring for a child under 18 living in the same home) on childcare. Standard errors are in parentheses. Asterisks indicate statistical significance at the *5 percent or the **1 percent level.

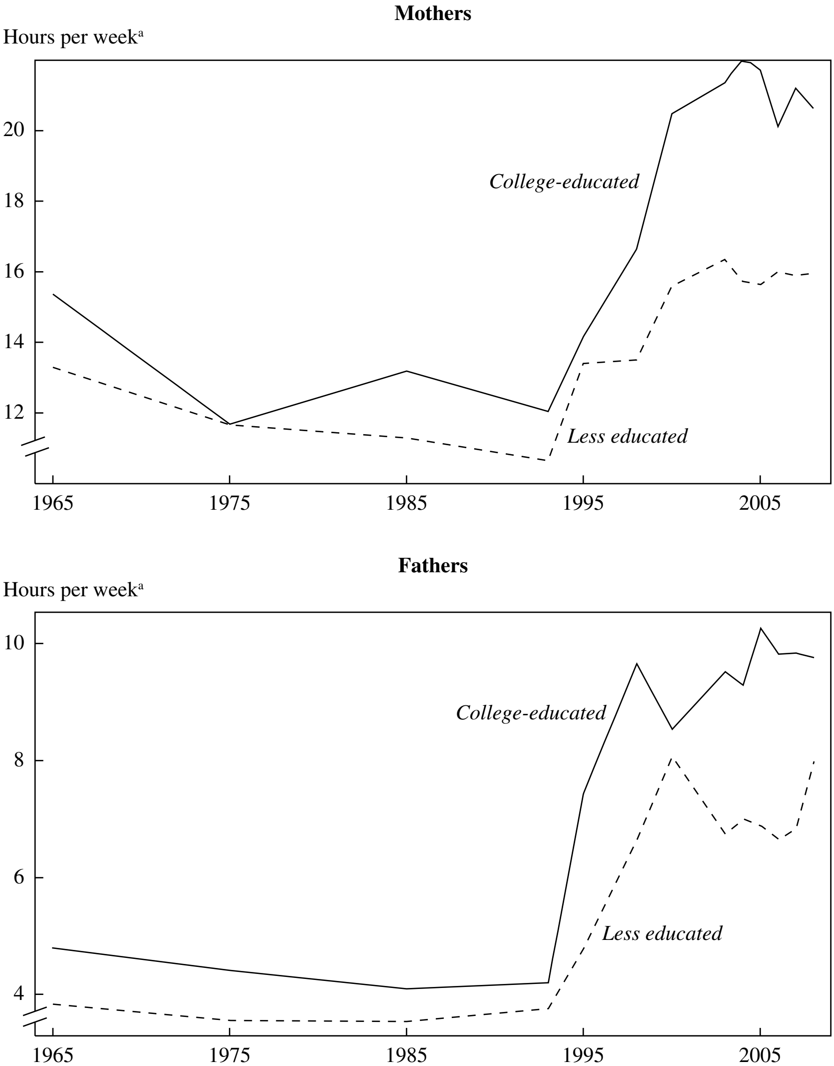
b. Year variables equal 1 when the observation is from a survey in the indicated year (the omitted year is 1975); "college" equals 1 when the parent is a college graduate.

c. Regressions include a dummy variable for the parent's age (equal to 1 if the parent is aged 18–24, 35–44, 45–54, or 55–64; the omitted category is age 25–34).

d. Regressions include, in addition to the age controls, a dummy for the parent's marital status, a quadratic in the number of children in the family, and a dummy for the age of the youngest child (1 or younger, 2, 3–5, 6–9, or 10–13; the omitted category is age 14–17).

Column 1-2 of table 1 compares childcare time spent by mothers in 1975, 2000, and 2003–08 using the more complete set of controls available for these 8 years. In addition to the age category of the mother, we control for marital status, the number of children in the household (using a quadratic), and the age category of the youngest child (using the ranges stated above). This full set of controls is useful for controlling for trends in fertility, such as the declining number of children per family and the rising maternal age at birth of the first child. In this specification we are seeking to identify differences across education levels among mothers with similar numbers and ages of children. The results are similar to those without the complete controls. The amount of time spent on childcare by less educated mothers rose by about 4 hours per week from 1975 to the 2000s, and time spent by college-educated mothers rose by about 8 hours per week. Thus even with the more complete set of controls, we find that college-educated mothers increased their amount of time spent in childcare by double the amount that less educated mothers did.

Figure 1. Time Spent on Childcare by Parents, by Educational Attainment, 1965–2008



Source: Authors' estimates based on regression results reported in table 1, columns 1-1 and 1-3.
a. Particular levels of hours shown are representative of the 25–34 age group.

Is the same true for men? Column 1-3 of table 1 reports the results of our benchmark specification for fathers, and the bottom panel of figure 1 plots the trends for fathers aged 25–34. Although fathers consistently spent much less time on childcare than mothers did, they, too, sharply increased their childcare time in the late 1990s and early 2000s. The final column of table 1 compares childcare time by fathers in 1975 with that in later years and includes the full set of controls. Again the results are similar to those without the additional controls. We conclude that time spent on childcare by both mothers and fathers increased beginning in 1995, and that this increase was significantly steeper among the college-educated.[4]

One might worry that the rise in reported childcare time might be the result of changing social norms causing parents to exaggerate their estimates of time spent with their children. However, corroborating evidence for these trends is provided by time-use studies of children. For example, John Sandberg and Sandra Hofferth (2005) link studies of time diaries kept by children in 1981 and 1997 to determine trends in time spent with parents. They find that children between the ages of 3 and 12 spent 18 more hours per week with one or both parents in 1997 than in 1981, corroborating the trends found using parents' time diaries.

I.C. Trends in Categories of Childcare

We now break down the trends documented above into trends for various categories of childcare. To produce consistent estimates of these trends over time, we limit the sample to the 1965, 1975, 1985, 1995, 1998, and 2000 surveys, because these six surveys used the same activity codes for subsets of childcare, distinguished between basic care of young children and care of older or mixed-age children, and included key variables that we could use as controls.[5] The controls include the parental age categories defined above, marital status, a quadratic in the number of children, and the number of young children.[6] We construct five categories of childcare:

4. Our results differ from those of Bianchi and others (2004), who do not find a statistically significant increase in the differential between college-educated parents and less educated parents from 1975 to 2000. On the other hand, a recent paper by Chalasani (2007) that studies married parents finds results similar to ours: a larger increase in childcare time among the college-educated between 1985 and 2003.

5. The reason we could not construct these categories for the BLS data is that they do not distinguish basic care of younger children from that of older children.

6. The definition of "young children" changed slightly across surveys. In 1965 it was "under 4 years of age," in 1998 it was "under 7 years of age," and in the rest it was "under 5 years of age."

general care of young children; general care of older or mixed-age children; playing with children; teaching children, which involves reading, helping with homework, disciplining, and conversing; and travel and activities, which includes travel related to childcare, coordinating children's activities, and picking up and dropping off children at school and other activities.[7]

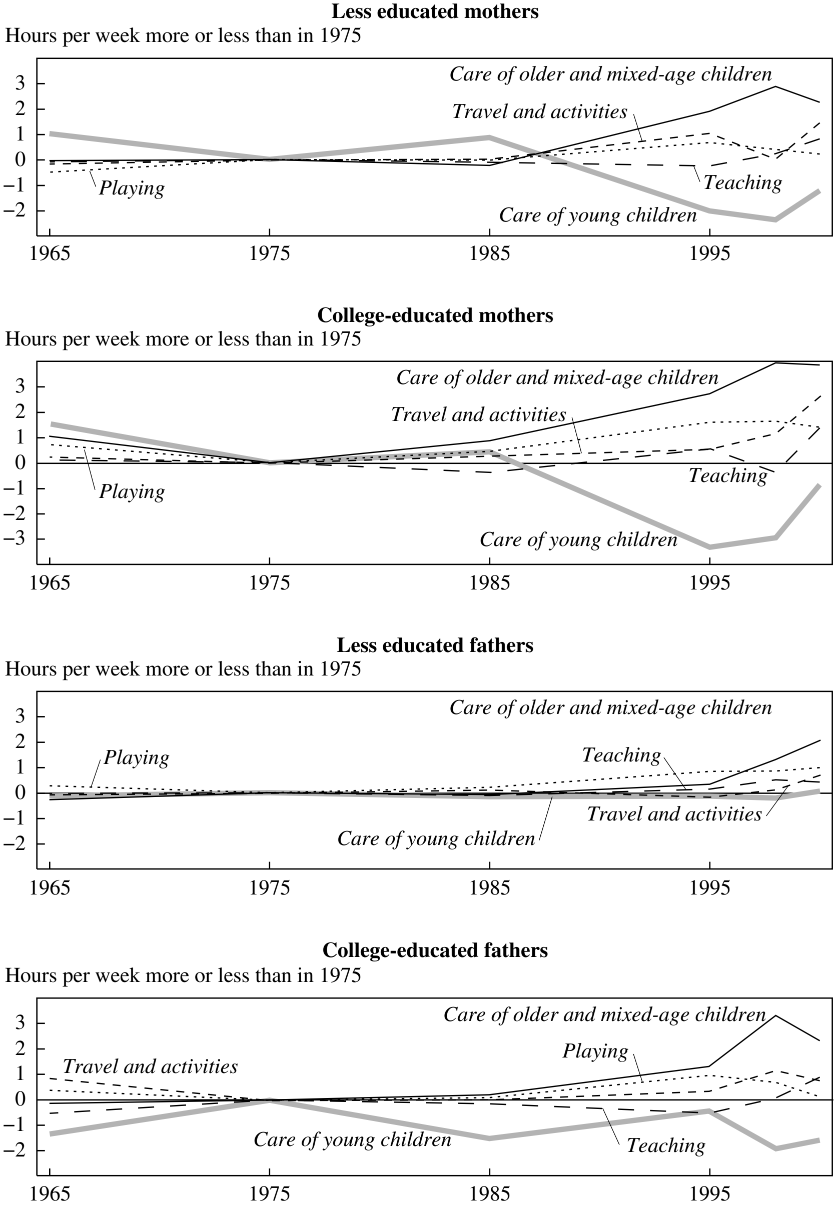
The top two panels of figure 2 show the results for each of the various care categories for mothers, and the bottom two panels for fathers. The estimates have been normalized to be zero in 1975. As the figure reveals, time spent in basic care of young children fell for all four gender-education groups, but it largely recovered after the mid-1990s for college-educated mothers. This recovery could be due in part to the increased propensity to breastfeed, as documented in Daniel Sacks and Betsey Stevenson's comment on this paper. However, Sacks and Stevenson also show that the gap in breastfeeding by education level decreased between the early 1990s and 2005–06. Thus breastfeeding cannot explain the increase in the gap in childcare time across education levels. The childcare category with the greatest increase for all four gender-education groups is general care of older or mixed-age children. Time spent in this category increased by 4 hours per week for college-educated mothers and by 3 hours per week for college-educated fathers. The category with the next-largest increase was travel and activities.

Our discussants compare trends in childcare time from the earlier surveys to the BLS surveys by comparing childcare time in households in which the youngest child is under 5 years old with that in households in which the youngest child is 5 years old or older. Based on the unconditional means from this cut of the data, they argue that much of the childcare time increase, and in particular the increase in the education differential, is attributable to households with young children. This result is interesting in itself, but it does not imply that time spent on care of young children accounts for the bulk of the increase. Childcare time in families with at least one child under 5 is distributed across all children in the household. Only the surveys through 2000 distinguish childcare by age of the child, and those surveys indicate that care of younger children is not what is driving the increase.

Although the BLS surveys do not distinguish time spent with younger children from time spent with older children, they contain other detailed

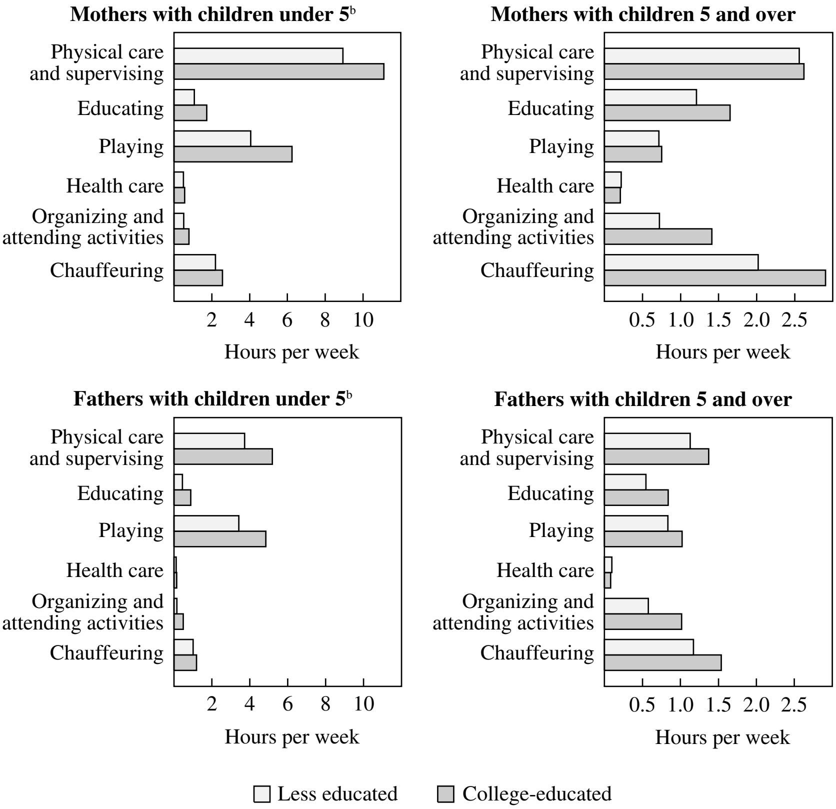
7. We omit medical care in order to make the graph clearer. There was no noticeable trend in time spent in medical care for any group.

Figure 2. Parental Time Spent on Childcare, by Educational Attainment and Type of Care, 1965–2000[a]



Source: Authors' estimates using pooled data from various time-use surveys from 1965 through 2000.
a. Estimates are averages for all parents in the indicated group and are normalized to be zero in 1975.

Figure 3. Parental Time Spent on Childcare, by Age of Child and Type of Care, 2003–08[a]



Source: Authors' estimates using data from various time-use surveys.
a. Results are averages for all parents in the indicated group aged 25–34.
b. Parents with children of mixed ages are classified in the "under 5" groups if their youngest child is less than 5 years of age.

categories of interest. We examine the following categories for the pooled sample from 2003 to 2008: physical care of children and supervision; educating and teaching children, including reading, helping with homework, and meeting with teachers; playing with children, including sports and arts and crafts; health care; organizing activities and attending children's events; and chauffeuring, which includes dropping off and picking up, waiting, and travel associated with childcare.

The two left-hand panels of figure 3 show time spent in each of these categories by parents whose youngest child is less than 5 years old. The bulk of the time spent by mothers is on physical care and supervision, followed by playing. College-educated mothers spend substantially more time per week on these two categories (11 hours in physical care and supervision and 6 hours in playing) than do less educated mothers (9 and 4 hours, respectively). Time spent on the other categories is much lower, less than 3 hours per week. The time spent by parental education level does not differ much for these other categories, with the exception of educational activities, where college-educated mothers spend noticeably more time. The story is similar for fathers, but at lower levels of hours. College-educated fathers spend more time in all categories than less educated fathers.

The two right-hand panels of figure 3 show time spent by parents whose youngest child is aged 5 or older. Time spent in physical care and supervision shrinks to about 2.5 hours per week for mothers (note the difference in scale between these and the left-hand panels). The most important category for college-educated mothers with older children is chauffeuring; physical care and supervision ranks second, and educational activities third. Also important is organizing and attending extracurricular activities. Fathers, regardless of education level, spend less time than mothers in all categories except playing. The two most important categories for fathers are chauffeuring and physical care and supervising, followed by organizing and attending activities and playing. Overall, college-educated parents of both sexes spend more time than less educated parents in each category except health care, a category in which all parents spend few hours and the difference between education groups is very small. The most important uses of the extra time spent by college-educated parents, however, are in chauffeuring and the educational and activities-related categories.[8]

The patterns revealed in figure 3 mirror the differences highlighted in other research. For example, Joseph Mahoney, Angel Harris, and Jacquelynne Eccles (2006) used pooled data from the 1997 and 2002 Child Development Supplement of the Panel Study of Income Dynamics (PSID) to show that children of white college-educated parents spend about 3 more hours per week on organized activities than children of less educated parents; however, there is no clear pattern for black families. Hofferth (2009) found an increase in time spent in academic activities from 1997 to 2003. Annette Lareau's (2003) ethnographic study, *Unequal Childhoods,* documents the dramatic differences in how educated and less educated parents raise their children. The children of less educated parents spend most of their free time playing with friends and relatives in

8. One should not infer from these results that pure travel time accounts for most of the increase in childcare time. Total travel time associated with childcare increased by approximately 1 hour from 1975 to the 2000s.

their neighborhood, unsupervised by adults. Lareau calls this the "natural growth" approach. More educated parents, she argues, take a "concerted cultivation" approach, which requires a significant commitment of parental time:

> Children's activities create substantial work for their parents. Parents fill out enrollment forms, write checks, call to arrange car pools, wash uniforms, drive children to events, and make refreshments. . . . Simply getting ready for an activity—collecting the equipment, organizing the children, loading the car—can be exhausting . . . in addition to the labor of preparing, there is the labor of watching. (p. 47)

In sum, it appears that college-educated parents with children aged 5 or over spend a good deal of their time on education and on children's organized activities.

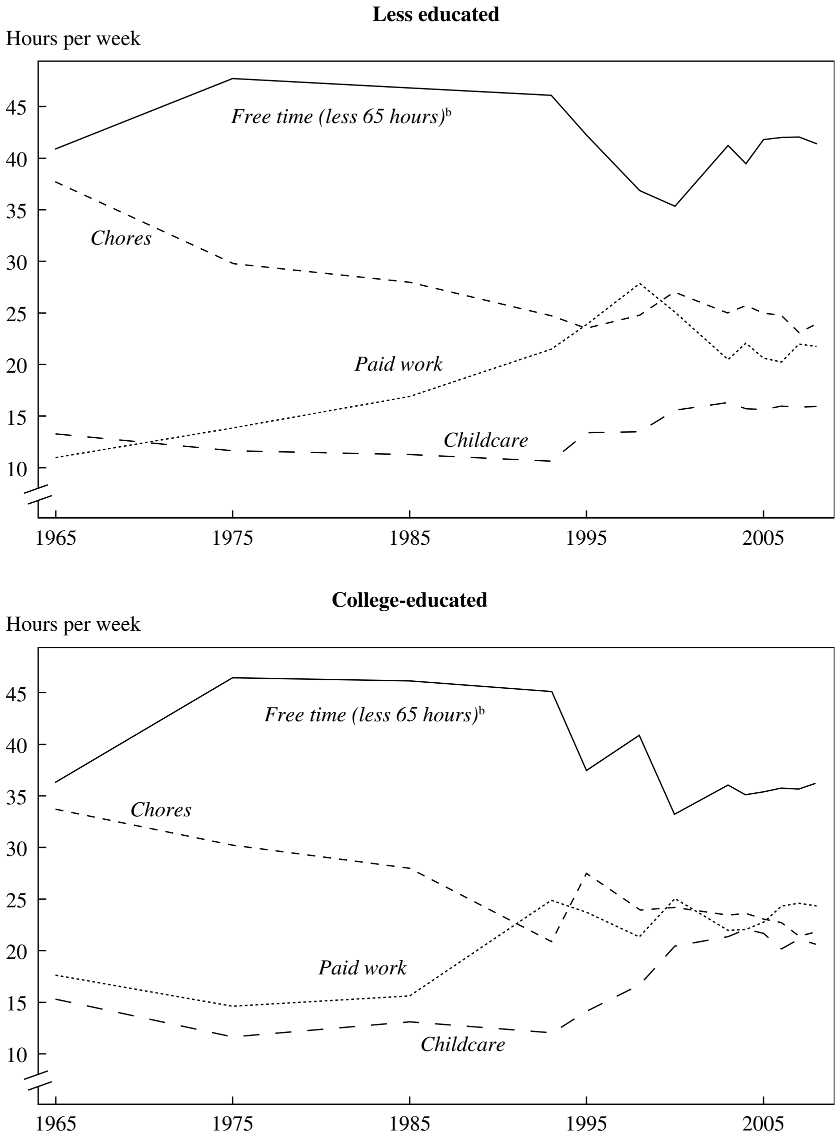
I.D. Trends in Overall Time Use by Mothers

Figure 4 sheds light on the sources of the extra time that mothers devote to their children. Here we categorize time spent other than in childcare into "work for pay," "chores," and "free and personal care time." "Chores" includes housekeeping, cooking, and shopping. "Free and personal care time" (hereafter "free time") includes any time not included in the other categories, such as sleeping, personal hygiene, and leisure activities. Data from all the time-use surveys are pooled, and the only controls are those for parental age. We report results for mothers aged 25–34. We subtract 65 hours from free time so that the magnitudes for the various categories are similar.

The figure shows that time spent in paid work by less educated mothers increased until the late 1990s and then fell somewhat. Work time for college-educated mothers increased between 1985 and the mid-1990s and then flattened out. Time spent on chores fell more or less steadily over the entire sample period for both education groups. Free time for both groups fell starting in the mid-1990s: for college-educated mothers it was 10 hours less per week in 2008 than in 1975 and 1985. Thus the decline in free time makes up for all of the increase in childcare time. In their comment, Sacks and Stevenson point out that college-educated parents are more likely to engage in childcare together. This behavior may be the result of having so little free time to spend together.

In sum, the evidence suggests that time spent in childcare has increased for all parents since 1975, but much more for more educated parents. Moreover, with the caveats about the 1992–94 survey in mind, it appears that these increases largely occurred within a single 10-year interval beginning in the mid-1990s, and an important part of the rising childcare

Figure 4. Time Spent by Mothers in All Activities, by Educational Attainment, 1965–2008[a]



Source: Authors' estimates using pooled data from various time-use surveys.

a. Particular levels of hours shown are representative of the 25–34 age group.

b. Includes personal care time. We subtract 65 hours from total free time to make the magnitudes of the categories similar.

differential between college- and less educated parents is attributable to travel and activities of older children. The trends we highlight are consistent with descriptions from popular publications, such as Judith Warner's book *Perfect Madness: Motherhood in the Age of Anxiety* (2005). The key question is, why have educated parents decided to spend their time in this way? The next section will evaluate various possible explanations.

II. Conventional Explanations

As discussed in the introduction, Bianchi and others (2006, p. 87) and other researchers have offered several possible explanations for the overall increase in time spent on childcare. We find, however, that none of these explanations is consistent with the evidence.

II.A. Selection Effects

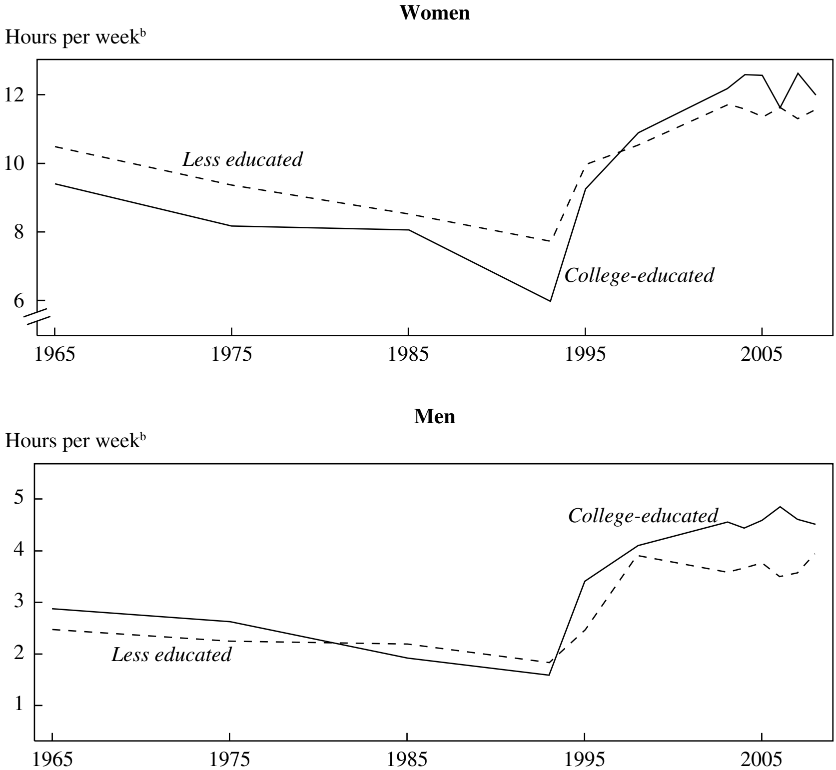
One possible explanation for this reallocation of time involves the declining incidence of parenthood over the same period. Since fewer individuals today choose to be parents, those who do might be those persons who enjoy childcare more. This selection effect could account for the observed trends in childcare time.

The easiest way to test for selection effects is to see how the results change when the universe of adults is included. If selection into parenthood explains the rise in childcare time, then childcare time averaged over both parents and nonparents (who presumably spend close to zero time on childcare) cannot be rising over time. If anything, total childcare time should decline over time, since the number of children per adult has declined. To test this argument, we obtain a new set of estimates using an expanded sample that includes all adults aged 18–64 who were not students. The results are presented in figure 5. The trends in total childcare and the college differential are clearly present among the general population of women; in particular, the rise in childcare beginning in 1995 continues to be much steeper for college-educated women. Thus our findings are not an artifact of selection into parenthood.

II.B. Income Effects

A second possible explanation invokes income effects. If childcare is a normal good, then increases in income should raise the time spent in childcare. However, since most income increases are from increases in wages, a

Figure 5. Average Childcare Time for All Adults,[a] by Educational Attainment, 1965–2008



Source: Authors' estimates based on regression results reported in online appendix A2.
a. Except those aged 65 and over and students.
b. Particular levels of hours shown are representative of the 25–34 age group.

substitution effect works in the opposite direction. The case is analogous to that of leisure, which has risen little over the last century, despite a dramatic rise in real wages, because the income and substitution effects largely cancel each other out. Thus the theoretical prediction for the impact of an increase in wages on childcare time is ambiguous.

We offer two types of evidence against income effects as a potential explanation. We first analyze the cross-sectional relationship between time spent on childcare and income and then use the estimates to determine whether they can explain the observed trends. According to our time-use surveys, average real annual household income in households with children rose from about $72,706 in 1975 to $98,608 in 2008, an increase of $25,902, or about 36 percent (in 2008 dollars).

We pool the BLS survey data from 2003 through 2008 and focus on parents. We use real household income in thousands of 2008 dollars. (The online data appendix contains more details on how we construct income.) Our regressions include dummy variables for survey year, parental age, and parental education (high school dropouts, college graduates, and those with a graduate degree), the full set of dummy variables for the age of the youngest child, a quadratic in the number of children, and a dummy for marital status. It is important to include controls for parents' education level because we do not want to attribute to income effects what are actually the effects of education itself on childcare. Less than 20 percent of the variation in income is explained by the education controls.

An unobserved preference for spending time with one's children could lead parents to spend less time working and more time with their children, resulting in a negative correlation between income and time spent with children. Thus in some specifications we also include controls for usual hours of work of the respondent as well as the respondent's spouse. Also, to try to estimate the pure income effect of childcare, we report results for a sample limited to nonworking mothers.

We consider a quadratic in income. Table 2 shows that although most of the estimated income coefficients are statistically different from zero, all of them are minuscule in their economic impact. The third data column combines the estimated cross-sectional coefficients with the actual change in income from 1975 to 2008 to see how much of the increase in childcare could have been induced by an increase in real income. Every number in the column is a mere fraction of an hour. Particularly damaging to the income hypothesis is the fact that the pure income effects for nonworking mothers (regression 2-3) are very small. Thus, for the set of mothers who decide not to work, household income has a very small effect on time spent in childcare after controlling for the mother's education. The cross-sectional evidence therefore implies that rising incomes cannot explain the increase in childcare time.[9]

Not only are the measured effects of income on childcare time small, but the observed trends in childcare time do not match chronologically

9. We find similarly small coefficient estimates if we use earnings instead of total income, as Kimmel and Connelly (2007) did. In his comment on this paper, Erik Hurst reports a positive correlation between childcare time and GDP across countries. We suspect that this correlation is due to education differences rather than income differences.

Table 2. Regressions Explaining Childcare Time by Parental Income

Regression	*Regression coefficient*[a] *Income*[b]	*Income squared*	*Income-induced increase in childcare, 1975–2008 (hours per week)*[c]	*Sample*	*Usual hours of work included?*[d]
Mothers					
2-1	0.0105 (0.0050)	–0.000012 (0.000010)	0.22	All mothers	No
2-2	0.0313 (0.006)**	–0.000057 (0.000013)**	0.56	All mothers	Yes
2-3	0.0348 (0.010)**	–0.000066 (0.000022)**	0.61	Nonworking mothers only[e]	No
Fathers					
2-4	0.023 (0.005)**	–0.000045 (0.000011)**	0.40	All fathers	No
2-5	0.025 (0.006)**	–0.000049 (0.000012)**	0.44	All fathers	Yes

Source: Authors' regressions using pooled data from Bureau of Labor Statistics (BLS) time-use surveys from 2003 to 2008.

a. The dependent variable is hours per week spent by parents on childcare. All regressions include controls for survey year, parent's age (see table 1 for categories), full controls for children's ages (see table 1), number of children in family, parent's marital status, and parent's educational attainment (high school dropout, college degree, or advanced degree; the omitted category is high school graduate).

b. In thousands of 2008 dollars.

c. Calculated by applying the cross-sectional income coefficients to the actual change in average family income from 1975 to 2008.

d. "Yes" indicates that the regression includes variables for the usual hours worked per week by the respondent and by the spouse.

e. Sample excludes mothers who work for pay outside the home.

with published Census tabulations of trends in household income. We focus here on households with married parents. (The online data appendix discusses the data sources.) According to our time-use data, for both married mothers and fathers (of any education level), time spent on childcare was flat or slightly decreasing through 1985. It began to rise in the mid-1990s, increasing by almost 7 hours per week for mothers and 4.5 hours per week for fathers by the mid-2000s. In contrast, inflation-adjusted median household income for married parents grew at approximately the same annual rate from 1969 to 1990 as it did from 1990 to 2008, just over 1 percent per year. If income were the driving force, one would expect childcare time to have risen from 1965 to 1990. In fact, it did not.

In sum, neither the cross-sectional evidence nor the time-series evidence provides any support for the hypothesis that rising incomes can explain the observed trends in childcare time.

II.C. Safety Concerns

Bianchi and others (2006) suggest that heightened concerns about safety may have induced parents to accompany their children in their activities more often and to substitute structured activities for the free, unaccompanied play on neighborhood streets that was the norm in earlier times. This explanation is problematic for two reasons. The first is that the trends again do not align chronologically. Nationwide, the incidence of violent crime rose from 200 per 100,000 population in 1965 to a peak of 758 in 1991 and then began declining again (U.S. Census, *2010 Statistical Abstract: Historical Statistics,* table HS-23), reaching 467 in 2007. Thus the violent crime rate has moved *inversely* with time spent in childcare, which is contrary to the hypothesis of a positive crime-childcare link.

Of course, what matters is parents' *perceptions* of safety. However, the evidence suggests that today's parents worry less than parents 20 years ago did about numerous safety issues. Safe Kids USA (2008) reported the results of polls in 1987 and 2007 that asked parents about their major concerns in raising kids. Among the categories were concerns about children being involved in accidents, kidnapped by strangers, influenced adversely by friends, and exposed to street drugs. In every case, parents were significantly *less* concerned in 2007 than they were in 1987 (Safe Kids USA 2008, p. 9). Thus, trends in safety perceptions cannot be the source of the observed trends in childcare time.

A second reason that this explanation is problematic is that families of higher socioeconomic status tend to perceive the neighborhoods they live in to be safer (Wildon and others 2004). Thus, if the explanation suggested by Bianchi and others (2006) were important, one would expect educated parents to spend *less* time on childcare than less educated parents, which is inconsistent with the evidence.

II.D. Increasing Enjoyment of Childcare

A fourth possible explanation is that parents now experience greater enjoyment from childcare. However, empirical studies that have measured the enjoyment of various activities do not indicate rising enjoyment of most types of childcare. Robinson and Geoffrey Godbey (1999) report enjoyment ratings for various activities from the 1985 survey. In this survey, which covered both men and women, basic childcare ranked below work and cooking, but above housework. Alan Krueger and others (2008) report measures of enjoyment of various activities by women in 2005. According to their table 8.3, basic childcare ranked below both cooking and housework. Thus there is no evidence that basic childcare has become more pleasant. Additional evidence against the increased enjoyment hypothesis is the lack of an increase in the fertility rate or in family size: if today's parents enjoyed childcare much more than did parents in earlier years, one would expect them to choose to have more children.

One caveat is that playing with children has always ranked high in terms of enjoyment. We have followed the standard practice of including time spent in play in our measure of childcare time, because play is often considered crucial for investment in children's human capital. However, it might alter the interpretation of the results if one believes that the increase in childcare time is simply a redirection of time from one high-enjoyment activity to another.

To investigate this possibility, we reestimate the regressions for parents, this time excluding time spent playing with children from our measure of childcare time. Only parents' ages are used as controls. The first two columns of table 3 show the results for mothers. Column 3-1 reproduces column 1-1 of table 1, and column 3-2 reports the corresponding results using the restricted childcare variable. Omitting time spent playing with children reduces the increase in total childcare time by about 1 hour for less educated mothers and by about 3 hours for college-educated mothers. Nevertheless, most of the increase over time and across education levels remains. The results for fathers (columns 3-3 and 3-4) are similar.

II.E. More Flexible Work Schedules

Yet another possible explanation is that parents now have more flexible work schedules and can thus reallocate their time so as to spend more time with their children even while working. Unfortunately, we do not have measures of work schedule flexibility, either in our time-use data or in the aggregate. However, one implication of this hypothesis is that the biggest increases in childcare time should be among working mothers rather than stay-at-home mothers.

To test this implication, we split the sample by work status and rerun our basic regressions. We find that the increase in childcare time from 1975 to 2008 was 4.8 hours for less educated working mothers, 5.5 hours for less educated stay-at-home mothers, 7.1 hours for college-educated working mothers, and 16.3 hours for college-educated stay-at-home mothers. Thus the increase in childcare time over this period is greater for those

Table 3. Regressions Identifying Trends in Time Spent on Childcare Excluding Play[a]

	Mothers		*Fathers*	
Independent variable[b]	*Including play 3-1*[c]	*Excluding play 3-2*[d]	*Including play 3-3*[c]	*Excluding play 3-4*[d]
1965	1.636	1.919	0.290	–0.187
	(0.613)**	(0.538)**	(0.512)	(0.413)
1985	–0.369	–0.443	0.005	–0.144
	(0.689)	(0.602)	(0.583)	(0.471)
1992–94	–1.103	–1.247	0.210	–0.172
	(0.552)	(0.485)**	(0.516)	(0.417)
1995	1.744	0.770	1.232	0.441
	(0.883)*	(0.775)	(0.821)	(0.663)
1998	1.842	1.182	3.102	2.143
	(0.933)*	(0.819)	(0.856)**	(0.692)**
2000	3.928	3.482	4.522	3.439
	(0.640)**	(0.562)**	(0.579)**	(0.468)**
2003	4.676	3.584	3.184	2.242
	(0.398)**	(0.349)**	(0.340)**	(0.274)**
2004	4.071	3.013	3.444	2.307
	(0.435)**	(0.382)**	(0.367)**	(0.296)**
2005	3.992	3.115	3.327	2.472
	(0.436)**	(0.382)**	(0.372)**	(0.300)**
2006	4.324	2.879	3.104	1.982
	(0.443)**	(0.389)**	(0.375)**	(0.303)**
2007	4.227	2.993	3.277	2.165
	(0.452)**	(0.396)**	(0.377)**	(0.305)**
2008	4.288	3.182	4.44	2.850
	(0.450)**	(0.395)**	(0.382)**	(0.309)**
College	0.026	0.193	0.854	0.718
	(0.900)	(0.790)	(0.579)	(0.468)
1965 × college	2.048	1.124	0.093	0.007
	(2.174)	(1.908)	(1.229)	(0.993)
1985 × college	1.873	1.488	–0.321	–0.399
	(1.671)	(1.467)	(1.142)	(0.923)
1992–94 × college	1.373	0.452	–0.422	–0.434
	(1.264)	(1.110)	(0.972)	(0.785)
1995 × college	0.741	–0.021	1.799	1.387
	(2.153)	(1.890)	(1.642)	(1.327)
1998 × college	3.117	1.705	2.134	1.978
	(2.052)	(1.801)	(1.666)	(1.346)
2000 × college	4.868	3.840	–0.406	0.144
	(1.479)**	(1.298)**	(1.098)	(0.887)
2003 × college	4.999	3.788	1.913	1.344
	(1.015)**	(0.890)**	(0.688)**	(0.556)*
2004 × college	6.344	4.110	1.425	0.803
	(1.062)**	(0.932)**	(0.733)	(0.592)
2005 × college	6.038	4.183	2.514	1.055
	(1.073)**	(0.942)**	(0.746)**	(0.603)

(continued)

Table 3. Regressions Identifying Trends in Time Spent on Childcare Excluding Play[a] *(Continued)*

	Mothers		*Fathers*	
Independent variable[b]	*Including play 3-1*[c]	*Excluding play 3-2*[d]	*Including play 3-3*[c]	*Excluding play 3-4*[d]
2006 × college	4.109	2.968	2.296	1.625
	(1.072)**	(0.941)**	(0.741)**	(0.599)**
2007 × college	5.291	3.193	2.138	1.032
	(1.073)**	(0.942)**	(0.755)**	(0.611)
2008 × college	4.659	3.487	0.900	0.846
	(1.074)**	(0.942)**	(0.745)	(0.602)
Constant	11.656	9.787	3.565	2.206
	(0.342)**	(0.300)**	(0.300)**	(0.2143)**
No. of observations	24,342	24,342	17,806	17,806
Adjusted R^2	0.09	0.07	0.06	0.04

Source: Authors' regressions using data from various time-use surveys.

a. The dependent variable is hours per week spent by parents (defined as all persons aged 18–64, except students, living in households with a child under the age of 18) on total childcare (columns 3-1 and 3-3) or on childcare excluding time spent playing with children (columns 3-2 and 3-4). Standard errors are in parentheses. Asterisks indicate statistical significance at the *5 percent or the **1 percent level.

b. See table 1 for variable definitions.

c. Childcare measure includes all time spent on childcare, in hours per week; regression includes dummy variables for parent's age (same regression as reported in column 1-1 or column 1-3 of table 1).

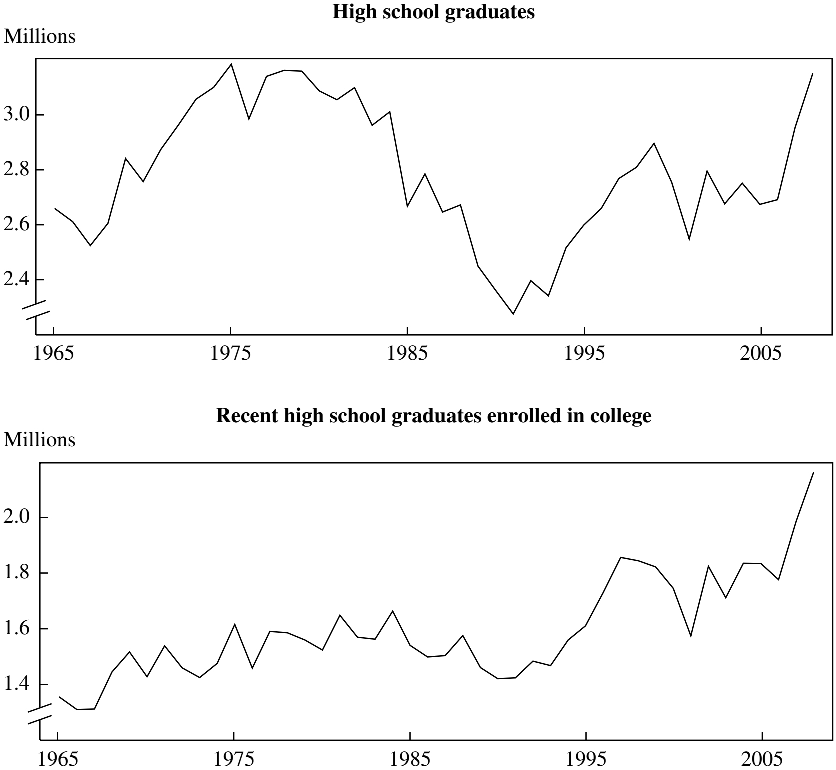
d. Same regression as reported in previous column except that the childcare measure excludes time spent playing with children.

mothers who do *not* work. One might worry about selection effects, since the labor force participation rate of mothers has increased since 1975, our usual comparison year. However, even when we instead compare with 1985, a year when the labor force participation rate for college-educated mothers was higher than in 1975, we find that the increase in childcare time among college-educated nonworking mothers was around 11 hours per week, compared with 6.8 hours per week for college-educated working mothers. Thus some other factor must have been at play to lead even nonworking mothers to increase their childcare time so much.

III. A New Explanation: The Rug Rat Race

We now offer a new explanation for the trends in time spent in childcare, one that is tied to increased competition for college admissions, particularly among the children of college-educated parents. Our claim is that a steep rise in demand for college admissions, together with a relatively con-

Figure 6. High School Graduates and College Enrollments, 1965–2008



Source: National Center for Education Statistics, *2009 Digest of Education Statistics,* table 200.

stant number of slots at the more attractive colleges, has resulted in "cohort crowding" for college admissions, which in turn has spurred competition among parents for attractive slots. This more intense competition manifests itself in parents spending more time preparing their children for college. We dub this expenditure of childcare time in dissipative rivalry the "rug rat race."

III.A. Shifts in Demand and Supply for College

Empirical trends in demand for college admissions display a remarkable agreement with trends in time spent on childcare. The top panel of figure 6 shows the number of new high school graduates each year since 1965, from the 2009 *Digest of Education Statistics.* The initial large hump is due to the baby-boom generation. The number of graduates fell to a low around 1990 but has since spiked upward as a result of the “baby boom echo.” Projections by the National Center for Education Statistics indicate that the number of high school graduates in the echo boom peaked in 2009.

The bottom panel of figure 6 traces the number of recent high school graduates enrolled in college. The pattern differs somewhat from that in the top panel because of the long-run upward trend in the propensity for high school graduates to go to college. As the figure reveals, after declining from 1980 to 1990, this number increased dramatically during the 1990s, fluctuated around a constant level, and then jumped again in 2007.

Meanwhile the supply of college admissions—the number of slots at the more attractive colleges—has not expanded commensurately with demand. Between 1990 and 2005, total enrollment in college by recent high school graduates increased by 30 percent, but the number of full-time-enrolled freshmen increased by less than 13 percent at the 10 elite universities of the “Ivy Plus” and by only 10 percent at the top 25 liberal arts colleges as ranked by *US News and World Report* in 2006.[10] Bound and Turner (2007) show that this “cohort crowding” extends to public institutions as well. Using variation in cohort size across states, they show that the elasticity of undergraduate enrollment with respect to the age-18 population is well below unity. According to table 4 of their paper, 2-year community colleges have the highest elasticity, 0.82, followed by nonflagship public universities with an elasticity of 0.56, and flagship public universities with an elasticity of only 0.2. This evidence indicates that the number of slots becomes much less responsive to enrollment pressure as the quality of the institution increases.

III.B. Evidence on Competition for College Slots

For the last several years, the popular press has been filled with stories of unprecedented competition for college. Some of the perceived increase in competition is simply a statistical mirage: the average student now applies to more colleges, both because of the increased ease of filling out applications and because of perceived greater uncertainty about getting into a given college. However, there is ample evidence that part of the increase in competition is real. Within the University of California (UC) system, mean grade point averages and standardized test scores of admitted students increased from 1994 to 2007 for seven of the eight campuses

10. These numbers are based on our calculations using data extracted from the Integrated Postsecondary Education Data System.

that admit undergraduates.[11] At UC Santa Barbara, which ranks fifth among the campuses in selectivity, the average GPA of admitted students rose by 0.3 point on a 4.0 scale. Bound and others (2009) document many other facets of the increase in competition. For example, they show that the test scores of entering students are higher now on average, particularly at the top-ranked schools. They also show that the percentage of high school graduates applying to a 4-year institution has increased over time, and that the probability of acceptance to a 4-year college for a student of given ability has declined significantly since 1982.[12]

Although many colleges still accept most applicants, there is evidence to suggest that parents and children today pay more attention to selective colleges. Caroline Hoxby (1997) documents that the market for higher education has changed from a collection of local markets to a nationally integrated market. Hoxby (2009) surveys the evidence and concludes that there are higher returns to attending a more selective college. According to the *New York Times,*[13] "The preoccupation with the top universities, once primarily a phenomenon in the Northeastern United States, has become a more countrywide obsession." Observers note that college admissions anxiety has spread to the Midwest and the Sun Belt; in the latter, enrollment in SAT/ACT preparation classes has grown more than seven times the overall national growth rate over the last 5 years.[14] The National Association for College Admission Counseling (2008, p. 18) documents that the 257 four-year colleges that accept fewer than 50 percent of applicants account for 18 percent of total full-time enrollment but receive 31 percent of all applications.

Numerous other disciplines, such as developmental psychology, pediatrics, and sociology, have drawn an explicit link between competition for

11. Our calculations are based on data available at www.universityofcalifornia.edu and exclude UC Merced, which opened in 2005.

12. Hoxby (2009) shows that in the aggregate, test scores of students admitted to U.S. colleges fell from the 1960s to the 1970s and 1980s but rose again in the last couple of decades. She also shows that the colleges with higher average test scores in the 1960s experienced an increase in the test scores of admitted students, whereas colleges with lower test scores in the 1960s experienced a decrease. These trends were noted earlier by the various studies surveyed by Davies and Hammack (2005). Based on this evidence, Hoxby argues that overall college selectivity has not increased. However, Hoxby bases her argument entirely on standardized test scores, as opposed to the controlled experiment run by Bound and others (2009). Nor does her analysis take into account the widespread belief that colleges now put greater emphasis on extracurricular activities.

13. Alan Finder, "Ivy League Admissions Crunch Brings New Cachet to Next Tier," *New York Times,* May 16, 2007, p. A14.

14. Justin Pope, "Admissions Anxiety Reaches New Regions," Associated Press, October 22, 2006.

college slots and the increase in time devoted to academics and extracurricular activities. For example, Suniya Luthar and Bronwyn Becker (2002) and Lareau (2003) argue that many middle- and upper-middle-class parents see building up their children's "after-school résumés" as absolutely necessary because of the competition for college admission. The American Academy of Pediatrics (Ginsburg 2007) cites the increase in competition in college admissions as a key reason for the decrease in free play time and increase in scheduled activities among children nationwide.[15]

Perhaps the most direct evidence in support of our hypothesis is presented by Hilary Levey. Her study asks, "What explains the increase in children's participation in activities outside of the home, structured and monitored by their parents, when family time is so scarce?" (Levey 2009, p. i). After 16 months of fieldwork involving 172 interviews of middle- and upper-middle-class parents, children, coaches, and teachers, Levey concludes that parents believe that extracurricular activities are essential for obtaining the credentials their children need to gain admittance to "good" colleges, which is seen as a necessary and sufficient condition for the children's future economic welfare. She specifically notes parents' perceptions of the increased competition to get into college and the "race towards college admissions" (Levey 2009, p. 11).

III.C. A Theoretical Model of the Rug Rat Race

The evidence presented above is consistent with the idea that an increase in rivalry for scarce college slots has induced parents to increase the time they spend preparing their children for college. In this section we develop a simple theoretical model that shows how shifts in the demand and supply for college, of the sort documented above, might account for the observed behavior of childcare hours through increased rivalry for college slots. The model is able to link the coincident S-shaped patterns of childcare hours and college demand documented in figures 1 and 6, respectively.

Our model posits that parents compete for college slots by investing in their children's college preparation. Each parent is assumed to have a single child. The parent has either a college degree or less education, represented by c and l, respectively. The numbers of college-educated and less educated parents are denoted by m_c and m_l, respectively, and $m = m_c + m_l$ is their total. Children's college attendance is restricted by the availability of slots. We assume that there are k_1 slots available at first-tier colleges, and

15. An ongoing debate among child development experts asks whether the dramatic increase in extracurricular activities helps or hurts children. See, for example, Rosenfeld, Wise, and Coles (2001) and Mahoney and others (2006).

k_2m slots available at second-tier colleges, where $k_1 + k_2m < m$. The demographic shifts depicted in figure 6 are modeled as an increase in the parameters m_c and m_l. When this occurs, college slots become scarcer overall, and the first-tier slots become relatively scarcer, in line with the evidence.

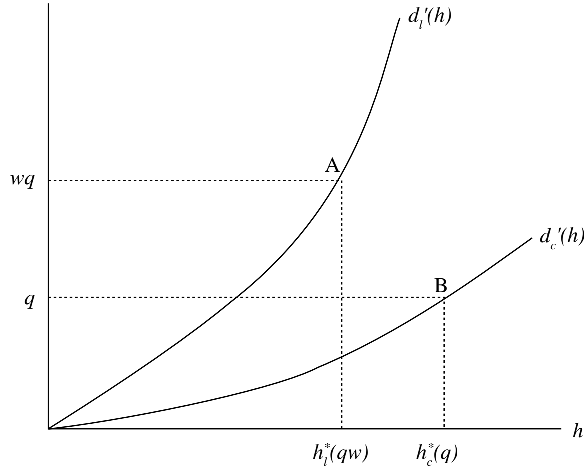
We assume that a child's preparation for college depends on her parent's time spent in childcare, denoted by h. The college admissions process operates as follows. Parents choose h, and colleges at the same time observe the value of h for each child. The colleges then fill their slots in descending order of h. This acceptance rule may be rationalized in a number of ways. For example, children may in later life contribute a proportion of their wealth, which increases in h, to their alma maters, and admissions decisions may be made so as to maximize total contributions. Since first-tier slots are most valuable, they will be filled first. In equilibrium, a threshold h_1 will exist such that children with $h \geq h_1$ are accepted to first-tier colleges, and there are exactly k_1 such children. The second-tier slots are filled next: there is a threshold h_2 such that children with $h \in [h_2, h_1)$, numbering k_2m, are accepted to second-tier colleges. The remaining $m - k_1 - k_2m$ children with $h < h_2$ do not attend college.

If a child goes to college, her ultimate wealth is given by wqh, where $w > 1$ reflects the college wage premium and q reflects the quality of the college attended by the child. The parameter q is meant to capture both pecuniary and psychic benefits from college attendance. For example, parents may value the prestige of sending their children to more elite institutions. Moreover, q may change over time across all quality levels, reflecting generalized changes in the value of a college education. Let q_1 and q_2 denote the quality parameters for first- and second-tier colleges, respectively, where $q_1 > q_2 > 0$. If a child does not attend college, then wealth is assumed to be q_0h, where $q_2 > q_0 > 0$.

Parents choose h so as to maximize their children's wealth net of their own disutility. A less educated parent incurs a disutility of $d_l(h)$ from choosing h, and a college-educated parent incurs $wd_c(h)$; note that a rise in w leads to greater disutility for the college-educated parents. The disutility functions satisfy, for $s = l, c$, $d_s', d_s'' > 0$, $d_s(0) = d_s'(0)$, and $d_s'(\infty) = \infty$. Moreover, we assume that college-educated parents incur lower marginal disutility in the absence of a wage premium, that is, $d_c'(h) < d_l'(h)$.[16]

16. Instead of assuming differences in marginal disutility across parental education levels, the model could specify college preparation as an increasing function of childcare hours, $p_s(h)$, $s = c, l$, where an hour spent by a college-educated parent has a higher productivity effect, so that $p_c(h) > p_l(h)$. This would yield the same comparative advantage for college-educated parents in preparing their children for college as in the specification we use.

Figure 7. Choosing Time Spent in College Preparation



Source: Authors' model described in the text.

The objective function of a less educated parent is $wqh - d_l(h)$. For a college of quality q, the unconstrained optimal level of preparation, $h_l^*(q)$, is determined by

$$d_l'\left[h_l^*(q)\right] = wq.$$

For a college-educated parent, the objective function is $wqh - wd_c(h)$, and the unconstrained optimal preparation level, $h_c^*(q)$, satisfies

$$d_c'\left[h_c^*(q)\right] = q.$$

The effect of a parent's schooling on optimal preparation time is considered in figure 7. For less educated parents the optimal decision occurs at point A, where the marginal return wq equals the marginal disutility d_l'. A parent's college education shifts the marginal disutility locus down to d_c'. This captures a *productivity effect* in preparing children for college. Countering this is an *opportunity cost effect,* whereby a given quantity of time commands a higher market wage. The marginal return to preparation, adjusted for opportunity cost, drops to q, and the optimal decision occurs at point B. We assume that the productivity effect dominates the opportunity cost effect, so that $h_c^*(q) > h_l^*(q)$ holds for every level of q. Thus college-educated parents have a comparative advantage in investing in college preparation.

We first consider the case in which m_c is small, in the sense that there are enough first-tier slots to accommodate all the children of college-educated parents. Proposition 1 of the theoretical appendix characterizes the outcome for the case where $m_c < k_1$. In equilibrium, college-educated parents exploit their comparative advantage in college preparation to get their children into first-tier colleges. The children of less educated parents take up the remaining first-tier slots along with all of the second-tier slots. The acceptance thresholds h_1 and h_2 make the latter parents just indifferent among first-tier, second-tier, and no college. The thresholds are distorted upward relative to the corresponding unconstrained optimal preparation levels $h_l^*(q_1)$ and $h_l^*(q_2)$, reflecting rivalry among these parents for scarce slots.

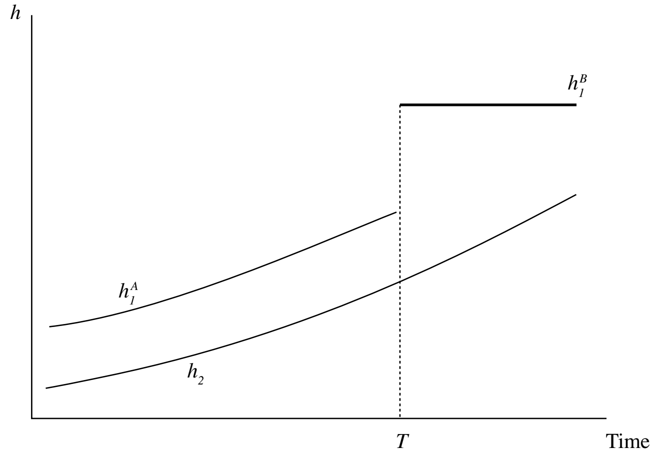
Now suppose that m_c rises to the point where there are too few first-tier slots for all the children of college-educated parents. As shown in proposition 2 of the theoretical appendix, once $m_c > k_1$, the focus of rivalry shifts from the less educated to the college-educated parents. The children of the less educated parents are driven completely from the first tier, as the acceptance threshold h_1 jumps to a level that makes the college-educated parents indifferent between the first and second tiers. This new level is distorted upward relative to unconstrained optimal preparation, $h_c^*(q_1)$. In this way, the change in the competition for college slots is directly linked to the increase in childcare hours among college-educated parents.

Figure 8 illustrates the time paths of college preparation choices when m_c and m_l rise gradually, with $m_c = k_1$ occurring at time T. We interpret T as corresponding to a point of time in the mid-1990s. The wage premium w is also assumed to rise over time, leading to gradually increasing paths of h_1 and h_2.[17] Up to time T, the growth of m_c gradually squeezes the children of less educated parents out of the first-tier colleges, shifting their parents' preparation choices from h_1 to h_2. The average level of h nevertheless increases if growth in w is sufficiently rapid. The preparation choices of college-educated parents also rise if $h_1 > h_c^*(q_1)$. At time T these parents jump to a discretely higher level of college preparation, while the choices of the less educated parents continue to rise with h_2.[18]

17. The increasing paths h_1 and h_2 could also be induced by a rise over time in the college quality parameters q_1 and q_2, due to greater psychic benefits from attending a quality college, for example.

18. The segment of the tier 1 acceptance threshold following time T is flat because we have assumed that changes in the wage premium have exactly offsetting effects on the costs and the benefits of preparation by college-educated parents. Thus the wage premium does not affect their preparation incentives.

Figure 8. College Preparation Choices over Time with Rising Cohort Size and Rising Wages



Source: Authors' model described in the text.

The model shows how rivalry for ever-scarcer slots can fuel a rug rat race among parents, where rivalry is manifested in higher college preparation requirements. Following a sharp increase in demand for college slots, rivalry among the college-educated parents intensifies greatly, driving up their time spent in childcare relative to that of the non-college-educated. This matches the coincident S-shaped patterns of childcare time and college demand seen in the U.S. data.[19]

III.D. A Comparison of Trends in the United States and Canada

Our theory links changes in childcare hours to rivalry for scarce college slots. This rivalry is tied in turn to the competitive admissions procedures used at U.S. colleges. The theory would predict a different path of childcare hours in a country such as Canada, where college admissions are determined in a much less rivalrous fashion. Thus, as a test of our theory, we compare trends in childcare time in the United States and Canada.

The comparison of these two countries is ideal for our purposes. Because of their geographic proximity and similarity of language and cul-

19. Akerlof (1976) introduced the first "rat race" model in economics. In his model, imperfect information causes workers to work faster in order to signal their underlying ability. In our model there is perfect information, but the scarcity of college slots causes parents to exert greater preparation effort in order to capture slots for their children. In other words, our model is based on rivalry rather than signaling.

ture, one would expect that changes in childcare time caused by such broad factors as knowledge diffusion and social fads would affect both countries similarly. In fact, the two countries differ significantly in the nature of the competition for admission to their universities and colleges.

Scott Davies and Floyd Hammack (2005) document the similarities and differences in the higher education systems of the United States and Canada. The countries are similar in that just over 60 percent of high school graduates in both countries pursue a postsecondary education. Both systems have decentralized governance, and both have experienced similar patterns of rising enrollment and increased scarcity of college slots.

However, as Davies and Hammack also document, the nature of the competition for college admission in the two countries is very different. They argue that whereas the Canadian system consists of formally equal public universities, the U.S. system is distinguished by a steep prestige hierarchy across colleges nationwide, leading to a distinctive form of competition that has increased over the last decade. In contrast, there is no national market for higher education in Canada; few Canadians go to college outside their home province. Thus it is not surprising that there is no Canadian equivalent to the SAT and that extracurricular activities are irrelevant for admission to Canadian colleges. Instead, competition in Canada occurs within the postsecondary system and takes the form of competition to enter lucrative majors. Davies and Hammack argue, "Where one studies is seen as more important in the U.S., while what one studies dominates in Canada." In fact, many Canadian college students who cannot get into their desired programs at 4-year colleges often transfer to community colleges in order to pursue their chosen field.

The lower competition to secure favored slots within a hierarchy of colleges suggests that there should be less pressure on educated Canadian parents to invest time in preparing their children to get into college as slots become scarcer. We test this prediction by studying trends in childcare time in Canada.

To this end, we use microdata on English-speaking Canadian parents from the 1986 (cycle 2), 1992 (cycle 7), 1998 (cycle 12), and 2005 (cycle 19) cycles of the Time Use Survey from Canada's General Social Survey (conducted by Statistics Canada) to construct trends in childcare time by parental education level.[20] However, there was a significant change in a def-

20. We use the Statistics Canada microdata file of the four cycles, which contain anonymized data. All computations on these microdata were prepared by the Nova Scotia Department of Finance; the responsibility for the use and interpretation of these data is entirely that of the authors.

inition between surveys: the two earlier surveys counted as childcare any care of children 18 years or younger, whereas in the two later surveys the cutoff age was 14. Using the standard childcare variables would therefore bias the trends downward over time. Instead we use the survey's measure of total family care (children and adults), a more consistent measure. The online data appendix shows that the change implied by this measure lies between a lower bound using the available childcare variable and an upper bound based on some imputations we made. Otherwise, our analysis of the data is similar to that for the U.S. data. We estimate the same baseline regressions as before, controlling for the age group of the parent and including interacted year and education variables.[21]

Table 4 and figure 9 show the results. Columns 4-1 and 4-3 of table 4 show the results with baseline controls, and columns 4-2 and 4-4 the results when controls for marital status and the age group of the youngest child are added to the regression. Figure 9 plots the results for the baseline regressions. We have normalized the hours to be zero in the base year in each country, 1985 in the United States and 1986 in Canada.

Figure 9 shows that time spent in care (childcare in the United States, family care in Canada) by less educated parents has increased by about the same amount in both countries since the mid-1980s: about 4 to 5 hours per week for mothers and about 3 hours per week for fathers. But whereas time spent by college-educated mothers rose by almost 9 hours per week in the United States, it rose by only 1 hour per week in Canada. For college-educated fathers, the increases were 6 and 2 hours per week, respectively. Thus, as our theory would predict, Canada did not experience the big increases in time spent on care among college-educated parents. In fact, the educational gap grew in the United States but shrank in Canada.

Our findings are broadly consistent with other analyses of trends in the Canadian data. Gilles Pronovost (2007) finds that the amount of time that parents spent in the presence of their children fell from 1986 to 2005. Similarly, Martin Turcotte (2007) finds a decrease in time spent with family members over the period 1986–2005.

III.E. Comparisons across Ethnic Groups in the United States

As an additional test of our theory, we compare childcare time across ethnic groups within the United States. One group that should not have felt as much increased competition to enter college is minorities who are

21. Because of data limitations in some years, the youngest age category includes those aged 20–24 rather than 18–24 as in the U.S. data.

Table 4. Regressions Identifying Trends in Time Spent on Family Care in Canada[a]

	Mothers		Fathers	
Independent variable[b]	*Age controls only 4-1[c]*	*Full set of controls 4-2[d]*	*Age controls only 4-3[c]*	*Full set of controls 4-4[d]*
1992	2.235	1.967	1.258	1.020
	(0.606)**	(0.561)**	(0.509)*	(0.491)*
1998	3.652	3.404	3.183	3.134
	(0.5491)**	(0.547)**	(0.508)**	(0.491)**
2005	5.050	4.894	2.453	2.476
	(0.533)**	(0.501)**	(0.442)**	(0.435)**
College	7.307	3.037	2.736	1.758
	(1.151)**	(1.068)**	(0.841)**	(0.807)*
1992 × college	–3.669	–1.346	–0.712	–0.397
	(1.585)**	(1.463)	(1.180)	(1.130)
1998 × college	–3.141	–2.099	–0.784	–0.993
	(1.536)*	(1.417)	(1.117)	(1.070)
2005 × college	–4.229	–2.091	–0.553	–0.278
	(1.307)**	(1.206)	(0.974)	(0.933)
Constant	14.209	–2.836	6.508	–4.215
	(0.464)**	(1.110)**	(0.407)**	(1.155)**
Controls for marital status and age of youngest child?	No	Yes	No	Yes
No. of observations	6,548	6,548	4,671	4,671
Adjusted R^2	0.12	0.25	0.06	0.14

Source: Authors' regressions using data from the General Social Survey of Canada.

a. The dependent variable is hours per week spent by parents (defined as all persons aged 20–64 living in a home with his or her own child under the age of 19) on family care (care of all family members, children and adults). Standard errors are in parentheses. Asterisks indicate statistical significance at the *5 percent or the **1 percent level.

b. See table 1 for variable definitions. The omitted year is 1986.

c. Regressions include a control for the parent's age range (equals 1 if the parent is aged 20–24, 35–44, 45–54, or 55–64; the omitted category is age 25–34).

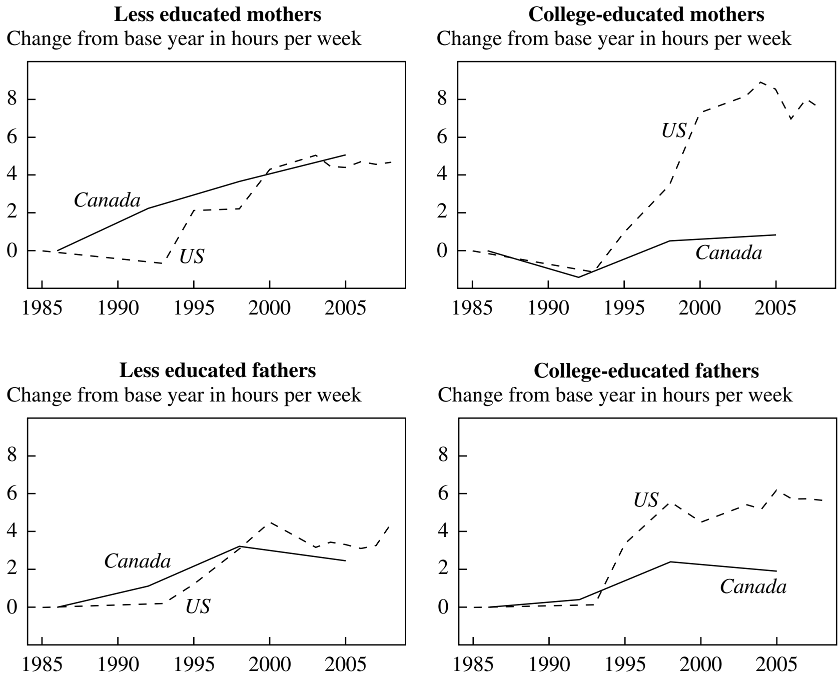
d. Regressions include, in addition to the age controls, a dummy for the parent's marital status and a dummy for the age range of the youngest child (4 or younger, 5–9, or 10–14; the omitted category is age 15–18).

underrepresented in U.S. colleges.[22] Despite the overall increase in applicants, most colleges are still eager to admit underrepresented minorities.

Unfortunately, in this comparison we cannot compare trends over time because the samples before 2003 are too small to allow a meaningful distinction across racial groups, particularly by education level. Instead, we make a cross-sectional comparison using pooled 2003–08 data and estimate regressions separately on less educated and college-educated groups.

22. We are indebted to Daniel Hamermesh for suggesting this test to us.

Figure 9. Cumulative Changes in Time Spent on Care in Canada and the United States, 1985–2008[a]



Source: Authors' estimates from regression results reported in table 1, columns 1-1 and 1-3, and table 4, columns 4-1 and 4-3.

a. Care is childcare in the United States, total family care (care of children and adults) in Canada. Hours are normalized to zero for each group in the initial year, 1985 for the United States and 1986 for Canada.

We include the full set of controls as well as controls for parents who are high school dropouts (within the less educated group) and parents with graduate degrees (within the college-educated group), because of potential differences across racial groups.

Table 5 shows that among less educated mothers, black and Hispanic mothers spend about 3 hours less per week in childcare than other mothers. Among college-educated mothers, black mothers spend 3 hours less and Hispanic mothers about 2 hours less than other mothers. Among less educated fathers, black and Hispanic fathers spend 1.4 to 1.9 hours less than other fathers. Among college-educated fathers, black fathers spend half an hour less and Hispanic fathers 2 hours less. These results are consistent with our hypothesis that underrepresented minorities feel less pressure to compete for college slots and hence spend less time in childcare.

Table 5. Differences in Childcare Time Spent by Minority and Nonminority Parents[a]

	Less educated		*College-educated*	
	Blacks	*Hispanics*	*Blacks*	*Hispanics*
Mothers				
Difference in hours[b]	−3.090	−2.893	−3.182	−1.768
	(0.375)**	(0.328)**	(0.705)**	(0.716)**
No. of observations	1,409	2,235	341	350
Fathers				
Difference in hours[b]	−1.404	−1.890	−0.478	−2.272
	(0.392)**	(0.314)**	(0.671)	(0.647)**
No. of observations	571	1,409	184	225

Source: Authors' regressions based on pooled data from BLS time-use surveys from 2003 to 2008.

a. Regressions are estimated separately on samples of less-than-college-educated and college-educated mothers and fathers. All regressions include controls for parent's age category and full controls for children's age category, number of children in family, parent's marital status, whether the parent is a high school dropout, and whether the parent has an advanced degree.

b. Compared with nonblack, non-Hispanic mothers or fathers with the same educational attainment.

III.F. Correlations with Measures of Admissions Competition across U.S. States

Although national integration of the U.S. college market has increased, regional factors still play an important role in determining competition for college. Bound and others (2009), using their new index of college competition by state, provide evidence that both levels of and trends in competition vary across states. As a further test of our theory, we can use this index to test whether greater college competition within a state is associated with more hours of childcare among college-educated parents who live in that state. Here we interpret our theory as applying to state-level college markets.

Ideally, we would compare trends in childcare by state over time to changes in the index over time. Unfortunately, sample sizes in the time-use surveys before 2003 are too small to provide information by state. Instead we make cross-state comparisons using the later surveys. To construct childcare time by state, we use pooled data from the 2003–08 surveys to estimate state-level childcare time measures by education and sex, using the full set of controls: parent's age, parent's marital status, and child number and age variables. The Bound and others (2009) index of competition by state is constructed by summing the following variables at the state level: the PSAT (Preliminary SAT) participation rate, the Advanced Placement examination participation rates, the fraction of students reporting

Table 6. Regressions Explaining Interstate Differences in Childcare Time, by Intensity of Competition for College Admissions[a]

	Mothers		*Fathers*	
	Less than college-educated	*College-educated*	*Less than college-educated*	*College-educated*
Coefficient on college competitiveness index[b]	1.504 (0.747)**	1.987 (0.825)**	−0.658 (0.563)	0.940 (0.682)
Adjusted R^2	0.085	0.117	0.030	0.041

Source: Authors' regressions.

a. Average time spent in childcare is estimated for each state by combining data from the 2003–08 American Time Use Surveys. Each estimate is a residual after controlling for parents' ages (dummies for ages 18–24, 25–34, 35–44, 45–54, and 55–64), parents' marital status, a quadratic in the number of children, and dummies for the age range of the youngest child in the family (age 1 year or less, 2, 3–5, 6–9, and 10–14). All regressions have 46 observations. Standard errors are in parentheses. Asterisks indicate statistical significance at the *5 percent and **1 percent level.

b. The index, from Bound, Hershbein, and Long (2009), is calculated for each state as the sum of the PSAT participation rate, the Advanced Placement participation rate, the fraction of students reporting 10 or more homework hours per week, the fraction applying to five or more colleges, and the fraction using private test preparation in 2004.

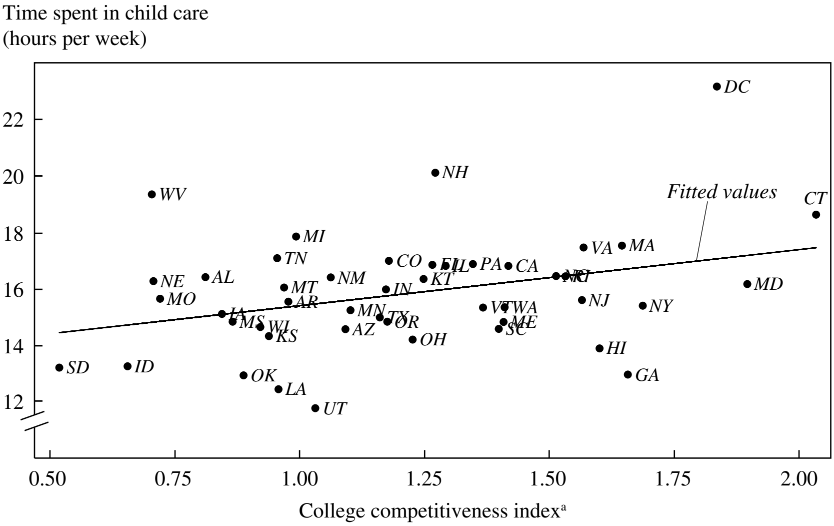
10 or more homework hours per week, the fraction applying to five or more colleges, and the fraction using private test preparation.

Table 6 shows the results of regressing state childcare time on Bound and others' index. Three of the four coefficients suggest a positive correlation, and the coefficients for both less educated and college-educated mothers are significantly positive, consistent with our hypothesis. Figure 10 shows a scatterplot of states on the two measures. South Dakota had the lowest index of competition for college in 2004 and Connecticut the highest. College-educated mothers spend an average of 13.2 hours per week on childcare in South Dakota (after controlling for family size and age characteristics) and 18.6 hours in Connecticut. The regression coefficient on the Bound and others index indicates that 3 hours of this difference may be related to differences in college competition. Thus for mothers there is evidence that greater competition at the state level is associated with greater time spent on childcare.

III.G. Discussion

Our theoretical model emphasizes the role of college preparation as a motive for increased childcare. According to the model, parents perceive a link between attendance at a good college and the accumulation of valuable human capital, and they exert effort, in the form of time spent in childcare, to influence their children's chances of admission to a good

Figure 10. Competition for College Admissions and Childcare Time Spent by College-Educated Mothers, by State



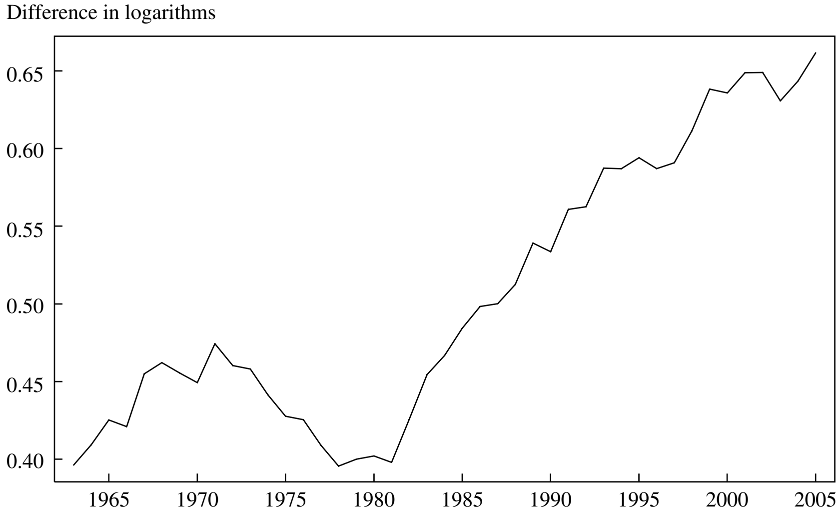
Source: Bound and others (2009), various time-use surveys, and authors' regression.
a. The index is defined in table 6, note b.

college. As we have shown, numerous studies from other disciplines have made the link between increased parental effort and the desire to get one's children into good colleges. The evidence we have presented is consistent with the idea that greater scarcity of desirable college slots has induced increases in childcare effort, especially by college-educated parents, who possess a comparative advantage in college preparation.

One question that arises is whether this factor alone can explain the magnitude of the increase in childcare time. It is possible that the rise in competition for college slots, triggered by the increase in the college-eligible population, provided the underlying impulse, but that other forces worked to amplify it. For example, if a subset of parents in a neighborhood get caught up in the "rug rat race" and enroll their children in multiple organized activities, other families with less concern about college competition might follow suit simply because there are fewer children available for unstructured play in the neighborhood. One could also envision models with "social contagion" that serve to amplify the effects.

From a broader perspective, any factors that alter the returns to college may enter parents' calculations and thus have the potential to affect childcare decisions. Changes in the college wage premium, in particular, may

Figure 11. College–High School Wage Gap, 1963–2005



Source: Data from Autor, Katz, and Kearney (2008).

have an important influence on college preparation incentives. Figure 11 traces the college wage premium, measured as the log difference in wages between college and high school graduates over the last half century (the data are from Autor, Katz, and Kearney 2008). Although the wage premium has risen steadily since the early 1980s, the rate of increase has slowed: the average annual change fell from 1.3 percent (measured in log points) over 1980–90 to 0.85 percent over 1990–2005. The path of the wage premium does not closely fit the S-shaped pattern of childcare hours for any of the four groups depicted in figure 1, even after adjusting for possible downward bias in the 1992–94 time-use survey. It seems very difficult to rationalize the sharp upward movement in the childcare hours of college-educated parents beginning in the mid-1990s as a response to the much smoother secular upward trend of the college wage premium.

The Canadian evidence provides further perspective on the role of the wage premium. Brahim Boudarbat, Thomás Lemieux, and Craig Riddell (2006) show that in Canada between 1980 and 2000 the wage premium of a college graduate relative to that of a high school graduate rose by about 10 percentage points for men and 6 percentage points for women. This contrasts with an average rise over both sexes of 25 percentage points in the United States over the same period (Autor and others 2008). Thus the returns to college have increased much more in the United States. However, as figure 9 showed, childcare hours for less educated parents display very similar behavior across the two countries over this period. Moreover, these parents ought to be more sensitive to changes in the wage premium, since college-educated parents experience a relatively greater increase in the opportunity cost of childcare time as the wage premium rises. In summary, the evidence does not point to changes in the college wage premium as an important factor in the behavior of childcare hours.

Nonpecuniary benefits of college, such as prestige or general "well-roundedness," may also have important effects on parents' calculations. The increased focus on selective colleges may reflect changes in societal attitudes that have raised the relative demand for admission to prestigious institutions. This motivation is complementary to our rivalry theory: not only are more children chasing a relatively constant supply of desirable college slots, but these slots may have themselves become relatively more desirable.

It is possible to imagine a plethora of alternative theories based on various parental motivations for investing time in children in order to increase their general human capital. To be persuasive, such theories must be capable of accounting for the key aspects of the evidence that we highlight. The first is the timing: we have shown that weekly hours spent in childcare in the United States have followed a pronounced S-shaped pattern, with almost all of the growth concentrated in a 10-year period beginning in the mid-1990s. The second is the composition: the increase in hours is much greater for college-educated parents and consists chiefly of increases in time spent caring for older children, and in travel and activities—the categories that relate most directly to the college application process. The third is the difference between the United States and Canada: we do not observe an increase in childcare time among college-educated parents in Canada, despite the similarity of the two countries' economic and cultural environments. Theories that rely on forces that unfold broadly and gradually will not easily explain this evidence.[23]

IV. Conclusion

This paper has documented a dramatic increase in time spent in childcare by college-educated parents since the mid-1990s. Although time spent in

23. For example, Stevenson (forthcoming) shows that participation in sports raises educational attainment and wages, but we are not aware of any evidence suggesting that the return to sports or other organized activities has increased over time.

childcare rose for all parents, the rise was far more pronounced for college-educated parents. Since the mid-1990s, less educated mothers have reallocated over 4 hours per week to childcare, but college-educated mothers have reallocated more than 9 hours per week. This reallocation occurred at the same time that competition to get into college intensified, as demographic forces led to a surge in the demand for college slots. In contrast, time spent in childcare by educated parents in Canada, where college competition is much lower, changed very little over this period.

We have explained these trends using a model in which the rise in time devoted to childcare is the optimal response to the increase in rivalry for scarce college slots. We postulate that college-educated parents have a comparative advantage in preparing their children for college, which they exploit to get their children into the most attractive colleges. When slots are plentiful relative to demand, the required amount of child preparation is relatively low. However, when demand rises, rivalry among the college-educated parents drives the required preparation upward.

In this paper we have focused on explaining observed trends in time use, but our results also have implications for socially efficient time allocation. To the extent that the private costs and benefits of college preparation reflect social costs and benefits, the intense rivalry for college slots in recent years implies wasteful overinvestment in preparation. Such overinvestment might be mitigated by expanding the number of slots at attractive colleges or by modifying their acceptance rules to place greater emphasis on criteria that parents cannot directly influence. In a broader context, however, parents may not fully internalize the social benefits of preparing their children, which raises the possibility that the rug rat race provides a useful stimulus to human capital investment and thus more closely aligns the private and the social benefits. These issues warrant closer investigation in future work.

THEORETICAL APPENDIX

Proposition 1. If $m_c < k_1$, then the equilibrium acceptance threshold h_1 is uniquely determined by $h_1 = h_1^A > h_l^*(q_0)$ and

$$\text{(A1)} \qquad wq_1h_1^A - d_l\left(h_1^A\right) = h_l^*\left(q_0\right) - d_l\left[h_l^*\left(q_0\right)\right],$$

and the equilibrium threshold h_2 is uniquely determined by $h_2 > h_l^*(q_2)$ and

$$\text{(A2)} \qquad wq_2h_2 - d_l\left(h_2\right) = h_l^*\left(q_0\right) - d_l\left[h_l^*\left(q_0\right)\right].$$

Moreover, $h_1^A > h_2$, and

—college-educated parents choose $h = \max\{h_c^*(q_1), h_1^A\}$

—less educated parents divide themselves among $h = h_1^A$, $h = h_2$, and $h = h_l^*(q_0)$, where $h_l^*(q_0)$ is the optimal preparation choice when a child does not attend college.

Proof. Let $G_l(h|q) = wqh - d_l(h)$ and $G_c(h|q) = wqh - wd_c(h)$ represent the objective functions of less educated and college-educated mothers, respectively. Under our assumptions, these functions are strictly concave in h and $G_l[h_l^*(q_0)|q_0/w]$, there is a unique point $h_1^A > h_l^*(q_1)$ satisfying $G_l(h_1^A|q_1) = G_l[h_l^*(q_0)|q_0/w]$. Similarly, $G_l[h_l^*(q_2)|q_2] > G_l[h_l^*(q_0)|q_0/w]$ implies that there is a unique point $h_2 > h_l^*(q_2)$ satisfying $G_l(h_2|q_2) = G_l[h_l^*(q_0)|q_0/w]$. Furthermore, $G_l(h_2|q_1) > G_l(h_2|q_2) = G_l[h_l^*(q_0)|q_0/w]$ implies $h_2 < h_1^A$.

Consider the h choices of college-educated mothers when $h_c^*(q_1) \geq h_1^A$. Clearly, $h_c^*(q_1)$ is optimal among $h \geq h_1^A$. Moreover, for all $h \in [h_2, h_1^A)$, $G_c[h_c^*(q_1)|q_1] > G_c[h_c^*(q_2)|q_2] \geq G_c(h|q_2)$, and for all $h < h_2$, $G_c[h_c^*(q_1)|q_1] > G_c[h_c^*(q_0)|q_0/w] \geq G_c(h|q_0/w)$. Thus $h_c^*(q_1)$ is strictly preferred to any other h.

Next suppose $h_1^A > h_c^*(q_1)$. Let $\hat{h} = \max\{h_c^*(q_2), h\}$. Note that $h_c^*(q_2) < h_c^*(q_1)$ and $h_2 < h_1^A$ imply $\hat{h} < h_1^A$. Moreover, $\hat{h} \geq h_2$ implies $G_l(\hat{h} \mid q_2) \leq G_l[h_l^*(q_0)|q_0/w]$. Thus,

$$0 \geq G_l\left(\hat{h}|q_2\right) - G_l\left[h_l^*\left(q_0\right)|q_0/w\right] = G_l\left(\hat{h}|q_2\right) - G_l\left(h_1^A|q_1\right)$$

$$= wq_2\hat{h} - wq_1h_1^A + \int_{\hat{h}}^{h_1^A} d_l'\left(h\right)dh > wq_2\hat{h} - wq_1h_1^A + \int_{\hat{h}}^{h_1^A} wd_c'\left(h\right)dh$$

$$= G_c\left(\hat{h}|q_2\right) - G_c\left(h_1^A|q_1\right),$$

where the strict inequality comes from the fact that $h_c^*(q) > h_l^*(q)$ for all q implies $wd_c'(q) < d_l'(q)$ for all q. Thus $G_c(h_1^A|q_1) > G_c(\hat{h}|q_2)$, and it follows that $G_c(h_1^A \mid q_2) > G_c(h \mid q_2)$ for all $h \in [h_2, h_1^A)$, since $\hat{h}$ maximizes $G_c(h|q_2)$ over this range of h. Finally, consider $h < h_2$. If $\hat{h} = h_c^*(q_2)$, then $G_c(\hat{h} \mid q_2) > G_c[h_c^*(q_0) \mid q_0/w] \geq G_c(h \mid q_0/w)$ for all $h < h_2$, whereas $\hat{h} = h_2$ implies $\hat{h} > h_c^*(q_0)$, and

$$0 = G_l\left[h_l^*\left(q_0\right)|q_0\right] - G_l\left(\hat{h}|q_2\right) > G_l\left[h_c^*\left(q_0\right)|q_2\right] - G_l\left(\hat{h}|q_2\right)$$

$$= h_c^*\left(q_0\right) - wq_2\hat{h} + \int_{h_c^*(q_0)}^{\hat{h}} d_l'\left(h\right)dh > h_c^*\left(q_0\right) - wq_2\hat{h} + \int_{h_c^*(q_0)}^{\hat{h}} wd_c'\left(h\right)dh$$

$$= G_c\left[h_c^*\left(q_0\right)|q_0/w\right] - G_c\left(\hat{h}|q_2\right),$$

so that $G_c(\hat{h}|q_2) > G_c(h|q_0/w)$ for all $h < h_2$.

Now consider the h choices of the less educated mothers. Because of strict concavity and because $h_1^A > h_l^*(q_1)$, $G_l(h|q_1) < G_l(h_1^A|q_1)$ for all $h > h_1^A$. Similarly, $G_l(h\,|\,q_2) < G_l(h_2\,|\,q_2)$ for all $h \in [h_2, h_1^A)$. Since $h_l^*(q_0) < h_2$ and $h_l^*(q_0)$ maximizes $G_l(h|q_0/w)$, it follows that $G_l(h\,|q_0) < G_l[h_l^*(q_0)\,|q_0/w]$ for all $h < h_2$, $h \neq h_l^*(q_0)$. Thus, the choices h_1^A, h_2, and $h_l^*(q_0)$ are strictly preferred to any others, and by construction these three are equally preferred. Q.E.D.

Proposition 2. If $k_1 < m_c$, then the equilibrium acceptance threshold h_1 is uniquely determined by $h_1 = h_1^B > h_c^*(q_1)$ and

$$\text{(A3)} \qquad wq_1h_1^B - wd_c\left(h_1^B\right) = q_0\hat{h} - wq_0d_c\left(\hat{h}\right),$$

where $\hat{h} = \max\{h_c^*(q_0), h_2\}$, and h_2 is determined as in proposition 1. Moreover, $h_1^B > h_1^A$, and

—college-educated parents divide themselves between $h = h_1^B$ and $h = \hat{h}$, and

—less educated parents divide themselves between $h = h_2$ and $h = h_l^*(q_0)$.

Proof. Equation A3 can be expressed as $G_c(h_1^B|\,q_1) = G_c(\hat{h}\,|\,q_2)$. Moreover, $G_c[h_c^*(q_1)\,|\,q_1] > G_c[h_c^*(q_2)\,|\,q_2] \geq G(\hat{h}\,|\,q_2)$. Since $h_1^B > h_c^*(q_1)$, it follows that h_1^B is uniquely defined, and $G_c(\hat{h}|q_1) > G_c(\hat{h}|q_2)$ implies $h_1^B > \hat{h}$.

We now verify that $h_1^B > h_1^A$:

$$0 = G_c\left(\hat{h}|q_2\right) - G_c\left(h_1^B|q_1\right) = wq_2\hat{h} - wq_1h_1^B + \int_{\hat{h}}^{h_1^B} wd_c'\left(h\right)dh$$

$$< wq_2\hat{h} - wq_1h_1^B + \int_{\hat{h}}^{h_1^B} d_l'\left(h\right)dh = G_l\left(\hat{h}|q_2\right) - G_l\left(h_1^B|q_1\right).$$

This implies $G_l(h_1^B|q_1) < G_l(\hat{h}|q_2) \leq G_l(h_2\,|q_2)$, and comparison with equations A1 and A2 shows that $h_1^B > h_1^A$.

Consider the h choices of college-educated mothers. Since $h_1^B > h_c^*(q_1)$, $G_c(h\,|q_1) < G_c(h_1^B|q_1)$ for all $h > h_1^B$. If $h_c^*(q_2) \geq h_2$, then $h_c^*(q_2)$ is strictly preferred to any other $h \in [h_2, h_1^B)$, whereas if $h_2 > h_c^*(q_2)$, then $G_c(h\,|\,q_2) < G_c(h_2|q_2)$ for all $h \in (h_2, h_1^B)$. Thus $G_c(\hat{h}\,|\,q_2)$ maximizes $G_c(h\,|\,q_2)$ over $h \in [h_2, h_1^B)$. The argument from the proof of proposition 1 shows that $G_c(\hat{h}|q_2) > G_c(h|q_0/w)$ for all $h < h_2$. Thus, the choices h_1^B and $\hat{h}$ are strictly preferred to any others, and they are equally preferred by construction.

For the less educated mothers, $h \geq h_1^B$ implies $G_l(h\,|\,q_1) \leq G_l(h_1^B\,|\,q_1) < G_l(h_2|q_2)$, so h_2 is strictly preferred over any such h. The arguments from the proof of proposition 1 establish that h_2 and $h_l^*(q_0)$ are strictly preferred to any other $h < h_1^B$, and they are equally preferred by construction. Q.E.D.

ACKNOWLEDGMENTS Valerie Ramey gratefully acknowledges financial support from National Science Foundation grant SES-0617219 through the National Bureau of Economic Research. We thank Kimberly Fisher for help with the AHTUS, Tarek Harchaoui of Statistics Canada for helping us gain access to microdata files on Canadian time use, and John Robinson for providing his 1995 U.S. time-use data. Brad Hershbein kindly provided the 2004 index of state-by-state college competitiveness. We thank Julie Cullen, Nora Gordon, Erik Hurst, Daniel Hamermesh, Daniel Sacks, David Schkade, Joel Sobel, Betsey Stephenson, Randall Wright, the editors, and participants at the Brookings Panel conference and numerous seminars for helpful comments. Christopher Nekarda and Ben Backes provided excellent research assistance. Conclusions drawn from the use of AHTUS and other data are the authors' own and may not reflect the views of the creators or funders of AHTUS or the collectors of the original surveys harmonized in the AHTUS dataset.

References

Aguiar, Mark, and Erik Hurst. 2007. "Measuring Trends in Leisure: The Allocation of Time over Five Decades." *Quarterly Journal of Economics* 122, no. 3: 969–1006.

Akerlof, George. 1976. "The Economics of Caste and of the Rat Race and Other Woeful Tales." *Quarterly Journal of Economics* 90, no. 4 (November): 599–617.

Allard, Mary Dorinda, Suzanne Bianchi, Jay Stewart, and Vanessa R. Wight. 2007. "Comparing Childcare Measures in the ATUS and Earlier Time-Diary Studies." *Monthly Labor Review* 130, no. 5 (May): 27–36.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *Review of Economics and Statistics* 90, no. 2 (May): 300–23.

Bianchi, Suzanne M. 2000. "Maternal Employment and Time with Children: Dramatic Change or Surprising Continuity?" *Demography* 37, no. 4 (November): 401–14.

Bianchi, Suzanne, and John Robinson. 2005. "National Survey of Parents, 2000–2001." ICPSR04247-v1 (computer file). Ann Arbor, Mich.: Inter-University Consortium for Political and Social Research.

Bianchi, Suzanne, Philip N. Cohen, Sara Raley, and Kei Nomaguchi. 2004. "Inequality in Parental Investment in Child-Rearing: Expenditures, Time, and Health." In *Social Inequality,* ed. by Kathryn M. Neckerman. New York: Russell Sage Foundation.

Bianchi, Suzanne M., John P. Robinson, and Melissa A. Milkie. 2006. *Changing Rhythms of American Family Life.* New York: Russell Sage Foundation.

Boudarbat, Brahim, Thomás Lemieux, and W. Craig Riddell. 2006. "Recent Trends in Wage Inequality and the Wage Structure in Canada." In *Dimensions of Inequality in Canada,* edited by David A. Green and Jonathan R. Kesselman. UBC Press.

Bound, John, and Sarah Turner. 2007. "Cohort Crowding: How Resources Affect Collegiate Attainment." *Journal of Public Economics* 91, nos. 5–6: 877–99.

Bound, John, Brad Hershbein, and Bridget Terry Long. 2009. "Playing the Admissions Game: Student Reactions to Increasing College Competition." *Journal of Economic Perspectives* 23, no. 4 (Fall): 119–46.

Bryant, W. Keith, and Cathleen D. Zick. 1996. "Are We Investing Less in the Next Generation? Historical Trends in Time Spent Caring for Children." *Journal of Family and Economic Issues* 17, nos. 3–4: 365–92.

Bureau of Labor Statistics. 2010. *American Time Use Survey, 2003–2008.* Washington.

Chalasani, Satvika. 2007. "The Changing Relationship between Parents' Education and Their Time with Children." *International Journal of Time Use Research* 4, no. 1 (September): 93–117.

Davies, Scott, and Floyd M. Hammack. 2005. "The Channeling of Student Competition in Higher Education: Comparing Canada and the U.S." *Journal of Higher Education* 76, no. 1 (January/February): 89–106.

Fisher, Kimberly, Muriel Egerton, Nuno Torres, Andreas Pollmann, and Jonathan Gershuny, with contributions by John Robinson and Anne Gauthier. 2006. *American Heritage Time Use Survey.* University of Essex. www.timeuse.org/ahtus/.

Ginsburg, Kenneth R. 2007. "The Importance of Play in Promoting Healthy Child Development and Maintaining Strong Parent-Child Bonds." *Pediatrics* 119, no.1: 182–91.

Guryan, Jonathan, Erik Hurst, and Melissa Kearney. 2008. "Parental Education and Parental Time with Children." *Journal of Economic Perspectives* 22, no. 3 (Summer): 23–46.

Hofferth, Sandra L. 2009. "Changes in American Children's Time, 1997 to 2003." *International Journal of Time Use Research* 6, no. 1: 26–47.

Hoxby, Caroline M. 1997. "How the Changing Market Structure of U.S. Higher Education Explains College Tuition." Working Paper no. 6323. Cambridge, Mass.: National Bureau of Economic Research (December).

———. 2009. "The Changing Selectivity of American Colleges." *Journal of Economic Perspectives* 23, no. 4 (Fall): 95–118.

Kimmel, Jean, and Rachel Connelly. 2007. "Mothers' Time Choices: Caregiving, Leisure, Home Production, and Paid Work." *Journal of Human Resources* (Summer): 643–81.

Krueger, Alan B., Daniel Kahneman, David Schkade, Norbert Schwarz, and Arthur A. Stone. 2008. "National Time Accounting: The Currency of Life." Princeton Industrial Relations Section Working Paper no. 523. Princeton University.

Lareau, Annette. 2003. *Unequal Childhoods: Class, Race, and Family Life.* University of California Press.

Leibowitz, Arleen. 1974. "Home Investments in Children." *Journal of Political Economy,* Part 2, 82, no. 2: S111–31.

Levey, Hilary Leigh. 2009. "Playing to Win: Raising Children in a Competitive Culture." Dissertation. Princeton University (September).

Luthar, Suniya S., and Bronwyn E. Becker. 2002. "Privileged but Pressured?: A Study of Affluent Youth." *Child Development* 73, no. 5: 1593–1610.

Mahoney, Joseph L., Angel L. Harris, and Jacquelynne S. Eccles. 2006. "Organized Activity Participation, Positive Youth Development, and the Over-Scheduling Hypothesis." *Social Policy Report* 20, no. 4: 3–31.

National Association for College Admission Counseling. 2008. *State of College Admission 2008.* Arlington, Va.

Pronovost, Gilles. 2007. "Le Temps dans Tous Ses États: Temps de Travail, Temps de Loisir et Temps pour la Famille à l'Aube du XXIe Siècle." Enjeux Publics IRPP vol. 8, no. 1. Montreal: Institute for Research on Public Policy (February).

Robinson, John P., and Geoffrey Godbey. 1999. *Time for Life: The Surprising Ways Americans Use Their Time,* 2nd ed. Pennsylvania State University Press.

Robinson, John, Suzanne Bianchi, and Stanley Presser. 2001. "Family Interaction, Social Capital, and Trends in Time Use (FISCT), 1998–1999: [United States]."

Computer file, ICPSR version. Ann Arbor, Mich.: Inter-University Consortium for Political and Social Research.

Rosenfeld, Alvin, Nicole Wise, and Robert Coles. 2001. *The Over-Scheduled Child: Avoiding the Hyper-Parenting Trap.* New York: St. Martin's Griffin.

Safe Kids USA. 2008. "Report to the Nation: Trends in Unintentional Childhood Injury Mortality and Parental Views on Child Safety." April 2008. sk.convio.net/site/DocServer/STUDY-SKWK-08-LRApril26.pdf?docID=9561.

Sandberg, John F., and Sandra L. Hofferth. 2005. "Changes in Children's Time with Parents: A Correction." *Demography* 42, no. 2 (May): 391–95.

Sayer, Liana C., Suzanne M. Bianchi, and John P. Robinson. 2004. "Are Parents Investing Less in Children? Trends in Mothers' and Fathers' Time with Children." *American Journal of Sociology* 110, no. 1 (July): 1–43.

Stevenson, Betsey. Forthcoming. "Beyond the Classroom: Using Title IX to Measure the Return to High School Sports." *Review of Economics and Statistics.*

Turcotte, Martin. 2007. "Time Spent with Family during a Typical Workday, 1986–2005." Canadian Social Trends no. 83. Statistics Canada.

U.S. Department of Agriculture, Agricultural Research Administration, Bureau of Human Nutrition and Home Economics. 1944. *The Time Costs of Homemaking—A Study of 1,500 Rural and Urban Households.* Washington.

Warner, Judith. 2005. *Perfect Madness: Motherhood in the Age of Anxiety.* New York: Riverhead Books.

Wildon, Dawn K., Karen A. Kirtland, Barbara E. Ainsworth, and Cheryl L. Addy. 2004. "Socioeconomic Status and Perceptions of Access and Safety for Physical Activity." *Annals of Behavioral Medicine* 28, no. 1: 20–28.

Comments and Discussion

COMMENT BY

ERIK HURST The introduction of the American Time Use Survey (ATUS) has led to a resurgence in research by economists interested in understanding how Americans allocate their time. By linking the ATUS with time-use surveys from earlier periods, researchers have also been exploring the changing nature of time use during the last few decades. Given that time is an important input into market work, nonmarket work, child development, and leisure, understanding the changing nature of time use allows researchers to better understand household preferences, home production technologies, and the changing nature of well-being.

This paper by Garey Ramey and Valerie Ramey tackles the important question of how parents have changed the amount of time they spend in childcare activities, with particular attention paid to the changing patterns by level of parental education. The paper has two distinct components. First, it sets out to document a set of facts about changes in time spent in childcare by highly educated and less educated parents over the last 40 years. The key result is that although time spent in childcare has increased for both groups, the increase was much greater for the highly educated parents. Second, the paper sets out to explain the primary mechanism that drives the patterns in the data. Ramey and Ramey propose a new explanation—the "rug rat race"—for the changing trends in time use between education groups. The crux of their argument is that as slots in elite postsecondary institutions have become scarcer, perhaps because demand for college attainment has increased, parents have responded by investing more in their children so that they appear more desirable to college admissions officers. Where before good grades were enough to gain admission, elite colleges are now seeking applicants with outstanding grades or lots of extracurricular activities, or both.

In the first part of the paper, the authors convincingly document that the amount of time spent by parents in childcare has changed, and changed differently in different educational groups, over the last 40 years. These facts stand by themselves and, I expect, will be an important input into future research. I am less convinced about the importance of their preferred explanation for the trends they document. Although I do think that a rug rat race could conceivably exist, very little in the paper directly supports its existence, and nothing in the paper allows one to reject other prominent stories that might explain the divergence in time spent in childcare between highly educated and less educated parents.

Throughout my discussion I will highlight many additional facts that make one think hard about whether or not the rug rat race can be driving the relationships observed in the data. In particular, I will show that nearly all the increase in educational differences in time use documented in recent surveys is driven by households with young children, and not by households with preteen or teenage children. Moreover, I will propose and empirically explore an additional test of the rug rat race hypothesis. As I will show, this test is rejected by the data. I will also show that an educational or income gradient in time spent in childcare is an important feature of the data within and across many different types of countries. Lastly, I will argue that the authors never convincingly reject alternative explanations—such as a pure income effect story—that might explain both the recent patterns in the United States and patterns using a variety of data from other countries. In the end, I am not sure what is driving the facts documented in the first part of the paper. Nothing in the paper tells me anything conclusive one way or another about the cause of the finding.

THE ROBUST RELATIONSHIP BETWEEN INCOME (AND EDUCATION) AND PARENTAL TIME SPENT ON CHILDCARE. The relationship between parental education (or income) and time spent in childcare is found in many datasets and across many time periods. For example, Russell Hill and Frank Stafford (1974) use data from 1965 to show that "high socioeconomic status" mothers spend between two and three times as much time in preschool childcare as do "low socioeconomic status" mothers. Some of these patterns can be seen in Ramey and Ramey's figure 1. In both 1965 and 1985, highly educated mothers spent between 2 and 4 hours more per week on childcare than less educated mothers (my calculations using similar data by slightly different age groups).

In my work with Jonathan Guryan and Melissa Kearney (Guryan, Hurst, and Kearney 2008), we document two additional facts that suggest that parental education and parental income are strongly associated with time spent in childcare. First, our analysis of time diaries collected around 2000 for a number of developing and developed economies finds a very strong positive relationship across countries between the average time spent by mothers in childcare and GDP per capita. This relationship became stronger when we controlled for demographic differences across the countries. Mothers in richer countries tend to spend more time in childcare than mothers in poorer countries. Second, within every country we studied, highly educated mothers spend more time with their children than less educated mothers.

These results suggest a persistent relationship between educational attainment and parental time spent with children. What drives this relationship? Many explanations are possible, but two stories immediately come to mind. First, it is possible that parental time spent in childcare has a high income elasticity in the parental utility function. As people get richer, they may take more of their utility in the form of time with their children. In that sense, time spent in childcare is a relative luxury good. Second, it is also possible that in a world where there are large differences in earnings between high- and low-skilled parents, the returns to investing in children are higher. If high-skilled parents are more adept at investing in their children than low-skilled parents, a change in the skill premium could yield greater investments by high-skilled parents relative to low-skilled parents. Note that this argument is similar in spirit to the rug rat race hypothesis. In the former, parents are responding directly to the change in returns from broadly investing in their children, whereas in the latter, parents are responding only to the increased competition to get their children into elite schools. I view the rug rat race as a special case of the broader possibility that high-skilled parents are responding to the increased returns to having children with higher skills, whether cognitive or noncognitive, or both.

If these two hypotheses explain the myriad facts outlined above, then the increase in income inequality between highly educated and less educated households in the United States during the last 25 years should have produced a greater increase in time spent with children among the former. Does the paper rule out these explanations? It attempts to, but not, I think, convincingly, as I will argue below.

FURTHER DECOMPOSING THE EDUCATIONAL TRENDS IN TIME SPENT WITH CHILDREN. Ramey and Ramey cut the data in many interesting ways and show many interesting trends with respect to parental time spent in child care. I want to highlight one additional cut of the data, using data from a short book I wrote with Mark Aguiar (Aguiar and Hurst 2009). As in that book, I restrict attention to mothers aged 25 through 44 where at least one child

Table 1. Mothers' Time Spent in Childcare, by Mother's Education and Age of Youngest Child, 2003–05[a]
Hours per week

Age of youngest child	*Less educated mothers*	*College-educated mothers*[b]	*Difference*
Less than 5	18.3	27.7	9.4
Between 6 and 12	10.1	12.5	2.4
Between 13 and 17	4.7	5.7	1.0

Source: Author's calculations using data from the 2003–05 ATUS.
a. Data are for mothers aged 25–44 and are the same as those used in Aguiar and Hurst (2009). The definition of time spent in childcare is similar to that in Ramey and Ramey (this volume) and in Aguiar and Hurst (2009). All data are unconditional and weighted using ATUS core sample weights.
b. College degree or more.

under the age of 17 is living in the home. My measure of total time spent on childcare is roughly equivalent to that used by Ramey and Ramey.

I cut the data in two ways. In table 1, I look only at data from the 2003–05 ATUS and compare time spent in childcare across education groups and by age of the youngest child in the household. The table shows that the biggest gap in time spent in childcare by educational attainment is in households where the youngest child is under the age of 5. For households with only teenagers who are on the cusp of college entry, time spent in childcare by highly educated and less educated parents is very similar. If the rug rat race hypothesis is correct, why is the biggest difference found in households with young children, and why is there so little difference between households with children closer to college age?

In table 2, I focus on the change in time spent in childcare between 1985 and 2003–05 for households where the youngest child is less than 5 years of age. Because the 1985 survey has only an indicator variable for whether the youngest child is less than 5, I cannot explore the detailed breakdown by age of youngest child as I did in table 1. The results nonetheless indicate that the *change* in childcare time spent by highly educated parents relative to less educated parents is driven mostly by the changing behavior of highly educated parents with young children in the household.

Ramey and Ramey also explore time spent on primary care activities separately for children under and over the age of 5. This distinction between children of different ages was not made, however, in the 2003–05 ATUS. The results I have presented indicate that the gaps between 1985 and 2003 with respect to increasing childcare time spent by highly educated parents relative to less educated parents are driven primarily by parents with young

Table 2. Change in Mothers' Time Spent in Childcare, by Education and Age of Youngest Child, 1985 to 2003–05[a]
Hours per week

Age of youngest child	*1985*	*2003–05*	*Difference*
Less than 5			
Less educated mothers	16.2	18.9	2.7
College-educated mothers[b]	18.0	25.6	7.6
Difference	1.8	6.7	4.9
6 or older			
Less educated mothers	5.6	9.3	3.7
College-educated mothers[b]	6.1	11.9	5.8
Difference	0.5	2.6	2.1

Source: Author's calculations using data from the 1985 and 2003–05 ATUS.

a. Data are for mothers aged 25–44 and are the same as those used in Aguiar and Hurst (2009). The definition of time spent in childcare is similar to that in Ramey and Ramey (this volume) and in Aguiar and Hurst (2009). All data are unconditional and weighted using ATUS core sample weights.

b. College degree or more.

children. Do these results prove the rug rat race does not exist? Not necessarily. They do, however, cast doubt on the hypothesis that it is parents' desire to get their children into elite colleges that is driving the bulk of the differential trends by education in parental time spent in childcare.

ANOTHER DIRECT TEST OF THE RUG RAT RACE HYPOTHESIS. One simple and direct way to test the rug rat race hypothesis is to look at the educational gaps (or the changes in educational gaps) in time spent in childcare across U.S. states, where the pressure of the rug rat race may differ. Ramey and Ramey look at the relationship between the average time spent in childcare by more educated mothers within a state and a measure of competition for college admission in that state (as computed by Bound, Hershbein, and Long 2009).

I propose another test: highly educated individuals who live in states with elite state universities should be less responsive to the pressures of the rug rat race than highly educated individuals who live in states without such universities. The reason is that elite state universities have a preference (some of which is statutory) to admit students from within the state. This gives students who live in a state with an elite state university greater access to an elite school than otherwise similar individuals in other states. Parents in these other states have to work harder to get their children into an elite school as admission to elite schools becomes more competitive.

To test this hypothesis, I divide states into three groups: those with a state university ranked in the top 30 national universities (public and private)

Table 3. Mothers' Time Spent in Childcare, by Mother's Education and Presence or Absence of an Elite State University[a]
Hours per week

Mother's educational attainment	*States with elite state universities*[b]	*States with near elite state universities*[c]	*All other states*
Less educated	14.2	12.8	13.9
College-educated[d]	18.6	18.9	19.3

Source: Author's calculations using data from the 2003 ATUS.

a. Data are for mothers aged 25–44 and are the same as those used in Aguiar and Hurst (2009). The definition of time spent in childcare is the same as that used in Aguiar and Hurst (2007). All data are unconditional and weighted using ATUS core sample weights.

b. California, Michigan, North Carolina, and Virginia.

c. Florida, Georgia, Illinois, Maryland, Minnesota, Ohio, Pennsylvania, Texas, Washington, and Wisconsin.

d. College degree or more.

by *U.S. News and World Report* in 2009, those with a state university ranked between 30th and 61st, and all others. The states in the first group (which I will refer to as those with elite state universities) are California, Michigan, North Carolina, and Virginia. Those in the second group, with what I will call near elite (very good) state universities, are Florida, Georgia, Illinois, Maryland, Minnesota, Ohio, Pennsylvania, Texas, Washington, and Wisconsin. Table 3 summarizes time spent with children by mothers aged 25 through 44 in each of these three tiers of states, using data from the 2003 ATUS.

According to the table, the difference in time spent in childcare between highly educated and less educated parents varies only a little between states with elite state universities and those with state universities outside of the top 60 national universities: the differences are 4.4 hours per week and 5.4 hours per week, respectively. Moreover, even these results are misleading because they do not control for persistent differences within a state across time. Using 1992–94 time-use data (which also include state identifiers), one can explore the change in the educational gap between states with elite state universities and all other states over time. The change between 1992–94 and 2003 was greatest in states with near elite state universities (8.2 hours per week) and lowest in the "other" states (1.7 hours per week). The change in states with elite state universities was 2.8 hours per week. These results are at odds with the implications of the rug rat race being a primary driver of the change in the educational gap in time spent in childcare. If the rug rat race were important, states where households did not have preferred access to elite or near elite state universities would have recorded the greatest increase in childcare time.

ALTERNATIVE STORIES: CAN INCOME EFFECTS BE RULED OUT? As discussed above, my first guess to explain the greater increase in childcare time among highly educated parents would be the widening of the income gap between education groups over the last 25 years. Ramey and Ramey offer two pieces of evidence to support their claim that this is not a primary driver of their results. First, they argue, the timing does not match up exactly. Undermining this claim, however, is the sporadic coverage of the time-use data and the fact that the 1992–94 data have been shown to be of lower quality and less reliable than those from the other time-use surveys. More important, to fully assess the timing issues, one must also consider differences in timing between entry into the labor market (where highly educated workers start to earn the skill premium) and fertility. When one takes into account that peak fertility among high-skilled workers begins 5 to 7 years after entry into the labor market, the timing of wage changes and the timing of the changes in childcare time line up more closely.

Second, Ramey and Ramey try to estimate income effects directly from the cross-sectional data by looking at how time spent in childcare by nonworking mothers varies with the income of their husbands. Even in this sample, they do find evidence that income effects are important. Nonworking women with high-income husbands spend more time with their children than nonworking women with lower-income husbands. However, Ramey and Ramey show that the income effects implied from this regression are not large enough to explain the time-series trends in the educational gap in time spent in childcare. This regression, however, is inherently flawed. Women face a time constraint. If they wish to spend more time with their children, they must spend less time on something else. One of the major ways women can allocate more time to their children is by leaving the labor force. By restricting their analysis to those women who have chosen not to work, the authors are missing a major margin available to women who wish to spend more time with their children. The problem is analogous to the well-documented criticisms of estimating female labor supply elasticities on a sample of working women only, which ignores adjustments on the extensive labor supply margin.

Ramey and Ramey do report some evidence that time spent in childcare responds to changes in the skill premium. I particularly liked the comparison with Canada. Canada has seen small changes in the skill premium and no increase in the educational gap in time spent in childcare. The United States, in contrast, has witnessed big changes in the skill premium and big changes in the educational gap. I am curious whether this cross-country pattern would hold up if more countries were analyzed. It would have been nice to see whether countries with larger increases in the skill premium had growing differences in time spent in childcare by level of parental education. The results contrasting the United States and Canada point in that direction, but it is hard to draw any conclusions from just two countries. I acknowledge that the U.S.-Canada comparison does not provide strong evidence for the story that income constraints are an important explanation for the documented patterns, but neither does it provide strong evidence for the importance of the rug rat race.

CONCLUSIONS There is a lot to like about this paper. The data work compiled by Ramey and Ramey is novel, well done, and interesting, and the paper was stimulating to read and think about. However, I remain unconvinced that the rug rat race is the primary (or even a secondary) explanation for the trend that the authors document. From my reading, nothing in the paper supports the contention that the rug rat race is a first-order explanation of the data, and nothing in their analysis rules out other potential explanations. I am not sure what story would best explain the facts documented in the paper—which means this is definitely a worthwhile area for future research.

REFERENCES FOR THE HURST COMMENT

Aguiar, Mark, and Erik Hurst. 2007. "Measuring Trends in Leisure: The Allocation of Time over Five Decades." *Quarterly Journal of Economics* 22, no. 3: 969–1006.

Aguiar, Mark, and Erik Hurst. 2009. *The Increase in Leisure Inequality: 1965–2005.* Washington: American Enterprise Institute for Public Policy Research.

Bound, John, Brad Hershbein, and Bridget Terry Long. 2009. "Playing the Admissions Game: Student Reactions to Increasing College Competition." *Journal of Economic Perspectives* 23, no. 4 (Fall): 119–46.

Guryan, Jonathan, Erik Hurst, and Melissa Kearney. 2008. "Parental Education and Parental Time with Children." *Journal of Economic Perspectives* 22, no. 3: 23–46.

Hill, C. Russell, and Frank P. Stafford. 1974. "Allocation of Time to Preschool Children and Educational Opportunity." *Journal of Human Resources* 9, no. 3: 323–41.

COMMENT BY

DANIEL W. SACKS AND BETSEY STEVENSON Gary Ramey and Valerie Ramey establish the important fact that time spent by parents in childcare, after remaining stable for decades, rose dramatically during the 1990s and reached a plateau in the 2000s. As Ramey and Ramey document, these changes are large by any metric, comparable to the decline in hours worked during the Great Recession. Moreover, this increase in childcare time was uneven: college-educated mothers' childcare time grew by 9 hours per week, while less educated mothers' time grew by 4 hours per week.

Ramey and Ramey argue that the only viable explanation for this relative growth among well-educated parents is a college rug rat race. Demographic pressures and increasing returns to higher education in the 1990s squeezed college-educated parents, whose children had in the past been virtually guaranteed a seat at selective colleges. Competition for increasingly scarce slots at these colleges drove parents to spend ever more time in childcare as they attempted to separate their own children from the pack.

This is a creative explanation that resonates with many contemporary accounts of parenthood and higher education. If correct, the college rug rat race has stark policy implications: the relative growth in childcare time does nothing to improve well-being; it is the result of parents hoping to transfer "college surplus" from other children to their own. This wasteful activity potentially comes at an enormous cost: Ramey and Ramey estimate that the rise in childcare represents over $300 billion in forgone wages annually.

To assess the social value of this $300 billion increase in childcare, it is useful to consider the alternative explanations that might have generated it. We posit three categories of explanations, each with different implications for social welfare: the rug rat race, investment-based explanations, and consumption-based explanations. If the rise in childcare hours is due exclusively to a rug rat race—which need not necessarily derive from college competition—the additional time spent in childcare is wasteful.[1] Investment-based explanations, in contrast, suggest that the increase in childcare is generating valuable returns for the next generation. Parents may be increasingly eager to invest in their children because of perceived increases in labor market returns to cognitive and noncognitive skills. Although socially beneficial, childcare-as-investment also means that college-educated parents' children, advantaged to begin with, will enjoy even greater levels of human capital than their peers, and thus the rising gap in childcare hours by education may portend widening inequality.

1. Ramey and Ramey note that parents may underinvest in children in the absence of a rug rat race if investment in children generates positive externalities. In this case the rug rat race may operate like a Pigouvian tax, correcting the underinvestment, and therefore be socially efficient.

Consumption-based explanations are more benign, suggesting that the relative rise in childcare among the well educated is no more alarming than a relative rise in expenditure on cars or computers would be. Instead, childcare-as-consumption points to other possible causes for the increase in childcare, for example changes in the structure of American families and the rise of hedonic marriages (Stevenson and Wolfers 2007, 2008).

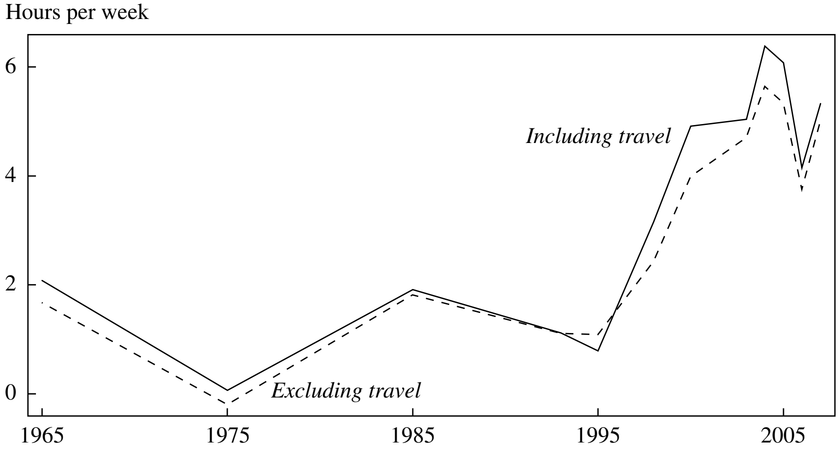
In this discussion we first offer some comments on Ramey and Ramey's evidence for the rug rat race as a driver of the increase in time spent in childcare. We show that, quantitatively, their evidence suggests the rug rat race can account for at most a modest fraction of this rise. We then present evidence on the importance of investment and consumption explanations in the rise of childcare time.

Ramey and Ramey present an impressively broad array of evidence for the rug rat race. First, they show that the trends in high school cohort size and in competition for college match the trends in childcare during the 1990s. Second, they point to the large rise in time spent transporting children and caring for older children. Third, they compare trends in childcare time by race and nationality. Finally, they use cross-sectional evidence on competition for college admission across states to demonstrate that college-educated mothers spend more time on childcare where college competition is greater.

Taken together, this evidence suggests that increases in college competition are playing a role in the rise in childcare, but there are some important caveats. First, although the childcare and cohort crowding time series match nicely in the 1990s, the two series are negatively correlated between 1965 and 1985, a period when childcare time was flat and graduation rates rose and then declined. Overall, the correlation between the size of the graduating cohort and the difference in childcare time between college-educated and less well educated mothers during the period Ramey and Ramey are studying is a negligible –0.06.

The second piece of evidence concerns relative changes in the components of childcare time and the importance of time spent in transportation. Ramey and Ramey emphasize that the rug rat race explains the rise in the differential trend between college-educated and less educated parents, and that an important component of the increase in the differential is "time spent on *older* children . . . transporting them to their activities" (emphasis in original). We find, however, that the increase in the differential trend was steeper among mothers with young children than among mothers without young children, and that the increase in time spent in transportation occurred nearly equally for college-educated and less educated parents.

Figure 1. Difference in Childcare Time between Well-Educated and Less Educated Parents, Including and Excluding Travel Time, 1965–2008[a]



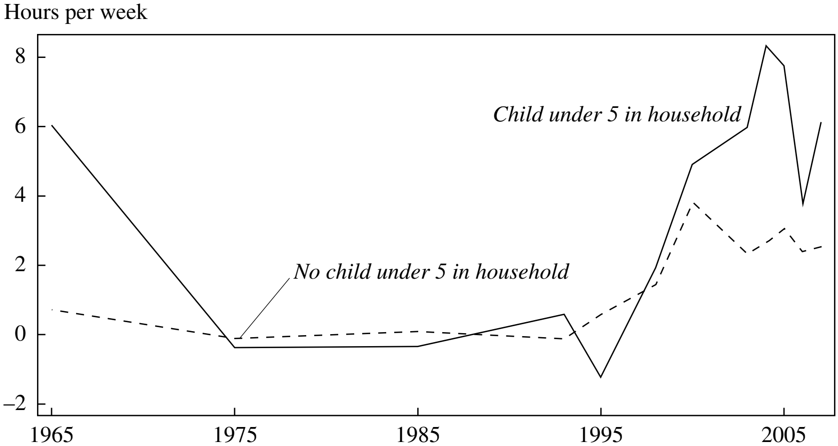
Source: Authors' calculations.

a. Figure plots the coefficients on the interaction between the year dummy and a dummy for whether the parent completed college, using the descriptive model and data in Ramey and Ramey (this volume) and including or excluding travel time in total childcare time.

To assess the role of older children and travel time in the rise in the differential trend, we reestimate Ramey and Ramey's basic regression but exclude time spent on travel. We plot the college × year interactions estimated from this exercise in figure 1, along with the original coefficients. The two patterns are very similar, and the regression results suggest that travel time may explain around 10 percent of the relative rise. Increases in travel and chauffeuring children constitute about 10 percent of the overall rise in childcare, but this increase in chauffeuring is unlikely to be a direct result of an increase in competition for college. In particular, college admissions offices do not directly observe parents chauffeuring children, and parents with a high value of time could simply make other transportation arrangements—hiring a driver or sending their child by taxi. Yet many parents have stated that such outsourcing would generate disapproval from other parents.[2] This social pressure for "homemade" transportation may reflect a rat race, but not a college-centric one.

2. Valerie Ramey noted this during the Brookings Panel discussion. Judith Warner (2005) has also pointed to the tendency of mothers to judge and ostracize each other.

Figure 2. Difference in Childcare Time between Well-Educated and Less Educated Parents, by Presence of Young Children, 1965–2008[a]



Source: Authors' calculations.

a. Figure plots the coefficients on the interaction between the year dummy and a dummy for whether the parent completed college, using the descriptive model and data in Ramey and Ramey (this volume) and including or excluding travel time in total childcare time.

In figure 2 we reestimate Ramey and Ramey's basic regression separately for mothers with children younger than 5 and mothers whose youngest child is 5 or older. College-educated mothers with young children experienced a much sharper increase in childcare time: the educational differential in childcare time rose by 6.6 hours per week for these mothers compared with 2.6 hours for mothers without young children.[3] This comparison suggests that college competition may not be responsible for a large share of the rise in childcare time.

Ramey and Ramey's third line of evidence comes from comparing trends in the United States with those in Canada, and, within the United States, between blacks and whites. They show that despite a similar upward trend in time spent in childcare among Canadian parents without a college degree, there has been no relative increase among college-educated Canadian mothers. The Canadian data illustrate the robustness of the upward trend in childcare documented by Ramey and Ramey. Canadian parents, like U.S. parents, are now spending several more hours a week

3. These differentials are the rise in the difference in childcare time for college-educated versus less educated mothers between the periods 1965–95 and 1998–2007, adjusting for age, for parents with and without young children.

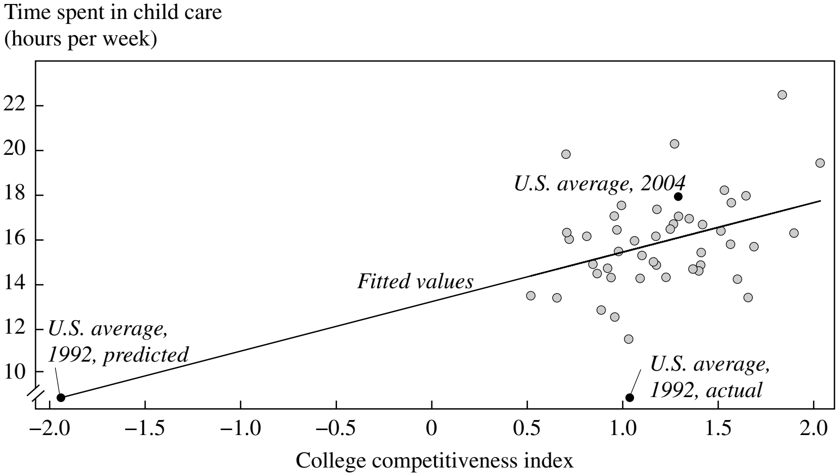
interacting with their children. However, the college gap in childcare is more difficult to interpret. Before the rise in childcare in the 1990s, the gap in hours spent in childcare between college- and non-college-educated mothers was less than an hour in the United States, but over 7 hours in Canada. The gap has since widened in the United States to around 5 hours, while falling in Canada to around 3 hours. Ramey and Ramey point to this fact as evidence of an increase in college competition in the United States. However, this explanation says little about why the educational gap shrank in Canada or why it was so much bigger in Canada to begin with.

Finally, Ramey and Ramey find evidence for the rug rat race in the relationship at the state level between competition for college admission in 2004 and parental childcare time in 2003–07. The results show that well-educated parents provide more childcare in states where college competition is greater, as their model predicts, but so do less well educated parents. The coefficients on childcare time are similar for both groups, as judged by either statistical or economic significance. These findings suggest that competition for college does potentially drive parental time in childcare, but they offer inconclusive evidence as to whether such competition affects college-educated parents' behavior differently.

The authors' cross-state evidence provides a useful benchmark for assessing the magnitude of the relationship between competition for college and time spent in childcare. We extend their results to assess how much of an increase in childcare time is implied, given the increase in competition for college that has occurred nationally. Figure 3 replicates Ramey and Ramey's figure 10, with three additional data points. The first is the national level of college competition in 2004, which we estimate by taking a population-weighted average of competition at the state level. The second is the national level of college competitiveness in 1992, estimated similarly using state-level data from Bound, Hershbein, and Long (2009). The third point shows what the level of competitiveness would have had to be to explain the difference in childcare time between 1992 and 2004 for college-educated mothers. Bound, Hershbein, and Long's index of college competitiveness implies an increase of 0.25 between 1992 and 2004. Applying this increase to Ramey and Ramey's cross-sectional results implies that college-educated mothers' childcare time should have risen by about half an hour per week.

In sum, the college rug rat race is a potential source of the increase in childcare time both relatively and absolutely, but it appears able to explain only a modest share of the rise. We now consider investment- and consumption-based explanations.

Figure 3. College Competitiveness and Childcare Time Spent by College-Educated Mothers, by State, 2004



Source: Figure 10 in Ramey and Ramey (this volume), Bound and others (2009), and authors' regression.

Some of the most productive investments parents make in their children occur at young ages, as parents lay the foundation for future learning by inculcating strong habits and cognitive skills in their children. Since parental time is an essential input to these investments, and since the returns to cognitive skills rose over the 1990s (Cunha and Heckman 2008), increases in parental investment in children could be an important part of the relative and absolute rise in childcare time.

Indeed, most of the relative increase in childcare time has been concentrated among families with young children, as figure 2 shows. To further investigate time spent on young children, we study trends in breastfeeding, which likely represents an investment and certainly does not contribute (directly) to college success.

In its 1988–91, 1991–94, and 2005–06 waves, the National Health and Nutrition Examination Survey (NHANES) asked respondents whether the youngest child in the household was breastfed. We study the percent of children that were ever breastfed and the percent that were breastfed through 6 months of age, by education of the mother. Table 1 shows that between the late 1980s and the early 1990s, college-educated mothers were increasingly likely to breastfeed their children, and that the gap between college- and non-college-educated parents widened and then narrowed

Table 1. Share of Mothers Breastfeeding, by Education of Mother, 1988–2006[a]

Mother's education	*Percent of mothers who ever breastfed*			*Percent of mothers breastfeeding at 6 months*		
	1988–91	*1991–94*	*2005–06*	*1988–91*	*1991–94*	*2005–06*
Completed high school	37	35	61	44	38	44
Completed college	74	86	88	43	64	61
Difference	37	51	27	–1	26	17
Standard error of the difference	5***	4***	4***	7	6***	5***

Source: Authors' calculations.

a. Estimated using the 1988–94 and 2005–06 waves of the NHANES, restricted to female respondents with no missing information. All numbers adjust for differences in the age distribution of college-educated and less educated women. Asterisks indicate statistical significance at the ***1 percent level.

between 1991–94 and 2005–06. The same is true for the percent still breastfeeding at 6 months. At that age, children of college-educated moms born in the early 1990s were 50 percent more likely to be receiving breast milk than children in the previous decade, and more than half of all college-educated mothers were still breastfeeding.

Could the rise in breastfeeding reflect a broader emphasis on childcare? If breastfeeding requires little effort, then trends in breastfeeding might not reveal much about overall attitudes toward childcare, whereas if breastfeeding is time intensive, then an increase in breastfeeding might indeed reflect an increased parental emphasis on childcare. Surprisingly few data exist on the time cost of breastfeeding. To estimate this cost, we conducted a survey and advertised it on the website facebook.com and the *New York Times'* Freakonomics blog.[4] The survey generated 2,099 responses. In no sense are these data necessarily representative of the general population, but they provide rough evidence on time spent breastfeeding.

We measured breastfeeding rates and intensity over time by asking about the breastfeeding of the latest child and noting when he or she was born. As in the NHANES, breastfeeding rates have increased over time; more interesting, so has breastfeeding intensity, which we report in table 2. Mothers spend a considerable amount of time breastfeeding, well over 2 hours per day on average. Breastfeeding thus requires an enormous time commitment;

4. The survey can be seen (and taken) here: wharton.qualtrics.com/SE?SID=SV_e2NOBSudMtPELFW&SVID=Prod. We used responses collected between March 12 and April 5, 2010.

Table 2. Average Time Spent Breastfeeding during Child's Infancy, by Education of Parent[a]
Hours per day

	Year of birth of youngest child				
Mother's education	*Before 1980*	*1980s*	*1990s*	*2000–05*	*2006–10*
High school diploma or less	1.3	1.7	2.8	3.0	3.1
	(0.4)	(0.5)	(0.5)	(0.5)	(0.3)
Bachelor's degree	2.5	2.5	2.8	2.8	3.0
	(0.5)	(0.3)	(0.2)	(0.2)	(0.1)
Master's degree	2.6	2.5	2.5	2.6	2.9
	(0.5)	(0.3)	(0.2)	(0.2)	(0.1)
Professional degree or Ph.D.	2.3	2.6	2.2	2.6	3.2
	(0.4)	(0.3)	(0.2)	(0.2)	(0.1)

Source: Authors calculations' based on data obtained via online polling. The survey can be seen (and taken) at wharton.qualtrics.com/SE?SID=SV_e2NOBSudMtPELFW&SVID=Prod. Table is based on responses collected between March 12 and April 5, 2010.

a. Standard errors are in parentheses.

mothers who breastfeed may be more likely to spend a great deal of time on other childcare activities as the child ages. These results hint at the potential for a cohort-based explanation for the rise in childcare time, since college-educated mothers of infants in the early 1990s were the first wave of parents to experience the rise in childcare time, and these mothers may have developed habits from their time-intensive breastfeeding activity. If such a habit formation cohort-based model explains the trends identified by Ramey and Ramey, it offers a clear prediction for the future: the gap in time spent in childcare by parental education should narrow over the next decade, since the breastfeeding differential has narrowed in recent years.

To understand the rise in breastfeeding better, we asked mothers why they breastfed. The most common response, chosen by 96 percent of mothers, was for the health of the baby. Thirty-eight percent said they breastfed to improve their baby's intelligence. (Mothers could give more than one reason.) These answers clearly indicate an investment motive for breastfeeding, but breastfeeding also has a consumption component: two-thirds of mothers breastfed to bond with their infant, and half breastfed for enjoyment.

Even if the absolute and relative rise in childcare time does not represent investment, it need not be wasteful if parents enjoy the time they spend with their children, that is, if the increase in childcare time represents consumption. We present two pieces of evidence that point to the importance of consumption motives in explaining the absolute and relative rise in childcare time: first, changes in time spent in play account for a nonnegligible fraction of both the absolute and the relative rise; and second, college-educated parents spend relatively more time on childcare with their spouses. The increasing importance of childcare as a source of consumption is consistent with both the rise of hedonic marriage among college-educated couples and the apparent geographic dispersion in children's college attendance.

Ramey and Ramey note that parents enjoy playing with their children and that increases in this activity may be indicative of rising consumption. "Playing with children," they write, "has always ranked high in terms of enjoyment." Their table 3 shows that after excluding play time, the relative differential remains large but has fallen by about a quarter, suggesting that time spent in play accounts for an important fraction of the overall relative rise in childcare time. In fact, play time accounts for about a quarter of the overall rise as well as of the relative rise.

The second piece of evidence supporting a consumption-based explanation is that college-educated parents spend a great deal of childcare time together, relative to less well educated parents.[5] The 2003–07 ATUS allows researchers to identify with whom the respondent performs a given activity, and we use this information to count up the hours of childcare time. If the rise in childcare time is due to college preparation, then one would not expect well-educated spouses to spend childcare time together, but if the rise in childcare time is about enjoying family time, then one might well expect well-educated parents to engage in childcare time together.

College-educated mothers spend much more time in childcare with their spouse, and college-educated fathers somewhat more, than do less educated mothers and fathers, respectively. We regress childcare time with spouse against a set of parental education indicators and Ramey and Ramey's age dummies; the results (table 3) show that college-educated parents spend 4.6 hours of childcare time per week with their spouse, compared with 2.5 hours for less educated parents. College-educated mothers spend 2.5 more hours in childcare with their spouse, and college-educated fathers spend 1.5 more hours. These differences are large, equal to nearly half the relative rise in childcare time.

5. We look at the difference at a point in time, rather than changes in this difference, because older surveys lack the "with whom" data necessary to see whether parents spend their childcare time together.

Table 3. Regression-Based Estimates of Time Spent in Childcare with Spouse, by Education[a]
Hours per week

Group	*College-educated parents*	*Less educated parents*	*Standard error of the difference*
All parents	4.6	2.5	0.1***
Mothers	5.8	3.3	0.1***
Fathers	3.0	1.5	0.1***

Source: Authors' regression using data from the 2003–07 ATUS.
a. Estimates obtained by regressing number of hours of childcare time per week on indicators for age categories as well as for college attainment from Ramey and Ramey (this volume). Sample restricted to nonstudent respondents aged 18–64 with a child younger than 18 present in the household. Asterisks indicate statistical significance at the ***1 percent level.

Adam Isen and Stevenson (2010) discuss the rise of hedonic marriages and the fall of "productive marriages," particularly among well-educated couples. Whereas a traditional view of marriage emphasizes the gains from specialization in the production of children and household goods (Becker 1981), hedonic marriages generate "marital surplus" in the form of complementarities in consumption: spouses enjoy spending time with each other. To the extent that spouses enjoy spending time in childcare with each other, the rise in hedonic marriage can explain a substantial fraction of the relative rise in childcare time. Since college admissions offices do not observe how many parents engage in these activities with their children, it is unlikely that college competition is driving the increase in joint parenting time. These results are only suggestive, however, because it is not known how time spent in childcare with one's spouse has trended over time.

An additional potential explanation for why parents, particularly college-educated parents, may be spending more time with their children relates to the increased distance that students now travel to attend college. The children of college-educated parents are more likely to travel long distances to college. Table 4 shows that the distance traveled to college rises steeply with the education of the parents. In response, these parents may spend more time with their children while they are small, either to make up for lost time in the future or to build a relationship with them, to ensure that they come home for visits. Although data on changes in the distance traveled to college over time are unavailable, Caroline Hoxby (2009) has argued that college has become more of a national market, and thus that college entrants, particularly the children of college-educated parents, may be increasingly likely to attend college far from home.

In sum, Ramey and Ramey find an important shift in parents' time allocation, but their preferred explanation, the college rug rat race, can

Table 4. Distance Traveled to College, by Education of Parents

Parental education	*Median distance traveled (miles)*
Less than high school	28
Completed high school	68
Completed college	102
Graduate or professional degree	130

Source: Mattern and Wyatt (2009).

explain only part of this shift. We propose two additional categories of explanations—consumption and investment—and find evidence of behavior consistent with both types of explanation that appears to explain some portion of the rise in childcare time. A key priority for future work is to distinguish among these explanations. The difficulty in explaining the rise in childcare time is that it happened so suddenly, whereas the possible driving forces, such as competition for college, the rise of hedonic marriage, and the increasing importance of cognitive and noncognitive skills, have all changed gradually (Hoxby 2009, Cunha and Heckman 2008). The sudden rise in childcare time suggests a tipping dynamic. This might arise because the value of childcare time depends on what other children are doing. If every child can play in the park, then no parent need closely watch her own child, because of safety in numbers.[6] But if other children are with their parents, then no parent can unilaterally send her child to the park unattended. As parents grow increasingly vigilant, the equilibrium level of childcare could shift rapidly upward. If tipping dynamics are important, then any of the possible driving forces could be important in explaining the growth in childcare time.

REFERENCES FOR THE SACKS AND STEVENSON COMMENT

Becker, Gary S. 1981. *A Treatise on the Family.* Harvard University Press.

Bound, John, Brad Hershbein, and Bridget Terry Long. 2009. “Playing the Admissions Game: Student Reactions to Increasing College Competition.” *Journal of Economic Perspectives* 23, no. 4: 119–46.

Cunha, Flavio, and James J. Heckman. 2008. “Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation.” *Journal of Human Resources* 43, no. 4: 738–82.

Hoxby, Caroline M. 2009. “The Changing Selectivity of American Colleges.” *Journal of Economic Perspectives* 23, no. 4: 95–118.

Isen, Adam, and Betsey Stevenson. Forthcoming. “Women’s Education and Family Behavior: Trends in Marriage, Divorce and Fertility.” In *Demography and the Economy,* edited by John B. Shoven. University of Chicago Press.

6. Christopher Sims pointed this out at the Brookings Panel discussion.

Mattern, Krista, and Jeff N. Wyatt. 2009. "Student Choice of College: How Far Do Students Go for an Education?" *Journal of College Admission* (Spring), no. 203: 18–29.

Stevenson, Betsey, and Justin Wolfers. 2007. "Marriage and Divorce: Changes and Their Driving Forces." *Journal of Economic Perspectives* 21, no. 2: 27–52.

———. 2008. "Marriage and Market." *Cato Unbound* (January 8).

Warner, Judith. 2005. *Perfect Madness: Motherhood in the Age of Anxiety.* New York: Riverhead Books.

GENERAL DISCUSSION Robert Gordon recalled that when he was a child, parents did none of the constant chauffeuring that today's parents do, and he wondered whether that was because it was safer in those days for young children to travel by themselves.

Susan Collins wondered whether the authors had tried to look at differences across regions with diverse characteristics, in particular with respect to traffic and travel times. It might be that much of the increase in chauffeuring is associated with changes in family lifestyles as commuting times have risen in urban and suburban areas. Collins also raised the problem of reporting issues in the measurement of time; perhaps, for example, parents today report dinner as time spent with children but did not do so before. She also raised the question of what increased childcare time was substituting for. If parents, particularly nonworking parents, are spending more time on child care, what are they spending less time on?

Finally, Collins suggested an alternative interpretation of the increase in childcare time. The past decade or so has seen a significant increase in at least some parts of society of angst around "mommy wars"—among women who have made different choices related to work outside the home once they became parents. In the United States, this sociological phenomenon tends to be concentrated among more highly educated women, who are more likely able to make such choices, but her perception was that it is less evident in other cultures—which may help to explain the differences reported by the authors between the United States and Canada. In particular, selection issues may arise among highly educated women who are not working, because these are precisely the women one would most expect to be spending—and to report spending—more time with their children.

Steven Davis found the evidence in the paper on the safety hypothesis completely unpersuasive, for two reasons. First, the concept of safety in the sense of shielding one's children from crime should be broadened to parents' fears of access to drugs, bad influences from peers, corruption by the media, and so on. Second, and along the lines of what Susan Collins had suggested, the key thing is perceptions. There are many reasons why the fears that parents have might differ quite a bit from the actual danger, at least as measured by crime rate data alone. Davis also found it entirely plausible that investments in noncognitive skills are complementary to intensive investments in the cognitive skills associated with higher education and more highly educated parents. It is also plausible, perhaps because of changes in the nature of the workforce, that the reward to these noncognitive skills has increased over time.

Steven Landefeld, like Collins, wondered about possible reporting biases. In household expenditure surveys, for example, "sin" goods tend to be underreported, and in a lot of time-use surveys there is concern about overreporting of uses of time that might be viewed more favorably by others. As a result, whether childcare is reported as the primary or a secondary activity might vary across parents with different levels of education. TV watching, ironing, or any other activity done while children are present could be reported primarily as time with children, if parents are conscious of wanting to spend more quality time with them. Total time devoted to childcare might not have changed much, but time primary-coded as childcare might have increased as a result of such reporting bias.

Kristin Forbes suggested another alternative explanation for the increase. Highly educated women who are spending more time with their children may fall into two quite different categories: those who are still working, and those—an increasingly large group—who have dropped out of the labor force. The latter, one would expect, are the ones driving the majority of the increase in time spent with their children. If so, the real question then becomes why more women have been dropping out of the labor force, allowing them to spend more time with their children. The answer could be a combination of explanations related to how jobs are changing, especially jobs available to highly educated women. For example, with the proliferation of Blackberrys and the expectation that people are "working" around the clock even when not in the office, it may be harder for highly educated women to balance high-powered jobs with raising a family.

Alan Krueger agreed that the paper overemphasized the competition for selective colleges. The evidence on the payoff of going to an elite school is weaker and more controversial than the paper acknowledged. On the other hand, the perception is there. But it is not necessary to model parents' behavior as being motivated by a desire to see their children attend a top college. The majority of college-educated parents are thinking about their kids going on to college, but not necessarily an elite college. Nonetheless, there has been an increase in the payoff from higher education during the period, which may be sufficient to explain these trends, because graduating from college has become more valuable generally.

Krueger suggested that although childcare time is defined somewhat subjectively in the surveys, one might be able to circumvent this problem by looking at it more objectively, that is, by measures of how much of parents' time is spent with their children. Finally, Krueger noted that care of aging parents has risen over the same period that childcare has. No one claims there is a rat race to get one's parents into college, so a broader explanation of both trends would be worthwhile.

Robert Hall, echoing Collins's comment, remarked that the hypervigilance of parents today is absolutely stunning. He was puzzled, however, by the discussion of whether an income effect contributed to the rise in childcare time. There is one primary factor in the U.S. economy that matters, and that is labor. The only reason income rises is because the value of time rises, and there is a strong presumption that the income and substitution effects of spending more time working offset each other. Whether they do or not is not clear, but in any event the income effect should not be invoked without mentioning the substitution effect that presumably offsets it.

Christopher Sims called the Panel's attention to an externality that had thus far been ignored. When he was a child, he played football in the park with other children in the neighborhood. No parent today, however, can simply send her child to the park to play with the other kids, because no one is there—all the other kids are in organized sports teams. Thus any parent who thinks her child should play sports has no option but to spend time getting her child involved in the organized teams.

Matthew Shapiro wondered about the social costs of "bowling alone" with one's family. If people are investing more time with their immediate family, where are they spending less time? Is the time coming out of community work or political activity? It would be worth investigating whether the increased focus on investing in the human capital of one's own children comes at the cost of investments in social capital.

Erik Hurst highlighted the choice between market and nonmarket inputs into child care. If parents feel pressured to spend time chauffeuring their children, there is always the option of paying somebody else to do it. As to where the increased childcare time is coming from, Hurst reported that for highly educated women, time spent on childcare has been found to be coming out of home production and, to some extent, out of market work. Total leisure time has been flat. For less educated women, however, and for both highly educated and less educated men, leisure is also rising. Almost all of the increase in childcare time is coming from nonworking women, and the gap by education has not increased as much among this group as it has among women generally.

David Cutler noted that women's wages tend to rise until they have children, which carries the implication that once their own occupational upgrading ceases, and with it the growth in their wages, they may decide to substitute time away from work. Activities associated with work become less rewarding, whereas time spent with children becomes relatively more rewarding, causing women to opt out of the labor force and invest more time in their children.]

ALAN GREENSPAN
Greenspan Associates

The Crisis

ABSTRACT Geopolitical changes following the end of the Cold War induced a worldwide decline in real long-term interest rates that, in turn, produced home price bubbles across more than a dozen countries. However, it was the heavy securitization of the U.S. subprime mortgage market from 2003 to 2006 that spawned the toxic assets that triggered the disruptive collapse of the global bubble in 2007–08. Private counterparty risk management and official regulation failed to set levels of capital and liquidity that would have thwarted financial contagion and assuaged the impact of the crisis. This woeful record has energized regulatory reform but also suggests that regulations that require a forecast are likely to fail. Instead, the primary imperative has to be increased regulatory capital, liquidity, and collateral requirements for banks and shadow banks alike. Policies that presume that some institutions are "too big to fail" cannot be allowed to stand. Finally, a range of evidence suggests that monetary policy was not the source of the bubble.

I. Preamble

The bankruptcy of Lehman Brothers in September 2008 precipitated what, in retrospect, is likely to be judged the most virulent global financial crisis ever. To be sure, the contraction in economic activity that followed in its wake has fallen far short of the depression of the 1930s. But a precedent for the virtual withdrawal, on so global a scale, of private short-term credit, the leading edge of financial crisis, is not readily evident in our financial history. The collapse of private counterparty credit surveillance, fine-tuned over so many decades, along with the failure of the global regulatory system, calls for the thorough review by governments and private risk managers now under way.

The central theme of this paper is that in the years leading up to the crisis, financial intermediation tried to function on too thin a layer of capital, owing to a misreading of the degree of risk embedded in ever-more-complex financial products and markets. Section II of the paper reviews the causes of the crisis. In section III the nature of financial intermediation is probed. In section IV a set of reforms is proposed that, I trust, address the shortcomings of the existing regulatory structure. In section V the role of monetary policy in the crisis is examined. I offer some conclusions in section VI.

II. Causes of the Crisis

II.A. The Arbitraged Global Bond Market and the Housing Crisis

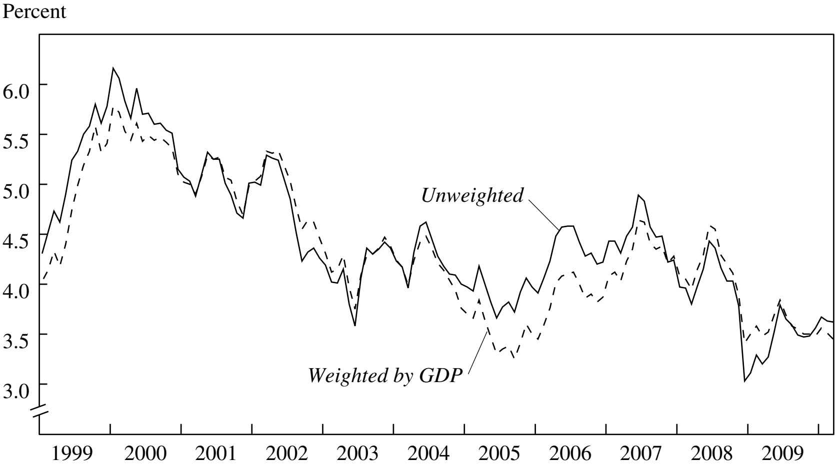
The global proliferation of securitized, toxic U.S. subprime mortgages was the immediate trigger of the crisis. But the origins of the crisis reach back, as best I can judge, to the aftermath of the Cold War.[1] The fall of the Berlin Wall exposed the economic ruin produced by the Soviet bloc's economic system. In response, competitive markets quietly, but rapidly, displaced much of the discredited central planning so prevalent in the Soviet bloc and the then Third World.

A large segment of the erstwhile Third World nations, especially China, replicated the successful export-oriented economic model of the so-called Asian Tigers (Hong Kong, Singapore, South Korea, and Taiwan): fairly well educated, low-cost workforces, joined with developed-world technology and protected by increasingly widespread adherence to the rule of law, unleashed explosive economic growth.[2] The International Monetary Fund (IMF) estimated that in 2005 more than 800 million members of the world's labor force were engaged in export-oriented and therefore competitive markets, an increase of 500 million since the fall of the Berlin Wall.[3] Additional hundreds of millions became subject to domestic competitive forces, especially in the former Soviet Union. As a consequence, between 2000 and 2007 the real GDP growth rate of the developing world was almost double that of the developed world.

Consumption in the developing world, however, restrained by culture and inadequate consumer finance, could not keep up with the surge of income, and consequently the saving rate of the developing world soared from 24 percent of nominal GDP in 1999 to 34 percent by 2007, far out-

1. For a more detailed explanation see Greenspan (2007, chapter 20).
2. Foreign direct investment in China, for example, rose gradually from 1980 to 1990, but then rose 39-fold by 2007.
3. IMF, *World Economic Outlook,* April 2007, chapter 5, p. 162.

Figure 1. Nominal Yields on 10-Year Government Debt, Average for 15 Countries, 1999–2010[a]



Source: Various country sources.
a. The countries are Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, the Netherlands, Norway, Spain, Sweden, Switzerland, the United Kingdom, and the United States.

stripping its investment rate. With investment elsewhere in the world slow to take up the slack, the result was a pronounced fall from 2000 to 2005 in global long-term interest rates, both nominal (figure 1) and real.

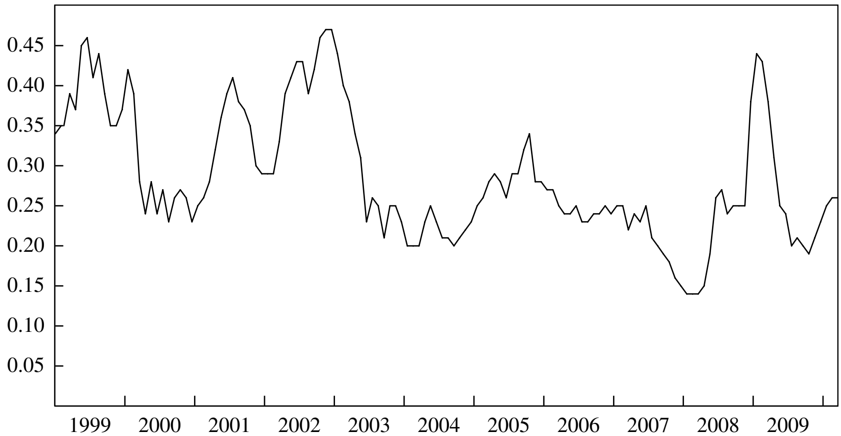
Although the decline in global interest rates indicated, of necessity, that global saving intentions were chronically exceeding global intentions to invest, ex post global saving and investment rates in 2007, overall, were only modestly higher than in 1999, suggesting that the uptrend in the saving intentions of developing economies tempered declining investment intentions in the developed world.[4] Of course, whether it was a glut of intended saving or a shortfall of investment intentions, the conclusion is the same: real long-term interest rates had to fall.

Inflation and long-term interest rates in all developed economies and the major developing economies had by 2006 converged to single digits, I believe for the first time ever. The path of the convergence is evident in the unweighted average variance of interest rates on 10-year sovereign debt of 15 countries: that average declined markedly from 2000 to 2005 (figure 2).[5]

4. That weakened global investment was a major determinant in the decline of global real long-term interest rates was also the conclusion of a March 2007 Bank of Canada study (Desroches and Francis 2007).

5. The variance of the logarithms of the 15 long-term interest rates exhibits similar trends.

Figure 2. Variance of Interest Rates: 10-Year Government Debt in 15 Countries, 1999–2010[a]



Source: Various country sources.
a. Unweighted average for the 15 countries in figure 1.

Equity and real estate capitalization rates were inevitably arbitraged lower by the fall in global long-term real interest rates. Asset prices, particularly home prices, accordingly moved dramatically higher.

The Economist's surveys document the remarkable convergence of nearly 20 individual nations' home price rises during the past decade.[6] Japan, Germany, and Switzerland (for differing reasons) were the only important exceptions. U.S. home price gains, at their peak, were no more than the global peak average.[7] In short, geopolitical events ultimately led to a fall in long-term mortgage interest rates that in turn led, with a lag, to the boom in home prices globally.

II.B. Securitization of Subprimes: The Crisis Story Unfolds

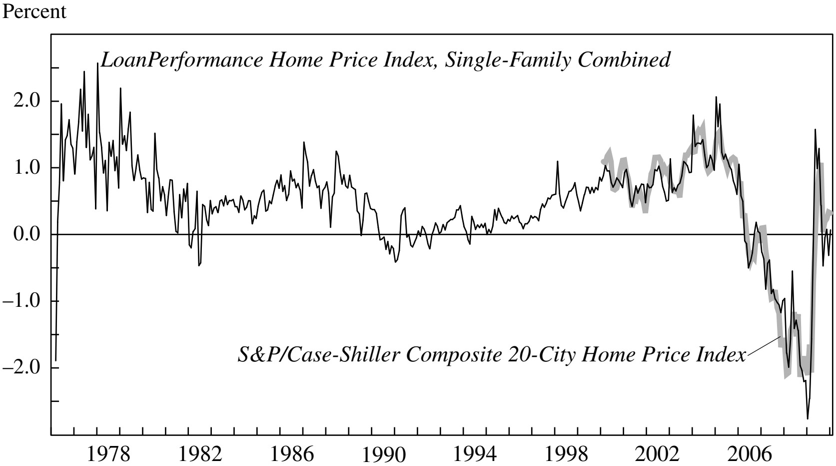
The subprime mortgage market that developed in the 1990s was a small but generally successful market of largely fixed-rate mortgages. It serviced mainly those potential homeowners who could not meet the down payment requirement of a prime loan, but still had income adequate to handle a fixed-rate mortgage.[8] Only a modest amount had been securitized, but with

6. For example, *The Economist,* "Finance and Economics: Houses Built on Sand," September 15, 2007, p. 104.

7. IMF, *World Economic Outlook,* April 2008, chapter 3, p. 113.

8. As recently as 2002, subprime mortgages accounted for 7 percent of total originations.

Figure 3. Monthly Changes in Home Prices, 1976–2010[a]



Source: Author's calculations based on data from LoanPerformance and Standard & Poor's.
a. Both series are seasonally adjusted.

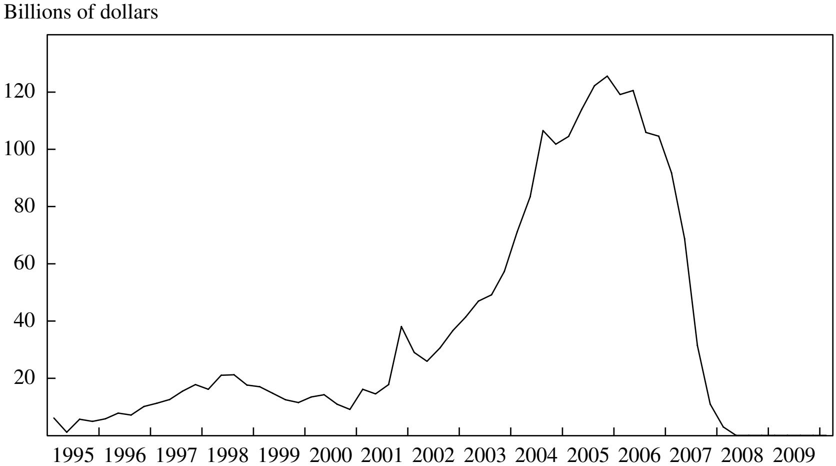
home prices having risen at a quickening pace since 1997 (figure 3), subprime lending was seen as increasingly profitable to investors.

Belatedly drawn to this market, financial firms, starting in late 2003, began to accelerate the pooling and packaging of subprime mortgages into securities (figure 4). The firms clearly had found receptive buyers. Heavy demand from Europe,[9] in the form of subprime mortgage-backed collateralized debt obligations, was fostered by attractive yields and a foreclosure rate on the underlying mortgages that had been in decline for 2 years.

An even heavier demand was driven by the need of Fannie Mae and Freddie Mac, the major U.S. government-sponsored enterprises (GSEs), pressed by the Department of Housing and Urban Development and

9. That many of the investors were European was confirmed by the recent heavy losses on U.S. mortgages reported by European investors. Euro-area banks, for example, exhibit a very high ratio of residential mortgage-backed securities write-downs to the residential mortgage loans they hold (IMF, *Global Financial Stability Report,* October 2009, p. 10). The size of the buildup of subprime securities holdings abroad during the bubble years is unclear. The U.S. Treasury's annual Foreign Holdings Survey reports that by mid-2006, foreign investors held $341 billion of privately issued U.S. mortgage-backed securities, some of which were commercial mortgage-backed securities. The less detailed mid-2002 survey reported a total for all asset-backed securities of $169 billion, compared with $594 billion in mid-2006.

Figure 4. Issuance of Subprime Mortgage-Backed Securities, 1995–2010[a]



Source: *Inside Mortgage Finance.*
a. Quarterly data, seasonally adjusted.

Congress to meet expanded "affordable housing goals."[10] Given the size of the GSEs' expanded commitments to fund low- and moderate-income housing, they had few alternatives but to invest, wholesale, in subprime securities. The GSEs accounted for an estimated 42 and 49 percent of all newly purchased subprime mortgage securities (almost all at adjustable interest rates) retained on investors' balance sheets during 2003 and 2004, respectively (table 1).[11] That was more than five times their estimated share in 2002.

Increasingly, the extraordinary demand pressed against the limited supply of qualified potential subprime borrowers. To reach beyond this limited population, securitizers unwisely prodded subprime mortgage originators to offer adjustable-rate mortgages (ARMs) with initially lower monthly pay-

10. In October 2000 HUD finalized a rule "significantly increasing the GSEs' affordable housing goals" for each year from 2001 to 2003. In November 2004 the annual housing goals for 2005 and beyond were raised still further (Office of Policy Development and Research 2001).

11. Federal Housing Finance Agency, *2008 Annual Report to Congress* (revised), Historical Data Tables 5b, Part 2, and 14b, Part 2 (originally published May 18, 2009, and updated to include a significant reclassification effective September 3, 2009). Before the revision, I estimated the share at less than 30 percent. Data newly reclassified by Fannie Mae account for almost all the revision.

Table 1. Holdings and Market Shares of Subprime Mortgage-Backed Securities by Fannie Mae and Freddie Mac, 2000–08

			Fannie Mae and Freddie Mac single-family private-label mortgages retained in their portfolios[a]			
					Change from end of previous year	
Year	*Total subprime MBSs outstanding (billions of dollars)*	*Change in total subprime MBSs outstanding (billions of dollars)*	*Billions of dollars*	*As percent of total subprime MBSs outstanding*	*Billions of dollars*	*As percent of change in total subprime MBSs outstanding*
2000	88.7					
2001	119.0	30.3	19.0	16.0		
2002	186.2	67.2	24.7	13.3	5.7	8.5
2003	280.8	94.6	64.9	23.1	40.1	42.4
2004	456.5	175.7	150.6	33.0	85.8	48.8
2005	644.3	187.8	179.2	27.8	28.6	15.2
2006	800.9	156.6	169.0	21.1	–10.2	–6.5
2007	774.0	–26.9	133.4	17.2	–35.6	132.5
2008	605.4	–168.6	99.4	16.4	–34.0	20.1

Sources: Federal Housing Finance Agency, *2008 Report to Congress* (revised), LoanPerformance data, and author's calculations.

a. Fannie Mae publishes its subprime securities holdings for each year from 2002 to 2008. For 2001 the sum of subprime and alt-A holdings is approximately reported, with the division between them guided by shares of total outstanding subprime and alt-A mortgage-backed securities (MBSs). Freddie Mac publishes similar data for 2006–08. With minor assumptions, estimates are made for the sum of subprime and alt-A holdings for earlier years. The separation is made essentially to reflect the ratio of total outstanding subprime and alt-A MBSs.

ments. As loan underwriting standards deteriorated rapidly, ARMs soared to nearly 62 percent of first-mortgage subprime originations by the second quarter of 2007.[12] By 2005 and 2006,[13] subprime mortgage originations had swelled to a bubbly 20 percent of all U.S. home mortgage originations, almost triple their share in 2002.

By the first quarter of 2007, virtually all subprime mortgage originations were being securitized, compared with less than half in 2000,[14] and subprime mortgage securities outstanding totaled more than $800 billion, almost seven times their level at the end of 2001. The securitizers, profitably packaging this new source of paper into mortgage pools and armed with what turned out, in retrospect, to be grossly inflated credit ratings, were able to sell seemingly unlimited amounts of these securities into what appeared to be a vast and receptive global market.

II.C. A Classic Euphoric Bubble Takes Hold

As a measure of how far the appetite for risk taking beyond the securitized mortgage market had gone, long-sacrosanct debt covenants were eased as a classic euphoric global bubble took hold.[15] By 2007, yield spreads in debt markets overall had narrowed to a point where there was little room for further underpricing of risk. Our broadest measure of credit risk, the yield spread of bonds rated CCC or lower and 10-year Treasury notes, fell to a probable record low in the spring of 2007, although only marginally so (figure 5). Almost all market participants of my acquaintance were aware of the growing risks, but also cognizant that risk had often remained underpriced for years. I had raised the specter of "irrational exuberance" over a decade before (Greenspan 1996), only to watch the dot-com boom, after a one-day stumble, continue to inflate for 4 more years, unrestrained by a cumulative increase of 350 basis points in the federal funds rate from 1994 to 2000. Similarly in 2002, I expressed my concerns before the Federal Open Market Committee (FOMC) that ". . . our

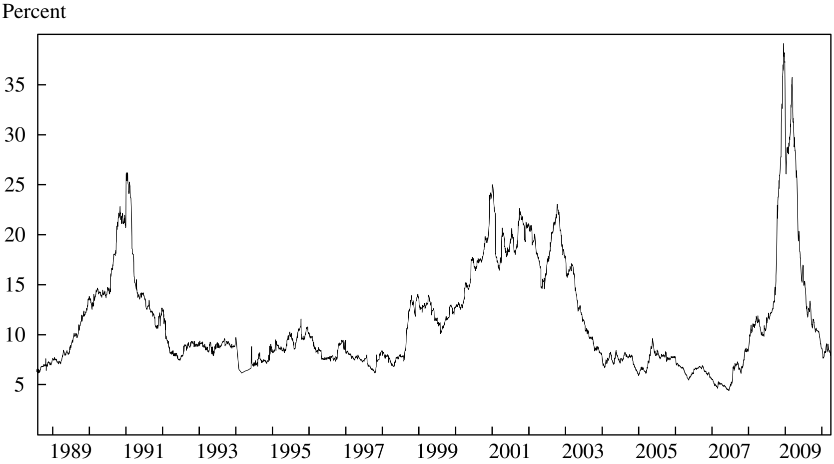
12. Data are from the Mortgage Bankers Association (Haver Analytics).

13. We at the Federal Reserve were aware earlier in the decade of incidents of some highly irregular subprime mortgage underwriting practices. But regrettably, we viewed it as a localized problem subject to standard prudential oversight, not the precursor of the securitized subprime mortgage bubble that was to arise several years later.

14. Inside Mortgage Finance Publications, *The 2009 Mortgage Market Statistical Annual,* vol. I, p. 4, and vol. II, p. 13.

15. These covenants are restrictions put on a borrower by a lender that might, for example, restrict other borrowings, the level of working capital, or debt service cover.

Figure 5. Yield Spread of Bonds Rated CCC and Lower over 10-Year Treasury Notes, 1988–2010[a]



Source: Bank of America Merrill Lynch, Federal Reserve.
a. Average yield on Bank of America Merrill Lynch high-yield cash pay bonds rated CCC and lower minus yield on 10-year Treasury notes at constant maturity.

extraordinary housing boom . . . financed by very large increases in mortgage debt, cannot continue indefinitely." It lasted until 2006.[16]

Clearly, with such experiences in mind, financial firms were fearful that should they retrench too soon, they would almost surely lose market share, perhaps irretrievably. Their fears were given expression in Citigroup chairman and CEO Charles Prince's now-famous remark in 2007, just before the onset of the crisis: "When the music stops, in terms of liquidity, things will be complicated. But as long as the music is playing, you've got to get up and dance. We're still dancing."[17]

The financial firms accepted the risk that they would be unable to anticipate the onset of crisis in time to retrench. They believed, however, that the seemingly insatiable demand for their array of exotic financial products would enable them to sell large parts of their portfolios without loss. They

16. The failure to anticipate the length and depth of the emerging bubble should not have come as a surprise. Although we like to pretend otherwise, policymakers, and indeed forecasters in general, are doing exceptionally well if we can get market projections essentially right 70 percent of the time. But that means we get them wrong 30 percent of the time. In 18½ years at the Federal Reserve, I certainly had my share of the latter.

17. Michiyo Nakamoto and David Wighton, "Citigroup Chief Stays Bullish on Buy-Outs," *Financial Times,* July 9, 2007.

were mistaken. They failed to recognize that the conversion of balance sheet liquidity to effective demand is largely a function of the degree of risk aversion.[18] That process manifests itself in periods of euphoria (risk aversion falling below its long-term, trendless average) and fear (risk aversion rising above its average). A lessening in the intensity of risk aversion creates increasingly narrow bid-asked spreads, in volume—the conventional definition of *market,* as distinct from balance sheet, liquidity.

In this context I define a bubble as a protracted period of falling risk aversion that translates into capitalization rates falling measurably below their long-term, trendless averages.[19] Falling capitalization rates in turn propel one or more asset prices to unsustainable levels. All bubbles burst when risk aversion reaches its irreducible minimum, that is, when credit spreads approach zero, although success at timing the onset of the deflation has proved elusive.

Some bubbles burst without severe economic consequences—the dot-com boom and the rapid run-up of stock prices in the spring of 1987, for example. Others burst with severe deflationary consequences. That class of bubbles, as Carmen Reinhart and Kenneth Rogoff (2009) demonstrate,

18. I am defining risk aversion more broadly here than the standard economic definition, which states it in terms of utility over different outcomes. Risk aversion, as I use the term, encompasses all factors that govern individuals' willingness to engage in risky actions. Most notably, it encompasses not only their preferences toward risk, but also their perceptions of risk.

Risk aversion is the primary human trait that governs the pricing of income-earning assets. When people become uncertain or fearful, they disengage from perceived risk. When their uncertainty declines, they take on new commitments. Risk aversion, by definition, ranges from zero to full.

The extremes of zero and full risk aversion, of course, are outside all human experience. Zero risk aversion—that is, the absence of any aversion at all to engaging in risky actions—implies that an individual does not care about, or cannot discriminate among, objective states of risk to life and limb. Such individuals cannot (or do not choose to) recognize life-threatening events.

To acquire food, shelter, and the other necessary contributors to life requires action, that is, the taking of risks, either by an individual or by others on the individual's behalf. Eschewing all objective risk is not consistent with life. Thus full risk aversion, like zero risk aversion, is a hypothetical state that is never observed in practice.

Day-by-day existence occurs well within these outer boundaries of risk aversion and can be very approximately measured by credit risk spreads. Credit spreads that very approximately track changing risk aversion exhibit little to no long-term trend. Prime railroad bonds of the immediate post–Civil War years reflect spreads over U.S. Treasuries that are similar to the post–World War II experience.

19. Yields on long-term Treasuries, a proxy for riskless capitalization rates, are essentially trendless. Real yields in recent years are not far from the nominal Treasury bond yields of 1900, when long-term inflation expectations (under the gold standard) were effectively zero.

appears to be a function of the degree of leverage in the financial sector, particularly when the maturity of debt is less than the maturity of the assets it funds.

Had the share of financial assets funded by equity been significantly higher in September 2008, it seems unlikely that the deflation of asset prices would have fostered a default contagion much, if at all, beyond that of the dot-com boom. It is instructive in this regard that since the start of the crisis, no unaffiliated hedge fund has defaulted on its debt, despite very large losses that often forced fund liquidation.

II.D. Why Did the Boom Reach Such Heights?

Why did the 2007 bubble reach century-rare euphoria? The answer, I believe, lies with the dot-com bubble, which burst with very little footprint on global GDP and, in the United States, produced the mildest recession in the post–World War II period. The previous U.S. recession, in 1990–91, was the second most shallow. Coupled with the fact that the 1987 stock market crash left no visible impact on GDP, this experience led the Federal Reserve and many a sophisticated investor to believe that future contractions would also prove no worse than a typical postwar recession.

The need for large bank capital buffers appeared increasingly less pressing in this period of Great Moderation. As late as April 2007, the IMF noted that "global economic risks [have] *declined* since . . . September 2006. . . . The overall U.S. economy is holding up well . . . [and] the signs elsewhere are very encouraging" (emphasis in original).[20] The banking regulations adopted internationally under the Basel Accords did induce a modest increase in capital requirements leading up to the crisis. But the debates in Basel over the pending global capital accord that emerged as Basel II were largely over whether to keep bank capital requirements unchanged or to reduce them. Leverage accordingly ballooned.

It is in such circumstances that we depend on our highly sophisticated global system of financial risk management to contain market breakdowns. How could it have failed on so broad a scale? The paradigm that spawned several Nobel Prize winners in economics—Harry Markowitz, Robert Merton, and Myron Scholes (and Fischer Black, had he lived)—was so thoroughly embraced by academia, central banks, and regulators that by 2006 it had become the core of the global regulatory standards embodied in Basel II. Many quantitative investment firms whose number crunching sought to expose profitable market trading principles were successful so

20. IMF, *World Economic Outlook,* April 2007, p. xii.

long as risk aversion moved incrementally (which it did much of the time). But crunching data that covered only the last 2 or 3 decades did not yield a model that could anticipate a crisis.

Mathematical models that calibrate risk, however, are surely better guides to risk management than the "rule of thumb" judgments of a half century ago. To this day it is hard to find fault with the *conceptual* framework of our models, as far as they go. Black and Scholes' elegant option pricing proof is no less valid today than a decade ago. The risk management paradigm nonetheless harbored a fatal flaw.

In the growing state of high euphoria, risk managers, the Federal Reserve, and other regulators failed to fully comprehend the underlying size, length, and impact of the negative tail of the distribution of risk outcomes that was about to be revealed as the post-Lehman crisis played out. For decades, with little to no data, most analysts, in my experience, had conjectured a far more limited tail risk. This assumption, arguably, was the major source of the critical risk management system failures.

Only modestly less of a problem was the vast and, in some cases, virtually indecipherable complexity of the broad spectrum of financial products and markets that developed with the advent of sophisticated mathematical techniques to evaluate risk.[21] In despair, investment managers subcontracted an inordinately large part of their task to the "safe harbor" risk designations of the credit rating agencies. No further judgment was required of investment officers who believed they were effectively held harmless by the judgments of these government-sanctioned rating organizations. But despite their decades of experience, the analysts at the credit rating agencies proved no more adept at anticipating the onset of crisis than the investment community at large.

Even with the breakdown of our sophisticated risk management models and the failures of the credit rating agencies, the financial system would have held together had the third bulwark against crisis—our regulatory system—functioned effectively. But under crisis pressure, it too failed. Along with the vast majority of market participants, regulators failed to anticipate the onset of crisis.

The heavily praised U.K. Financial Services Authority was unable to anticipate, and thus to prevent, the bank run that threatened one of that country's largest commercial banks, Northern Rock. The venerated credit

21. I often maintained that because of this complexity, policymakers had to rely on an international "invisible hand" to bring equilibrium to such undecipherable markets. The high level of market liquidity appeared, erroneously, to confirm that the system was working.

rating agencies bestowed ratings that implied triple-A future smooth sailing for many a highly toxic derivative product. The Basel Committee on Banking Supervision, representing regulatory authorities from the world's major financial systems, promulgated a set of capital rules that failed to foresee the need that arose at the height of the crisis for much larger capital and liquidity buffers. The Federal Deposit Insurance Corporation had noted as recently as the summer of 2006 that "more than 99 percent of all insured institutions met or exceeded the requirements of the highest regulatory capital standards."[22] U.S. commercial and savings banks are extensively regulated, and even though for years our 10 to 15 largest banking institutions have had permanently assigned on-site examiners to oversee daily operations, many of these banks still were able to take on toxic assets that brought them to their knees.

III. Financial Intermediation

III.A. The Purpose of Finance

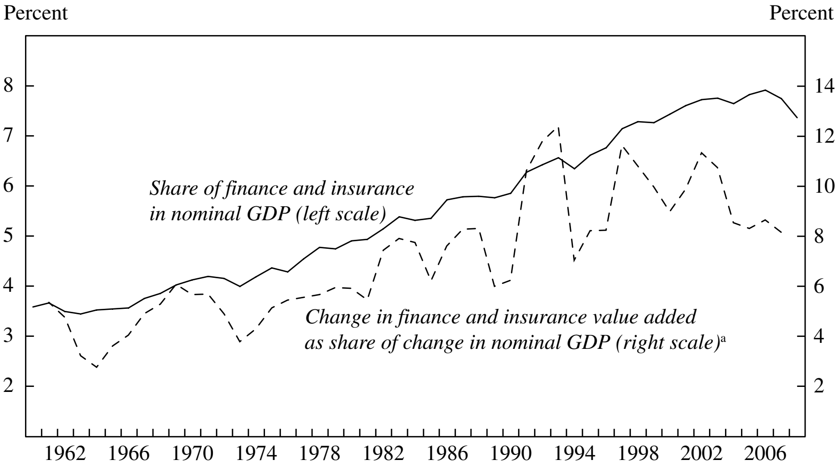
The ultimate goal of a financial system and its regulation in a market economy is to direct the nation's saving, plus any saving borrowed from abroad (the current account deficit), toward investments in plant, equipment, and human capital that offer the greatest increases in the nation's output per worker hour. Nonfinancial output per hour, on average, rises when obsolescent facilities (with low output per hour) are replaced with facilities that embody cutting-edge technologies (with high output per hour). This process improves average standards of living for a nation as a whole. In the United States, the evident success of finance in the decades before the crisis in directing scarce savings into real productive capital investments appears to explain the generous compensation that nonfinancial market participants had been willing to pay to the domestic producers of financial services.

The share of U.S. gross domestic product accruing as income to finance and insurance, according to the Bureau of Economic Analysis, rose fairly steadily from 2.3 percent in 1947 to 7.9 percent in 2006 (figure 6). Many other global financial centers exhibit similar trends.[23] Only a small part of the rise in the United States represented an increase in net foreign demand

22. *FDIC Quarterly Banking Profile,* 2nd Quarter 2006, p. 3.

23. Increased, but less pronouncedly so, financial shares are evident in the United Kingdom, the Netherlands, Japan, Korea, and Australia, among others. The world's most rapidly expanding (and increasingly market-oriented) economy, China, reports a rise in financial intermediaries' share of GDP from 1.6 percent in 1980 to 5.4 percent in 2008.

Figure 6. Share of the Financial Sector in GDP, 1960–2008



Source: Bureau of Economic Analysis.
a. Three-year moving averages.

for U.S. financial and insurance services.[24] The decline in the share to 7.4 percent in 2008 reflects write-offs of savings previously presumed to be productively employed.[25]

Given the historic breakdown of the last 2 years, did nonfinancial market participants over the decades misread the efficiency of finance and inappropriately compensate this small segment of our economy? The prevalence of so many financial product failures certainly suggests so, for the decade leading up to the crisis. Nonetheless, it is difficult to make the same judgment in the face of the fairly persistent rise of finance's share for the previous half century. Moreover, finance's share of *growth* in nominal GDP has been largely trendless since 1990, averaging about 10 percent (figure 6).

The proportion of nonfarm employment accounted for by finance and insurance since 1947 has risen far less than the share of gross income originating in that sector, implying a significant upgrading of the skills attracted to finance and their compensation. A recent study (Philippon and Reshef 2009) finds a markedly above-average rise in the salaries of those

24. The net foreign demand for U.S. financial services has grown significantly but has been largely offset by net imports of insurance services.

25. The share of national income originating in a somewhat broadened measure of finance was little changed in 2009 from 2008.

employed in finance since 1980, presumably reflecting the greater skills drawn to finance in recent years. By 2007 a quarter of all graduates of the venerable California Institute of Technology were entering finance.[26]

What are we to make of these extraordinarily persistent and stable uptrends? Are they wholly accidental? (After all, there is no evidence of such a trend in the prewar years.) It is not that the value of assets to be managed has been persistently rising relative to GDP.[27] The answer to this question matters a great deal.

In the context of financial reform, the critical issue that must be addressed is whether the growing share of financial services was happenstance, or evidence that a growing share of financial services was *required* to intermediate an ever more complex division of labor. I raise the issue because many recent policy recommendations would lower the share of financial services income in GDP. Would such policies affect the growth of U.S. nonfinancial productivity and our standards of living? More important, given the recent failures of risk management and regulation, would increased financial regulation at this time thwart or (through increased stability) enhance economic growth? We need a far deeper understanding of the role of financial intermediation in promoting growth to answer that question. How finance evolves in the postcrisis years should bring clarity to many of today's uncertainties.

III.B. Risky Financial Intermediation

As I noted earlier, the shape of the distribution of the extreme negative tail risk was unknown before the default of Lehman. Since tail risk, in principle at least, is open-ended,[28] there will always be *some* risk that bank capital cannot cover, and hence some, perhaps even many, banks will fail. But that need not become a systemic problem if equity capital and liquidity requirements are raised substantially and a significant part of an intermediary's debt takes the form of mandated contingent capital bonds (see section IV.F). Still, there will always be the possibility, however remote, of the private financial intermediary system faltering, requiring sovereign credit to keep vital intermediation functioning.

26. *The Economist,* "Number-Crunchers Crunched," February 13, 2010, p. 568.

27. Household net worth can be taken as a proxy for the net worth of the economy to be managed at a fee. The ratio of that net worth to disposable personal income was largely unchanged between 1952 and 1996. Since then it has been volatile, with recent quarters returning to the long-term average.

28. Tail risk would converge to zero only if risk aversion were to become absolute, an impossibility if life is to be sustained (see note 18).

Central bankers have long been aware of the potential for a breakdown in private financial markets. Indeed, in the United States as recently as 1991, in contemplation of the unthinkable and at the urging of the Federal Reserve Board of Governors, Section 13-3 of the Federal Reserve Act was reconsidered and amended by Congress. The section as revised grants virtually unlimited authority to the Board to lend in "unusual and exigent circumstances."

III.C. The Hundred-Year Flood

A decade ago, addressing that issue, I noted,

> There is a . . . difficult problem of risk management that central bankers confront every day, whether we explicitly acknowledge it or not: How much of the underlying risk in a financial system should be shouldered [solely] by banks and other financial institutions? . . . [Central banks] have all chosen implicitly, if not in a more overt fashion, to set our capital and other reserve standards for banks to guard against outcomes that exclude those once or twice in a century crises that threaten the stability of our domestic and international financial systems.
>
> I do not believe any central bank explicitly makes this calculation. But we have chosen capital standards that by any stretch of the imagination cannot protect against all potential adverse loss outcomes. There is implicit in this exercise the admission that, in certain episodes, problems at commercial banks and other financial institutions, when their risk-management systems prove inadequate, will be handled by central banks. At the same time, society on the whole should require that we set this bar very high. Hundred-year floods come only once every hundred years. Financial institutions should expect to look to the central bank only in extremely rare situations. (Greenspan 2000a)

At issue is whether the crisis that arrived a few years later is that "hundred-year flood." At best, once-in-a-century observations yield results that are scarcely robust. But recent evidence suggests that what happened in the wake of the Lehman collapse is likely the most severe global financial crisis ever. In the Great Depression, of course, the collapse in economic output and the rise in unemployment and destitution far exceeded the current and, in the view of most, prospective future state of the global economy. And of course, the widespread bank failures markedly reduced short-term credit availability. But short-term financial markets continued to function.

Financial crises are characterized by a progressive inability to float first long-term debt and eventually short-term and overnight debt as well. Long-term uncertainty and therefore risk are always greater than near-term risk, and hence risk spreads almost always increase with the maturity of the financial instrument in question.[29] The depth of a financial crisis is

29. Yields on riskless longer maturities can fall below short-term riskless rates if tight money persuades investors that future inflation will be less.

properly measured by the degree of collapse in the availability of short-term credit.

The evaporation of the global supply of short-term credits within hours or days of the Lehman failure is, I believe, without historical precedent. A run on money market mutual funds, heretofore perceived to be close to riskless, was under way within hours of the announcement of Lehman's default.[30] Within days, the withdrawal of trade credit set off a spiral of global economic contraction, and the Federal Reserve had to move quickly to support the failing commercial paper market. Even the almost sacrosanct, fully collateralized repurchase agreement market encountered severe and unprecedented difficulties.

One has to dig very deep into peacetime financial history to uncover similar episodes. The market for call money, the key short-term financing vehicle of a century ago, shut down at the peak of the 1907 panic, "when no call money was offered at all for one day and the [bid] rate rose from 1 to 125%" (Homer and Sylla 1991, p. 340). Even at the height of the 1929 stock market crisis, the call money market functioned, although annual interest rates did soar to 20 percent. In lesser financial crises, availability of funds in the long-term market disappeared, but overnight and other short-term markets continued to function.

The withdrawal of overnight money represents financial stringency at its maximum. Investors will be willing to lend overnight before they feel sufficiently protected by adequate capital to reach out for more distant, and hence riskier, maturities.

The evaporation in September 2008 of short-term credits was global and all encompassing. But it was the same process we had previously observed at a more micro level.[31]

IV. Regulatory Reform

IV.A. Principles of Reform

Given this apparently unprecedented period of turmoil, by what standard should proposals for reform of official supervision and regulation be judged? I know of no form of economic organization based on a division

30. Hugo Bänziger, "Money Market Funds Need New Global Standards," *Financial Times,* November 5, 2009. Bänziger was chief risk officer at Deutsche Bank at the time.

31. As the credit of New York City, for example, became suspect in the mid-1970s, the first failure of issuance was evident in long-term municipal bonds, followed by failures in progressively shorter maturities, until even overnight markets started to crumble. A similar progression led up to the Mexican financial crisis of 1994–95.

of labor, from unfettered laissez-faire to oppressive central planning, that has succeeded in achieving both maximum sustainable economic growth and permanent stability. Central planning certainly failed, and I strongly doubt that stability is achievable in capitalist economies, given that always-turbulent competitive markets are continuously being drawn toward, but never quite achieving, equilibrium (and that it is precisely this process that leads to economic growth).

People acting without forethought cannot be productive except by chance. Identification of effective innovation is, of necessity, a rational act. Hence, regulation, by inhibiting irrational behavior when it can be identified, can be stabilizing, as recent history has demonstrated. But there is an inevitable cost of regulation in terms of economic growth and standards of living when it imposes restraints beyond containing unproductive behavior.

Regulation by its nature imposes restraints on competitive markets. The elusive point of balance between growth and stability has always been a point of contention, especially when it comes to financial regulation.

Throughout the postwar years in the United States, with the exception of a limited number of bank bailouts (Continental Illinois in 1984, for example), private capital proved adequate to cover virtually all provisions for lending losses. As a consequence, there was never a definitive test of what then constituted conventional wisdom, namely, that an equity capital–to–assets ratio of 6 to 10 percent on average, the range that prevailed between 1946 and 2003, was adequate to support the U.S. banking system.

Risk managers' assumption of the size of the negative tail of the distribution of credit and interest rate risk was, as I noted earlier, of necessity conjectural, and for generations we never had to test those conjectures. Most of the shape of the distribution of perceived risk was thoroughly documented in the precrisis years, as "moderate" financial crises and euphorias traced out their relevant parts of the curve. But since modern financial data compilation began, we had never had a "hundred-year flood" that exposed the full intensity of negative tail risk.

Risk managers, of course, knew in earlier decades that an assumption of normality in the distribution of risk was unrealistic, but as a first approximation that greatly facilitated calculation, it prevailed. The mathematics implied by fat tails was also well understood, but our number crunching capabilities fell far short of making the required calculations to guide actions, except at prohibitive cost. That is no longer the case.

Clearly what we experienced in the weeks following the Lehman default is exactly the type of market seizure that tail risk conjecture was supposed to capture, and did not. Having experienced Lehman, risk managers will be far more cautious in evaluating future risk—at least for a while.

Many investment firms are constructing probability distributions of outcomes employing, as the negative tail, data based on the experience of the last 2 years. Using Monte Carlo simulations or other techniques, they have concluded, not unexpectedly, that a financial crisis as severe as the one that followed the Lehman default would have been predicted to occur far more often than indicated by models in which risk is distributed normally. Such evidence suggests the onset of a "hundred-year flood" somewhat more often than once in a century.

Indeed, the aftermath of the Lehman crisis traced out a startlingly larger negative tail than almost anybody had earlier imagined. At least partly responsible may have been the failure of risk managers to fully understand the impact of the emergence of shadow banking, a development that increased financial innovation but, as a result, also increased the level of risk. The added risk was not compensated by higher capital.

When risk premiums are low over a protracted period, as they were, for example, from 1993 to 1998 and from 2003 to 2007, investors' willingness to bid for all types of financial assets, especially the high-risk tranches of collateralized debt obligations, creates an illusion of permanent market liquidity that in the latest episode turned out to be intoxicating. It led several major investment banks to attempt to weather the financial storm with only a thin veneer of tangible capital.

The most pressing reform, in my judgment, in the aftermath of the crisis is to fix the level of regulatory risk-adjusted capital, liquidity, and collateral standards required by counterparties. Private market participants are now requiring economic capital and balance sheet liquidity well in excess of the yet-to-be-amended Basel II requirements. The shadow banks that survived the crisis are now having to meet significantly tighter market standards, with respect to capital, liquidity, and collateral, than existed before the crisis. These are major changes that need to be reflected in the new set of regulatory requirements and standards currently undergoing global review.

One major fallout of the crisis is a marked rise in the degree of moral hazard (see note 41), which requires that all financial intermediaries be subject to maximum leverage ratios. These ratios, as with all risk-adjusted capital adequacy measures, need to be based on more realistic risk adjustment factors applied to their assets and on the proportion of their liabilities funded with overnight or other short-term debt. Precrisis regulatory capital requirements, although based on decades of experience, were clearly too lax: for example, they erroneously designated pools of self-amortizing home mortgages as among the safest of private instruments. And a surprisingly and unfortunately large proportion of investment portfolio decisions were, by law, accorded "safe harbor" status if they adhered to the credit risk judgments (or rather, misjudgments) of the credit rating agencies.

To ensure that financial intermediaries have adequate cash to meet ongoing commitments in the event of a shutdown in external funding, international bank liquidity regulation should match the tightening already evident in private risk management paradigms (Basel Committee on Banking Supervision 2009). Collateral has shown itself particularly subject to rapid recapture. Bear Stearns had nearly $20 billion in pledgeable liquid funds a week before it collapsed. Morgan Stanley lost more than a half trillion dollars of pledgeable collateral during the height of the crisis. In the United States, to lower the risk of a "run on the broker," the amount of customer assets (collateral) held by broker-dealers that cannot be commingled with their own assets needs to be increased. That would decrease the amount of funds that can "run." However, such action must be measured and coordinated with other global regulators to avoid regulatory arbitrage (see French and others forthcoming).

Unaffiliated hedge funds have weathered the crisis—as extreme a real-life stress test as one can construct—without taxpayer assistance or, as I noted earlier, default. Although hedge funds are only lightly regulated, much of their leveraged funding comes from more heavily regulated banks. Moreover, as Sebastian Mallaby (2010) writes, "Most hedge funds make money by driving prices *away* from extremes and toward their rational level." In so doing, they supply much-needed liquidity to financial markets when other competitors have withdrawn. Regulations that inhibit the ability of hedge funds to supply such services are counterproductive.

Capital, liquidity, and collateral, in my experience, address almost all of the financial regulatory structure shortcomings exposed by the onset of the crisis. In retrospect, there has to be a level of capital that would have prevented the failure of, for example, Bear Stearns and Lehman Brothers. (If not 10 percent, think 40 percent.) Moreover, generic capital has the regulatory advantage of not having to forecast which particular financial products are about to turn toxic. Certainly investors did not foresee the future of subprime securities or the myriad other broken products. Adequate capital eliminates the need for an unachievable specificity in regulatory fine tuning.

The jerry-built regulatory structure that has evolved over the decades in the United States has become much too complex. Policymakers failed to recognize, during the debates that led to legislation resulting in a badly needed opening up of financial competition (the Gramm-Leach-Bliley Act of 1999), that increased competition, especially through shadow banking, also increased negative tail risk. And increased negative tail risk necessitates higher capital requirements.

IV.B. Upward Revisions of Bank Economic Capital

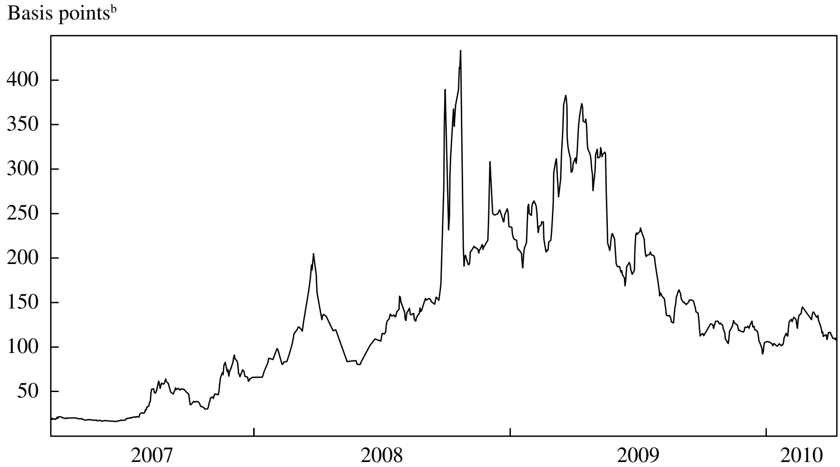
How much capital is currently being required of financial institutions by their counterparties will strongly influence the upcoming revisions in *regulatory* capital requirements. It is too soon to have definitive answers. But very rough approximations for U.S. commercial banks can be inferred from the response of bank credit default swaps (CDSs), a measure of bank insolvency risk, to postcrisis events.[32] Movements in the CDS market should also give us some direct insight into when the banking system is perceived to have overcome the market's fear of widespread insolvency—and beyond that, to when markets perceive that banks will feel sufficiently secure to return to the free lending of the precrisis years.

Starting late in 2008 and accelerating into the first quarter of 2009, the U.S. Treasury, through its Troubled Asset Relief Program (TARP), added $250 billion to bank equity, the equivalent of adding approximately 2 percentage points to the equity capital-to-assets ratio. Its impact was important and immediate.

As the financial crisis took hold and deepened, the unweighted average price of 5-year CDSs of six major U.S. banks—Bank of America, JP Morgan, Citigroup, Goldman Sachs, Wells Fargo, and Morgan Stanley—rose from 17 basis points in early 2007 (for 5-year contracts, the average annual price of insurance was 0.17 percent of the notional amount of the underlying swap instruments) to 170 basis points just before the Lehman default on September 15, 2008. In response to the Lehman default, the 5-year CDS average price rose to more than 400 basis points by October 8. On the *day* the TARP was announced (October 14), the price fell to approximately 200 basis points, or essentially by half (figure 7). That a 2-percentage-point addition to the banks' book equity capital-to-assets

32. The seller of a CDS insures the holder of a particular debt instrument against loss in the event of default. Prices of CDSs are thus the most sensitive measure of the probability of bank default.

Figure 7. Price of Five-Year Credit Default Swaps[a]



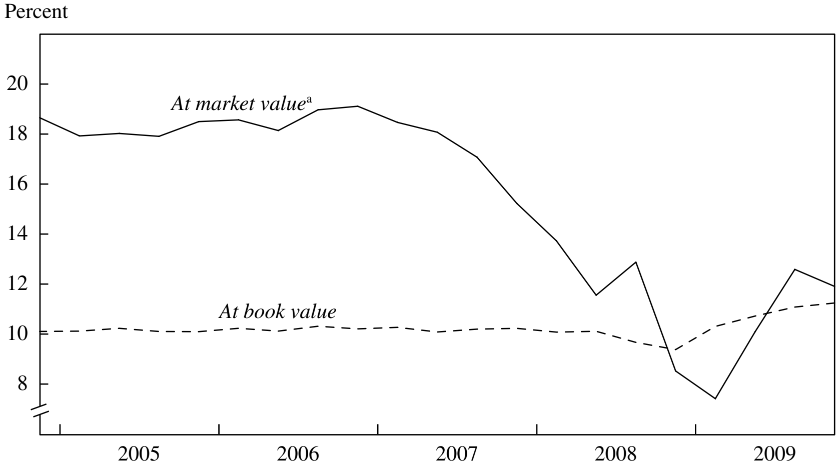
Source: Author's calculations; Bloomberg.
a. Unweighted average prices of CDSs issued by Bank of America, Citigroup, Goldman Sachs, JPMorgan, Wells Fargo, and Morgan Stanley.
b. Hundredths of a percent of the notional value of the underlying swap contract.

ratio reversed roughly half the crisis surge in the price of 5-year CDSs implies an overall additional 4-percentage-point rise (from 10 percent in mid-2007 to 14 percent) in the equity capital cushion required by market participants to fund the liabilities of banks. That, of course, assumes linear extrapolation, an admittedly herculean assumption, and, of course, presumes that the probability of a TARP before the Lehman default was de minimis. The abruptness of the market reaction to the TARP announcement appears to confirm such a presumption, however.

Current book equity–to–assets ratios are still far from 14 percent. The average ratio for commercial banks (as reported by the Federal Deposit Insurance Corporation, FDIC) was 10.9 percent on March 31, 2010, compared with 10.1 percent in mid-2007. But unacknowledged loan losses were estimated by the IMF last October (they are now less) to be in the hundreds of billions of dollars. Trends in relevant liquidity are less readily measured but are assumed to parallel changes in capital.

That banks still have more equity capital to add is also indicated by the fact that the 5-year CDS price of March 31, 2010 (and since) remains over 100 basis points, still significantly elevated relative to the 17 basis points that prevailed in early 2007, when 10 percent capital was apparently

Figure 8. Equity-to-Assets Ratios at FDIC-Insured Commercial Banks, 2004–09



Source: Federal Deposit Insurance Corporation.
a. Averages constructed from Bloomberg data on the book and market equity value of 24 leading banks.

enough to virtually eliminate the threat of default and induce loan officers to lend freely.

There is little doubt that the TARP's cash injection markedly reduced the fear of bank default through early 2009. More difficult to judge is the impact on bank CDSs of the dramatic increase in bank equity at *market* value relative to bank assets at market value. That ratio rose 4.5 percentage points from the end of March 2009 to the end of December, from 7.4 percent to 11.9 percent (figure 8). There can be little doubt that this has materially increased the solvency of banks, although apparently less effectively, dollar for dollar, than the more permanent change in book-value equity.[33]

Much of the repayment of TARP investments to the U.S. Treasury was doubtless financed by new equity issuance, made possible by a more than one-half-trillion dollar increase in U.S. commercial bank equity at market value, and by borrowings made much easier (and cheaper) by the increased equity buffer engendered by gains in market-valued bank equity. The parceling of relative contributions of the TARP and of capital gains to bank solvency and willingness to lend may not be fully clear even in retrospect.

33. Between the end of March and the end of December 2009, the average CDS price fell from 369 to 104 basis points, while the ratio of the market value of equity to the market value of assets rose 450 basis points.

Table 2. CDS and LIBOR-OIS Spreads at Various Maturities, September 2009 and March 2010
Basis points

Maturity	*September 15, 2009*	*March 31, 2010*
CDS		
10 years	129	111
5 years	125	107
3 years	129	88
1 year	123	61
LIBOR-OIS		
3 months	12	11
1 month	7	8

Sources: British Bankers' Association, Bloomberg, Reuters, and Haver Analytics.

The TARP not only inserted capital but also induced market participants to infer that the U.S. Treasury would, at least for a while, stand behind the liabilities of the banking system. This may explain the divergence since mid-September 2009 between short-term (1- and 3-month) LIBOR-OIS spreads (an alternative to CDS spreads as a short-term measure of the likelihood of bank default) and 5- and 10-year CDS spreads. Short-term LIBOR-OIS spreads had returned to their precrisis level by the end of September 2009. Long-maturity CDS prices are only partway back (table 2). The 1-year LIBOR-OIS spread falls in between. Clearly, either markets are discounting some of the bank capital cushion at market value 5 and 10 years hence, owing to the volatility of stock prices, and/or they question the political willingness, or ability, of the U.S. government, after markets return to normal, to initiate another bank bailout.[34]

Given the foregoing set of fragile assumptions and conclusions (and they are all we have), I would judge that regulatory equity capital requirements in the end will be seen to have risen from the 10 percent precrisis level (in terms of book value) to 13 or 14 percent by 2012, and liquidity and collateral requirements will toughen commensurately.

IV.C. What Regulation Can Do

What, in my experience, supervision and examination *can* do as a backup to capital requirements and counterparty surveillance is promulgate

34. As fear of contagion from the European sovereign debt crisis mounted in the spring of 2010, CDS and LIBOR-OIS spreads rose markedly.

rules that are preventative and *do not require anticipating an uncertain future.* Supervision

—can audit and enforce capital and liquidity requirements[35]

—can require that financial institutions issue some debt that will become equity should equity capital become impaired (see section IV.F)

—can, and has, put limits or prohibitions on certain types of concentrated bank lending

—can prohibit complex affiliate and subsidiary structures whose sole purpose is tax avoidance or regulatory arbitrage

—can inhibit the reconsolidation of affiliates previously sold to investors, especially structured investment vehicles (SIVs)[36]

—can require "living wills" in which financial intermediaries indicate, on an ongoing basis, how they can be liquidated expeditiously with minimum impact on counterparties and markets.

IV.D. Some Lessons of Regulatory Capital History

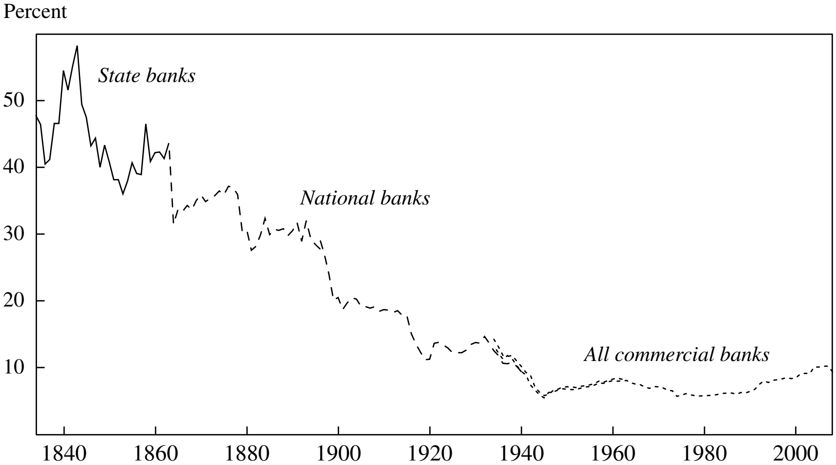
In the late 19th century, U.S. banks required equity capital of 30 percent of assets to attract the liabilities required to fund their assets. In the pre–Civil War period, that figure topped 50 percent (figure 9). Given the rudimentary nature of 19th-century payment systems and the poor geographical distribution of reserves in what was then an agricultural economy, competition for bank credit was largely local. It enabled national banks on average to obtain returns (net income) on their assets of well over 200 basis points in the late 1880s, and probably more than 300 basis points in the 1870s (compared with 70 basis points a century later).

Increasing efficiency of financial intermediation, owing to consolidation of reserves and improvements in payment systems, exerted competitive pressure on profit spreads to narrow and allowed capital-to-assets ratios to decline. In marked contrast, the annual average net income rate of return on *equity* was amazingly stable, rarely falling outside a range of 5 to 10 percent, measured annually, during the century from 1869 to 1966 (figure 10). That meant that net income as a percentage of assets and the degree of leverage were approximately inversely proportional during that century.

35. Increased capital requirements can go a long way toward containing large compensation packages. The recent higher profits will be needed to fulfill the capital requirements, especially if global competitors have similar capital requirements.

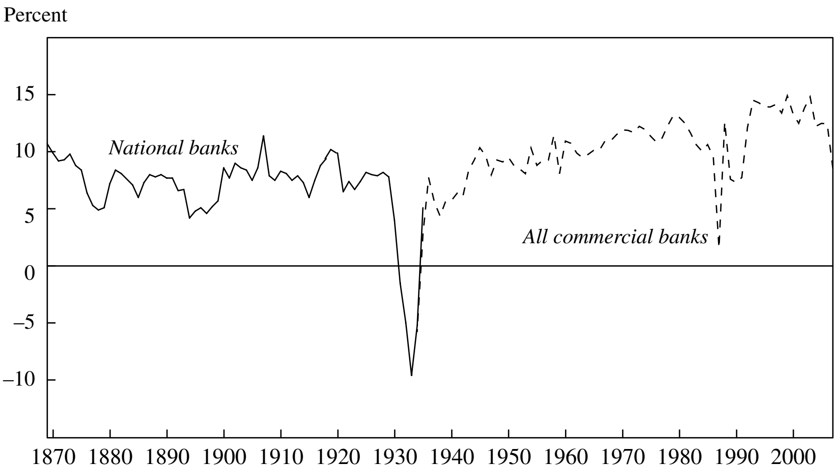
36. When, during the crisis, such assets appeared about to fail, sponsoring companies, fearful of reputation risk (a new insight?), reabsorbed legally detached affiliates at subsequent great loss.

Figure 9. Ratio of Equity Capital to Assets in the Banking Sector, 1834–2009



Source: Federal Deposit Insurance Corporation and Office of the Comptroller of the Currency.

Figure 10. Ratio of Net Income to Equity in the Banking Sector, 1869–2007



Source: Federal Deposit Insurance Corporation and Office of the Comptroller of the Currency.

Rates of return on assets and equity (despite the decline in leverage that resulted from rising Basel capital requirements) moved modestly higher during 1966–82 owing to a rapid expansion in noninterest income, for example from fiduciary activities, service charges and fees, and securitizations (and later from expansion into investment banking and brokerage). Noninterest income rose significantly between 1982 and 2006, increasing net income to nearly 15 percent of equity, as a consequence of a marked increase in the scope of bank powers. That increase in part reflected the emergence in April 1987 of court-sanctioned and Federal Reserve–regulated "Section 20" investment banking affiliates of bank holding companies.[37] The transfer of such business is clearly visible in the acceleration of gross income originating in commercial banking relative to that in investment banking starting in 2000 (Bureau of Economic Analysis).[38]

I tentatively conclude that the historical relative stability of average net income–to–equity ratios dating back to the post–Civil War years reflects an underlying ex ante market-determined rate of return on intermediation.

In summary, the crisis will leave in its wake a significantly higher capital-to-assets ratio requirement, both economic and regulatory, that must be reached if intermediation is to be restored to the point where banks and other financial institutions are confident they have a sufficiently secure capital cushion to lend freely.

IV.E. Limits to Regulatory Capital Requirements

If we accept as a benchmark the remarkable stability of the ratio of bank net income to equity capital (ranging between 5 and 15 percent) that has prevailed, with rare exceptions, since the end of the Civil War (figure 10), we can infer the highest average ratio of capital to assets that a banking system can tolerate before a significant number of banks are required to raise their margin, or shrink their size, or both. I assume a 5 percent annual

37. This development meant that the repeal, under the Gramm-Leach-Bliley Act, of the 1933 Glass-Steagall Act, which had separated commercial and investment banking, changed very little. From the enactment of Gramm-Leach-Bliley in 1999 to the Federal Reserve's acceptance of Goldman Sachs and Morgan Stanley as financial services holding companies at the height of the crisis, no applications to employ the greater powers were forthcoming. That forbearance apparently reflected a desire to stay clear of the Federal Reserve's regulatory embrace.

38. Rates of return crashed during the first half of 2009, with declines matched (on an annual basis) only by those in the depression years 1932–34. Both cases reflected a rare sharp breakout from the historical range, resulting mostly from large write-offs on previously extended loans.

average rate of return (the lower limit of the range) as a proxy for the full distribution of the thousands of banks that would make up the average. Accordingly, for this exercise it is employed as the ex ante competitively required average minimum return on intermediation. I assume as a first approximation that all variables are independent. If so, the *highest* ratio of capital to assets that the U.S. banking system can tolerate and still supply the nonfinancial sector with adequate financial service capacity can be inferred from the following identity:

$$\frac{\pi}{C} = \frac{\pi}{A} \times \frac{A}{C},$$

where π is net income, C is equity capital, and A is total assets. If $\pi/C = 0.05$, then $\frac{C}{A} = 20 \times \frac{\pi}{A}$.

It can be shown that $\pi/A = (r_r - r_p - k)w + n - e - \alpha$, where r_r is the rate of interest received from earning assets, r_p is the interest rate paid on earning assets, k is the ratio of losses to earning assets, w is the ratio of earning assets to total assets, n is the ratio of noninterest income to assets, e is the ratio of noninterest expense to total assets, and α is the ratio of taxes and minor other adjustments to total assets. As can be seen from table 3, virtually all of the rise in π/A and π/C for U.S. banks as a group since 1982 is due to the marked rise in noninterest income.

In the years immediately before the onset of the crisis, π/A averaged 0.012, and therefore the inferred maximum average regulatory capital, C/A, as a first approximation, was 0.24. A rate higher than 0.24, all else equal,[39] would put the average rate of return on equity below the critical 5 percent level. If π/A were to revert back to its average for 1950–75 (0.0074), then $C/A = 0.15$, marginally above the 12 to 14 percent presumed market-determined capital requirement that would induce banks to lend freely.

These calculations, as I noted, assume a static model in which all variables are independent. But clearly the required rate of return on equity cannot be independent of the capital-to-assets ratio. Increased capital

39. I do not deny that all else is not equal, and hence such conclusions are more illustrative than explanatory. A dynamic model is beyond the scope of this paper. Net interest income has enough of a history to effectively model, but noninterest income arguably does not.

Table 3. Accounting for Net Income in the U.S. Banking System

Level or change from previous period	*Net income divided by total assets* π/A	*Interest rate spread × share of earning assets in total assets* $(r_r - r_p) \times w$	*Loan-loss provisions and allocated transfer risk divided by earning assets × share of earning assets in total assets* $k \times w$	*Noninterest income divided by total assets* n	*Noninterest expense divided by total assets* e	*Taxes plus minor items, net, divided by total assets* α
Period average						
1962–66	0.766	2.580	0.079	0.566	1.965	0.335
1978–82	0.728	3.035	0.265	0.774	2.546	0.269
1992–96	1.092	3.673	0.426	1.949	3.617	0.488
2002–06	1.276	3.048	0.399	2.296	3.106	0.563
Change						
From 1962–66 to 1978–82	–0.038	0.455	0.186	0.207	0.581	–0.067
From 1978–82 to 1992–96	0.363	0.638	0.161	1.176	1.071	0.219
From 1992–96 to 2002–06	0.184	–0.624	–0.027	0.347	–0.511	0.075

Source: Author's calculations based on Federal Deposit Insurance Corporation data.

reduces the risk of the balance sheet and hence will attract equity investors despite a lower rate of return. This implies that owing to the recent rise in π/A, the actual regulatory capital ceiling can thus readily exceed the static ceiling of $C/A = 0.24$. In any event, increased capital requirements will surely reduce the marginal lending that occurred in recent decades owing to the failure to fully fund tail risk. Much of that marginal lending was in effect being subsidized by taxpayers. That subsidy became fully funded in 2008 by sovereign credit. Removing the subsidy through higher capital requirements will, of course, shrink financial intermediary balance sheets. Much of this lending was evidently nonproductive, and its loss is not apt to be a problem for our complex economy's required level of intermediation.

IV.F. Too Big to Fail

Beyond significantly increased capital requirements is the necessity of addressing the problem of some financial firms being "too big to fail" or, more appropriately, "too interconnected to be liquidated quickly." The productive employment of the nation's scarce saving is threatened when financial firms at the edge of failure are supported with taxpayer funds and designated as systemically important institutions. I agree with Gary Stern, the former president of the Federal Reserve Bank of Minneapolis, who has long held that "creditors will continue to underprice the risk-taking of these financial institutions, overfund them, and fail to provide effective market discipline. Facing prices that are too low, systemically important firms will take on too much risk" (Stern 2009, p. 56). These firms absorb scarce savings that need to be invested in cutting-edge technologies, if output per hour and standards of living are to continue to rise.

After wallowing in the backwaters of economics for years, "too big to fail" has arisen as a major, visible threat to economic growth. It finally became an urgent problem when Fannie Mae and Freddie Mac were placed into conservatorship on September 7, 2008. Before then, U.S. policymakers (with fingers crossed) could point to the fact that Fannie and Freddie, by statute, were not backed by the "full faith and credit of the U.S. government." Market participants however, did not believe the denial, and they consistently afforded Fannie and Freddie a special credit subsidy (Passmore, Sherlund, and Burgess 2005). On September 7, 2008, market participants were finally vindicated.

Fannie Mae and Freddie Mac need to be split up into smaller companies, none of them "too big to fail," and then reconstructed into stand-alone securitizers. Their future solvency (and the threat of contagion) requires that these GSEs be prohibited from accumulating large portfolios of assets that add no useful backing to the process of securitization or the mortgage markets more generally. Those portfolios' sole purpose is to profit from the subsidy that market participants grant to these GSEs (Greenspan 2004b).

One highly disturbing consequence of the too-big-to-fail problem that has emerged since the September 2008 federal takeover of Fannie Mae and Freddie Mac is that market players will now believe that *every* significant financial institution, should the occasion arise, is subject to being bailed out with taxpayer funds. It is going to be very difficult for legislators to persuade future investors otherwise.

Businesses that are subject to being bailed out have competitive market and cost-of-capital advantages, but not necessarily efficiency advantages, over firms not thought to be systemically important. For years the Federal Reserve was concerned about the ever-growing size of our largest financial institutions. Federal Reserve research had been unable to find economies of scale in banking beyond a modest size (Berger and Humphrey 1994, p. 7; see also Berger 1994). A decade ago, citing such evidence, I noted that "megabanks being formed by growth and consolidation are increasingly complex entities that create the potential for unusually large systemic risks in the national and international economy should they fail" (Greenspan 1999). Regrettably, we did little to address the problem.

How to deal with systemically threatening institutions is among the major regulatory problems for which there are no good solutions. Early resolution of bank problems under the Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA) appeared to have worked with smaller banks during periods of general prosperity. But the notion that risks can be identified in a sufficiently timely manner to enable the liquidation of a large failing bank with minimum loss proved untenable during this crisis, and I suspect will prove untenable in future crises as well.[40]

The solution that, in my judgment, has at least a reasonable chance of reversing the extraordinarily large "moral hazard" that has arisen over the

40. The FDIC has experienced large losses in the value of assets taken over in resolution during the last 2 years.

past year and more[41] is to require banks and possibly all financial intermediaries to issue contingent capital bonds, that is, debt that is automatically converted to equity when equity capital falls below a certain threshold. Such debt will, of course, be more costly on issuance than simple debentures.

However, should contingent capital bonds prove insufficient, we should allow large institutions to fail and, if assessed by regulators as too interconnected to liquidate quickly, be taken into a special bankruptcy facility, whereupon the regulator would be granted access to taxpayer funds for "debtor-in-possession financing" of the failed institution. Its creditors (when equity is wholly wiped out) would be subject to statutorily defined principles of discounts from par ("haircuts"), and the institution would then be required to split up into separate units, none of which should be of a size that is too big to fail. The whole process would be administered by a panel of judges expert in finance.

I assume that some of the newly created firms would survive, and others fail. If, after a fixed short period of time, no viable exit from the bank-

41. Moral hazard, in an economic context, arises when an institution is not debited with the full costs of its actions and therefore will tend, in part at least, to act contrary to how it would act were it pressured solely by unfettered competition, where the externalities of potential bailout costs are fully internalized by competitors. The institution accordingly requires other parties to suffer some of the costs of its actions.

An interesting speculation is whether the crisis that emerged in August 2007 from the extraordinary leverage (as much as 20 to 30 times tangible capital) taken on by U.S. investment banks would have occurred had these firms remained the partnerships that they were up to a quarter century ago. The 1970 ruling that allowed broker-dealers to incorporate and gain permanent capital seemed sensible at the time. Nonetheless, as partnerships, Lehman Brothers and Bear Stearns almost surely would not have departed from their historically low leverage. Before incorporation, fearful of the joint and several liability to which general partnerships are subject, those entities shied away from virtually any risk they could avoid. Their core underwriting of new issues rarely exposed them for more than a few days.

To be sure, the senior officers of Bear Stearns and Lehman Brothers lost hundreds of millions of dollars from the collapse of their stocks. But none, to my knowledge, has filed for personal bankruptcy, and their remaining wealth allows them to maintain much of their previous standard of living.

Replicating the incentive structure of partnerships should be a goal whenever feasible in future reform. That goal will doubtless not be always met given that the corporate structure is seen as required to raise capital on a scale perceived as necessary in today's global market. To eliminate moral hazard, it should not be necessary to follow Hugh McCulloch, our first Comptroller of the Currency in 1863, who went somewhat over the edge in proposing that the National Bank Act "be so amended that the failure of a national bank be declared prima facie fraudulent, and that the officers and directors, under whose administration such insolvency shall occur, be made personally liable for the debts of the bank, and be punished criminally, unless it shall appear, upon investigation, that its affairs were honestly administered." Under such a regime, moral hazard surely would not exist.

ruptcy appears available, the financial intermediary should be liquidated as expeditiously as feasible.

IV.G. Regulations Embodying a Forecast Fail with Regularity

The crisis has demonstrated that neither bank regulators nor anyone else can consistently and accurately forecast whether, for example, subprime mortgages will turn toxic, or to what degree, or whether a particular tranche of a collateralized debt obligation will default, or even whether the financial system as a whole will seize up. A large fraction of such difficult forecasts will invariably be proved wrong. Regulators can readily identify underpriced risk and the existence of bubbles, but most important, they cannot, except by chance, effectively time the onset of crisis.[42] This should not come as a surprise.

A financial crisis is defined as an abrupt and sharp decline in the price of income-producing assets, usually induced by a dramatic spike in the discount rate on expected income flows as market participants swing from euphoria to fear. Implicit in any sharp price change is that it is unanticipated by the mass of market participants, for were it otherwise, the price imbalances would have been arbitraged away.

Indeed, for years leading up to August 2007, it was widely expected that the precipitating event of the "next" crisis would be a sharp fall in the dollar, as the U.S. current account deficit, starting in 2002, had increased dramatically. The dollar accordingly came under heavy selling pressure. The rise in the euro-dollar exchange rate from around 1.10 in the spring of 2003 to 1.30 at the end of 2004 appears to have *gradually* arbitraged away the presumed dollar trigger of the "next" crisis. The U.S. current account deficit did not play a prominent direct role in the timing of the 2007 crisis, although because of that, it may in the next.

In the years ahead, forecasters will readily identify risks that are underpriced—or at least priced at less than their historical average. But in instance after instance, as I noted earlier, risk has remained underpriced for years. Forecasters as a group will almost certainly miss the onset of the next financial crisis, as they have so often in the past, and I presume any newly designated "systemic regulator" will also.

Many analysts argue that forecasting is not required. A systemic regulator, they hold, could effectively fine-tune capital and liquidity require-

42. There has been confusion on the issue, to which I may have been a party. With rare exceptions it *has* proved impossible to identify the point at which a bubble will burst, but its emergence and development *are* visible in credit spreads.

ments to match the stage of the business cycle. Properly calibrated, such requirements presumably could be effective in assuaging imbalances. But cycles are not uniform. In real time, where we are in the cycle is a forecast, and cycles vary. For example, the low of the unemployment rate at cyclical peaks (as identified by the National Bureau of Economic Research) since 1948 has ranged between 2.6 and 7.2 percent. Would we have judged a turn in the business cycle when, for example, the unemployment rate rose to 5.8 percent in April 1995, up from 5.4 percent in March? In the event, the unemployment rate soon reversed itself and continued to fall for 5 more years.

It is best to fix regulatory parameters and let monetary policy carry the discretionary load. The Federal Reserve will tighten if it observes rising euphoria that signals mounting inflationary pressures (as it did in February 1994 and June 2004) or if risk premiums fall inordinately.

Moreover, discretionary regulatory rules would raise uncertainties that could undesirably curb investment. Thus, in the current environment of complexity, I see no ready alternative to significantly increasing—and fixing—regulatory capital requirements and liquidity and beefing up individual banks' counterparty risk surveillance.

The Federal Reserve has been concerned for years about the ability of regulatory supervisors and examiners to foresee emerging problems that have eluded internal bank auditing systems and independent auditors. I remarked in 2000 before the American Bankers Association, "In recent years rapidly changing technology has begun to render obsolete much of the bank examination regime established in earlier decades. Bank regulators are perforce being pressed to depend increasingly on greater and more sophisticated private market discipline, the still most effective form of regulation. Indeed, these developments reinforce the truth of a key lesson from our banking history—that private counterparty supervision remains the first line of regulatory defense" (Greenspan 2000b). Regrettably, that first line of defense failed.

A century ago, examiners could appraise individual loans and judge their soundness.[43] But in today's global lending environment, how does a U.S. bank examiner judge the credit quality of, say, a loan to a Russian bank, and hence of the loan portfolio of that bank? That in turn would

43. In 1903, O. Henry (W. S. Porter), who had more than a passing relationship with banking shenanigans, wrote in "A Call Loan" about a fictional bank examiner from the Office of the Comptroller of the Currency who was obsessed with the collateral backing a $10,000 loan. Such detailed scrutiny is exceptionally rare in today's larger banks.

require vetting the Russian bank's counterparties and those counterparties' counterparties, all to judge the soundness of a single financial transaction. In short, a bank examiner cannot, and neither can a credit rating agency. How deep into the myriad layers of examination is enough for certification?

The complexity of our financial system in operation spawns, in any given week, many alleged pending crises that, in the event, never happen, and innumerable allegations of financial misconduct. To examine each such possibility at the level of detail necessary to reach meaningful conclusions would require an examination force many multiples larger than those now in place in any of our banking regulatory agencies. Arguably, at such levels of examination, sound bank lending and its necessary risk taking would be impeded.

The Federal Reserve and other regulators were, and are, therefore required to guess which of the assertions of pending problems or allegations of misconduct should be subject to full scrutiny by a regulatory workforce with necessarily limited examination capacity. But this dilemma means that in the aftermath of an actual crisis, we will find highly competent examiners failing to have spotted a Bernie Madoff. Federal Reserve supervision and evaluation is as good as it gets, even considering the failures of past years. Banks still have little choice but to rely upon counterparty surveillance as their first line of crisis defense.[44]

V. The Role of Monetary Policy

V.A. Monetary Policy and Home Price Bubbles

The global home price bubble of the last decade was a consequence of lower interest rates, but it was *long-term* interest rates that galvanized home asset prices, not the overnight rates of central banks, as has become the seeming conventional wisdom. In the United States, the bubble was driven by the decline in interest rates on fixed-rate long-term mortgage loans,[45] relative to their mid-2000 peak, 6 months before the FOMC began easing the federal funds rate in January 2001.

44. Having served on JP Morgan's board for a decade just before my joining the Federal Reserve, I had an extended insight into the effectiveness of that company's counterparty surveillance of Citicorp, Bank of America, Wells Fargo, and others, relative to the regulatory surveillance by Federal Reserve banks.

45. Their average maturity is more than 26 years (Federal Housing Finance Agency).

Between 2002 and 2005, the monthly fixed-rate mortgage rate closely tracked changes in U.S. home prices 11 months earlier (as measured by the 20-city S&P/Case-Shiller home price index), with an adjusted R^2 on the regression of 0.500 and a *t*-statistic of –6.93. Thus long-term mortgage rates were a far better indicator of home prices than the federal funds rate: a regression of home prices on the latter exhibits an adjusted R^2 of 0.205 and a *t*-statistic of –3.62 with only an 8-month lead.[46] Regressing home prices on *both* the fixed-rate mortgage (with an 11-month lead) and the federal funds rate (with an 8-month lead) yields a highly significant *t*-statistic for the mortgage rate of –5.20, but an insignificant *t*-statistic for the federal funds rate of –0.51.

This should not come as a surprise. After all, the prices of long-lived assets have always been determined by discounting the flow of income (or imputed services) using interest rates on assets of comparable maturity. No one, to my knowledge, employs overnight interest rates—such as the federal funds rate—to determine the capitalization rate of real estate, whether it be the cash flows of an office building or the imputed rent of a single-family residence.

It is understandable why, before 2002, the federal funds rate would have been perceived as a leading indicator of many statistics that in fact are driven by longer-term interest rates. The correlation between the federal funds rate and the rate on fixed-rate mortgage loans from 1983 to 2002, for example, had been a tight 0.86.[47] Accordingly, during those years, regressions with home prices as the dependent variable would have seemingly worked equally well with either long-term rates or overnight rates as the explanatory variable.

46. Both regressions, however, especially that using the funds rate, exhibit significant serial correlation, suggesting that the *t*-statistics are likely too high.

47. As a consequence, the Federal Reserve assumed that the term premium (the difference between long- and short-term rates) was a relatively stable, independent variable. The failure in 2004 and 2005 of the 325-basis-point rise in the funds rate to carry the yield on the 10-year Treasury note along with it (as historically it almost invariably had) was deemed a "conundrum." That episode has dramatically changed the long-held view that U.S. long-term interest rates were significantly influenced, if not largely determined, by monetary policy.

The emergence of globally arbitraged long-term rates has largely delinked U.S. long-term rates from Federal Reserve policy. It has accordingly changed the "conundrum" from why the 10-year Treasury note yield unexpectedly failed to respond to changes in the funds rate in 2004, to why the interest rate term structure was so stable through the latter part of the 20th century. Any notion that the Federal Reserve had of that stability being a fundamental characteristic of U.S. finance was dashed with the emergence of globally arbitraged long-term rates.

But the fixed-rate mortgage clearly delinked from the federal funds rate in the early part of this century. The correlation between them fell to an insignificant 0.10 during 2002–05, the period when the bubble was most intense, and as a consequence, the funds rate exhibited little, if any, influence on home prices.

The funds rate was lowered from 6½ percent in early 2001 to 1¾ percent in late 2001, and then eventually to 1 percent in mid-2003, a rate that held for a year. The Federal Reserve viewed the lowering to 1 percent as an act of insurance against the falling rate of inflation in 2003, which had characteristics similar to the Japanese deflation of the 1990s. We thought the probability of deflation small, but the consequences, should it occur, dangerous. On the other hand, we recognized that a funds rate held too low for too long might encourage *product* price inflation. I thought at the time that the rate decrease nonetheless reflected an appropriate balancing of risks. I still do.

To my knowledge, that lowering of the federal funds rate nearly a decade ago was not considered a key factor in the housing bubble. Indeed, as late as January 2006, Milton Friedman, historically the Federal Reserve's severest critic, evaluating monetary policy from 1987 to 2005, wrote, "There is no other period of comparable length in which the Federal Reserve System has performed so well. It is more than a difference of degree; it approaches a difference of kind."[48]

It thus came as somewhat of a surprise when, in August 2007, Stanford University's John Taylor (with whom I rarely disagree) argued that Federal Reserve policy in the aftermath of the dot-com bubble was the principal cause of the emergence of the U.S. housing bubble. According to Taylor (2007), had the funds rate followed his eponymous rule, housing starts would have been significantly lower and the U.S. economy would have avoided "much of the housing boom" and price bubble. His conclusion, often copied and repeated, seems, I fear, to have become close to conventional wisdom.[49]

As evidence, Taylor notes first the "significant" inverse correlation, with a lag, from mid-1959 to mid-2007 between the federal funds rate and

48. Milton Friedman, "The Greenspan Story: 'He Has Set a Standard,' " *Wall Street Journal,* January 31, 2006.

49. For example, a recent survey by the *Wall Street Journal* (Jon Hilsenrath, "Bernanke Challenged on Rates' Role in Bust," January 14, 2010) found that 78 percent of Wall Street and business economists surveyed and 48 percent of academic economists surveyed thought, "Excessively easy Fed policy in the first half of the decade helped cause a bubble in house prices."

housing starts and argues that according to his rule (a useful first approximation to a central bank's monetary policy stance), the Federal Reserve had set an inappropriately low funds rate during 2002–05.[50] As a consequence, he claims, "housing starts jumped to a 25-year high. . . . The surge in housing demand led to a surge in housing price inflation. [The] jump in housing price inflation then accelerated the demand for housing in an upward spiral" (Taylor 2007).

Taylor postulates housing starts as the primary driver of home prices. The evidence, however, suggests that it is not starts that drive prices and initiate the "upward spiral," but the other way around (figure 11). Home price changes, with a 6-month lead, have significant explanatory power for single-family starts from mid-1976 to 2009: the adjusted R^2 is 0.36, and the t-statistic is 15.0. American home builders, in my experience, respond to home price changes, not the federal funds rate, to determine how many "homes for sale" they start. And the home price change, as I noted earlier, is a function of lagged long-term mortgage rates.

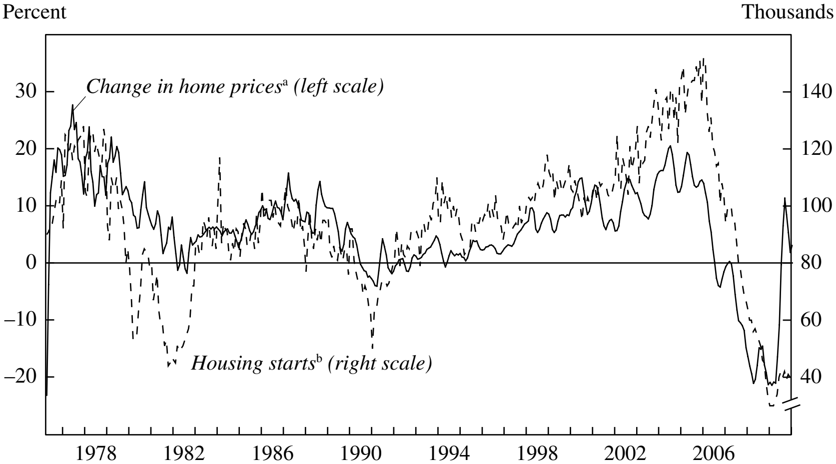
Housing starts, in any event, should be extraneous to Taylor's explanation of the bubble. It is employed because the Taylor rule by itself is structured to indicate a proper federal funds rate to balance the trade-off between inflation and unemployment. There are no asset price inputs, especially home prices, called for in the Taylor rule. Home prices cannot be substituted willy-nilly for the consumer price index (CPI) or the core personal consumption expenditures (PCE) price index in the Taylor paradigm. The CPI could stand as a proxy for home prices if the correlation between the two were very high. But it is not. The correlation between home prices and consumer prices, and between asset prices in general and product prices, is small to negligible, and on occasion negative. The Taylor rule clearly cannot be applied to asset prices, especially when benign product price inflation is almost surely a necessary condition for an income-producing-asset price bubble.[51]

The correct interpretation of a Taylor rule as applied to the period 2002–05 that stipulates that the federal funds rate is too low is that *product*

50. The Taylor rule indicated, according to a chart in Taylor (2007), that the funds rate should have been set at an average of 3.7 percent during 2002–05, compared with an actual average rate of 1.8 percent. Taylor's calculations employ the consumer price index as the inflation variable. Employing the core personal consumption expenditures price index, the Federal Reserve's preferred measure, narrows the gap significantly.

51. Moreover, the usual culprits behind either asset or product price inflation were missing. Growth in the M2 measure of the money stock, for example, was well behaved during 2002–05.

Figure 11. Home Prices and Housing Starts, 1976–2009



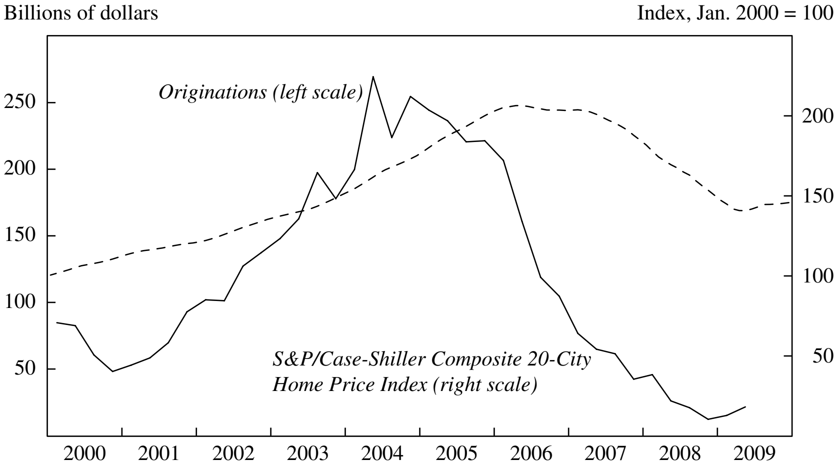
Source: Standard & Poor's, LoanPerformance, and Bureau of the Census.
a. Three-month moving average of seasonally adjusted monthly data. Before December 1999, LoanPerformance Single-Family Combined Home Price Index; from December 1999 onward, S&P/Case-Shiller Composite 20-City Index.
b. Starts of single-family homes, seasonally adjusted monthly data.

price inflation (the core implicit PCE deflator in the Federal Reserve's case) is threatening, and rate hikes to meet it are indicated. But inflation did not threaten. Indeed, core PCE averaged a modest annual inflation rate of only 1.9 percent during that period. Thus not only was the Taylor rule inappropriate for assessing the causes of asset price increases; it also gave a false signal for policy to stabilize the core PCE price.

The believers in Federal Reserve "easy money" policy as the root of the housing bubble correctly note that a low federal funds rate (at only 1 percent between mid-2003 and mid-2004) lowered interest rates for ARMs. That, in turn, they claim, increased demand for homes financed by ARMs and hence was an important contributor to the emergence of the bubble.

But in retrospect, it appears that the decision to buy a home most likely preceded the decision of how to finance the purchase. I suspect (but cannot definitively prove) that during that period of euphoria, a large majority of homebuyers who ended up financing with ARMs would have instead funded their purchases with fixed-rate mortgages had ARMs not been available. How else can one explain the peaking of originations of ARMs 2 years *before* the peak in home prices (figure 12)? Market demand obviously did not need ARM financing to elevate home prices during the last 2 years of the expanding bubble.

Figure 12. Home Prices and Originations of Adjustable-Rate Mortgages, 2000–09[a]



Source: Mortgage Bankers Association and Standard & Poor's.
a. Both series are quarterly data, seasonally adjusted.

Taylor (2009), confronted with evidence that the housing bubble was global, alludes to a seemingly tight relationship in a number of European countries between the degree of deviation from the Taylor rule and the size of the bubble. But a recent study by Federal Reserve staff (Dokko and others 2009), using a broader sample of countries, notes that deviations from the Taylor rule do not seem to be correlated with changes in home prices. They conclude (p. 31) that the relationship is "statistically insignificant (and relatively weak in economic terms as well)."

Moreover, Taylor does not buy the global saving-investment explanation of the decline in real long-term interest rates (which he foreshortens into the "saving glut") as the trigger of the global housing bubble. He succinctly states,

> Some argue that the low interest rates in 2002–4 were caused by global factors beyond the control of the monetary authorities. If so, then the interest-rate decisions by the monetary authorities were not the major factor causing the boom. This explanation—appealing at first glance because long-term interest rates remained low for a while after the short-term federal funds rate began increasing—focuses on global saving. It argues that there was an excess of world saving—a global saving glut—that pushed interest rates down in the United States and other countries. The main problem with this explanation is that there is no actual evidence of a global saving glut. On the contrary . . . the global saving rate—world saving as a fraction of world GDP—was low in the 2002–4 period, especially when compared with the 1970s and 1980s. (Taylor 2009, p. 6)

Here Taylor is employing ex post data to refute analysis based on ex ante saving and investment intentions (see section II.A above), an argument most economists should find puzzling.

V.B. Could the Breakdown Have Been Prevented?

Could the breakdown that so devastated global financial markets have been prevented? Given inappropriately low financial intermediary capital (that is, excessive leverage) and two previous decades of virtually unrelenting prosperity, low inflation, and low long-term interest rates, I very much doubt it. Those economic conditions are the necessary, and likely the sufficient, conditions for the emergence of a bubble in income-producing assets. To be sure, central bank monetary tightening has the capacity to break the back of any prospective cash flow that supports bubbly asset prices, but almost surely at the cost of a severe contraction of economic output, with indeterminate consequences. The downside of that trade-off is open-ended.[52]

But why not tighten incrementally? There are no examples, to my knowledge, of a successful incremental defusing of a bubble that left prosperity intact. Successful incremental tightening by central banks to gradually defuse a bubble requires a short-term feedback response.[53] But policy affects an economy with long and variable lags of as much as 1 to 2 years.[54] How does the FOMC, for example, know *in real time* if its incremental tightening is affecting the economy at a pace the policy requires? How much in advance will it have to tighten to defuse the bubble without disabling the economy? But more relevant, unless incremental tightening significantly raises risk aversion (and long-term interest rates) or disables the

52. Tight regulations on mortgage lending—for example, down payment requirements of 30 percent or more, the removal of the mortgage interest tax deduction, or elimination of home mortgage nonrecourse provisions—would surely severely dampen enthusiasm for homeownership. But they would also limit homeownership to the affluent, unless ownership by low- and moderate-income households were fully subsidized by government. Since January 2008 the subprime mortgage origination market has virtually disappeared. How will HUD's affordable housing goals (see note 10) be achieved in the future?

53. Some econometric models imply such capability for asset prices in general and home prices in particular. They achieve this by assuming a stable term structure, which, of necessity, yields a tight relationship between the federal funds rate and long-term rates. The latter is then employed to capitalize a flow of income (imputed housing services in the case of homes).

54. See, for example, Alan S. Blinder, "The Case for Optimism on the Economy," *Wall Street Journal,* December 16, 2009.

economy enough to undercut the cash flow that supports the relevant asset prices, I see little prospect of success.

The Federal Reserve's one attempt at incremental tightening failed. In early 1994 we embarked on a 300-basis-point tightening to confront what we perceived at the time as growing inflationary pressures. It was a policy that could have been just as easily read by the market as an incremental tightening to defuse the then-incipient dot-com bubble.

We not only failed to defuse the nascent stock market bubble that was evident in late 1993, but arguably enhanced it. The ability of the economy to withstand a severe monetary tightening in 1994 inadvertently demonstrated that the emerging boom was stronger than markets had anticipated and, as a consequence, raised the equilibrium level of the Dow Jones Industrial Average.[55] This suggested that a tightening far greater than the 1994 episode or the tightening in 2000 would have been required to quash the bubble. Certainly a funds rate far higher than the 6½ percent that was reached in mid-2000 would have been required.

At some rate, monetary policy can crush any bubble. If 6½ percent is not enough, try 20 percent, or 50 percent for that matter. But the state of prosperity will be an inevitable victim.[56] In 2005 we at the Federal Reserve did harbor concerns about the possible resolution of the housing bubble euphoria that gripped the nation. In 2005 I noted, "History has not dealt kindly with the aftermath of protracted periods of low risk premiums" (Greenspan 2005, p. 7).

However, we at the Federal Reserve never had a sufficiently strong conviction about the risks that could lie ahead. As I noted earlier, we had been lulled into a state of complacency by the only modestly negative economic aftermaths of the stock market crash of 1987 and the dot-com bust. Given that history, we believed that any decline in home prices would be gradual. Destabilizing debt problems were not perceived to arise under those conditions.

For guidance, we looked to the policy response to the unprecedented one-day stock-bubble bust of October 19, 1987, and the 2000 bear market. Contrary to prior experience, large injections of Federal Reserve liquidity

55. For details see Greenspan (2004a).

56. Such actions would obviously provoke an extreme political response. Although the decisions of the FOMC are not subject to legal reversal, the range of monetary policy choices has been politically constrained to what constitutes conventional wisdom in academia. As recent evidence reaffirms, the Federal Reserve's degree of policy independence is fixed by statute, and it can be altered or eliminated by statute.

apparently did help stabilize the economy—previously such crashes had led to economic retrenchment.

Unless there is a societal choice to abandon dynamic markets and leverage for some form of central planning, I fear that preventing bubbles will in the end turn out to be infeasible. Assuaging their aftermath seems the best we can hope for. Policies, both private and public, should focus on ameliorating the extent of deprivation and hardship caused by deflationary crises. But if an effective way, other than substantial increases in capital, to defuse leveraged bubbles without a major impact on economic growth were discovered, it would be a major step forward in organizing our market economies.

VI. In Summary

In this paper I have endeavored to trace the powerful economic forces that emerged in the aftermath of the Cold War and led to a dramatic decline and convergence of global real long-term interest rates. That in turn engendered, first, a dramatic global home price bubble heavily leveraged by debt, and second, a delinking of monetary policy from long-term interest rates.[57]

The global bubble was exacerbated by the widespread packaging of U.S. subprime and alt-A mortgages into securities, which found willing buyers at home (especially the GSEs) and abroad, many encouraged by grossly inflated credit ratings. More than a decade of virtually unrivaled global prosperity, low inflation, and low long-term interest rates reduced global risk aversion to historically unsustainable levels.

The bubble started to unravel in the summer of 2007. But unlike in the "debt-lite" deflation that followed the earlier dot-com boom, heavy leveraging set off serial defaults, culminating in what is likely to be viewed as the most virulent financial crisis ever. The major failure of both private risk management (including credit rating agencies) and official regulation was to significantly misjudge the size of the tail risks that were later exposed in the aftermath of the Lehman default. Had capital and liquidity provisions to absorb losses been significantly higher going into the crisis, contagious defaults surely would have been far less.

This paper has argued accordingly that the primary imperative going forward has to be increased regulatory capital, liquidity, and collateral

57. Whether the latter will continue with a less arbitrageable international bond market remains to be seen.

requirements for banks and shadow banks. I have also noted a number of less important reform initiatives that may be useful.

But the notion of an effective "systemic regulator" as part of a regulatory reform package is ill advised. The chronic sad state of economic forecasting should give governments pause on that issue. Standard models, except when heavily adjusted by ad hoc judgments, could not anticipate the current crisis, let alone its depth. Indeed, models rarely anticipate recessions, unless, again, the recession is artificially forced into the model structure.

In closing, let me reiterate that the fundamental lesson of this crisis is that, given the complexity of the division of labor required of modern global economies, we need highly innovative financial systems to ensure the proper functioning of those economies. But although, fortunately, most financial innovation is successful, much is not. And it is not possible in advance to discern the future success of each innovation. Only adequate capital and collateral can resolve this dilemma. If capital is adequate, then, by definition, no financial institution will default and serial contagion will be thwarted. Determining the proper level of risk-adjusted capital should be the central focus of reform going forward.

We can legislate prohibitions on the kinds of securitized assets that aggravated the current crisis. But markets for newly originated alt-A and adjustable-rate subprime mortgages, synthetic collateralized debt obligations, and many previously highly popular structured investment vehicles no longer exist. And private investors have shown no inclination to revive them. The next crisis will no doubt exhibit a plethora of innovative new assets, some of which will have unintended toxic characteristics that no one can forecast in advance. But if capital and collateral are adequate, losses will be restricted to those equity shareholders who seek abnormal returns but in the process expose themselves to abnormal losses. Taxpayers should not be at risk.

ACKNOWLEDGMENTS I wish to express my gratitude to the editors and to Gregory Mankiw, Jeremy Stein, and William Brainard for their most constructive comments and suggestions. The conclusions of this paper, of course, are my own.

References

Basel Committee on Banking Supervision. 2009. "International Framework for Liquidity Risk Measurement, Standards and Monitoring—Consultative Document." Basel: Bank for International Settlements (December).

Berger, Allen N. 1994. "The Relationship between Capital and Earnings in Banking." Working Paper 94-17. Wharton Financial Institutions Center (February).

Berger, Allen N., and David B. Humphrey. 1994. "Bank Scale Economies, Mergers, Concentration, and Efficiency: The U.S. Experience." Working Paper 94-25. Wharton Financial Institutions Center (June.)

Desroches, Brigitte, and Michael Francis. 2007. "World Real Interest Rates: A Global Savings and Investment Perspective." Bank of Canada Working Paper 2007-16. Ottawa (March).

Dokko, Jane, Brian Doyle, Michael T. Kiley, and others. 2009. "Monetary Policy and the Housing Bubble." Finance and Economics Discussion Series. Washington: Federal Reserve Board (December 22).

French, Kenneth R., Martin Neil Baily, John Y. Campbell, and others. 2010. *The Squam Lake Report: Fixing the Financial System.* Princeton University Press.

Greenspan, Alan. 1996. "The Challenge of Central Banking in a Democratic Society." Francis Boyer Lecture, American Enterprise Institute for Public Policy Research, Washington (December 5).

———. 1999. "The Evolution of Bank Supervision." Speech before the American Bankers Association, Phoenix (October 11).

———. 2000a. "Technology and Financial Services." Speech before the Journal of Financial Services Research and the American Enterprise Institute Conference in Honor of Anna Schwartz, Washington (April 14).

———. 2000b. "Banking Supervision." Speech before the American Bankers Association, Washington (September 18).

———. 2004a. "Risk and Uncertainty in Monetary Policy." Speech at the Meetings of the American Economic Association, San Diego (January 3).

———. 2004b. "Government-Sponsored Enterprises." Testimony before the Committee on Banking, Housing, and Urban Affairs, U.S. Senate, Washington (February 24).

———. 2005. "Reflections on Central Banking." Speech at a symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyo. (August 26).

———. 2007. *The Age of Turbulence: Adventures in a New World.* New York: Penguin.

Homer, Sidney, and Richard Sylla. 1991. *A History of Interest Rates,* 3rd ed. Rutgers University Press.

Mallaby, Sebastian. 2010. *More Money Than God: Hedge Funds and the Making of a New Elite.* London: Bloomsbury.

Office of Policy Development and Research, Department of Housing and Urban Development. 2001. "HUD's Affordable Lending Goals for Fannie Mae and Freddie Mac." Issue Brief no. 5. Washington (January).

Passmore, Wayne, Shane M. Sherlund, and Gillian Burgess. 2005. "The Effect of Housing Government-Sponsored Enterprises on Mortgage Rates." *Real Estate Economics* 33, no. 3: 427–63.

Philippon, Thomas, and Ariell Reshef. 2009. "Wages and Human Capital in the U.S. Financial Industry: 1909–2006." Working Paper 14644. Cambridge, Mass.: National Bureau of Economic Research (January).

Reinhart, Carmen M., and Kenneth Rogoff. 2009. *This Time Is Different: Eight Centuries of Financial Folly.* Princeton University Press.

Stern, Gary H. 2009. "Addressing the Too Big to Fail Problem." Statement before the Committee on Banking, Housing, and Urban Affairs, U.S. Senate, Washington (May 6).

Taylor, John M. 2007. "Housing and Monetary Policy." Speech at a symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyo. (August).

———. 2009. *Getting off Track: How Government Actions and Interventions Caused, Prolonged, and Worsened the Financial Crisis.* Stanford, Calif.: Hoover Press.

Comments and Discussion

COMMENT BY

N. GREGORY MANKIW This is a great paper. It presents one of the best comprehensive narratives about what went wrong over the past several years that I have read. If you want to assign your students only one paper to read about the recent financial crisis, this would be a good choice. There are some pieces of the analysis about which I am skeptical. But before I get to that, let me emphasize several important points of agreement.

Greenspan refers to recent events in the housing market as a "classic euphoric bubble." He is certainly right that asset markets can depart from apparent fundamentals in ways that are often hard to understand. This has happened before, and it will happen again. When the bubble bursts, the aftershocks are never pleasant.

Greenspan then points out that the political process, rather than reducing the risks associated with the bubble, actually contributed to them. In a footnote, he points out that in October 2000, in the waning days of the Clinton administration, the Department of Housing and Urban Development finalized rules that expanded the affordable housing goals of the government-sponsored enterprises (GSEs) Fannie Mae and Freddie Mac. As a result, the GSEs increased their holdings of subprime mortgages substantially. Although neither Greenspan nor I would suggest that the crisis was primarily the result of misguided housing policies, we both believe that these policies served to make a bad situation worse. This fact is important to keep in mind—not to assess blame; there is more than enough of that to go around. Rather, in judging how much policy can accomplish going forward, one should be mindful of how imperfect the political process is.

When considering what future regulation can do to reduce the likelihood of future crises, Greenspan emphasizes that whatever rules are promulgated cannot be premised on policymakers' ability to anticipate an uncertain future. In my view this is particularly wise. Some think the main cause of the recent crisis is that policymakers failed to anticipate the bursting of the housing bubble. If only we had central bankers with greater prescience, the argument goes, all this could have been avoided. In my view—and, I believe, Greenspan's as well—this is wishful thinking in the extreme. It indeed would be nice if somehow the individuals guiding the national economy had superhuman powers to see into the future. In reality, our economic leaders are mortals who share the same biases and flaws in perception as market participants.

What, then, can be done to make the financial system more crash-proof? Greenspan offers several good suggestions. First and most obvious, capital requirements should be raised. This is truer now than it has ever been. By bailing out almost every major financial institution that needed it, as well as a few that did not, the federal government raised the expectation of future bailouts, thereby turning the entire financial system, in effect, into a group of GSEs. Going forward, creditors to these institutions will view them as safe, and so they will lend to them too freely. The institutions, in turn, will be tempted to respond to their low cost of debt by leveraging to excess. Higher capital requirements are needed to counteract this newly expanded moral hazard.

Second, I like Greenspan's idea of "living wills," in which financial intermediaries are required to offer their own plans to wind themselves down in the event they fail. The advantage of this idea is that when future failures occur, as they surely will, policymakers will have a game plan in hand. How well these financial living wills will work, however, is hard to say. Like real wills, they may well be contested by "next of kin"—the counterparties to the institution's transactions. For living wills to work, they would need to be made public—say, by putting them on a centralized webpage—to discourage the counterparties from complaining after the fact that they thought they had more legal rights in the event of liquidation than they do.

Third, and perhaps most important, I like the idea of requiring financial firms to issue contingent debt that will turn into equity when some regulator deems that the firm has insufficient capital. Essentially, this debt would become a form of preplanned recapitalization in the event of a future financial crisis. But most important, the recapitalization would be done with private rather than public money. Because the financial firm would pay the cost of these contingent funds, rather than enjoying taxpayer subsidies, it would have an incentive today to lower its risk profile, for instance by reducing its leverage. The less risky the firm, the less likely it is that the contingency would be triggered, and the lower the interest rate the firm would pay on this contingent debt.

This brings me to the one conclusion of the paper with which I disagree—or, at least, I was not sufficiently persuaded. The issue concerns the importance of leverage to the viability of a financial intermediary. Greenspan proposes raising capital requirements and reducing leverage, but he suggests that there are limits to this. If leverage is reduced too much, he argues, financial intermediaries will be not be sufficiently profitable to remain viable. He offers some back-of-the-envelope calculations that purport to show how much leverage the financial system needs to stay afloat.

When I read this part of the paper, my first thought was, What about the Modigliani-Miller theorem? Recall that this famous theorem says that a firm's value as a business enterprise is independent of how it is financed. The debt-equity ratio determines how the risky cash flow from operations is divided among creditors and owners, but it does not affect whether the firm is fundamentally viable as a going concern. It seems to me that, as least as a first approximation, the logic of this theorem should apply to financial intermediaries as well as to other types of business. If not, why not?

I think it is clear where, from the Modigliani-Miller perspective, Greenspan's calculations go awry. He assumes that the rate of return on equity must be at least 5 percent. But as he notes, this number is actually endogenous to the degree of leverage. If a bank is less leveraged, its equity will be safer, and the required rate of return should fall.

Indeed, one can imagine a bank with almost no leverage at all. Suppose banks were required to hold 100 percent reserves against demand deposits, and that all bank loans had to be financed 100 percent with bank capital. A bank would then, in essence, be a marriage of a super-safe money market mutual fund with an unleveraged finance company. (Such a system would be similar to what is sometimes called "narrow banking.") It seems to me that a banking system operating under such strict regulations could well perform the crucial economic function of financial intermediation. No leverage would be required.

Such a system would, however, forgo the "maturity transformation" function of the current financial system, in which many banks and other intermediaries borrow short and lend long. The issue I am wrestling with is whether maturity transformation is a crucial feature of a successful financial system. The resulting maturity mismatch seems to be a central element of banking panics and financial crises. The open question is what value it has and whether the benefits of today's highly leveraged financial system exceed the all-too-obvious costs.

To put the point most broadly: The Modigliani-Miller theorem says leverage and capital structure are irrelevant, yet many bankers would surely claim they are central to the process of financial intermediation. A compelling question on the research agenda is to figure out who is right, and why.

COMMENT BY

JEREMY C. STEIN It is a pleasure to comment on this important and wide-ranging paper by Alan Greenspan. In light of the breadth of ground that it covers, I will have to focus my comments on just a couple of the issues that struck me as particularly interesting. The first of these concerns the central role of capital and liquidity requirements in any attempt to reform financial markets. As the paper states, "The most pressing reform, in my judgment, in the aftermath of the crisis is to fix the level of regulatory risk-adjusted capital, liquidity, and collateral standards required by counterparties." I agree with this view. Moreover, Chairman Greenspan makes a highly welcome contribution by taking this observation to the logical next step: he poses, and attempts to answer, the quantitative question of just how high capital requirements should be raised. This is a point on which most policymakers have thus far been conspicuously silent.

The paper argues for a regulatory minimum ratio of book equity to assets in the neighborhood of 14 percent. The argument has two parts. First, a rough calculation suggests that a 14 percent ratio would provide the banking sector with a buffer adequate to see it through a crisis equal in magnitude to that of the last few years. And second, another back-of-the-envelope exercise yields the conclusion that a 14 percent regulatory minimum would not be overly burdensome, in the specific sense that it would not prevent banks from earning a return on equity in line with historical averages.

In the same spirit of simple calibration, I would like to offer another approach to the second piece of the puzzle: the costs associated with raising capital requirements by several percentage points. My analysis is nothing more than an application of the standard weighted average cost of capital (WACC) machinery that is routinely taught to MBA students everywhere, which augments the Modigliani-Miller (1958) paradigm to take account of corporate income taxes. Suppose that equity capital requirements are raised very substantially—say, by 10 percentage points. Moreover, suppose that at the margin, this additional equity displaces long-term debt in the capital structure of the affected banks. According to Modigliani-Miller, the only net effect of this change on banks' WACC (and hence on the rate they charge for corporate or consumer loans, for example) comes from the lost tax deductions on the long-term debt that is eliminated. Thus, if the displaced debt yielded, say, 7 percent, then given a 35 percent corporate tax rate, a 10-percentage-point reduction in the debt tax shield would raise the WACC by $0.10 \times 0.07 \times 0.35 = 0.00245$, or about 25 basis points. Again, this is the impact of a very large increase in the equity capital ratio, equivalent to going from a low initial ratio of 4 percent all the way up to the level suggested by Greenspan of 14 percent.

Of course, this calculation comes with a number of caveats. First, and perhaps most important, it should be thought of as capturing the long-run steady-state costs of having to *hold* more equity on the balance sheet, while disregarding the transitional flow costs associated with *raising* the required new equity. Given the adverse selection problems associated with new equity issues (Myers and Majluf 1984), these flow costs may be significant. This implies that if higher capital requirements are phased in too abruptly—so that banks have to get there through large external equity issues, rather than by gradually accumulating retained earnings—the transitional impact on their lending behavior may be much higher than my 25-basis-point figure suggests.

Another caveat is that even in a long-run steady state, taxes may not be the only relevant violation of the idealized Modigliani-Miller conditions. To take one example, Gary Gorton and Andrew Metrick (2010) and Stein (2010) argue that banks like to issue collateralized short-term debt because this debt commands a "money-like" convenience premium based on its relative safety and the transactions services that safe claims provide. If one takes a crude upper bound on this convenience premium to be 1 percent, and if capital requirements have the effect of crowding out such short-term debt at the margin, as opposed to long-term debt, this would add another $0.10 \times 0.01 = 10$ basis points to the overall effect,[1] for a total of 35 instead of 25. This logic suggests that other sensible modifications are also likely to have only a relatively small effect.

All of this would therefore seem to reinforce—albeit with a quite different methodology—the broad conclusions in Greenspan's paper, namely, that although there are undoubtedly costs associated with significant increases in bank capital requirements, a crude estimate of these costs does not

1. Krishnamurthy and Vissing-Jorgensen (2010) estimate the convenience premium associated with Treasury securities to be on the order of 70 basis points, which suggests that my 100-basis-point number is probably a conservative upper bound.

suggest that they are prohibitive. Said differently, both his analysis and mine would appear to give significant comfort to those who worry that plausibly higher capital requirements will make bank loans much more expensive.

And yet there would seem to be an obvious tension here. Banks manifestly care a great deal about optimizing their capital structures, and they show a persistent tendency to gravitate toward high leverage. In contrast, most nonfinancial firms, many of which operate with dramatically lower leverage, seldom appear to be nearly as strongly drawn toward any fixed target capital structure. So although the Modigliani-Miller-plus-taxes paradigm may be adequate for capturing the relatively small benefits of debt for nonfinancial firms, one wonders, in light of their very different behavior, whether the same paradigm does not leave out something of first-order importance when it comes to financial firms. Put simply: if higher capital ratios have only a small impact on the WACC for financial firms, why do they—unlike their nonfinancial counterparts—resist them so forcefully?

My own attempt at reconciling this tension goes as follows. Perhaps the substitution of equity for debt finance does in fact have the same small effects on the WACC for financial and nonfinancial firms—say, 25 basis points for a 10-percentage-point change in the equity ratio. But what is different about financial firms are the *competitive implications* of a small cost-of-capital disadvantage. An auto manufacturer or a software firm is unlikely to be driven out of business over a 25-basis-point cost-of-capital difference; so many other factors—the quality of its product, the loyalty of its customer base, and so on—are so much more important that it can fail to fully optimize on the cost-of-capital dimension and still survive. In contrast, for a financial firm, cheap capital is the single dominant input, and it simply cannot afford to cede a 25-basis-point edge to its competitors. In this sense, high leverage is for financial firms like what a performance-enhancing drug is for elite sprinters: even if the drug is harmful to health and cuts only a few hundredths of a second from their times, with all else so closely matched, they may not feel they can afford not to take it.

On the one hand, the drug analogy makes much stricter capital regulation seem like a no-brainer: if it can stop a systemically unhealthy form of competition with only a minimal impact on performance (in this case, on the cost of loans to corporations and households), then it would seem highly desirable from a social perspective. The hitch, however, is that, much like with drug testing, the same competitive forces create a powerful motive for evading the regulation. One important channel for this evasion is migration of credit creation from the regulated banking sector to the less regulated shadow banking sector. For example, instead of keeping a consumer loan on its balance sheet, subject to the more stringent capital rules, a bank can bundle the loan with other, similar loans into a security, which winds up, say, in the portfolio of a hedge fund, which in turn finances its purchase of the security largely with overnight repos and only a very thin slice of capital.

Although such migration may leave the banks themselves safer, it is much less clear that it leaves the financial system in better shape should a crisis occur. One of the most dramatic features of the subprime crisis was the complete collapse of the market for asset-backed securities—and not just those related to subprime mortgages, but also those based on auto loans, credit card receivables, student loans, and other assets. This market collapse, which was arrested only by the Federal Reserve's intervention with the Term Asset-Backed Securities Loan Facility (TALF), played an important role in deepening the credit crunch.

The bottom line is that I do not worry too much about the effects of higher capital requirements on the cost of loans to households and firms. Based on the sorts of calculations sketched above, my best estimate is that these effects will be relatively muted. At the same time, I worry a great deal about the effects on *how and by whom* credit is provided, and the potential implications of these changes for overall systemic stability.

To be clear, I do not at all mean to suggest that capital requirements for banks should not be significantly higher. Indeed, if forced to pick a number for the required capital ratio, I might well come out somewhere in the same range as Greenspan. However, the danger of competition leading to evasion of the capital requirement suggests that the focus should not be just on banks, or even just on all bank-like institutions. Rather, an effort must be made to impose similar capital standards *across a given asset class,* no matter who winds up holding the asset. This will not be an easy task, but one tool that might be helpful is broad-based regulation of "haircuts" (that is, minimum margin requirements) on asset-backed securities that trade in the shadow banking market. Returning to the previous example, this regulation might stipulate that whoever holds a tranche of a consumer loan securitization, be it a hedge fund, a pension fund, or anybody else, would be required to post a minimum haircut against that tranche. The value of the haircut would depend on the seniority of the tranche, the underlying collateral, and so forth. If these haircut requirements are well structured, they could go a long way toward achieving harmonization across organizational forms, in that there would be no obvious advantage based on avoidance of regulation to moving the consumer loans off the balance sheets of banks and into the shadow banking sector.

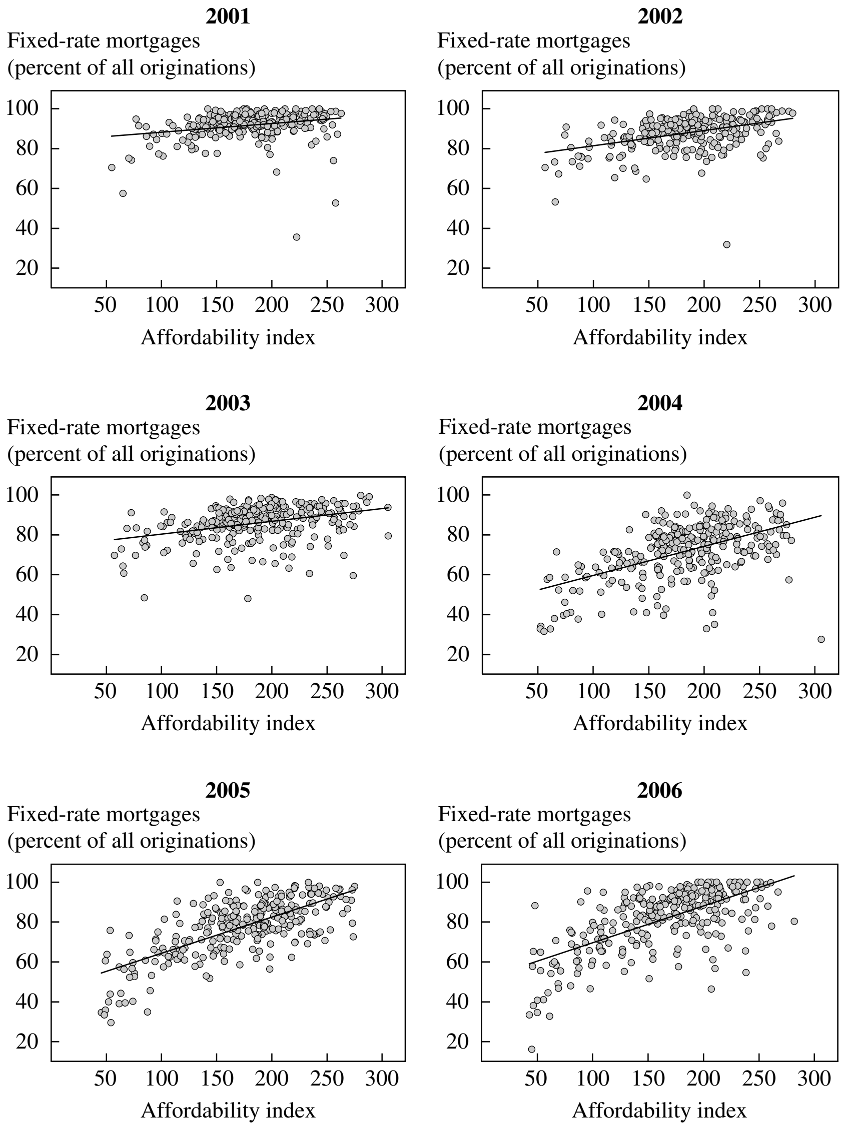
My reading of Greenspan's paper is that he is fundamentally sympathetic to this approach, and indeed that he has something very much along these lines in mind when he refers to the need to "fix the level of . . . collateral standards required by counterparties." If so, I hope that other policymakers will pay careful attention to his advice.

On a different note, I am inclined to be more skeptical of Greenspan's analysis when he downplays the role of low short-term interest rates in the initial years of the housing boom. He writes, "The global home price bubble . . . was a consequence of lower interest rates, but it was *long-term* interest rates that galvanized home asset prices, not the overnight rates of central banks, as has become the seeming conventional wisdom" (emphasis in original). My own suspicion is that short-term rates did play an important independent role, by reducing the required monthly payments for borrowers taking out adjustable-rate mortgages (ARMs), whose rates are tied to short-term market rates. This hypothesis presumes that some of these borrowers were either myopic or liquidity constrained, so that their initial monthly payment—as opposed to the expected payments over the life of the loan—was decisive in their choice. Although this presumption may not accurately characterize the behavior of the majority of borrowers in normal real estate markets, perhaps it rings more true as a description of the recent subprime boom.

In any case, although I do not have conclusive evidence for my hypothesis, I can offer one suggestive set of plots. Figure 1 plots, for each year from 2001 through 2006, the share of fixed-rate mortgages in total mortgages initiated in each of 269 metropolitan statistical areas (MSAs) against an affordability index for that MSA. The affordability index is from Moody's Economy.com and is based on the median family income in an MSA relative to the monthly mortgage payment on a median-priced home in that MSA (assuming a conventional fixed-rate mortgage loan). Higher values of the index correspond to greater affordability, that is, to higher ratios of incomes to home prices.

The figure conveys two key messages. First, throughout the period, ARM use is more prevalent in more expensive cities, where liquidity constraints are presumably more likely to be binding on homebuyers. Second, this relationship becomes strikingly more pronounced between 2002 and 2004, when the federal funds rate was bottoming out and home prices began to rise dramatically. This latter effect is consistent with the key mechanism underlying my hypothesis, namely, that the short-term rate works through its ability to reduce the monthly payments for income-constrained borrowers who finance their homes with ARMs.

Figure 1. Share of Fixed-Rate Mortgages and Housing Affordability in 269 MSAs, 2001–06



Source: Benjamin Iverson, Harvard Business School, and James Vickery, Federal Reserve Bank of New York, using data from Moody's Economy.com and the Monthly Interest Rate Survey from the Federal Housing Finance Agency.

a. Each observation is for a single MSA. The affordability index is based on the ratio of median family income in an MSA to the monthly payment on a conventional fixed-rate mortgage for a median-priced home in that MSA. Higher values of the index indicate greater affordability (that is, a higher ratio of median income to mortgage payment). Lines are fitted regression lines.

Again, this evidence is only suggestive, and more work would be required to support the story I have in mind with any real degree of confidence. Nevertheless, at a minimum, I believe that the role of short-term rates in the recent housing bubble remains an important open question.

REFERENCES FOR THE STEIN COMMENT

Gorton, Gary B., and Andrew Metrick. 2010. "Securitized Banking and the Run on Repo." Working Paper 15223. Cambridge, Mass.: National Bureau of Economic Research.

Krishnamurthy, Arvind, and Annette Vissing-Jorgensen. 2010. "The Aggregate Demand for Treasury Debt." Working paper. Northwestern University.

Modigliani, Franco, and Merton H. Miller. 1958. "The Cost of Capital, Corporation Finance and the Theory of Investment." *American Economic Review* 48, no. 3: 261–97.

Myers, Stewart C., and Nicholas C. Majluf. 1984. "Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not Have." *Journal of Financial Economics* 13, no. 2: 187–221.

Stein, Jeremy C. 2010. "Monetary Policy as Financial-Stability Regulation." Working paper. Harvard University.

GENERAL DISCUSSION Several panelists expressed thanks to Chairman Greenspan for his service to the nation and for his candor in stating that the events of the last few years had led him to revise some long-held views.

Gregory Mankiw agreed with Jeremy Stein that in the presence of taxes there is a preference for debt over equity, to which banks may well respond more than other firms. If that is the case, then the policy prescription is clear: reform the tax code to eliminate the preference for debt.

Alan Blinder pointed to what seemed an incipient consensus on there being two types of bubble, although this may oversimplify what might really be a continuum. Bubbles of the first type, which includes the tech stock bubble of the late 1990s, are based on equity rather than leverage and credit, whereas those of the second, which include the recent crisis, are based on excessive leverage. The Greenspan-Bernanke mop-up-after view of how to deal with crises continues to make sense for the first type but not for the second. One reason is that the Federal Reserve has informational advantages in the credit system, especially the banking part. If the Federal Reserve were designated an explicit systemic risk regulator over the entire financial system, that informational advantage would become even larger.

Blinder agreed with Greenspan that more equity capital in the financial structure of financial institutions has to be part of the solution, but he was unsure whether it is the whole solution. There is potentially an intermediate course of action between forced bankruptcy, as in the case of Lehman Brothers, and bailout, as in the case of AIG. The government can step in and grab hold of the reins, fire management, wipe out the shareholders, impose some losses on creditors, and then take the failed firm into either receivership or conservatorship. Authority to undertake such resolutions will be an important part of any reform—not because it will prevent bubbles, but because it will mitigate the fallout and the cost to taxpayers when they happen. Related to this is the need to require more collateral behind purchases of over-the-counter derivatives—the capital markets analogue to increasing bank capital. One could go further and more or less force derivatives transactions onto organized exchanges, by imposing a higher capital requirement for derivatives not traded on exchanges. Reform should also include doing something about the go-for-broke incentives that were rampant in financial markets in the run-up to the crisis.

Benjamin Friedman observed that strengthening capital requirements is also about accounting reform. Often what matters is not just the specific percentage by which assets must be backed by equity, but also the specification of the asset total by which that percentage gets multiplied. For example, the chief problem at Citibank was the $100 billion in assets that were off the balance sheet, and therefore against which the bank held zero capital. The comparable off-balance-sheet amount at Lehman Brothers was $50 billion. In each case the needed reform is not choosing a new percentage but requiring that capital be held against a much more inclusive specification of the firm's assets.

Friedman also thought the Modigliani-Miller perspective, which had been suggested by both discussants, was interesting but led to a troubling conclusion. The standard Modigliani-Miller theorem assumes not only no taxes, but also no bankruptcy. If the banking system as a whole is operating at one level of leverage, then any one bank that is forced to make do with less leverage is at a competitive disadvantage. If the banking system requires some minimum amount of leverage to do business, this implies some probability of any bank (or even all of them) failing, and this in turn requires either a public sector subsidy or the possibility of a taxpayer bailout. If that is so, it means there cannot be a banking sector unless the banks collectively have a leverage ratio high enough to put the taxpayer at risk.

Finally, Friedman posed a question for Greenspan on the choice between regulation by public institutions and regulation by creditors. Before the crisis, Greenspan had argued forcefully and articulately that the latter was superior. The paper, however, was as sharply critical of one as of the other. Friedman therefore wondered whether the experience of the crisis had changed Greenspan's thinking on the relative advantages of the two.

Olivier Blanchard followed up on Mankiw's remarks regarding maturity transformation. In the aggregate, most savers probably have a longer horizon than the firms to which they lend. Much of this saving is for retirement or other long-term purposes, whereas much physical capital has a life of about 10 years. Thus, at the macro level, the transformation of short-term saving into long-term investment does not seem that important, yet many institutions are involved in precisely that process.

Martin Baily laid out two views prevalent among noneconomists of what caused the crisis: one is that it was all about greedy bankers, whose actions produced a market failure of the worst kind. The other is that it was a government failure, either of the regulators or of housing policy. For those who think the culprit was federal policy, the answer is to change the policy—to get the government out of the way and let the market work. For those who think the problem was market failure, the answer is to strengthen regulation. But in Baily's view the crisis was caused by both market failure and government failure, and therefore to some extent both things have to be done—some mix of less government in some areas and more government in others is needed.

Baily agreed that bubbles cannot be forecast precisely, but that does not mean that nothing can or should be done when one sees a bubble forming. If you know you have high cholesterol, you may not know whether or when you will have a heart attack, but it is still a good idea to take anti-cholesterol medication. When policymakers—both financial regulators and monetary authorities—observe a highly leveraged increase in asset prices, they should do something, even though they risk being wrong in their diagnosis. It is worth taking out the insurance policy of at least leaning against that particular wind. It would also be a good idea for the Federal Reserve to have another tool that it presently lacks, namely, the ability to adjust margin requirements or capital requirements of all kinds—for example, to set minimum down payments for mortgages in the event of an incipient mortgage bubble.

Baily agreed, to some degree, that a large moral hazard had been created. There were good reasons to protect debtors in the heat of the crisis, but doing so also created a danger: investors might believe that the regulators will not regulate the next time either. On the other hand, the moral hazard problem can be overstated. The managers of financial institutions have certainly taken a hit: almost all the institutions that got into trouble have replaced their managers. Shareholders have taken a big hit as well. The problem is mainly on the debtors' side, and it needs to be dealt with through the living wills and other resolution mechanisms that Greenspan mentioned, to make sure that the debtors cannot walk away unscathed.

Christopher Carroll called the Panel's attention to the fact that Robert Shiller, in a December 1996 speech at the Federal Reserve, had warned of a bubble emerging in the stock market, and that in January 2004, speaking again at the Federal Reserve, Shiller had warned of a bubble emerging in the housing market. Perhaps when Robert Shiller enters the precincts of the Federal Reserve Board, he takes on supernatural powers that give him intuition on this subject that others lack. But if one or at least a few respected economists have strong intuitions that a bubble is in the process of forming, that does seem an appropriate time for regulators to think about becoming more vigilant.

Christopher Sims cited the paper's observation that the private sector did not seem to price systemic risk very well leading up to the crisis. This suggests the presence of an externality: just because private agents do not take account of the risks they impose on the system does not mean that systemic risk does not exist. That the markets did not seem to react to this risk in advance raises the question of whether regulators could do better. There is some chance that the right kind of regulators could do better, through aggressive information collection or examination of accounting practices, for example. Then the question becomes how to avoid regulatory capture. In the years leading up to the crisis, it had become politically difficult to suggest tighter regulation. One argument for assigning more responsibility for systemic regulation to the Federal Reserve rather than some other agency is that the Federal Reserve has a dedicated revenue source and its governors serve 14-year terms. These things go a long way toward making regulators independent and allowing them to avoid capture.

George von Furstenberg interpreted the paper's message to private financial institutions as "Go ahead and spill it—we will mop it up." Yet this policy, he argued, has already led to enormous underpricing of risk and subsequent socialization of enormous losses. This was indeed the opposite of central planning—it was central bungling. In other words, what produces a deviation from market models is to let markets be perceived as failing in an egregious way. Therefore, it is important to take precautions and buffer the system against the destructive effects of bubbles. If bubbles are not preventable, then much greater precautions are necessary. Some of these have absurdly kicked in after the fact: now the Federal Housing Administration and the GSEs have increased their lending standards; now firms like AIG are subject to special margin requirements. There are many things that can be done to reduce the vulnerability of the system to bubbles, if bubbles there must be. If you know that hailstorms exist but cannot predict them, you do not have to stand outdoors bareheaded. There are ways to reduce your exposure. Certainly the only course is not just to mop up afterward—a strategy that in this case has been anything but market directed, and has been very incomplete and extremely costly to the taxpayer.

Richard Cooper agreed with Greenspan that banks, at a minimum, should be subject to higher capital requirements, and perhaps other financial institutions should as well. He wondered why that general point is not even more broadly applicable—why not, for example, impose minimum down payment requirements on homebuyers? The Federal Reserve had full authority to require the institutions it regulated to impose such requirements; in the spirit of Baily's anti-cholesterol metaphor, imposing such requirements in 2003 and 2004 might have prevented the heart attack, given what was known at the time.

Cooper also agreed with Friedman on the need for accounting reform, and specifically for bringing structured investment vehicles and the like onto banks' consolidated balance sheets. He was at least as interested, however, in the principles governing the valuation of assets and liabilities, particularly when regular markets do not exist (for venture capital, for example) or when they have frozen, as happened in late 2008. In this country, accounting rules are left to an entirely private body called the National Accounting Standards Board, which operates under the principle that all information that can conceivably be brought onto the financial statements should be. Transparency and bringing things onto the balance sheet are two different things, however. Cooper wondered whether the process for setting accounting standards in general ought to be reviewed, or whether a set of regulatory accounting standards ought to be established that would be used for setting capital requirements, rather than relying on mark-to-market rules, particularly when the market valuations have to be artificially simulated or taken from a few distress transactions.

Robert Hall noted that Martin Baily had prescribed anti-bubble medication as if a best-selling medication of that type already existed. But the paper's argument was that the interest rate controlled by the central bank is not an effective anti-bubble medication, at least with respect to real estate. Prospective investors capitalizing the value stream from a piece of real estate look far into the future, yet the central bank's influence is limited to a relatively short horizon. Hall pointed out that many countries that did not have ARMs—and the great majority do not—also had huge housing bubbles. The evidence points to low long-term interest rates as what matter when valuing housing, and therefore suggests that it was low global long-term rates, not short-term rates, that caused what became a worldwide bubble.

Hall went on to note that the other anti-bubble medication, suggested by Richard Cooper, involves introducing frictions into financial markets by regulating down payments or margins. There is nothing intrinsically wrong or dangerous about making risky loans, provided that the institutions holding the loans are not huge, highly leveraged, and systemically important. Getting a much more robust financial system is the solution to this problem, not anti-bubble medication. The economy rode through the equity bubble that popped in 2000 without any financial crisis. It should be able to ride through a real estate bubble just as well.

Justin Wolfers noted that the paper was largely silent on the shadow banking system and hoped for more discussion of that topic in the final draft.

OLIVIER J. BLANCHARD
International Monetary Fund
Massachusetts Institute of Technology

HAMID FARUQEE
International Monetary Fund

MITALI DAS
International Monetary Fund

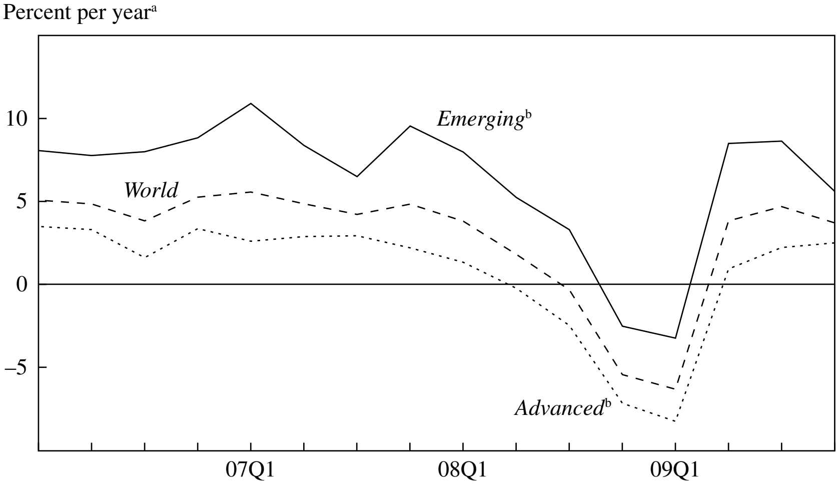
The Initial Impact of the Crisis on Emerging Market Countries

ABSTRACT To understand the diverse impact of the crisis across emerging market countries, we explore the role of two shocks—the collapse in trade and the sharp decline in financial flows—in the transmission of the crisis from the advanced economies. We first develop a simple open economy model, which allows for imperfect capital mobility and potentially contractionary effects of currency depreciation due to foreign debt exposure. We then look at the cross-country evidence. The data suggest a strong role for both trade and financial shocks. Perhaps surprisingly, the data give little econometric support for a central role of either reserves or exchange rate regimes. We end by presenting case studies for Latvia, Russia, and Chile.

One of the striking characteristics of the financial crisis that originated in the United States is how quickly and how broadly it spread to the rest of the world. When the crisis intensified, first in the United States and then in Europe, in the fall of 2008, emerging market countries thought they might escape more or less unharmed. There was talk of decoupling. This was not to be.

Figure 1 shows growth rates of GDP for a group of advanced economies and a group of emerging market countries from the first quarter of 2006 through 2009. The two series have moved largely in tandem. In the fourth quarter of 2008 and the first quarter of 2009, economic growth in the advanced group averaged −7.2 percent and −8.3 percent, respectively (at annual rates). In the same two quarters, growth in the emerging market countries was −1.9 percent and −3.2 percent, respectively. As the figure shows, the better numbers for the emerging market countries reflect their

Figure 1. Growth in GDP in Advanced and Emerging Market Economies, 2006–09



Sources: IMF, Global Data Source, and IMF staff estimates.

a. Quarter over quarter at an annual rate. Series are averages weighted by GDP at purchasing power parity (PPP).

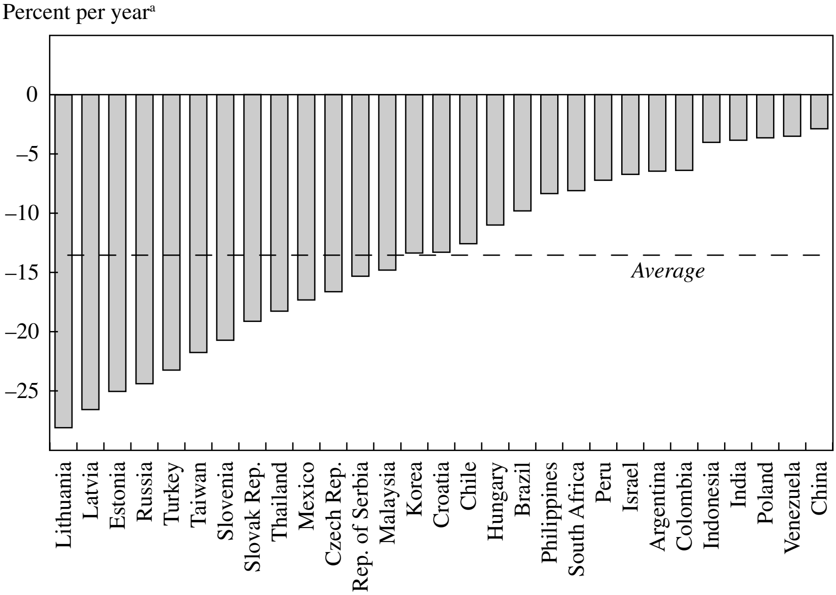
b. The figure is based on 17 advanced economies (including the euro area as a single economy) and 25 emerging market countries. See footnote 1 for the list of emerging market countries.

higher underlying average growth rate. Growth rates for both groups during those two quarters were roughly 10 percentage points below their 2007 value.

The parallel performance of the two groups in figure 1 hides substantial heterogeneity within each group. Figure 2 shows, for a sample of 29 emerging market countries, the actual growth rate for the semester composed of the two quarters with large negative growth, 2008Q4 and 2009Q1, minus the April 2008 International Monetary Fund (IMF) forecast growth rate over the corresponding period—"unexpected growth" in what follows.[1] All the countries in the sample had negative unexpected growth, but with

1. The countries and their abbreviations are as follows: Argentina (ARG), Brazil (BRA), Chile (CHL), China (CHN), Colombia (COL), Croatia (HRV), Czech Republic (CZE), Estonia (EST), Hungary (HUN), India (IND), Indonesia (IDN), Israel (ISR), Republic of Korea (KOR), Latvia (LVA), Lithuania (LTU), Malaysia (MYS), Mexico (MEX), Peru (PER), Poland (POL), Philippines (PHL), Russia (RUS), Republic of Serbia (SER), Slovak Republic (SVK), Slovenia (SVN), South Africa (ZAF), Taiwan Province of China (TWN), Thailand (THA), Turkey (TUR), and Venezuela (VEN). In figure 1, the series for emerging market countries includes Bulgaria, Pakistan, Romania, and Ukraine (not in our sample) but excludes HRV, CZE, ISR, SER, SVK, SVN, and TWN. Some of the emerging market countries listed here are classified as "advanced economies" in the IMF's *World Economic Outlook.*

Figure 2. Unexpected Growth in GDP in Emerging Market Countries, 2008Q3–2009Q1



Sources: IMF, Global Data Source and World Economic Outlook; Eurostat.

a. Actual growth in GDP over the two quarters 2008Q4 and 2009Q1, seasonally adjusted at an annual rate, minus April 2008 IMF forecast for the same period.

considerable variation across them. In seven countries, including some as diverse as Latvia and Turkey, growth was lower than forecast by more than 20 percentage points (again at an annual rate); at the same time, in five countries, China and India most notable among them, the unexpected growth shortfall was smaller than 5 percentage points. (Looking at growth rates themselves, or at deviations of growth rates from trend, gives a very similar ordering.)

Figure 2 motivates the question we take up in this paper, namely, whether one can explain the diverse pattern of growth across emerging market countries during the crisis. The larger goal is an obvious one: to better understand the role and the nature of trade and financial channels in the transmission of shocks in the global economy.

We focus on emerging market countries. We leave out low-income countries, not on the basis of their economic characteristics, but because they typically lack the quarterly data we think are needed for an informed analysis of the impact effects of the crisis. We focus only on the acute phase of the crisis, namely, 2008Q4 and 2009Q1. Looking at later quarters, which in most countries are characterized by positive growth and recovery, would be useful, including for understanding what happened in the acute phase. But for reasons of data and scope, we leave this to further research.[2]

We start in section I by presenting a simple model. It is clear that emerging market countries were affected primarily by external shocks, mainly through two channels. The first was a sharp decrease in their exports and, in the case of commodity producers, a sharp drop in their terms of trade. The second was a sharp decrease in net capital flows. Countries were exposed in various ways: some were very open to trade, others not; some had large short-term external debts or large current account deficits, or both, others not; some had large foreign currency debts, others not. They also reacted in different ways, most relying on some fiscal expansion and some monetary easing, some using reserves to maintain the exchange rate, others instead letting it adjust. The model we provide is little more than a placeholder, but it offers a useful framework for discussing the various channels and the potential role of policy, and for organizing the empirical work.

We then turn to the empirical evidence, which we analyze through econometrics, in section II, as well as case studies. We start with simple cross-country specifications, linking unexpected growth over the two quarters to various trade and financial variables. With at most 29 observations in each regression, econometrics can tell us only so much. But the role of both channels, trade and financial, comes out clearly. The most significantly robust variable is short-term external debt, suggesting a central role for the financial channel. Trade variables also clearly matter, although the relationship is not as tight as one might have expected. Starting from this simple specification, we explore a number of issues, such as the role of reserves. Surprisingly, we find little econometric evidence in support of the hypothesis that high reserves limited the decline in output in the crisis.

We turn finally in section III to case studies, looking at Latvia, Russia, and Chile. Latvia was primarily affected by a financial shock, Chile mostly by a sharp decrease in the terms of trade, and Russia by both strong financial and terms of trade shocks. Latvia and Russia suffered large declines in

2. Other studies that attempt to explain differences across countries in the impact of the crisis include Lane and Milesi-Ferretti (2009), Giannone and others (2009), Berkmen and others (2009), and Rose and Spiegel (2009a, 2009b). These studies typically use annual data, either for 2008 alone or for 2008 and 2009, and a larger sample of countries than we do. For differences across emerging European countries, see Bakker and Gulde (2009) and Berglof, Korniyenko, and Zettlemeyer (2009). A parallel and larger effort within the IMF (2010), with more of a focus on policy implications, is currently being conducted. We relate our results to the various published studies below.

output. The effect on Chile was milder. Together, the country studies provide a better understanding of the ways in which initial conditions, together with the specific structure of the domestic financial sector, the specific nature of the capital flows, and the specific policy actions, shaped the effects of the crisis in each country.

I. A Model

To organize our thoughts, we start with a standard short-run, open economy model, modified, however, in two important ways. First, to capture the effects of shifts in capital flows, we allow for imperfect capital mobility. Second, we allow for potentially contractionary effects of a depreciation stemming from exposure to foreign currency debt.

The model is shamelessly ad hoc, static, and with little role for expectations.[3] Our excuse for its ad hoc nature is that the micro foundations for all the complex mechanisms we want to capture are not yet available, and even if available would make for a complicated model. Our excuse for the lack of dynamics is that we focus on the effects of the shocks immediately upon impact, rather than on their dynamic effects. Our excuse for ignoring expectations is that the direct effect of lower exports and lower capital flows probably dominated expectational effects, but this excuse is admittedly poor; as we will show, an initial quasi peg on the exchange rate, coupled with anticipations of a future depreciation, initially aggravated capital outflows in Russia in the fall of 2008, making the crisis worse.

The model is composed of two relationships, one characterizing balance of payments equilibrium, and the other goods market equilibrium.

I.A. Balance of Payments Equilibrium

Balance of payments equilibrium requires that the trade deficit be financed either by net capital flows or by a change in reserves. Taking capital flows first, we consider three different interest rates:

—the policy (riskless) interest rate, denoted by r (given our focus on the short run, we assume constant domestic and foreign price levels, and thus zero domestic and foreign inflation, and so we make no distinction between nominal and real interest rates)

—the interest rate at which domestic borrowers (firms, people, and the government; we make no distinction among them in the model) can

3. A model in the same spirit as ours, but with more explicit micro foundations and a narrower scope, is developed in Céspedes, Chang, and Velasco (2004).

borrow, denoted by $\hat{r}$. Assume that $\hat{r} = r + x$, where x is the risk premium required by domestic lenders. Think of the United States as the foreign country, and thus of the dollar as the foreign currency. We assume that the exchange rate is expected to be constant, so $\hat{r}$ is also the domestic dollar interest rate.[4]

—the U.S. dollar interest rate, that is, the rate at which foreign investors can lend to foreign borrowers abroad, denoted r^*. $\hat{r} - r^*$ is usually referred to as the EMBI (emerging markets bond index) spread.

Assume that all foreign borrowing is in dollars, so that foreign investors can choose between foreign and domestic dollar-denominated assets. Let D be debt vis-à-vis the rest of the world, expressed in dollars. Assume then that net capital inflows (capital inflows minus capital outflows and interest payments on the debt), expressed in dollars and denoted by F, are given by

$$F = F\left[\hat{r} - r^* - (1+\theta)x, D\right], \delta F/\delta\left[\hat{r} - r^* - (1+\theta)x\right] > 0,$$
$$\delta F/\delta D < 0, \theta > 0.$$

Net capital inflows thus depend on the EMBI spread, adjusted for a risk premium. The assumption that θ is positive captures the home bias of foreign investors, who are assumed to be the marginal investors.[5] When risk increases, foreign investors, if they are to maintain the same level of capital flows, require a larger increase in the premium than domestic investors.

Net capital inflows also depend, negatively, on foreign debt. To think about the dependence of F on D, assume, for example, that a proportion a of the debt is short-term debt (that is, debt due this period) and that the rollover rate is given by b. Then, in the absence of other inflows, net capital flows are given by $-[a(1 - b) + \hat{r}]D$. Thus the higher the debt, or the higher the proportion of short-term debt, or the lower the rollover rate, the larger net capital outflows will be.

4. If the exchange rate were expected to change, then the domestic dollar rate would be given by $\hat{r}$ plus expected depreciation. This, in turn, would introduce a dependence of net flows, considered below, on the expected change in the exchange rate.

5. As the country studies will show, the increase in capital outflows by foreigners was sometimes offset by a symmetric increase in capital inflows by domestic residents (such as in Chile), and sometimes instead reinforced by an increase in capital outflows by domestic residents (such as in Russia). The case where the increase in capital outflows was more than offset by the increase in capital inflows can be captured in our model by assuming a negative value for θ. A more thorough analysis would require explicitly introducing gross flows by domestic and foreign investors separately, each group with its own perception of risks at home and abroad.

Using the relationship between $\hat{r}$ and r, net capital flows are given by

$$F = F(r - r^* - \theta x, D). \tag{1}$$

For a given policy rate and a given dollar interest rate, an increase in perceived risk or an increase in home bias reduces net capital flows.

We turn next to net exports. We normalize both the domestic and the foreign price levels, which we have assumed to be constant, to equal 1. Let e be the nominal exchange rate, defined as the price of domestic currency in dollars or, equivalently, given our normalization, the price of domestic goods in terms of U.S. goods. An increase in e then represents a (nominal and real) appreciation. Assume that net exports, in terms of domestic goods, are given by

$$NX = NX(e, Y, Y^*), \delta NX/\delta Y < 0, \delta NX/\delta Y^* > 0.$$

A decrease in domestic economic activity leads to a decrease in imports and an improvement in net exports; a decrease in foreign activity leads to a decrease in exports and thus a decrease in net exports. Although the Marshall-Lerner (ML) condition is likely to hold over the medium run, it may well not hold over the short run (again, we are looking at the quarter of the shock and the quarter just following the shock)[6]; thus we do not assign either a positive or a negative sign to the effect of a depreciation on net exports.

In a number of commodity-exporting countries, the adverse trade effects of the crisis took the form of a large decrease in commodity prices rather than a sharp decrease in exports; for our purposes, these shocks have similar effects. Thus we do not introduce terms of trade shocks formally in the model.

Let R be the level of foreign reserves, expressed in dollars, or equivalently, in terms of foreign goods. The balance of payments equilibrium condition is thus given by

$$F(r - r^* - \theta x, D) + eNX(e, Y, Y^*) = \Delta R. \tag{2}$$

6. The Marshall-Lerner condition holds that, given domestic and foreign output, a depreciation improves the trade balance. Some analytical results on the short-run effects of an exchange rate change on the trade balance are given in von Furstenberg (2003).

This implies that a trade deficit must be financed either through net capital inflows or through a decrease in reserves.

I.B. Goods Market Equilibrium

Assume that equilibrium in the goods market is given by

$$Y = A\left(Y, r + x, D/e\right) + G + NX\left(e, Y, Y^{*}\right), \tag{3}$$

where A is domestic private demand and G is government spending. A depends positively on income Y, negatively on the domestic borrowing rate $r + x$, and negatively on foreign debt expressed in terms of domestic goods D/e. This last term captures foreign currency exposure and balance sheet effects: the higher the foreign debt (which we have assumed to be dollar debt), the larger the increase in the real value of debt from a depreciation, and the stronger the adverse effect on output.

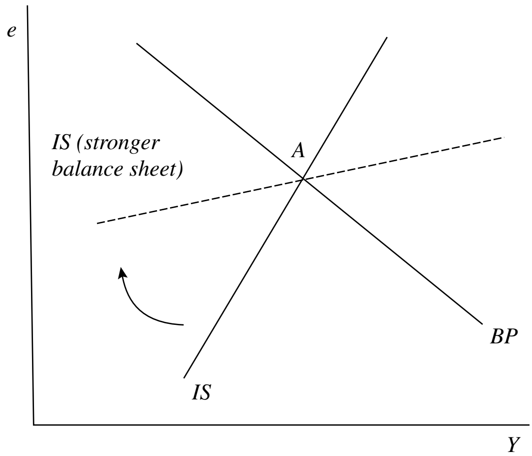
Note that the net effect of the exchange rate on demand is ambiguous. A depreciation may or may not increase net exports, depending on whether the ML condition holds. A depreciation decreases domestic demand, through balance sheet effects. If the ML condition holds and the balance sheet effect is weak, the net effect of a depreciation is to increase demand. But if the ML condition fails, or if it holds but is dominated by the balance sheet effect, the net effect of a depreciation is to decrease demand. A depreciation is then contractionary.

I.C. Equilibrium and the Effects of Adverse Financial and Trade Shocks

It is easiest to characterize the equilibrium graphically in the exchange rate–output space (figure 3). There are three possible configurations, depending on whether the ML condition is satisfied (this determines the slope of the balance of payments curve, BP), and whether, even if the ML condition is satisfied, the net effect of a depreciation is expansionary or contractionary (this determines the slope of the goods market curve, IS). We draw the BP and IS curves in figure 3 under the assumptions that the ML condition is satisfied but that the net effect of a depreciation is contractionary. We discuss the implications of the other cases below.

For given exogenous variables, the balance of payments equation implies a negative relationship between the exchange rate e and output Y. As capital flows depend neither on e nor on Y, for unchanged reserves ($\Delta R = 0$) the BP relationship implies that the trade balance must remain constant. Under the assumption that the ML condition is satisfied, the BP curve is downward sloping: an increase in output, which leads to a deterio-

Figure 3. Output and the Exchange Rate in Equilibrium



Source: Authors' model described in the text.

ration of the trade balance, must be offset by a depreciation, which improves the trade balance.[7]

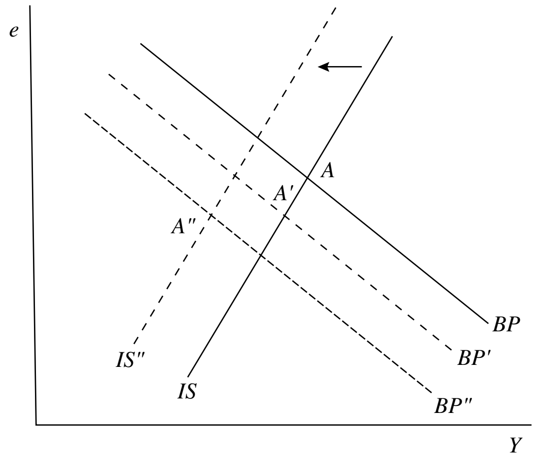
For given exogenous variables, the goods market equilibrium equation implies a positive relationship between the exchange rate e and output Y. Under our assumption that the positive effect of a depreciation on net exports is dominated by the adverse balance sheet effect on private domestic demand, a depreciation leads to a decrease in output. The IS curve is thus upward sloping. The larger the foreign debt, the stronger the balance sheet effect and the stronger the adverse effect of a depreciation on output, and thus the flatter the IS curve.

Equilibrium is given by point A in figure 3. Having characterized the equilibrium, we can now look at the effects of different shocks and the role of policy.

One can think of countries during the crisis as being affected through two main channels: a financial channel, either through an increase in the financial home bias of foreign investors θ, or through an increase in perceived risk x, or both; and a trade channel, through a sharp decrease in foreign output Y^*, and thus a decrease in exports. We consider each of these in turn.

7. Differentiation is carried out around a zero initial trade balance.

Figure 4. Effects of Financial Shocks on Output and the Exchange Rate



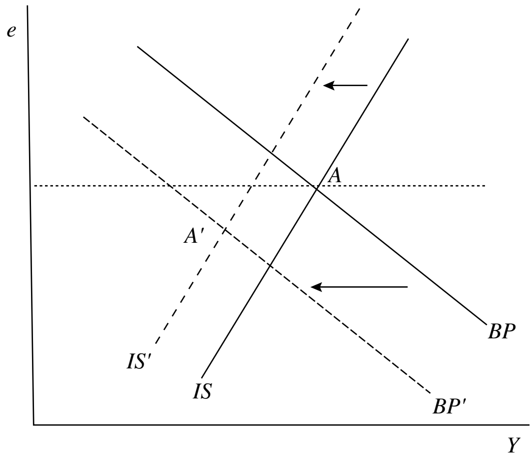
Source: Authors' model described in the text.

Consider first an *increase in home bias.* This was clearly a central factor in the crisis, as the need for liquidity led many investors and financial institutions in advanced economies to reduce their foreign lending. The effect of an increase in θ is shown in figure 4. For a given policy rate and unchanged reserves, net capital flows decrease, and so must the trade balance. This requires a decrease in output at a given exchange rate, and so the BP curve shifts to the left. The IS curve remains unchanged, and so the new equilibrium is at point A′. The currency depreciates (the exchange rate, as we have defined it, falls), and output decreases. The stronger the balance sheet effect, the flatter the IS curve, and thus the larger the decrease in output.

Consider next an *increase in perceived risk,* surely another important factor in the crisis.[8] Indeed, in many cases it is difficult to distinguish how much of the outflow was due to increased home bias and how much was due to an increase in perceived risk. The analysis is very similar in either case, with one difference: whereas an increase in home bias directly affects only net capital flows, an increase in perceived risk directly affects both net capital flows and domestic demand. A higher risk premium increases the domestic borrowing rate, leading to a decrease in domestic demand and, through that channel, a decrease in output. Thus both the IS and the

8. See, for example, Kannan and Köhler-Geib (2009).

Figure 5. Effects of a Trade Shock on Output and the Exchange Rate



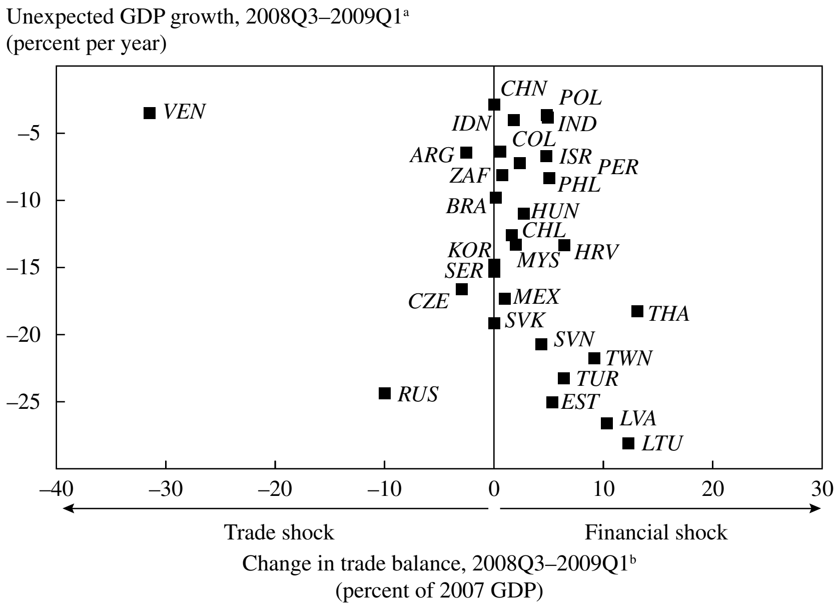
Source: Authors' model described in the text.

BP curves shift to the left, and the equilibrium moves from point A to point A″. Output unambiguously decreases, and the exchange rate may rise or fall. The higher the level of debt, the flatter the IS curve, and the larger the decrease in output.

Finally, consider an adverse trade shock, in the form of a *decrease in foreign output.* Again, sharp decreases in exports (and, for commodity producers, large adverse terms of trade shocks) were a central factor in the crisis. Under our stark assumption that net capital flows do not depend on the exchange rate and, at this stage, the maintained assumption of unchanged policy settings, the BP relationship implies that net capital flows must remain the same, and so, by implication, must net exports. At a given exchange rate, this requires a decrease in imports, and thus a decrease in output. The BP curve shifts to the left. The IS curve also shifts, and it is easy to verify that, for a given exchange rate, it shifts by less than the BP curve. In figure 5 the equilibrium moves from point A to point A′. Output is lower, and the exchange rate falls. Here again, the higher the debt level, the flatter the IS curve, and the larger the adverse effect of the trade shock on output.

Note that in this model both types of financial shock—an increase in home bias and an increase in risk or uncertainty—force an improvement in the trade balance. Under our assumptions and in the absence of any policy reaction, our model implies that trade shocks have no effect on the trade

Figure 6. Change in the Trade Balance and Unexpected GDP Growth in Emerging Market Countries



Sources: IMF, Global Data Source and *World Economic Outlook*; Eurostat.
a. Defined as in figure 2.
b. Seasonally adjusted, annualized change.

balance. More realistically, if we think that part of the trade deficit is financed through reserve decumulation, trade shocks do lead to a deterioration of the trade balance. This suggests a simple examination of the data, looking at the distribution of trade balance changes across countries. This is done in figure 6, which plots unexpected GDP growth over 2008Q3–2009Q1 against the change in the trade balance as a percentage of 2007 GDP. As crude as it is, the figure suggests a dominant role for financial shocks in most countries, in particular in some of the Baltic countries, with trade shocks playing an important role in Venezuela and Russia (in both cases more through terms of trade effects than through a sharp drop in net exports).

We have so far looked at only one of the equilibrium configurations. Next we briefly describe the other two.

Consider the case where the ML condition holds, so that a depreciation improves the trade balance, and the balance sheet effects are weak, so that a depreciation is expansionary.[9] In this case an increase in home bias actually *increases output.* The reason is simple: absent a policy reaction, lower capital flows force a depreciation, and the depreciation increases demand and output. This is a very standard result, but one that seems at odds with reality, probably because lower capital flows affect demand through channels other than the exchange rate. Indeed, if the adverse capital flows also reflect in part an increase in perceived risk, the effect on output becomes ambiguous: the favorable effects of the depreciation may be more than offset by the adverse effect of higher borrowing rates on domestic demand. Trade shocks, just as in the case examined above, lead to a decrease in output.

Consider finally the case where the ML condition does not hold, so that a devaluation leads to a deterioration of the trade balance, and the balance sheet effects are strong, so that a devaluation is contractionary.[10] In this case all the previous results hold, but the decrease in output and the depreciation effects are even stronger. Adverse shocks can lead to very large adverse effects on output, and very large depreciations. Indeed, a further condition, one that puts bounds on the size of the balance sheet effect and the violation of the ML condition, is needed to get reasonable comparative statics.[11]

I.D. The Role and the Complexity of Policies

The analysis so far has assumed unchanged policies. In reality, one of the characteristics of this crisis was the active use of monetary and fiscal policies. Our model allows us to think about the effects of interest rate and exchange rate policies—that is, of using the policy interest rate, or reserve decumulation, or both—and of fiscal policy. A full taxonomy of the effects of each policy in each of the configurations is beyond the scope of this paper. The main insights, and in particular a sense of the complexity of the situation confronting policymakers in this environment, can, however, be given easily.[12]

9. In this case both the IS curve and the BP curve are downward sloping. The IS curve is necessarily the steeper of the two.

10. In this case both the IS curve and the BP curve slope upward.

11. That condition (which is always satisfied if the ML condition holds) is the following: $NX_e < [(A_D D/e^2)NX_Y]/(1 - A_Y)$, where A is domestic private demand and NX is net exports. Graphically, with the exchange rate plotted on the vertical axis and output on the horizontal axis, this requires that the slope of the (upward-sloping) IS curve be less than that of the (upward-sloping) BP curve.

12. Much of this complexity will not surprise those familiar with the earlier Latin American and Asian crises.

Return to the case of an increase in perceived risk, which, in the absence of a policy response, leads to a decrease in net capital flows, a depreciation, and, we shall assume, a decrease in output (which we argued is the most likely outcome). One policy option is to increase the policy interest rate, thus reducing capital outflows but also adversely affecting domestic demand. If the elasticity of flows to the domestic dollar interest rate is small, which appears to be the case in financial crises, the net effect is likely to decrease rather than increase output. If reserves are available, using them to offset the decrease in capital flows, while sterilizing so as to leave the policy rate unchanged, can avoid the depreciation. If a depreciation would be contractionary, this is a good thing. But the direct effect of higher perceived risk on the domestic borrowing rate, and thus on domestic demand, remains, and so output still declines. Thus, to maintain output, sterilized intervention must be combined with expansionary fiscal policy.

Consider next a decrease in foreign output, which, in the absence of a policy response, leads to a depreciation at home and a decrease in domestic output. An increase in the policy rate, to the extent that it increases net capital flows, allows for a smaller depreciation and thus less adverse balance sheet effects. But a smaller depreciation also leads to lower net exports, and a higher policy rate leads to lower domestic demand. The net effect of these three forces may well be a larger decrease in output. To the extent that reserves are available, sterilized intervention avoids the adverse effect of a higher policy rate on output, but the lower net exports may still lead to a decrease in output. In that case, to maintain output, sterilized intervention needs again to be used in conjunction with fiscal policy.

If the policy implications seem complicated, it is because they are. Whether, when faced with a given shock, a country is better off maintaining its exchange rate depends, among other factors, on the tools it uses—the policy rate or reserve decumulation—and the strength of the balance sheet effects it is trying to avoid, and thus the level of dollar-denominated liabilities.

In this context it is useful to note that foreign debt affects the adjustment in two ways. We have focused so far on the first, through balance sheet effects on spending. What matters there is the total amount of foreign currency–denominated debt. The second is through the effects of the foreign debt on the change in capital flows. What matters here is the amount of debt that needs to be refinanced in the short run. The effect then depends on whether, for a given financial shock—be it an increase in home bias or an increase in uncertainty—a higher initial debt leads to a larger decrease in capital flows. Such a second, cross-derivative effect is indeed likely. Recall our earlier example, which showed how debt is likely to affect capital flows, and suppose that an increase in home bias leads investors to decrease the rollover rate. In this case the larger the debt, the larger will be the decrease in capital flows, and the more drastic the required trade balance adjustment. By a similar argument, the larger the current account deficit, and thus the larger the capital flows before the crisis, the larger the required trade balance adjustment.

To summarize: The model has shown how adverse financial and trade shocks are all likely to decrease output, while having different effects on the current account balance. Combinations of reserve decumulation and fiscal expansion can help reduce the decrease in output, but to what extent they can be used clearly depends on the initial level of reserves and on the fiscal room for maneuver. The model also suggests a number of interactions between initial conditions and the effects of the shocks on output. Larger foreign debt, in particular, both through its implications for net capital flows and through balance sheet effects, is likely to amplify the effects of the shocks. With the model and its implications as a rough guide, we now turn to the empirical evidence.

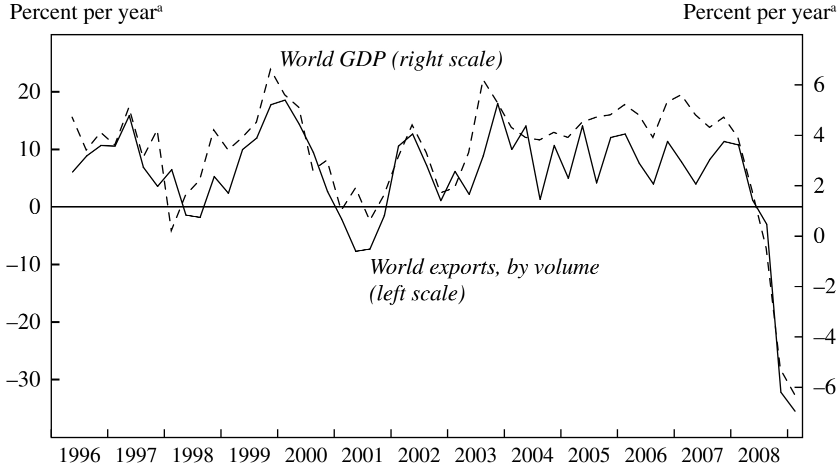
II. Econometric Evidence

The evidence points to two main shocks, to trade and to financial flows. Although our focus is on whether we can explain differences across countries, it is useful to start by looking at the global picture.

II.A. The Collapse of Global Trade and Capital Flows

Figure 7 plots growth in the volume of world exports alongside growth in world output from 1996Q1 to 2009Q2. It reveals in striking fashion the parallel collapse of both output and trade during the crisis, but also that their co-movement in the crisis is not unusual. This second observation has already been the subject of much controversy and substantial research. For the two quarters we are focusing on, growth of world output was –6 percent, and growth of world exports was –30 percent (both at annual rates), implying an elasticity of around 5. The question is whether this elasticity is unusually large, and if so, why. Historical evidence suggests that this elasticity has been increasing over time, from around 2 in the 1960s to close to 4 in the 2000s (using data up to 2005; Freund 2009, *World Economic Outlook* 2009). This suggests that the response of trade to output in this crisis was larger than expected, but not much larger.

Figure 7. Growth in World Output and World Trade, 1996–2009



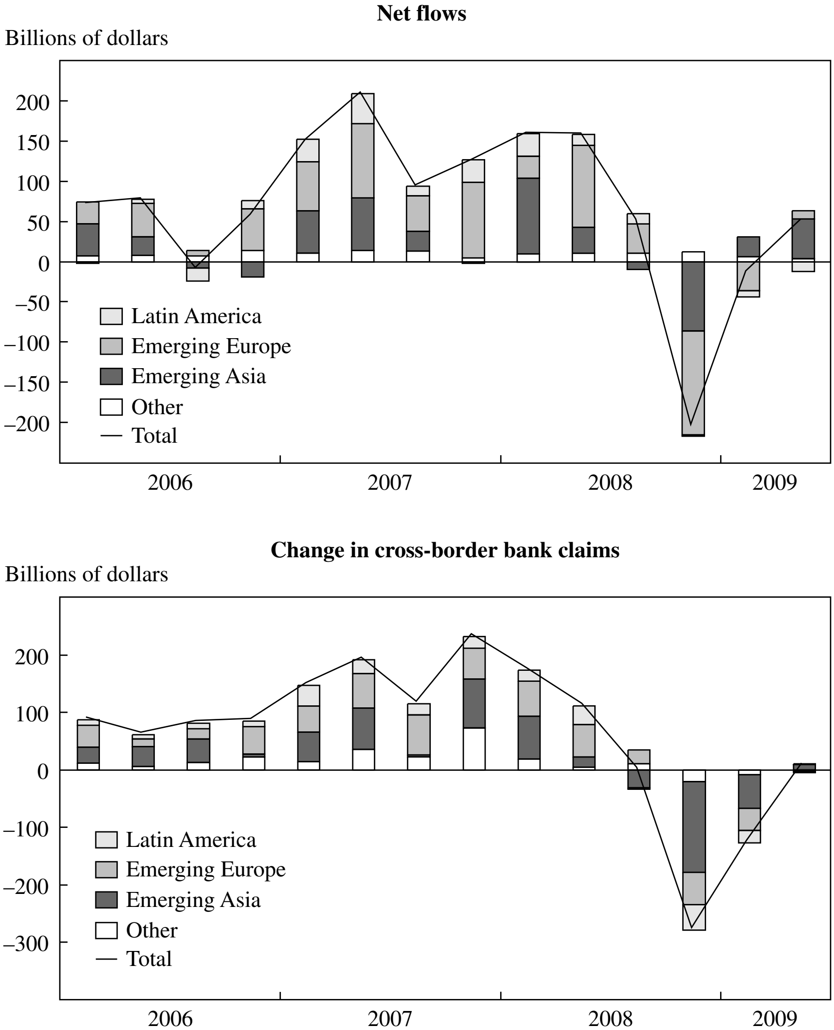
Sources: Netherlands CPB Trade Monitor; IMF, Global Data Source.
a. Quarter over quarter, at an annual rate.

Three main hypotheses for why the response was larger have been explored. The first invokes constraints on trade finance. The second involves composition effects: the large increase in uncertainty that characterized the crisis may have led to a larger decrease in durables consumption and in investment than in typical recessions. Because both of these components have a high import content, the effect on imports was larger for a given decrease in GDP. The third hypothesis relates to the presence of international production chains and the behavior of inventories. High uncertainty led firms to cut production and rely more on inventories of intermediate goods than in other recent recessions, leading to a larger decrease in imports.[13] We read the evidence as mostly supportive of the last two explanations.

The top panel of figure 8 plots net private capital flows, and the bottom panel the change in cross-border bank liabilities, for various regional subgroupings of emerging market countries, from 2006Q1 to 2009Q2. The figure documents the sharp downturn of net flows, from large and positive before the crisis to large and negative during the period we are focusing

13. On trade finance see Auboin (2009). On composition effects see Levchenko, Lewis, and Tesar (2009), Anderton and Tewolde (2010), and Yi, Bems, and Johnson (2009). On inventory adjustment see Alessandria, Kabosky, and Midrigan (2009).

Figure 8. Capital Flows to Emerging Market Countries, 2006–09[a]



Sources: IMF, Balance of Payments Statistics; Bank for International Settlements.
a. Excludes changes in reserves and IMF lending.

on. It also shows the sharp differences across regions, with the brunt of the decrease affecting emerging Europe, and to a lesser extent emerging Asia.

II.B. A Benchmark Specification: Growth, Trade, and Debt

Having documented the global pattern, we now turn to the heterogeneity of country outcomes. We focus on the same 29 emerging market countries as before. The sample is geographically diverse, covering parts of Central and Eastern Europe, emerging Asia, Latin America, and Africa.[14]

Our benchmark specification focuses on the relationship of unexpected growth (the forecast error for output growth during the semester composed of 2008Q4 and 2009Q1) to a simple trade variable and a simple financial variable. Using the unexpected component of growth allows us to separate out the impact of the crisis from domestic trends that were already in place leading up to 2008Q4.[15]

We consider two trade variables. The first captures trade exposure, defined as the export share of GDP (in percent) in 2007. More open economies are likely to be exposed to a larger trade shock. The second is unexpected partner growth, defined as the export-weighted average of actual growth in the country's trading partners, minus the corresponding forecast, scaled by the export share in GDP. For a given export share, the worse the output performance of the countries to which a country exports, the worse the trade shock.[16]

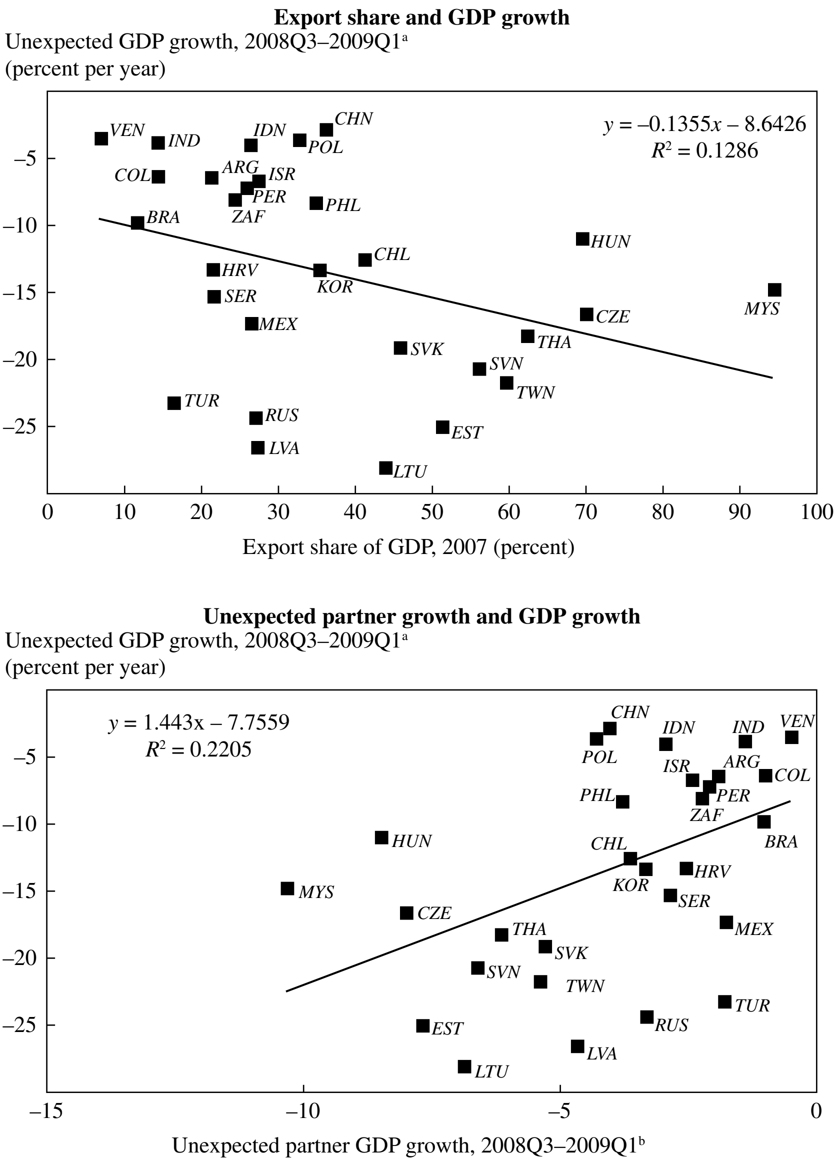
Figure 9 shows scatterplots of unexpected GDP growth against the export share (top panel) and against unexpected partner growth (bottom panel).

14. The sample is the union of all countries classified as "emerging and developing" in the *World Economic Outlook* (WEO) and those classified as either emerging markets or frontier markets in Standard & Poor's Emerging Markets Database (EMDB) for which we have quarterly GDP data and quarterly IMF forecasts of GDP.

15. We have also explored the relationship using two larger datasets. The first is a set of 33 emerging market countries for which quarterly data on GDP are available but forecasts are missing in some cases; in that exercise we used de-meaned growth as the dependent variable, constructed as growth minus mean growth over 1995–2007. The second is a set of 36 emerging market countries for which quarterly data on industrial production can be used to create an interpolated series for quarterly GDP. The results, available in an online appendix (www.brookings.edu/economics/bpea, under "Conferences and Papers"), are largely similar to those presented here.

16. A caveat: if exports to another country are part of a value chain, and thus later reexported, what matters is not so much the growth rate of the first importing country, but the growth rate of the eventual country of destination. That this is relevant is illustrated by the case of Taiwan, whose exports to China are largely reexported to other markets. The decrease in Taiwan's exports to China in 2008Q4 was 50 percent (at an annual rate), much larger than can be explained by the mild slowdown in growth in China during that quarter.

Figure 9. Unexpected GDP Growth, Export Share, and Unexpected Partner Growth in Emerging Market Countries, 2008Q3–2009Q1[a]



Sources: IMF, Global Data Source and *World Economic Outlook*; Eurostat. and authors' calculations.
a. Unexpected growth is defined as in figure 2.
b. Scaled by the export share of 2007 GDP; data are seasonally adjusted at an annual rate.

The fit with the export share is poor, but that with unexpected partner growth is stronger. A cross-country regression of the latter delivers an R^2 of 0.22 and implies that a decrease in unexpected partner growth by 1 percentage point is associated with a decrease in domestic unexpected growth of about 1.4 percentage points.[17]

We consider two financial variables, both of which aim at capturing financial exposure. The first is the ratio of short-term foreign debt to GDP in 2007. Short-term debt is defined as liabilities coming due in the following 12 months, including long-term debt with a remaining maturity of 1 year or less. The second is the ratio of the current account deficit to GDP for 2007. The rationale, from our model, is that the larger the initial short-term debt, or the larger the initial current account deficit, the larger the likely adverse effects of a financial shock.[18]

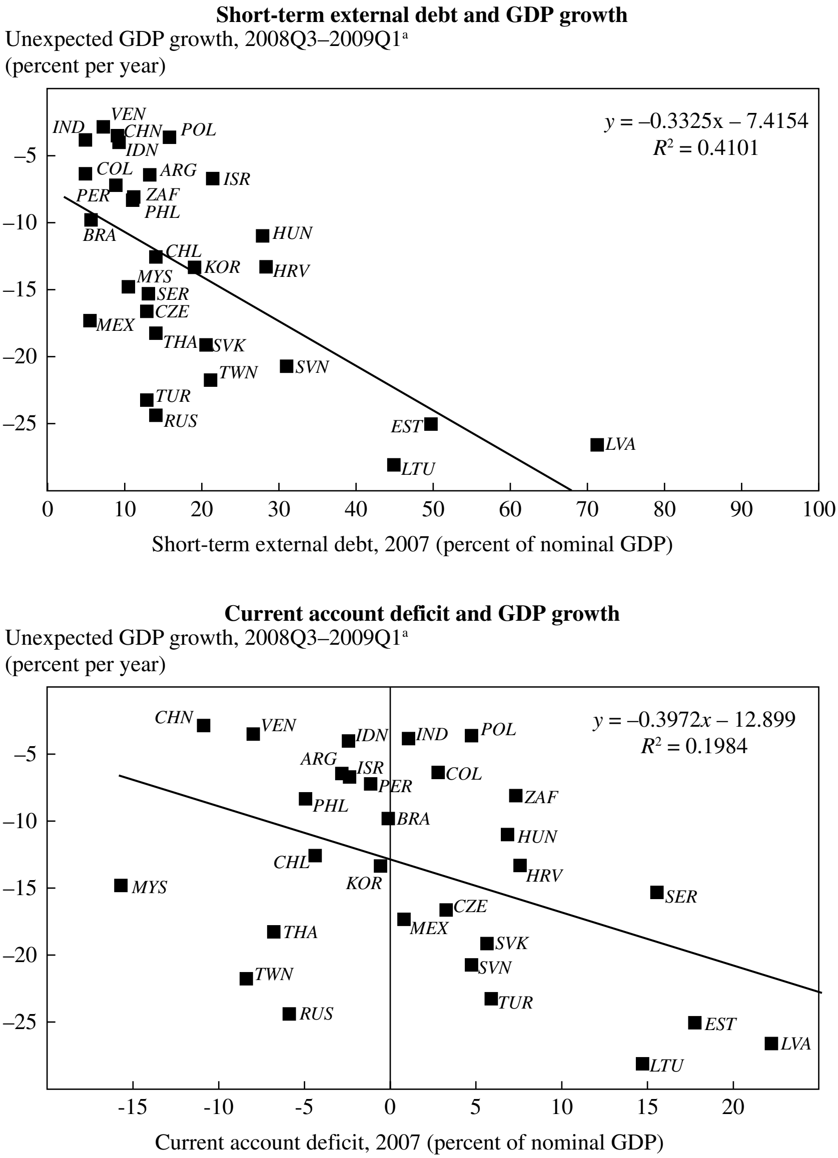
Figure 10 shows scatterplots of unexpected growth over 2008Q4–2009Q1 against short-term debt (top panel) and against the current account deficit (bottom panel), both in 2007. The relationship between short-term debt and unexpected growth is strong. A cross-country regression yields an R^2 of 0.41 and implies that an increase of 10 percentage points in the initial ratio of short-term debt to GDP decreases unexpected growth by 3.3 percentage points (at an annual rate; the relationship remains when the Baltic states are removed from the sample). There is also a relationship between unexpected growth and the initial current account deficit, but it is much weaker than that for short-term debt.

Bivariate scatterplots take us only so far. Table 1 shows the results of simple cross-country multivariate regressions in which unexpected growth is the dependent variable and one of the trade and one of the financial measures are independent variables. The export share, when included in the regression with short-term external debt (column 1-1), is signed as predicted but only weakly significant. Unexpected partner growth is also signed as predicted and significant in all regressions where it is included. Short-term debt is always strongly significant. When the current account deficit is introduced as the only "financial" variable, it has the predicted sign and is significant. When introduced in addition to short-term debt,

17. In our sample the means of unexpected growth, short-term debt to GDP, and unexpected partner growth, respectively, are −13.5 percent, 18 percent, and −4.2 percent, and the respective standard deviations are 7.8, 15, and 2.6.

18. Ideally, one would want to construct a variable conceptually symmetrical to that used for trade, namely, a weighted average of financial inflows into partner countries, using relative bilateral debt positions as weights and scaling by the ratio of foreign liabilities to GDP. Data on relative bilateral debt positions are not available, however.

Figure 10. Unexpected GDP Growth, Short-Term Debt, and the Current Account Deficit in Emerging Market Countries, 2008Q3–2009Q1



Sources: IMF, Global Data Source and *World Economic Outlook*; Eurostat; and authors' regressions.
a. Defined as in figure 2.

Table 1. Regressions Explaining Unexpected GDP Growth with Trade and Short-Term Debt[a]

	Regression				
Independent variable	*1-1*	*1-2*	*1-3*	*1-4*	*1-5*
Export share[b]	–0.09*				
	(0.04)				
Unexpected trading-partner growth[c]		0.73*	1.35***	0.84*	0.93**
		(0.38)	(0.40)	(0.42)	(0.37)
Short-term external debt[d]	–0.31***	–0.28***		–0.23**	
	(0.05)	(0.04)		(0.10)	
Current account deficit[e]			–0.37***	–0.11	
			(0.12)	(0.19)	
Short-term external debt + current account deficit[e]					–0.18***
					(0.03)
Constant	–4.67*	–5.46**	–7.51***	–5.82**	–6.13***
	(2.47)	(2.16)	(1.97)	(2.18)	(2.04)
No. of observations	29	29	29	29	29
Adjusted R^2	0.46	0.46	0.39	0.46	0.46

Source: Authors' regressions.

a. The dependent variable is GDP growth over 2008Q4 and 2009Q1, seasonally adjusted at an annual rate (SAAR), minus the April 2008 IMF forecast of GDP growth over the same period. Robust standard errors, corrected for heteroskedasticity, are in parentheses. Asterisks indicate statistical significance at the ***0.01, **0.05, or *0.1 level.

b. Nominal exports as a percent of nominal GDP in 2007.

c. Trade-weighted average of actual growth in trading partners over 2008Q4 and 2009Q1 minus corresponding forecast growth, SAAR, multiplied by the partner's export share of nominal 2007 GDP.

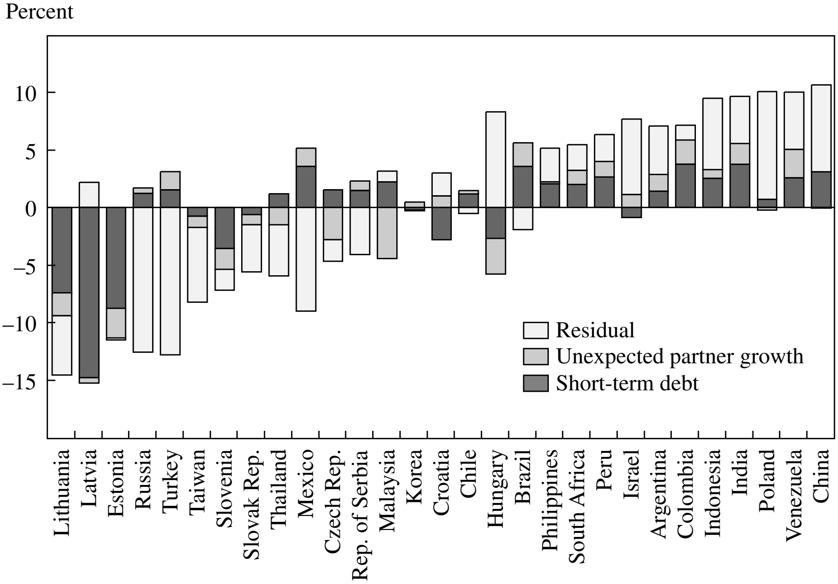
d. Debt with remaining maturity of less than 1 year in 2007, as a percent of 2007 nominal GDP.

e. As a percent of 2007 nominal GDP.

however, it is no longer significant. When the financial variable is the sum of short-term debt and the current account deficit (that is, the short-term financing requirement), the coefficient is less negative than that on short-term debt alone. The estimated constant (which should be zero if we assume that a country with no trade and no foreign debt would have been immune to the crisis) is negative and significant in all regressions. This suggests that some of the average unexpected output decline during the crisis is not explained by the right-hand-side variables.

Nevertheless, these baseline regressions suggest that trade and financial shocks can explain a good part of the heterogeneity in country outcomes. Using results from column 1-2 of table 1, figure 11 decomposes the variation across countries in unexpected growth (relative to the sample average)—similar to what is shown in figure 2—into variation explained by unexpected partner growth, variation explained by short-term debt, and the

Figure 11. Decomposition of GDP Growth in Emerging Market Countries, 2008Q3–2009Q1[a]



Sources: IMF, Global Data Source and *World Economic Outlook*; Eurostat; and authors' calculations
a. GDP semester growth is seasonally adjusted at an annual rate.

residual. Although, in general, countries with worse outcomes had larger debt (this is especially true of the Baltic states) and a larger decline in exports, it is clear that this regression leaves the outcome in some countries (Turkey and Russia, for example) largely unexplained.

In what follows we use the regression reported in column 1-2 of table 1, with unexpected partner growth and short-term debt as the explanatory variables, as our baseline. These results imply that an increase in the ratio of short-term debt to GDP of 10 percentage points leads to a decrease in unexpected GDP growth of 2.8 percentage points, and a decrease in unexpected partner growth of 1 percentage point leads to a decrease in unexpected GDP growth of 0.7 percentage point (much smaller than in the bivariate regression). The magnitude of the short-term debt effect appears to be consistent with that found in other studies.[19]

19. See, for example, Patillo, Poirson, and Ricci (2002). Their results are for the ratio of total debt, rather than just short-term debt, to GDP, and for actual rather than unexpected growth.

Next we explore alternative measures for both trade and financial variables, as well as the effects of institutions and policies. Given the small number of observations, one should be realistic about what can be learned. But as we shall show, some results are suggestive and interesting.

II.C. Alternative Trade Measures

We explored a number of alternative or additional trade measures. None emerges as strongly significant, and no specification obviously dominates our baseline regression.[20]

The trade variable we use in the baseline does not capture changes in the terms of trade. In many countries, however, the crisis was associated with a dramatic decline in the terms of trade. Oil prices, for example, dropped by 60 percent during the crisis semester relative to the previous semester. Thus we constructed a commodity terms of trade variable for each country, defined as the rate of change in the country's export-weighted commodity prices times the 2007 commodity export share in GDP, minus the rate of change in the country's import-weighted commodity prices times 2007 commodity imports as a percent of GDP. The variable ranges from –26 percent for Venezuela to +8 percent for Thailand; 11 countries experience a deterioration of their terms of trade by this measure, and 18 see an improvement.[21] When we add the variable to the baseline regression, its coefficient is close to zero and is not significant, and the coefficients on unexpected partner growth and on short-term debt are roughly unchanged.

The earlier discussion of the response of global trade to output suggests that the composition of exports may be relevant. And indeed, other work (Sommer 2009) has documented a striking relationship among a sample of *advanced* economies between the share of high- and medium-technology manufacturing in GDP and growth during the crisis. To test whether this was the case for our sample of emerging market countries, we constructed such a share for each country, relying on disaggregated data from the UN Industrial Development Organization. Again the coefficient is close to zero and not significant, and the other coefficients are little affected.

20. The full results from the set of alternative regressions described in this and the next subsection are available in the online appendix.

21. A better variable would be the unexpected change in the terms of trade. Unfortunately, forecasts of prices for all relevant commodities are not available. Given that most commodity prices follow a random walk, the use of the actual rather than the unexpected change in the terms of trade is unlikely to be a major issue.

Using the share of exports in GDP overstates the effect of the partner growth variable on demand if exports are part of a value chain, that is, if they are partly produced using imports as intermediate goods. One would like to measure the share of exports by the ratio of value added in exports to GDP, but the data are not available. Instead we constructed a proxy for this share by relying on the import content of exports for the 10 largest export industries (ranked by gross value) for each country, from the Global Trade Analysis Project. The adjustment is typically largest for the small countries of emerging Europe: for example, the export share is reduced by roughly half for Hungary.[22] The results of using this adjusted partner growth measure are similar to those in the baseline. As expected, the coefficient is somewhat larger than that obtained using the original share, but it is not significant, and the other coefficients are roughly unchanged.

The unexpected change in real exports is clearly the most direct measure of the trade shock. The reason for not using it in the baseline is that it is also likely to be partly endogenous, and thus subject to potential bias. We nevertheless ran a regression using the change in real exports (export forecasts do not exist, and therefore we used the actual change rather than the unexpected change). The results are largely similar to those using unexpected partner growth.[23]

II.D. Alternative Financial Measures

Our model suggests that both total foreign debt (through balance sheet effects) and short-term debt (through capital flows) should matter. We therefore explored a number of alternative measures for the financial variable.

We included total foreign liabilities as a percentage of GDP in 2007 as an additional explanatory variable in the baseline regression. This "financial openness" measure is not significant, and the coefficients on both short-term debt and trade are roughly unaffected. These results are consistent with those of Philip Lane and Gian Maria Milesi-Ferretti (2010).

22. This approach does not address another problem raised by value chains and discussed earlier in the context of Taiwan, namely, the fact that exports to another country may then be reexported and thus depend on growth in the ultimate rather than the initial importer country.

23. Taken literally, the coefficient on real exports, 0.43, can be interpreted as the domestic multiplier associated with real exports, whereas the coefficient on partner growth, 0.73, can be interpreted as the multiplier for real exports times the partner countries' average elasticity of imports to GDP.

A question that has been raised, in the context of emerging Europe in particular, is whether the composition of short-term debt, and especially the relative importance of bank debt, was an important factor in determining the effects of the crisis on output. Some have argued that given their problems at home, foreign banks were often one of the main sources of capital outflows. Others have argued that, to the contrary, banks played a stabilizing role in many countries. They point, for example, to the Vienna Initiative, in which a number of major Western banks have agreed to roll over their debt to a number of Central European economies. To explore this question, we decomposed short-term debt into that owed to foreign banks (that is, banks reporting to the Bank for International Settlements) and that owed to foreign nonbanks, both expressed as a ratio to GDP in 2007.[24] The coefficients on both types of debt are negative and significant. The coefficient on bank debt is less negative, suggesting that, other things equal, it was indeed an advantage to have a higher proportion of bank debt.

One might argue from the U.S. experience that the effects of the financial shock on other countries depended on the degree of regulation of their financial system. In a provocative paper, Domenico Giannone, Michele Lenza, and Lucrezia Reichlin (2009) have argued that, controlling for other factors, the "better" the regulation, at least as assessed by the Fraser Institute, the *worse* the output decline during the crisis.[25] Their result suggests that what was thought by some to be light, and thus good, regulation before the crisis turned out to make things worse doing the crisis. When we introduce this index as an additional regressor, it has the same sign as that found by Giannone and others but is not significant.

Finally, we explored the role of net capital (both bank and nonbank) flows directly as right-hand-side variables (instead of short-term debt). These are natural variables to use, but they cannot be taken as exogenous: worse shocks or worse institutions probably triggered larger net capital outflows. We therefore took an instrumental variables approach, using indexes

24. The decomposition is not clean. The numbers for total short-term debt include not only short-term debt instruments, but also longer-term debt maturing within the year. However, the numbers for foreign bank debt, which come from the Bank for International Settlements rather than the World Economic Outlook database, include only short-term debt instruments but not longer-term debt maturing within the year that is owed to foreign banks.

25. The index, which is part of an Index of Economic Freedom, is constructed from measures of the ownership of banks (the percentage of deposits held in privately owned banks), competition (the extent to which domestic banks face competition from foreign banks), extension of credit (the percentage of credit extended to the private sector), and the presence of interest rate controls. The highest value of the index for the countries in our sample is 9.6 for Lithuania, and the lowest is 6.1 for Brazil.

Table 2. Regressions Explaining Unexpected Growth with Reserves[a]

	Regression	
Independent variable	*2-1*	*2-2*
Unexpected trading-partner growth[b]	1.22***	0.53
	(0.43)	(0.44)
Ratio of reserves to short-term external debt, 2007[c]	2.68**	
	(1.15)	
Short-term external debt, as a percent of GDP, 2007[c]		–6.35***
		(1.62)
Reserves as a percent of GDP, 2007[c]		–0.24
		(1.51)
Constant	–21.61***	6.23
	(6.27)	(7.02)
No. of observations	29	29
Adjusted R^2	0.33	0.44

Source: Authors' regressions.

a. The dependent variable is GDP growth over 2008Q4 and 2009Q1, seasonally adjusted at an annual rate (SAAR), minus the April 2008 IMF forecast of GDP growth over the same period. Robust standard errors, corrected for heteroskedasticity, are in parentheses. Asterisks indicate statistical significance at the ***0.01, **0.05, or *0.1 level.

b. Trade-weighted average for the country's trading partners of projected GDP growth over 2008Q4 and 2009Q1 minus actual growth over the same period, SAAR, multiplied by the partner's export share of nominal 2007 GDP.

c. In logarithms.

of foreign bank access and of capital account convertibility (both indexes again from the Fraser Institute) as instruments, in addition to unexpected partner growth and short-term external debt. These plausibly affected growth during the crisis only through their effects on capital flows. The first-stage regressions suggest a strong negative effect of capital account convertibility on net flows: countries that were more open financially had larger net outflows. The second-stage regressions suggest that declines in net capital flows were indeed harmful to growth, more so for changes in bank flows. But these regressions were not robust to the specific choice of instruments.

II.E. The Role of Reserves

Many countries accumulated large reserves before the crisis, and one of the lessons many countries appear to have drawn from the crisis is that they may need even more. Our model indeed suggests that reserve decumulation can play a useful role in limiting the effects of trade and financial shocks on output.

Column 2-1 of table 2 shows that when unexpected partner growth is controlled for, the ratio of reserves to short-term debt is statistically and economically significant. (For reasons that will be made clear below, the reserves variable is entered in logarithmic form.) The coefficient implies that a 50 percent increase in the ratio increases unexpected growth by 1.3 percentage points. This would suggest a relevant role for reserves. The question is, however, whether this effect comes from the denominator or the numerator, or both. To answer it, column 2-2 enters the log of the ratio of short-term debt to GDP and the log of the ratio of reserves to GDP separately. The results are reasonably clear: the coefficient on short-term debt is large and significant, and the coefficient on reserves is incorrectly signed and insignificant.

We have explored this result at some length, using different controls and conditioning or not on the exchange rate regime, and found it to be robust. Although in some specifications the coefficient has the predicted sign, it is typically insignificant and much smaller in absolute value than the coefficient on short-term debt. The econometric evidence is obviously crude and is not the last word, but it should force a reexamination of the issue.[26] Anecdotal evidence suggests that even when reserves were high, countries were reluctant to use them, for fear of using them too early, or that the use of reserves would be perceived as a signal of weakness, or that financial markets would consider the lower reserve levels inadequate.[27]

II.F. The Role of the Exchange Rate Regime

The question of whether, other things equal, countries with fixed exchange rates did better or worse in the crisis is clearly also an important one. Our model has shown that the theoretical answer is ambiguous, depending, for given shocks, on whether the ML condition is satisfied or violated, on the strength of balance sheet effects, and on the policies used to maintain the peg, namely, the combination of policy rate increases and reserve decumulation.

We look at the evidence by dividing countries into two groups according to whether they had a fixed or a more flexible exchange rate regime in 2008. We adopt the classification system used at the IMF, which is based on an assessment of de facto rather than de jure arrangements. Thus the defini-

26. The result is consistent with other studies such as Berkmen and others (2009). Trivedi and Ahmed (2010) also find that the level of reserves did not directly affect output, although larger reserves buffers resulted in a lower rise in country risk premiums and a smaller fall in exchange rates.

27. For more on the "fear of losing international reserves," see Aizenman (2009) and Aizenman and Sun (2009).

Table 3. Regressions Explaining Unexpected Growth with the Exchange Rate Regime[a]

	Regression	
Independent variable	*3-1*	*3-2*
Unexpected trading-partner growth[b]	0.83**	0.91**
	(0.38)	(0.38)
Short-term external debt[c]	–0.22**	–0.10
	(0.08)	(0.24)
Exchange rate regime dummy[d]	–2.72	–0.56
	(3.50)	(5.38)
Exchange rate regime dummy × short-term external debt		–0.14
		(0.26)
Constant	–5.29**	–6.56*
	(2.26)	(3.23)
No. of observations	29	29
Adjusted R^2	0.47	0.48

Source: Authors' regressions.

a. The dependent variable is GDP growth over 2008Q4 and 2009Q1, seasonally adjusted at an annual rate (SAAR), minus the April 2008 IMF forecast of GDP growth over the same period. Robust standard errors, corrected for heteroskedasticity, are in parentheses. Asterisks indicate statistical significance at the ***0.01, **0.05, or *0.1 level.

b. Trade-weighted average for the country's trading partners of projected GDP growth over 2008Q4 and 2009Q1 minus actual growth over the same period, SAAR, multiplied by the partner's export share of nominal 2007 GDP.

c. Debt with remaining maturity of less than 1 year in 2007, as a percent of 2007 nominal GDP.

d. Equals 1 if the country had a fixed exchange rate regime in 2008, and zero otherwise.

tion of fixed-rate regimes we use covers countries with no separate legal tender (including members of currency unions), currency boards, narrow horizontal bands, and de facto pegs. Russia's exchange rate regime, for example, was reclassified from a managed float to a (de facto) fixed rate in 2008, as it tried to stabilize the value of its currency through heavy intervention and use of its ample foreign exchange reserves. We constructed a dummy variable equal to 1 if the country had a fixed exchange rate regime in 2008, and zero otherwise.

Under this classification, countries with fixed exchange rates saw unexpected declines in real output by an average of 18.6 percent (14.6 percent if one excludes the Baltic states) during the crisis semester, compared with 11.3 percent for the group with more flexible exchange rates. Although this appears to be evidence against fixed rates, it does not control for the size of the shock. This is what we do in table 3, starting from our baseline specification. Column 3-1 adds the exchange rate regime as a regressor. The resulting coefficient is negative and insignificant. Its value implies that, controlling for trade and short-term debt, a country with a fixed-rate regime had 2.7 percentage points lower growth. Our model also suggests adding an interaction term between foreign currency debt and the exchange rate. Although exploring the presence of interactions in samples of 29 observations is surely overambitious, column 3-2 introduces an interaction between the exchange rate and the ratio of short-term debt to GDP. The resulting coefficient is negative but insignificant. Taken at face value, it suggests that the adverse effects of short-term debt may have been stronger in countries with a fixed exchange rate.

We also explored the role of fiscal policy. Many countries, for example, India, reacted to the crisis with large fiscal stimuli. In most cases, however, given the decision and spending lags involved, their implementation started either at or after the end of the crisis semester. Nevertheless, we constructed a variable capturing the change in the cyclically adjusted primary fiscal balance from 2008 to 2009 as a ratio to GDP.[28] When added to the baseline regression, this variable was statistically insignificant over the initial period of the crisis. We leave it to further work to examine the effectiveness of fiscal stimulus over a longer period.

In summary, despite the limitations of a small sample, the econometrics suggest a number of conclusions. The most statistically and economically significant variable on a consistent basis is short-term foreign debt. There is some evidence that bank debt had less of an adverse effect than nonbank debt. Short-term debt does not appear to proxy for other variables. Trade, measured by trade-weighted growth in partner countries, also matters; its effect is economically but not always statistically significant. Alternative measures of trade, focusing on composition effects, do not appear to do better. Of the policy dimensions, the most interesting result is the weak role of reserves. Although the ratio of reserves to short-term debt is significant, its effect comes mostly from short-term debt rather than from reserves.

III. Country Studies

Econometrics cannot capture the richness and the complexity of the crisis in each country. Only studies of specific countries can give a sense of how the trade and the financial channels actually operated. For this reason, we turn next to case studies of three countries, Latvia, Russia, and Chile.

28. The use of an annual change is clearly not ideal. Quarterly data are available, however, for only a small number of countries in our sample.

Table 4. Latvia: Selected Macroeconomic Indicators, 2005–09

Indicator	*Average, 2005–07*	*2008 Q1*	*2008 Q2*	*2008 Q3*	*2008 Q4*	*2009 Q1*
GDP growth[a]	10.7	−10.2	−7.4	−6.1	−18.5	−38.4
Current account[b]	−19.0	−17.1	−15.6	−11.5	−7.4	−1.4
Consumer price inflation[c]	7.8	16.3	17.6	15.8	12.2	9.2
Real effective exchange rate[d]	94.8	109.2	112.8	112.4	113.8	120.3
Stock market capitalization[e]	1,829.0	1,814.2	1,828.4	1,480.0	1,166.4	1,051.6
Change in stock market capitalization[c]	32.3	9.5	−16.6	−38.4	−44.4	−40.3

Sources: IMF, Global Data Source and International Financial Statistics; Riga Stock Exchange.
a. Quarter over quarter, seasonally adjusted at an annual rate, percent.
b. Percent of GDP.
c. Year over year, percent.
d. CPI-based, 2000 = 100.
e. Millions of euros.

III.A. Latvia and the Role of Banks

No other country may be as emblematic of this crisis as Latvia. Output there declined at an annual rate of 18½ percent in 2008Q4 and of 38 percent in 2009Q1. (Table 4 provides some basic macroeconomic statistics for Latvia.) In contrast to most other countries, growth in Latvia is forecast to remain negative in 2010. The obvious question is why the output decline was so large.

In the case of Latvia, the right starting point is not the start of the crisis itself, but the boom that the economy experienced in the 2000s—before and after its accession to the European Union in 2004. GDP growth exceeded 6 percent each year from 2000 to 2007, reaching or exceeding 10 percent each year from 2005 to 2007. Inflation, low and stable until 2005, increased to 7 percent by 2006 and to 14 percent in 2007. Asset prices boomed. Stock market capitalization increased by 32 percent a year in nominal terms from 2005 to 2007. The evidence also suggests very large increases in housing prices: in Riga, housing prices increased by 367 percent from 2005 to 2007. The domestic currency, the lat, was pegged to the euro in 2005 (it had been pegged to the SDR previously), so that higher inflation led to a steady real appreciation.

The main cause of the boom was wider access to credit, largely through local subsidiaries of foreign banks, leading to very rapid domestic credit growth. From 2005 to 2007, annual domestic credit growth exceeded

50 percent, leading to high consumption and high investment, in particular residential investment. One result was steadily larger current account deficits, which in 2007 reached an astounding 24 percent of GDP. Capital inflows increasingly took the form of bank flows, from foreign parent banks to domestic subsidiaries. By the end of 2007, gross external debt had reached almost 135 percent of GDP, and short-term external debt was 58 percent of GDP. Foreign ownership of Latvia's banks, primarily by Nordic banks, was 60 percent. Foreign currency debt was 86 percent of the total. More than two-thirds of the loans were backed by real estate. Reserves were only 20 percent of GDP.

In short, Latvia was very much exposed to foreign financial shocks. A slowdown, however, preceded the crisis. By early 2007, signs of overheating and of an impending bust were starting to become apparent. House prices peaked in early 2007 and then started to decline sharply. In February, Standard & Poor's changed its outlook on Latvia from stable to negative. Growth decreased throughout the year and turned sharply negative in each of the first three quarters of 2008. Forecast growth for 2008Q4 and 2009Q1, from the April 2008 *World Economic Outlook,* was −1.5 percent at an annual rate. For the most part, it was the (un)natural end of a boom. Financial factors also played a role. Worried about the decrease in the value of real estate collateral and the likely increase in nonperforming loans, Swedish banks instructed their subsidiaries to decrease credit growth. The (reported) average rate charged by banks to domestic borrowers remained stable, however, until September 2008, suggesting that credit tightening played a limited role in the initial slowdown.

Until September, it appeared that Latvia was headed for a long period of stagnation, perhaps similar to that of Portugal after euro entry. The crisis, however, led to a dramatic decrease in output. Part of this was due to trade. But as figure 9 shows, the decline in GDP was much larger than could be explained by trade. The rest must be attributed to a combination of financial factors.

Despite problems at home, Nordic banks, for the most part, maintained their credit lines to their subsidiaries—but with a sharp deceleration from earlier high rates of credit growth. The reduced level of credit proved insufficient to finance Latvia's large current account deficit. Broad commitments by foreign banks to maintain credit lines were part of the IMF-supported program in December 2008.[29] But the same was not true of

29. These commitments were made more explicit later, in September 2009, through the so-called Vienna Initiative.

Table 5. Latvia: Current Account, Capital Flows, and Reserves, 2005–09
Millions of dollars

Indicator	*Average, 2005–07*	*2008 Q1*	*2008 Q2*	*2008 Q3*	*2008 Q4*	*2009 Q1*
Exports of goods and services[a]	10,524.9	3,843.4	4,265.1	4,341.7	3,507.5	2,816.6
Imports of goods and services[a]	−15,322.7	−5,313.4	−5,954.9	−5,745.2	−4,205.3	−2,853.9
Current account balance[b]	−4,312.8	−1,336.3	−1,397.7	−1,147.3	−610.7	77.1
Net bank flows	3,891.8	707.9	1,207.7	1,245.7	−1,230.4	−1,486.1
Net nonbank financial flows	1,369.0	1,276.2	4.1	−116.8	160.8	600.5
Financial account balance[c]	5,260.8	1,984.1	1,211.8	1,128.9	−1,069.6	−885.6
Exceptional financing from IMF and European Union					814.2	
Change in reserves[d]	966.8	446.3	110.9	−64.7	−979.2	−639.7

Source: IMF, International Financial Statistics.

a. Includes factor income flows.

b. Includes transfers.

c. Excludes changes in reserves and official (IMF) financing.

d. Differs from the sum of the current account balance, the financial account balance, and official financing due to errors and omissions (not shown).

domestic banks. One of them in particular, Parex, with assets equal to 20 percent of GDP and relying heavily on foreign depositors, suffered a run by foreign and then by domestic depositors. In November the Latvian treasury and the central bank stepped in, both to guarantee some of the debt and to provide liquidity. In the second semester, liquidity provision operations associated with Parex alone amounted to $1.1 billion, or more than 3 percent of GDP. Finally, worry about a possible devaluation led to a large-scale shift from lat to euro deposits by domestic residents.

The strategy of the central bank in reaction to these shocks was twofold: first, to avoid balance sheet effects and maintain the peg using reserves, and second, to provide liquidity to the financial system and maintain a low policy interest rate. The result was a large decrease in reserves. Table 5 reports Latvia's current account, financial account, and reserves during this period. (To keep these numbers in perspective, note that Latvian GDP was $33 billion in 2008.) Large net outflows from domestic banks led to large decreases in reserves, only partly compensated through exceptional financing from the European Union and the IMF. In the second half of 2008, the central bank lost roughly one-fourth of its initial reserves. However, the current account achieved a sharp turnaround, from a deficit of \$1.3 billion in 2008Q1 to a small surplus in 2009Q1, which limited further losses in reserves. This turnaround came from a sharp drop in imports, itself reflecting the sharp drop in domestic demand.

This drop in domestic demand raises an important puzzle. Given that the central bank was willing both to use reserves to maintain the exchange rate and to provide liquidity and maintain a low policy rate, why was the decrease in demand so dramatic? Why didn't the banks, which had relied on foreign credit, fully maintain credit by turning to the central bank for liquidity and to the foreign exchange market if they needed foreign currency? In other words, why wasn't sterilized intervention enough to prevent major effects on real activity? The answer is probably twofold.

First, as already noted, foreign banks gave instructions to their subsidiaries to reduce their domestic credit exposure. To the extent that the subsidiaries were limited in the amount of loans they could extend, they had no incentive to borrow at the policy (or at the interbank) rate. In other words, even generous liquidity provision by the central bank would not have led to greater extension of credit by the subsidiaries. In terms of our model, the shadow borrowing rate went up as credit was rationed. Second, doubts about the banks' solvency, coming from the initial shocks, the decrease in housing prices, and the associated decrease in the value of collateral, led, just as in the advanced economies, to a higher interbank rate and, in turn, to higher borrowing rates. The Rigibor, the Latvian equivalent of the LIBOR (London interbank offered rate), went up from 6 percent in August to 14 percent in December. The average rate on lat-denominated loans by banks went up from about 10 percent in August to almost 16 percent in December. In terms of our model, the crisis clearly increased x and thus $r + x$.

We draw two main lessons from the Latvian experience. The first concerns the complex role of banks in the transmission of financial shocks. On the one hand, foreign banks largely maintained their exposure, more so than other foreign investors and depositors. On the other, direct restrictions on credit limited the usefulness of liquidity provision by the central bank. The second, related, and more general lesson is that even when central banks are willing to use reserves and provide liquidity, the adverse output effects of capital outflows on credit and, in turn, on economic activity can still be very large.

III.B. Russia and the Role of Reserves

Aside from the Baltics, Russia is the country in our sample that suffered the largest output decline during the crisis. Although output declined by

Table 6. Russia: Selected Macroeconomic Indicators, 2005–09

Indicator	*Average, 2005–07*	*2008 Q1*	*2008 Q2*	*2008 Q3*	*2008 Q4*	*2009 Q1*
GDP growth[a]	7.9	9.6	5.1	−1.3	−8.8	−29.7
Current account[b]	8.9	7.3	6.4	7.1	3.5	0.9
Consumer price inflation[c]	10.5	12.9	14.8	15.0	13.8	13.8
Real effective exchange rate[d]	163.3	181.5	186.7	187.3	189.5	165.1
Stock market capitalization[e]	140.4	189.0	195.4	109.4	55.8	57.0
Change in stock market capitalization[c]	69.2	4.9	5.4	−39.2	−71.5	−69.9

Sources: IMF, Global Data Source and International Financial Statistics; Russian Trading System Stock Exchange.

a. Quarter over quarter, seasonally adjusted at an annual rate, percent.

b. Percent of GDP.

c. Year over year, percent.

d. CPI-based, 2000 = 100.

e. Billions of dollars.

only 9 percent at an annual rate in 2008Q4, it then declined by 30 percent in 2009Q1. The question, again, is why output fell so steeply.

To answer this question, one needs again to start long before the crisis. When the crisis came, the Russian economy had been booming for some time. Growth had averaged 7 percent per year from 2000 to 2007, and 8 percent from 2005 to 2007. (Table 6 gives basic macroeconomic numbers for 2005–07 and for each quarter from 2008Q1 to 2009Q1.) The boom was due in large part to the increase in the price of oil and the associated increase in oil export revenue, and the economy showed all the signs of a commodity price-led boom. In sharp contrast to the Baltics, however, Russia's boom was accompanied by large current account surpluses, running on average at 10 percent of GDP from 2000 to 2007 and at 8.9 percent of GDP from 2005 to 2007. Large fiscal surpluses reflected high oil revenues, and the public debt fell steadily. In 2007 the primary fiscal balance showed a surplus of 7.4 percent of GDP (the primary nonoil balance showed, however, a deficit of 3.3 percent of GDP), and the ratio of public debt to GDP fell below 10 percent. Oil revenue was partly allocated to two stabilization funds, to smooth the effects of fluctuating oil prices on spending. Inflation was high but stable at around 10 percent. Bank credit growth was extremely high, running at an annual rate of 40 percent from 2001 to 2007.

The current account surpluses, combined with large capital inflows, led to a large buildup of reserves. By December 2007, reserves (including the foreign asset positions of the two oil stabilization funds) had reached

$480 billion, equivalent to 36 percent of GDP. Total foreign debt was $464 billion, of which $114 billion reflected loans to banks, $42.6 billion foreign deposits in banks, and $210 billion loans to households and firms. Of this debt, $361 billion was denominated in foreign currency, and $100 billion was short-term debt.

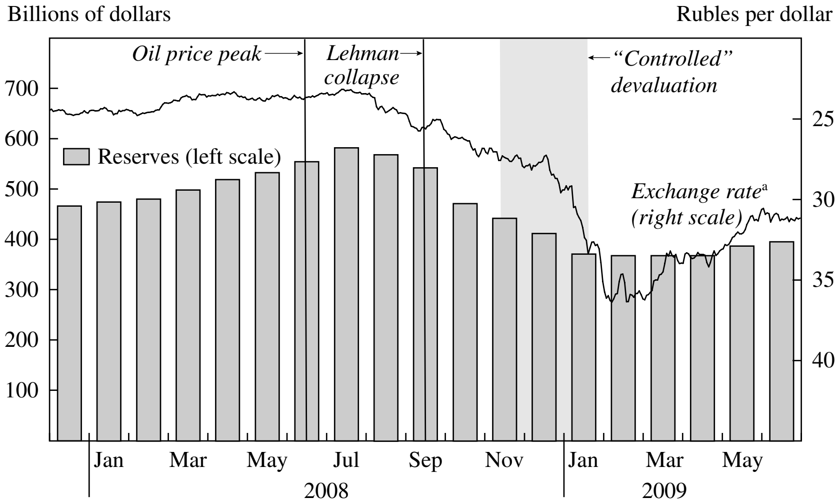
With a large current account surplus, a large fiscal surplus, a smoothing mechanism against oil price fluctuations, nearly no public debt, and a ratio of reserves to short-term debt equal to over 480 percent, one would have expected Russia to manage the crisis well. This was not the case.

The trade shock was severe. The dominant channel was not so much the decrease in export volumes as the decrease in oil prices, which fell from $138 per barrel in July 2008 to $44 per barrel in early 2009. The terms of trade for Russia's overall commodity exports, which accounted for a very large 22 percent of GDP, fell 36 percent during the crisis semester relative to the previous semester. The decline in our terms of trade variable was the third largest in our sample, exceeded by only Venezuela and Chile. The interesting question is whether, given the presence of stabilization funds, the terms of trade decrease had a large adverse effect on demand. Put another way, given that most of Russia's oil revenue goes to the state, was the decline in revenue reflected in fiscal tightening? The answer is not obvious. The increase in the fiscal deficit in 2008Q4 far exceeded the decrease in oil revenue. But this increase was followed by a sharp decrease in the deficit in 2009Q1, while oil revenue was decreasing further. This would suggest a positive effect on demand in 2008Q4 but a strongly adverse effect in 2009Q1, which could help explain the large decline in output in that quarter. What complicates the matter is that Russia typically experiences large fiscal deficits in the fourth quarter for seasonal reasons. Thus, the relevant question is whether the deficit was larger than expected, and this we cannot answer. A strong fiscal stimulus program was put in place in April 2009, too late to have an effect on the period under consideration.

The post-Lehman financial shock was not the first such shock experienced by Russia in 2008. The first, triggered by the war with Georgia, came in August: large portfolio withdrawals led to a 22 percent decline in the stock market from the start of the war in early August to just before the collapse of Lehman and gross outflows of $20 billion. The same happened after Lehman: the stock market declined by 17 percent in two days, after which the Russian authorities closed it for two days.

The initial reaction of the Russian central bank was twofold. First, it sought to use reserves to limit the size of the depreciation and avoid balance sheet effects. (Figure 12 shows the path of reserves and of the exchange

Figure 12. Russia: Reserves and the Exchange Rate



Sources: IMF, International Financial Statistics; Bloomberg L.P.
a. Daily data, inverted scale.

rate from December 2007 to June 2009.) The second was to provide ruble liquidity to banks, through a decrease in reserve requirements, the provision of uncollateralized loans to a larger set of banks, and the provision of $50 billion to the large state bank, VEB, to help firms repay their external debt. More exotic measures were taken as well, such as the allocation of roughly $5 billion from the National Reserve Fund to buy shares, in order to increase the value of the collateral (often their own shares) posted by firms.

Despite these measures, outflows continued at a rapid pace, and the Russian central bank steadily lost reserves: $25 billion in September, $72 billion in October, $29 billion in November, and $29 billion in December. (Table 7 reports the current account, the financial account, and reserves as averages for 2005–07 and for each quarter from 2008Q1 to 2009Q1.) Why were outflows so large? For the most part, because investors perceived that the rate of loss in reserves was too high to be sustained, and thus anticipated a larger depreciation to come. Domestic firms repaid their dollar loans. Domestic depositors shifted from ruble to dollar accounts; the share of foreign currency–denominated bank deposits increased from 14 percent in September to 27 percent in December. Domestic banks shifted from making domestic loans to buying dollar assets, in amounts beyond what was needed to hedge the change in the currency structure of

Table 7. **Russia: Current Account, Capital Flows, and Reserves, 2005–09**
Billions of dollars

Indicator	Average, 2005–07	2008 Q1	2008 Q2	2008 Q3	2008 Q4	2009 Q1
Exports of goods and services[a]	364.0	136.0	156.9	167.2	121.9	74.4
Imports of goods and services[a]	−276.5	−97.2	−130.6	−136.5	−112.2	−64.7
Current account balance[b]	85.4	38.0	26.2	29.7	8.5	9.3
Net bank flows	20.5	−11.3	22.1	−13.2	−51.4	0.5
Net nonbank financial flows	12.5	−14.2	12.8	3.5	−84.4	−32.7
Financial account balance[c]	33.0	−25.6	34.9	−9.8	−135.9	−32.2
Exceptional financing from IMF	−1.2					
Change in reserves[d]	71.4	6.4	64.2	15.0	−131.1	−30.5

Source: IMF, International Financial Statistics.
a. Includes factor income flows.
b. Includes transfers.
c. Excludes changes in reserves and official (IMF) financing.
d. Differs from the sum of the current account balance, the financial account balance, and official financing due to errors and omissions (not shown).

their liabilities. (In view of the expected depreciation, the demand for dollar loans was obviously low.) By mid-November the Russian central bank decided to widen the exchange rate band and allow for a faster depreciation. The ruble was devalued by about 20 percent in January 2009, largely ending the net outflows and reserve losses.

By then, however, it was too late to avoid an output decline. Despite the provision of liquidity, doubts about solvency had increased the interbank rate from 4 percent in July 2008 to 16 percent in January 2009. Over the same period, the shift by banks from domestic loans to dollar assets was reflected in an increase in the average interest rate charged to firms from 11 percent in July 2008 to 17 percent. Credit to households, which had grown by 3 percent monthly from January to September 2008, remained flat for the rest of the year and then decreased by 1 percent monthly from January 2009 on. Credit to firms, which had grown by 2.6 percent monthly from January to September 2008, actually increased further to 3.5 percent monthly from October to January, in some measure because of government pressure on state banks to increase credit, as well as a strong desire of firms to replace dollar debt with ruble debt. It then remained flat from January on, in part because firms began to repay debt they had assumed during the crisis, as the ruble began to appreciate.

In short, Russia was affected by two shocks, a terms of trade shock and a financial shock. One might have hoped that the existence of the stabilization funds for oil would limit the adverse effects on demand of the decrease in oil prices. One might also have hoped that the initial high reserves and low debt positions would limit the effects of the financial shocks. This was not the case, and the story has an interesting twist: the problems did not come so much from capital outflows by foreign investors as from a shift by domestic residents—households, firms, and banks—out of ruble and into dollar assets. In this sense Russia may be the country that most corresponds to the case considered by Maurice Obstfeld, Alan Taylor, and Jay Shambaugh (2010), who argue that the variable to which reserves should be compared is not short-term debt, but rather total liquid assets held by domestic residents. At the start of the crisis, short-term debt in Russia was about \$100 billion, but M2 was about \$430 billion, much closer to the number for reserves. Given the ease with which domestic residents could shift into dollar assets, this may be why it was rational to expect a depreciation, and the equilibrium was self-fulfilling.

Russia's experience also exemplifies the dangers of pegging (or, more accurately, of sharply limiting the decline in the currency) when other actors expect the policy to come to an end and the currency to depreciate. One can question whether, ex ante, Russia's policy was mistaken. Ex ante, it was plausible that the crisis would come to an end sooner, that oil prices would recover, and that reserves would prove more than sufficient. Also (and this is the other side of the same coin), the controlled depreciation allowed firms to decrease their foreign currency exposure and thus suffer smaller balance sheet effects when the depreciation actually came. One can also ask whether a Federal Reserve swap line like those extended to Mexico, Korea, and Brazil would have allowed Russia to credibly maintain its exchange rate and reduce its capital outflows.

III.C. Chile and the Role of Institutions

Like Russia, Chile depends very much on commodity exports—in Chile's case, copper—and is financially open. Yet it suffered a relatively small decline in output: −10 percent in 2008Q4 and −4 percent in 2009Q1 (again at annual rates). The question this time is why the decline was so modest.

Chile entered the crisis in strong macroeconomic shape. From 2005 to 2007 growth was steady, averaging 4.5 percent per year. This performance reflected in part Chile's strong dependence on copper—copper exports were 23 percent of GDP in 2007—and the doubling of the price of copper between 2005 and 2007. Strong copper exports led to large trade and current account surpluses. Inflation was stable, at least until 2008 when it started to increase, leading to a steady increase in the policy interest rate

Table 8. Chile: Selected Macroeconomic Indicators, 2005–09

Indicator	*Average, 2005–07*	*2008 Q1*	*2008 Q2*	*2008 Q3*	*2008 Q4*	*2009 Q1*
GDP growth[a]	4.5	6.7	6.5	−1.4	−9.8	−4.3
Current account[b]	3.5	0.5	0.4	−4.5	−5.7	0.0
Consumer price inflation[c]	3.6	8.0	9.0	9.3	8.5	5.9
Real effective exchange rate[d]	93.8	102.9	100.3	94.0	85.2	91.4
Stock market capitalization[e]	178.0	241.4	200.8	177.7	132.7	149.7
Change in stock market capitalization[c]	21.1	15.1	−15.7	−22.8	−41.3	−38.0

Sources: IMF, Global Data Source and International Financial Statistics; Santiago Stock Exchange.
a. Quarter over quarter, seasonally adjusted at an annual rate, percent.
b. Percent of GDP.
c. Year over year, percent.
d. CPI-based, 2000 = 100.
e. Billions of dollars.

from 5 percent in January to 8.15 percent in September. (Table 8 gives some basic macroeconomic numbers for Chile.)

The country's balance sheets, both public and private, were strong. The effects of copper prices on the fiscal balance, and thus on aggregate demand, were smoothed by a fiscal rule setting annual spending in line with medium-term revenue, including copper revenue, under a conservative copper price assumption. The surplus was accumulated in a stabilization fund, which by 2007 had accumulated \$15 billion. (GDP that year was \$164 billion.) Public debt, including debt of public enterprises, was a low 24 percent of GDP. For 2007 the primary balance showed a surplus of 8.8 percent of GDP, 0.2 percent excluding mining. Private foreign debt, owed mostly by individuals and firms rather than banks, was \$56 billion. The banking sector was heavily regulated and strong, reflecting lessons learned in earlier banking crises. Subsidiaries of foreign banks accounted for roughly half of total bank assets. Central bank reserves were \$24 billion, roughly 75 percent of the country's short-term debt. (Beginning in April 2008, in the face of higher global risk, the central bank had started a reserve accumulation program. By the time the program ended in September, it had accumulated \$5.75 billion.)

The main effect of the crisis was through the trade channel. The crisis was associated with a decrease in exports but also, and more important, with a sharp decline in the price of copper. The decline in our terms of trade measure for Chile was the second largest of the countries in our sample (after Venezuela), and only marginally larger than Russia's. Given the country's fiscal rule, the effect on disposable income and demand was lim-

Table 9. Chile: Current Account, Capital Flows, and Reserves, 2005–09
Billions of dollars

Indicator	*Average, 2005–07*	*2008 Q1*	*2008 Q2*	*2008 Q3*	*2008 Q4*	*2009 Q1*
Exports of goods and services[a]	67.9	23.5	22.7	20.7	16.4	15.1
Imports of goods and services[a]	–65.4	–22.6	–23.9	–24.3	–19.0	–14.5
Current account balance[b]	5.3	1.5	0.1	–2.9	–2.1	0.9
Net bank flows	0.2	1.6	1.2	0.1	–1.1	–2.1
Net nonbank financial flows	–4.1	–1.1	1.0	7.5	2.8	2.9
Financial account balance[c]	–3.9	0.5	2.2	7.6	1.7	0.8
Change in reserves[d]	1.2	0.4	2.4	4.6	–0.9	0.5

Sources: IMF, International Financial Statistics.
a. Includes factor income flows.
b. Includes transfers.
c. Excludes changes in reserves and official (IMF) financing.
d. Differs from the sum of the current account balance, the financial account balance, and official financing due to errors and omissions (not shown).

ited, however; instead the decrease showed up in a sharp decline in accumulations of the stabilization fund, from \$3 billion in 2008Q1 to \$1 billion in 2008Q4. In 2009Q1 the government put in place an additional fiscal stimulus program of \$4 billion; financing needs increased further later in the year by another \$4 billion.

On the financial side, what is most striking is that net capital flows actually remained positive in both 2008Q4 and 2009Q1. (Table 9 reports the current account, the financial account, and reserves as averages for 2005–07 and for each quarter from 2008Q1 to 2009Q1.) Thus, despite a sharp decrease in the current account balance, the decrease in reserves was small—\$1 billion in 2008Q4, followed by an increase of \$0.5 billion in 2009Q1—and associated with a moderate depreciation: the real exchange rate index fell from 100 in 2008Q2 to 85 in 2008Q4 and then recovered to 91 in 2009Q1.

This behavior of reserves and the exchange rate was probably due to two main factors. The first was the central bank's decision to allow the exchange rate to adjust rather than to use the policy interest rate or to rely on reserve decumulation. Only in January 2009, after inflation had substantially declined, was the policy rate lowered, by almost 500 basis points between January and March 2009. Starting at the end of September, the central bank made some dollar liquidity available to banks, but at a fairly large spread (300 basis points initially) over LIBOR.

The second factor was the behavior of gross capital flows. Gross outflows were only marginally higher during the two quarters of the crisis than before. Interestingly, gross inflows increased even more. These inflows came not only from the repatriation of funds by pension funds but also, indeed to a larger extent, from domestic firms and households. This is in sharp contrast to what happened, for example, in Russia, where capital outflows by foreign investors led to capital outflows by domestic residents. How much was due to the decision to let the peso depreciate (in contrast with Russia, which tried to maintain a peg despite the anticipation by investors of a future devaluation) and how much was due to the perception of Chile as a relatively safe financial haven is difficult to assess. The result, in any case, was only a small loss in reserves and a moderate depreciation.

Nevertheless, the trade shocks and the financial crisis had some effect on the real economy. The stock market fell by almost 15 percent from September to December, a small decrease relative to other emerging market countries. And although the interbank rate rose little relative to the policy rate, there was an increase in lending rates of roughly 4 percentage points from September to December, at a time when, in addition, inflation was falling, implying a larger increase in real interest rates.

The overall result was a decrease in demand and in output, but on a more limited scale than in many other countries. The fiscal rule, the framework for smoothing the effect of copper revenue, a strong financial sector, limited foreign currency exposure, and the decision early on to let the peso depreciate probably all played a role in the outcome.

IV. Conclusions

One can read the three preceding sections as first building the bone structure and progressively adding the flesh. The model presented in section I has allowed us to identify and analyze the effects of the main two shocks that affected emerging market countries during the crisis: a sharp decrease in exports (together with a sharp decrease in the terms of trade for commodity producers), and a sharp increase in capital outflows. It showed the dependence of the unexpected output losses on initial conditions, in particular on foreign debt. It showed the complexity of the decisions policymakers faced in this environment, and the effects of using the policy interest rate, the exchange rate, reserve decumulation, and fiscal policy.

The econometrics in section II provided a first pass at the data. Despite the limitations inherent in using a cross-sectional dataset with only 29 observations, our empirical analysis yielded strong evidence that both the trade and the financial channel played important roles. The differing effects of the shocks across countries, coming from different trade and financial exposures and the differing growth performances of countries' trading partners, explain much of the heterogeneity of growth performances during the crisis. When it comes to policy, our most interesting findings are two "nonresults." Countries with fixed exchange rate regimes fared, on average, much worse. However, when we control for other factors, in particular short-term debt, the direct effect of fixed exchange rates largely disappears. This finding is consistent with the ambiguous effect of exchange rates in our model: the outcome depends on the strength of expenditure switching and balance sheet effects. We did not find compelling econometric evidence that international reserves were important buffers in the crisis.

The case studies give a better sense of the many factors that shaped the effects of the crisis in each country, which cannot be captured by econometrics alone. The comparison between Russia and Chile is perhaps the most interesting. Both countries are large commodity producers, and both were hit by a large adverse trade shock. Both were financially open. Russia had larger reserves relative to its short-term debt than Chile. Yet Chile was much less affected by the crisis than Russia. The proximate reasons for Chile's relative success are probably twofold. First, Chile used its fiscal stabilization mechanisms more effectively than Russia did. Second, whereas Chile experienced small capital outflows by foreigners and more than offsetting capital inflows by domestic residents, Russia suffered large capital outflows by both foreigners and domestic residents. The deeper reasons for these differences in capital flows are probably the greater confidence in the macrofinancial structure in Chile than in Russia, and Chile's decision early on to let its currency depreciate, which compares favorably with Russia's initial decision, eventually abandoned, to maintain the parity, giving rise to speculative outflows.

ACKNOWLEDGMENTS We are indebted to Nese Erbil and David Reichsfeld for superb research assistance. We thank Brian Pinto, Irineu de Carvalho, Jorg Decressin, Kristin Forbes, Ayhan Kose, Hélène Rey, Matthew Shapiro, Linda Tesar, and the editors for comments, and Chris Rosenberg, Pablo García, Julie Kozack, and Bas Bakker for very useful information and discussions. The views expressed in this paper are those of the authors and do not necessarily represent those of the International Monetary Fund.

References

Aizenman, Joshua. 2009. "Reserves and the Crisis: A Reassessment." *Central Banking* 19, no. 3: 21–26.

Aizenman, Joshua, and Yi Sun. 2009. "The Financial Crisis and Sizable International Reserves Depletion: From 'Fear of Floating' to the 'Fear of Losing International Reserves'?" Working Paper no. 15308. Cambridge, Mass.: National Bureau of Economic Research (October).

Alessandria, George, Joseph P. Kaboski, and Virgiliu Midrigan. 2009. "The Great Trade Collapse of 2008–09: An Inventory Adjustment?" Working Paper no. 10-18. Federal Reserve Bank of Philadelphia.

Anderton, Robert, and Tadios Tewolde. 2010. "Financial Turmoil, the Global Trade Downturn, and Global Production Chains." Frankfurt: European Central Bank.

Auboin, Marc. 2009. "Restoring Trade Finance during a Period of Financial Crisis: Stock-Taking of Recent Initiatives." Staff Working Paper ERSD-2009-16. Geneva: World Trade Organization (December).

Bakker, Bas, and Anne Marie Gulde. 2009. "The Credit Boom in EU New Member States: Bad Luck or Bad Policies?" IMF Working Paper 130. Washington: International Monetary Fund.

Berglof, Erik, Yevgeniya Korniyenko, and Jeromin Zettlemeyer. 2009. "Understanding the Crisis in Emerging Europe." EBRD Working Paper no. 109. London: European Bank for Reconstruction and Development (November).

Berkmen, Pelin, Gaston Gelos, Robert Rennhack, and James P. Walsh. 2009. "The Global Financial Crisis: Explaining Cross-Country Differences in the Output Impact." IMF Working Paper 280. Washington: International Monetary Fund (December).

Céspedes, Luis Felipe, Roberto Chang, and Andrés Velasco. 2004. "Balance Sheets and Exchange Rate Policy." *American Economic Review* 94, no. 4: 1183–93.

Freund, Caroline. 2009. "The Trade Response to Global Downturns: Historical Evidence." World Bank Policy Research Working Paper no. 5015. Washington: World Bank.

Ghosh, Atish R., Jonathan David Ostry, and Chanalambos G. Tsangarides. 2010. "Exchange Rate Regimes and the Stability of the International Monetary System." IMF Occasional Paper 270. Washington: International Monetary Fund.

Giannone, Domenico, Michele Lenza, and Lucrezia Reichlin. 2009. "Market Freedom and the Global Recession." Paper presented at a conference on Economic Linkages, Spillovers and the Financial Crisis, organized by the International Monetary Fund and the Banque de France chair of the Paris School of Economics, Paris, January 29. www.imf.org/external/np/res/seminars/2010/paris/.

IMF. 2010. "How Have Emerging Markets Coped with the Crisis?" Strategy, Policy, and Review Department (forthcoming).

Kannan, Prakash, and Fritzi Köhler-Geib. 2009. "The Uncertainty Channel of Contagion." IMF Working Paper 219. Washington: International Monetary Fund.

Lane, Philip R., and Gian Maria Milesi-Ferretti. 2010. "The Cross-Country Incidence of the Global Crisis." Washington: International Monetary Fund (January).

Levchenko, Andrei A., Logan T. Lewis, and Linda L. Tesar. 2009. "The Collapse of International Trade during the 2008–2009 Crisis: In Search of the Smoking Gun." Working Paper no. 16006. Cambridge, Mass.: National Bureau of Economic Research.

Obstfeld, Maurice, Alan M. Taylor, and Jay C. Shambaugh. 2010. "Financial Stability, the Trilemma, and International Reserves." *AEJ Macro* 2, no.2: 57–94.

Patillo, Catherine, Hélène Poirson, and Luca Ricci. 2002. "External Debt and Growth." IMF Working Paper 69. Washington: International Monetary Fund.

Rose, Andrew, and Mark Spiegel. 2009a. "Cross-Country Causes and Consequences of the 2008 Crisis: Early Warning," NBER Working Paper no. 15357. Cambridge, Mass.: National Bureau of Economic Research.

———. 2009b. "Cross-Country Causes and Consequences of the 2008 Crisis: International Linkages and American Exposure." Working Paper no. 2009-18. Federal Reserve Bank of San Francisco.

Sommer, Martin. 2009. "Why Has Japan Been Hit So Hard by the Global Recession?" IMF Staff Position Note SPN/09/05. Washington: International Monetary Fund.

Trivedi, Kamakshya, and Swarnali Ahmed. 2010. "Reserves and the EM Crisis Experiences—A Mixed Blessing." *Global Economics Weekly* (March 10).

Von Furstenberg, George M. 2003. "Price Insurance Aspects of Monetary Union." *Journal of Common Market Studies* 41, no. 3: 519–39.

Yi, Kei-Mu, Rudolfs Bems, and Robert C. Johnson. 2009. "The Role of Vertical Linkages in the Propagation of the Global Downturn of 2008." Paper presented at a conference on Economic Linkages, Spillovers and the Financial Crisis, organized by the International Monetary Fund and the Banque de France chair of the Paris School of Economics, Paris, January 29. www.imf.org/external/np/res/seminars/2010/paris/.

Comments and Discussion

COMMENT BY

KRISTIN J. FORBES This paper by Olivier Blanchard, Mitali Das, and Hamid Faruqee asks a well-defined and extremely important question: how did the recent crisis affect emerging markets in late 2008 and early 2009? The answer has critical policy implications both for emerging markets and for the international financial institutions. To answer this question, the paper begins with an intuitive model that clearly lays out the main channels by which the crisis could affect emerging markets, and the effects of different policy responses. Then it reports a series of regressions to test the role of various channels in explaining the spread of the crisis, focusing on the role of trade versus that of finance and the impact of macroeconomic policies. The paper closes with several case studies, which provide important detail on the cross-country regression results—and show the challenges in generalizing about emerging market experiences during the crisis.

This paper should be required reading for anyone attempting to understand how emerging markets were affected during the peak of the crisis. It is straightforward to read and understand and does an excellent job of articulating a model to frame the issues and then evaluating the predictions of the model through cross-country analysis and more in-depth country studies. Both approaches clearly benefit from the authors' mix of academic knowledge and real-world experience. The regression analysis carefully tests a variety of alternative hypotheses and measures, and the results are surprisingly strong given the limited degrees of freedom available. The most robust findings are that the crisis spread to emerging markets through both the trade and the financial channels, but with a more important role for finance, as measured by countries' exposure to short-term external debt. This result is logical and supports anecdotal evidence gathered during the crisis as well as the more detailed analysis in the case studies. The results also suggest that neither exchange rates nor reserve accumulation had much of a direct effect in determining how the crisis affected emerging markets. These results have important policy implications.

The authors have also done an impressive job in addressing many of the concerns that were raised when they presented a draft of this paper at the Brookings Papers conference. My comments will therefore focus on only four issues: the dependent variable, the sample size, omitted variables, and the assumptions about capital flows. These issues are not new to the authors—indeed, they are very candid about the limitations of their data and analysis.

Let me begin by highlighting one important innovation that was already present in the conference version and has been further improved in this version. The earlier version did not focus on explaining growth in emerging markets during the whole of 2008 or 2009, although this is the standard measure used in other, related papers and would have been straightforward to measure. Instead it attempted to explain the difference between growth during the peak semester of the crisis (2008Q4 and 2009Q1) and trend growth (average growth from 1995 through 2007).

This measure of the dependent variable was better than that used in other work, not only because it focused on the change in growth versus the trend, but also because it focused on growth during the peak of the crisis rather than over an entire year. Growth in many countries was strong both at the start of 2008 and at the end of 2009, so that focusing on annual growth could have missed the full impact of the crisis. This measure, however, still had the shortcoming of overstating the impact of the crisis on countries that were already expected to have slower growth in 2008Q4 or 2009Q1 for reasons unrelated to the crisis. (For example, annual growth in Latvia was already expected to slow from a trend rate of 8.8 percent from 2000 to 2007 to 3.6 percent in 2008, before any effect of the crisis, according to IMF data.) The published version of the paper adjusts for this by focusing on "unexpected growth"—the forecast error for output growth during the 2008Q4–2009Q1 semester—instead of growth versus trend. This choice of measure should more accurately capture how the crisis changed growth in these countries, which is the key variable of interest.

One challenge resulting from this choice of measure of the growth shock, however, is that the available data are limited. Quarterly growth data are not available for many emerging markets and other developing countries, and several of the remaining countries lack the necessary forecast data, so the main regressions have a maximum of 29 observations. Many emerging markets are omitted from the sample, such as Bahrain,

Table 1. Regressions Explaining GDP Growth in Emerging Market Countries[a]

	Full sample			
Independent variable	*Errors not clustered*	*Errors clustered by region[b]*	*Sample omits Eastern Europe[c]*	*Sample omits Estonia, Latvia, and Lithuania*
Unexpected growth in partner countries	0.732*	0.732**	0.783	0.659*
	(0.374)	(0.184)	(0.454)	(0.371)
Short-term external debt	–0.279**	–0.279**	–0.463	–0.265*
	(0.044)	(0.041)	(0.322)	(0.135)
No. of observations	29	29	19	26
Adjusted R^2	0.46	0.46	0.24	0.21

Source: Author's regressions.

a. The dependent variable is unexpected GDP growth in 2008Q4 and 2009Q1, defined as the difference between actual growth and the International Monetary Fund's April 2008 forecast; all growth rates are annualized. Except where stated otherwise, the sample consists of the 29 countries included in the main regressions in Blanchard, Das, and Faruqee (this volume). Numbers in parentheses are standard errors. Asterisks indicate statistical significance at the *10 percent and the **5 percent level.

b. Regions are Eastern Europe, Asia, Latin America, and other.

c. The omitted countries are Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Serbia, Slovak Republic, and Slovenia.

Bangladesh, Bolivia, Botswana, the Dominican Republic, Ecuador, Egypt, El Salvador, Jamaica, Jordan, Kazakhstan, Kuwait, Morocco, Pakistan, Romania, Singapore, Uruguay, and Vietnam. Moreover, the sample is dominated by countries in Eastern Europe—just over one-third of the sample is from this region. In comparison, more traditional emerging market samples that do not rely on quarterly data generally have less than 20 percent of the sample from Eastern Europe. Moreover, this unbalanced sample is not random, because the overrepresentation of Eastern Europe results from requirements on EU members to report quarterly data.

The authors are candid about this shortcoming with the sample size and careful not to ask too much of the data, given the limited degrees of freedom. Nonetheless, the small sample raises questions about whether the results are driven by outliers or by patterns in Eastern Europe or other small groups of countries that may not apply to the full set of countries. My table 1 reports several tests to see whether this is important. I focus on the main regression results in column 1-2 of the authors' table 1, in which the trade channel is measured by unexpected growth in trading partners and the financial channel by short-term external debt. The first column in my table replicates the results in the paper. The second column clusters errors by region—which, one could argue, is the preferred method of estimation. This increases the significance of the trade variable, and the financial variable remains significant. The next column then drops Eastern

Europe from the sample. The coefficients on both the trade and the financial variables are now insignificant. This suggests that patterns in Eastern Europe may be driving the results, but because the sample size is now so small, it may be too much to expect statistically significant results. To maintain a larger sample and some representation of Eastern Europe, the last column drops just three countries—Estonia, Latvia, and Lithuania—that appear to be outliers when residuals are plotted. Now the coefficients on partner growth and short-term debt are both borderline significant (at the 10 percent level), suggesting that these three countries in Eastern Europe may be important in driving the results.

This series of results suggests that it may be worth expanding the sample size to ensure that the results are not driven by a small subset of countries or by the specific characteristics of Eastern Europe. Of course, this is much easier said than done. One solution would be to continue using quarterly growth data, but to add countries that are traditionally classified as developed even though they share some characteristics with countries in the emerging market sample. For example, why not include Greece, Iceland, Italy, Portugal, and Spain? Income per capita in each of these countries is about the same as in Israel, Slovenia, or Taiwan—all of which are in the sample and are generally classified as emerging markets. The challenge in including these Western European countries may be political, in the sense that they might not appreciate being classified as "emerging markets"—especially by a group of authors from the International Monetary Fund.

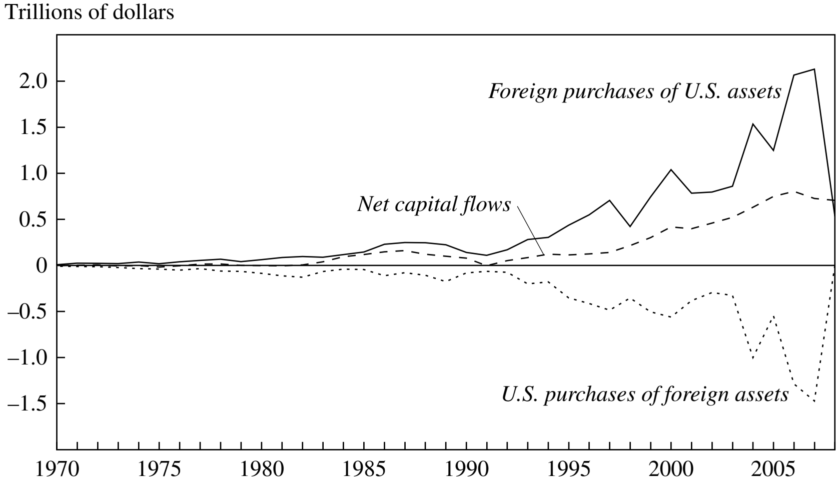
A related issue to consider when interpreting the results is the possibility of omitted variables. The literature on contagion suggests a number of other mechanisms by which the crisis could have affected emerging markets (see Claessens, Dornbusch, and Park 2001). For example, the paper interprets the significant negative coefficient on short-term debt as showing the importance of the financial channel in spreading the crisis. But is there an omitted variable, correlated with short-term debt, that actually drives this relationship? For example, are countries that are riskier and more vulnerable more likely to have higher short-term debt ratios? Probably. And wouldn't these more risky and vulnerable countries be more likely to experience a large growth slowdown during the crisis as risk aversion increases—independent of their share of short-term debt? Similarly, other work on contagion has discussed how trade can spread crises through different effects, for example by affecting import demand and competitiveness (see Forbes 2004). The paper tests its measures of the trade channel individually, but it should test them simultaneously along with the various financial measures. Of course, the challenge in controlling for many of these factors simultaneously is again the small sample size, which again underscores the importance of extending the sample to more countries.

My final comment relates to the authors' model and its relationship to the empirical results. In the model, net capital inflows depend on the EMBI spread adjusted for a risk premium and home bias. A key assumption is that an increase in perceived risk or an increase in home bias causes investors and financial institutions in developed countries to reduce their foreign lending and thereby reduce net capital flows to emerging markets. This assumption is critical for the analysis. The reduction in net capital flows that results from an increase in home bias (assuming a given policy rate and unchanged reserves) reduces the trade balance, causes the home currency to depreciate, and lowers output. The model yields similar results if there is an increase in risk aversion: net capital flows and output again decline, although the effect on the exchange rate is ambiguous.

But how valid is the assumption that when the crisis hits, the result is necessarily to reduce net capital flows? This has been the standard assumption in a large literature on "sudden stops," which argues that during crises, capital flows to emerging markets suddenly cease (see Calvo 1998). But there has been little formal testing of this hypothesis. The authors deserve credit for at least mentioning that this assumption may not hold in all cases, although they leave exploring the ramifications for the model and the empirical analysis for future work. Moreover, the case study on Chile provides a clear example of an emerging market where this assumption does not hold—a great example of the benefits of doing detailed case studies.

But is this pattern of increasing rather than decreasing net capital inflows unique to the Chilean experience, or is it a broader phenomenon? My figure 1 shows gross capital inflows and outflows and the resulting net capital flows for the United States during the crisis. (I focus on the United States because data distinguishing gross flows by domestic from those by foreign investors are readily available.) The figure shows that gross capital inflows from foreigners fell in late 2008. At the same time, however, gross capital outflows by domestic investors were negative, suggesting that they brought home large amounts of capital previously invested abroad. As a result, net capital flows into the United States actually increased during this period. Granted, the United States is not an ideal comparator, as it is a developed country with large and liquid capital markets, which may have become relatively more attractive to investors during the crisis. The example does show, however, that changes in investment by domestic residents

Figure 1. United States: Capital Inflows and Outflows, 1970–2008



Source: Bureau of Economic Analysis.

can easily overwhelm changes by foreigners and lead to a net increase instead of a net decrease in capital flows during a crisis.

Do any countries other than the United States and Chile exhibit this pattern? As a rough test, I examine a group of 101 countries to see whether net capital flows in 2008Q4 were larger or smaller than in 2007Q4.[1] Table 2 shows that in the full sample, net capital flows increased in 45 countries but decreased in 56. Many of the countries in which capital inflows increased, however, are developed countries. The last row of the table therefore looks at the patterns for emerging markets only; it shows that emerging markets were more likely to see a decrease in net capital flows than an increase during the crisis. This "sudden stop" is apparent in many of the major emerging markets, including Argentina, Brazil, Peru, Poland, Russia, South Africa, and Turkey. But the table also shows that the pattern of increasing instead of decreasing net capital inflows is not unique to Chile among developing countries. In fact, even many countries in the authors' sample—including Colombia, the Czech Republic, Israel, Mexico, and Thailand—experienced a net increase in net capital flows in 2008Q4 over 2007Q4, contradicting their model's key assumption.

1. The sample includes all countries for which data were available. I focus only on the fourth quarter because, as of this writing, data for 2009Q1 are not as widely available, and only by comparing similar quarters can one control for seasonal effects that can significantly affect capital flows.

Table 2. Countries in Which Net Capital Flows Increased or Decreased from 2007Q4 to 2008Q4

	No. of countries	
Sample	*Increase*	*Decrease*
Full sample	45	56
Developed countries	14	7
Emerging markets	31	49

Source: International Monetary Fund, International Financial Statistics.

Given that this key assumption of the model does not appear to hold for a number of countries, many of its key predictions might not apply to this subset of countries. For example, for countries with net capital inflows during the peak of the crisis, the financial channel would not be expected to have as large an effect. To test this, it would be straightforward to repeat the main regression analysis but split the sample into two groups: those with net capital inflows (or at least not large outflows), and those with large net capital outflows. Given the small sample size, this would certainly be pushing the degrees of freedom, but it could show very different effects of the crisis for these two subsamples of emerging markets.

To conclude, this paper addresses a very important question: how did the crisis spread to emerging markets? It does an excellent job of laying out the key issues and testing several different hypotheses. It takes pains to evaluate several different theories but is challenged by the very stark limitations of the data—especially the small sample size, which makes it difficult to control for various effects and relationships simultaneously. Nonetheless, the empirical results seem fairly robust, especially given the limitations of the exercise, suggesting that financial mechanisms were likely the most important factor in transmitting the crisis to emerging markets during late 2008 and early 2009. Although this paper may not be the last word on the issue, it presents convincing evidence on how the crisis spread and should provide an excellent resource for anyone seeking to understand why a crisis that started in the U.S. subprime housing market had such virulent effects in emerging markets around the world.

REFERENCES FOR THE FORBES COMMENT

Calvo, Guillermo A. 1998. "Capital Flows and Capital-Market Crises: The Simple Economics of Sudden Stops." *Journal of Applied Economics* 1, no. 1 (November): 35–54.

Claessens, Stijn, Rudiger Dornbusch, and Yung Chul Park. 2001. "Contagion: Why Crises Spread and How This Can Be Stopped." In *International Financial Contagion,* edited by Stijn Claessens and Kristin J. Forbes. Norwell, Mass.: Kluwer Academic Publishers.

Forbes, Kristin J. 2004. "The Asian Flu and Russian Virus: The International Transmission of Crises in Firm-Level Data." *Journal of International Economics* 63, no. 1: 59–92.

COMMENT BY

LINDA L. TESAR[1] The U.S. recession that began in late 2007 had significant spillover effects on the rest of the globe. This paper by Olivier Blanchard, Mitali Das, and Hamid Faruqee studies the impact of the U.S. financial crisis and the accompanying economic contraction on 29 emerging market countries in South America, the Middle East, Eastern Europe, and Asia. As figure 2 of the paper shows, the contraction experienced by emerging markets over the interval 2008Q4–2009Q1 was far from uniform. Lithuania, Latvia, Estonia, and Russia experienced "unexpected" economic growth rates (the difference between actual growth and the April 2008 IMF forecast) on the order of negative 25 percent, while Poland, Venezuela, and China experienced only mild declines. The objective of this paper is to explain the heterogeneity in these negative growth rates. In particular, the paper seeks to isolate which of two channels of transmission—openness to trade and openness to capital flows—is the more significant in accounting for cross-country differences in growth rates during the crisis.

This is a thought-provoking paper on an important and timely issue. It is well written and clear in its objective and in presenting its findings. The paper begins with a simple model of a small, open economy that trades with the rest of the world and has access to international credit markets. The model is a highly stylized IS-LM framework—one that abstracts from dynamics, expectations, and uncertainty—that the authors use to perform simple comparative static exercises. In this framework, a decrease in demand for a country's exports or a shift away from its assets will contract the aggregate budget constraint and, conditional on endogenous shifts in the exchange rate or adjustments in fiscal policy, will lead to a contraction in output.

The model motivates the regressions that are the core of the paper. In essence, the authors run a horserace between various measures of openness

1. I thank the authors for making the data used in their paper readily available, and my student Logan Lewis for his help in analyzing the data.

in the current and the capital accounts on the cross section of unexpected GDP growth rates in emerging markets during the two quarters of interest. The overall conclusion is that both channels played a role in global transmission, although the financial channel dominates in terms of statistical significance and magnitude. Using the authors' data, I was able to verify that the core results are robust to changes in the specification of right-hand-side variables, sample selection, and other factors. My comments therefore focus largely on the interpretation of the results and whether the takeaway from this paper is really as straightforward as the authors suggest.

THE THOUGHT EXPERIMENT. The premise of the paper is that emerging markets were the victims of a collapse in global demand for their goods and for their financial assets. The shock that hit emerging markets is assumed to be both external to the countries in the sample and common to all of them. The baseline regression implied by this thought experiment is a simple one: the dependent variable is unexpected growth in GDP in each country, and the independent variables include measures of each country's "exposure" to the shock: for example, trade as a share of GDP as a measure of the trade channel, and exposure to short-term debt for the financial channel. Other right-hand-side variables are tested, and in general, the financial variables come in significant and dominate the trade variables.

Of course, to conclude that the financial channel beats the trade channel, or even that the financial channel results are economically meaningful, one has to impose the all-else-equal assumption. As is clear even in this simple open economy model, the transformation of a fall in foreign demand for a country's exports or its assets into a contraction in output depends on a number of auxiliary assumptions about the structure of the economy. If countries differ in the strength of their financial institutions, in the degree of adjustment in goods prices or the exchange rate, or in elasticities of substitution between home and foreign goods and between home and foreign assets, to list just a few possibilities, the coefficients on the "trade" and the "finance" effects will differ across countries. In addition, there may be endogenous policy responses to the shock, which would mitigate its effects. Indeed, the bigger the exposure to the shock, the more likely other variables such as prices will adjust, and the more likely governments will react. What is effectively being estimated is the *net* effect of the shock on output, which results from a complicated mix of structural differences across countries and heterogeneous policy responses to shocks.

One could, in principle, control for some of these differences in order to isolate the "pure" trade and finance channels. The authors are well aware of the nature of the problem, and in a sense the model itself exposes the various pitfalls in the regression analysis. Some controls are added to the regressions to try to address the issue, but one can only do so much in a regression with 29 observations. Therefore, the results should be viewed as a set of correlations between changes in output and external balances and not as a set of causal relationships.

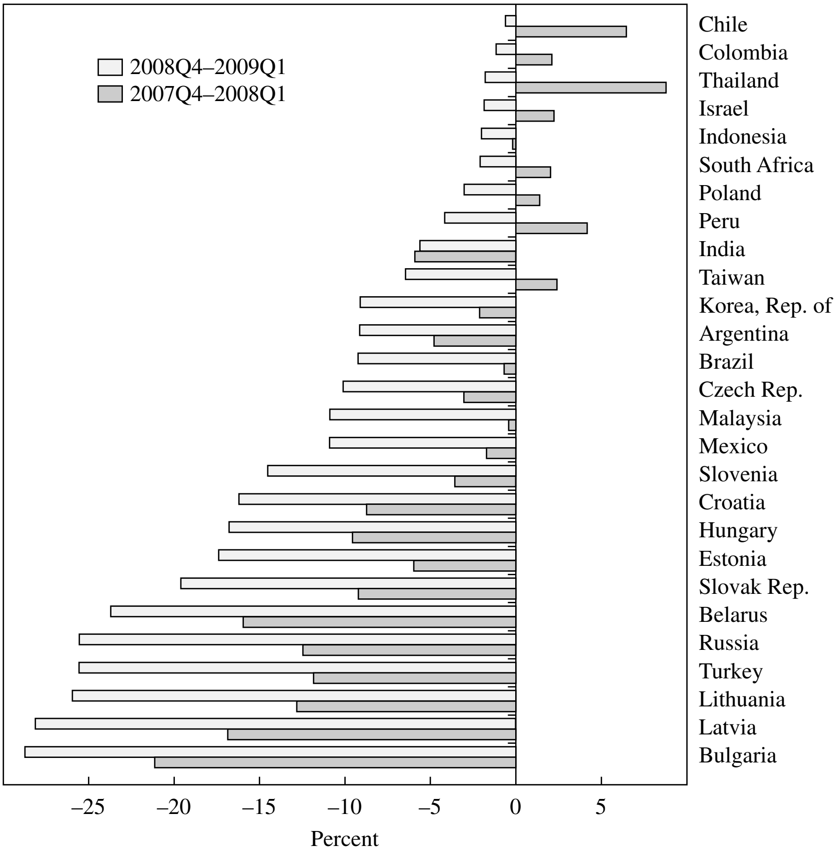
AN ALTERNATIVE INTERPRETATION. An alternative to the emerging-markets-as-victims scenario is that emerging markets, to a greater or lesser degree, rode the same credit boom that fueled the U.S. crisis. Low global interest rates, innovations in the banking sector, and rising real estate prices resulted in an easing of credit and a boom in both private and public expenditure in many countries. In this scenario the contraction in the second half of 2008 was triggered not so much by a collapse in global demand as by the global realization that the party was coming to an end.

The paper's description of the sequence of events in Latvia in 2007 and 2008 casts doubt on the emerging-markets-as-victims hypothesis. The case study of the Latvian crisis notes that "the right starting point is not the start of the crisis itself, but the boom that the economy experienced in the 2000s." Stock prices and real estate prices in Latvia soared in the mid-2000s, and despite rising domestic goods and services prices, the country maintained its peg to the euro. Access to credit, with real estate as collateral, resulted in high rates of consumption and investment growth. By early 2007, the paper notes, "signs of overheating and of an impending bust were starting to become apparent." In early 2008 GDP growth turned negative and asset prices began to fall—all of this well before the external shocks of mid-2008.

Perhaps not surprisingly, the Latvian financial sector increasingly had to shift to shorter lines of credit. Figure 10 of the paper shows that Latvia had the highest ratio of short-term external debt to GDP of any emerging market in the sample in 2007. This raises an important issue for the regression analysis. It is well known that as credit conditions tighten and risk assessments deteriorate, countries may become unable to borrow at long maturities. Short-term debt is then no longer an exogenous variable revealing a country's exposure to external credit market conditions, but an endogenous measure of its own creditworthiness. It is not clear then whether the correct specification is a regression of output growth on short-term debt or the other way around. Again, absent a more complete structural model and the imposition of plausible identifying assumptions, the best one can do is conclude that the two variables are correlated.

The Latvian case also suggests that in order to separate the "victim of external shocks" scenario from the "we got into the same trouble ourselves"

Figure 1. GDP Growth in 27 Emerging Market Countries, 2007Q4–2008Q1 and 2008Q4–2009Q1



Source: Author's calculations using International Monetary Fund data.

scenario, one can either use more country-specific information about the dynamics leading up to the contraction, or look carefully at the timing of the output collapse, or both. The collection of more country-specific information is beyond the scope of this paper, and certainly beyond the scope of this discussion. However, it is fairly easy to look at the patterns in output in the period preceding that studied in the paper.

I show in figure 1 GDP growth rates for 27 emerging market countries over two intervals: 2007Q4–2008Q1 and 2008Q4–2009Q1, the former being the semester one year before that on which the paper focuses, and the latter the crisis semester itself. The countries are ranked by their growth in GDP in the latter period, calculated using GDP volume data from the IMF. This differs a little from the dependent variable in the paper, which is the estimated deviation from the April 2008 IMF forecast. However, the variable used in the regressions and the GDP growth rates calculated here have a correlation coefficient of 0.73, so the message here should not be affected by the use of slightly different data. (The results of the basic regressions in the paper can also be replicated quite closely using GDP volume data rather than the deviations-from-forecast series.)

The figure suggests that the cross section of growth rates in the second semester of 2007 is highly correlated with the cross section of growth rates in 2008. In fact, the two sets of growth rates have a correlation coefficient of 0.93. This means that the countries with weak economic performance in the last half of 2008, after experiencing the "external shock," were the same set of countries with weak performance in the last half of 2007, before the shock. Growth rates across the board were certainly lower in the latter half of 2008 than in the latter half of 2007. But what the paper seeks to explain is the cross-sectional distribution of GDP growth—why some countries fared so much worse than others—not why some countries have persistently low growth rates. If this distribution is the same before and after the shock, then it appears that one should be looking for longer-run reasons for differences in growth rates across countries and not the differential impact of a shock specific to the end of 2008.

Indeed, when the baseline regression is run including the growth rate for the second semester of 2007 as a control, both the trade and the financial variables lose their significance. Depending on the specification, some appear with the opposite sign. I am not suggesting that this is the most appropriate test—a test symmetric to those in the paper would use the deviation of growth in 2007 from the forecast, and there are serious problems of endogeneity in my regression. However, the fact that the regression is not robust to including growth in 2007, together with the very high persistence of growth rates, casts doubt on the empirical evidence that either the trade or the financial channel is the primary explanation for the cross-sectional distribution of growth in emerging markets in the latter half of 2008.

Now, setting the empirical evidence in this paper aside, do I believe that emerging markets were affected by their openness to global markets? Absolutely. But I also believe that those economies benefited from access to those markets in the period leading up to the crisis. The challenge remains what it was in the aftermath of previous emerging market crises: to develop models capable of explaining the dynamics before, during, and after the crisis, and then, through the lens of those models, propose policy tools that can help countries manage their exposure, in good times and in bad.

GENERAL DISCUSSION George von Furstenberg raised three points. First, with respect to the specification of the risk premium, severe positive shocks to that premium were experienced by essentially all countries whether or not they had a collapsing housing bubble. Second, he was surprised that the authors were agnostic about whether the Marshall-Lerner condition holds in the long term for developing countries that generally are obliged to price their exports to market. Third, he thought the paper needed a better proxy for indebtedness effects.

Richard Cooper was troubled that the authors' sample was too small to allow for some necessary distinctions. He suggested thinking more aggressively about expanding the list of countries, to include, for example, smaller countries like Costa Rica. Given the constraint imposed by the need for quarterly GDP figures, he wondered whether the list could be enlarged by looking at industrial production for those countries that typically report monthly or quarterly industrial production data. From the estimated relationship between GDP and industrial production for the countries that have both sets of data, one could then simulate quarterly GDP data for those that do not.

Cooper also would have liked to see the paper distinguish between the impact of trade shocks that initially fall on the government—the case for most oil-exporting countries, as well as Russia and Chile, two of the three countries examined in the case studies—and that of shocks that initially fall on the private sector. He agreed with the authors' position on the Marshall-Lerner condition. Although von Furstenberg's point was valid, if a country has a large external debt denominated in foreign currency, then, starting from a current account deficit, it is very easy to imagine circumstances in which the Marshall-Lerner condition would not be met. Hence, the authors' agnosticism is warranted.

Susan Collins agreed with Cooper that there are often situations, especially in the short run, in which the prerequisites for the Marshall-Lerner condition are not satisfied. She encouraged the authors to devote more attention to the extent to which having accumulated reserves helped, given that their usefulness is currently such a huge issue in the literature and the policy debate. She also noted that for a variety of reasons it is important to think about the role of domestic investors. In the paper's case studies, domestic investors obviously mattered in both Russia and Chile, but in different ways. In the older literature on capital flight from developing countries, before there was a lot of investment by the foreign private sector, domestic investors were seen as the main source of net capital outflows. Not only are domestic investors important, but their role can differ across countries. Because they know the domestic economy better, foreign investors may look to their behavior when deciding whether to enter, stay, or leave.

Kathryn Dominguez agreed with Kristin Forbes that the paper needed to do more to take initial conditions into account. One way to do this might be to examine what the model would have expected for the emerging market countries in the sample when the financial crisis initially hit the developed countries. In the authors' regressions, both initial conditions and the crisis show up as significant factors in the results. As a consequence, countries whose initial conditions were poor and made worse by the crisis are indistinguishable from other countries that were doing well before the crisis but were hit particularly hard by it. These effects should be separated out.

Gregory Mankiw agreed with Forbes that the dataset ought to be expanded to include some developed countries whose income per capita is comparable to those of the richer emerging markets. Beyond that, he suggested including France, Italy, and some other higher-income countries as well. The important question is why the developed countries fared differently in the crisis from emerging market countries, and it seemed natural to at least make the comparison. Indeed, a future Brookings Paper might take the methodology one step further and apply it to U.S. states, whose performance in the crisis was also heterogeneous.

Alan Blinder noted that both discussants had raised the issue of timing, as had Dominguez. He thought it would be interesting to know whether the countries in the authors' sample had already decoupled before the fourth quarter of 2008. The paper gave the impression that there was decoupling, but that it ended with the shock; it would be interesting to see to what extent there was actually "coupling" before the shock. He also suggested exploring whether countries' level of external debt interacted with—and whether their outcomes differed depending on—the nature of the exchange rate regime. Finally, it would also be interesting to know whether the foreign currency composition of countries' debt on the eve of the crisis looked different than it had several years before, and whether countries differed in this respect. This might show to what extent countries had learned the lesson of 1997, which demonstrated the horrific wealth effects possible from issuing debt denominated in foreign currencies.

David Romer noted that although the case studies were interesting in themselves, they lacked a strong link to the rest of the paper. They provided interesting detail on the mechanism by which the shock was transmitted, and they suggested potentially important variables that had not been considered previously and for which good measures were lacking. The case studies might also provide evidence about whether the relationships found in the paper's regressions reflected omitted variables or causal effects. For example, Latvia is an influential observation in the short-term debt analysis, but the case study of that country suggested that its high short-term debt was really a symptom of an unsustainable boom. If Latvia's short-term debt had been lower while everything else remained the same, its outcome might have been closer to what the regression predicted. To some extent, short-term debt seemed to be proxying for other things.

Robert Gordon endorsed Richard Cooper's suggestion of expanding the sample by using quarterly interpolations for countries that publish only annual data. He recalled that his own very first paper had used quarterly data generated using the Chow-Lin method of interpolation, which is still the best technique available and automatically aligns quarterly estimates with annual figures. But any number of methods for interpolating monthly or quarterly data could be used, and indeed one could use different interpolators for different countries.

Valerie Ramey added that in the early postwar period the Economics and Statistics Administration, the predecessor agency of the Bureau of Economic Analysis, had published quarterly nominal GDP data going back to 1939, whereas the currently available data go back only to 1947. She had come across the earlier data and figured out how to create deflators to link them with a plausible series of quarterly real GDP. Her results lined up almost exactly with Gordon's interpolated quarterly real GDP series, especially at the important turning points around the beginning of World War II. If one could successfully do interpolations for the United States going back that far, it should also be possible for more recent low-frequency data from other countries, and the sample could probably be doubled.

Justin Wolfers asked Kristin Forbes whether her discussion implied that she thought that the paper's findings were not very robust. After all, the authors had rerun the regressions in different ways, testing for robustness and stability, and the coefficients had rarely moved by more than half of a standard error. What did change was their statistical significance. Forbes responded that the sample was so small that significance does vary depending on whether one includes or excludes one or two countries, or whether one includes or excludes an additional control, but she thought that with the addition of more countries, the robustness and the results would probably hold up.

Christopher Sims was skeptical of the short-term debt variable, which he saw as basically an endogenous variable that may not be that useful. Both short-term debt and reserves ought to be thought of as endogenous, and the authors' case study of Chile showed that what really matters is the credibility of monetary and fiscal policy. A country with a credible monetary and fiscal regime can borrow if it runs out of reserves; less credible countries cannot. He read the regression results as showing that the regression coefficient on the short-term debt variable did fall when Latvia was taken out of the sample, demonstrating that it was not just the statistical significance that changed.

TOMAS J. PHILIPSON
University of Chicago

DANA P. GOLDMAN
University of Southern California

SETH A. SEABURY
RAND Corporation

DARIUS N. LAKDAWALLA
University of Southern California

LEE M. LOCKWOOD
University of Chicago

Geographic Variation in Health Care: The Role of Private Markets

ABSTRACT The Dartmouth Atlas of Health Care has documented substantial regional variation in health care utilization and spending, beyond what would be expected from such observable factors as demographics and disease severity. However, since these data are specific to Medicare, it is unclear to what extent this finding generalizes to the private sector. Economic theory suggests that private insurers have stronger incentives to restrain utilization and costs, while public insurers have greater monopsony power to restrain prices. We argue that these two differences alone should lead to greater regional variation in utilization for the public sector, but either more or less variation in spending. We provide evidence that variation in utilization in the public sector is about 2.8 times as great for outpatient visits ($p < 0.01$) and 3.9 times as great for hospital days ($p = 0.09$) as in the private sector. Variation in spending appears to be greater in the private sector, consistent with the importance of public sector price restraints.

There is considerable variation in health care utilization and spending across geographic areas in the United States, but little evidence of corresponding differences in health outcomes or satisfaction with care.[1] This variability is often cited as evidence that current levels of health care spending reflect "flat-of-the-curve" medicine, that is, treatment for which the marginal benefit of an additional unit of care is approximately zero.

1. The main data source used to document regional variations is the Dartmouth Atlas of Health Care, which can be found at www.dartmouthatlas.org/ (accessed January 15, 2010).

Interpreted this way, these findings have dramatic implications for the potential to increase the productivity of health care spending, and for this reason they have figured prominently in the policy debate.

However, the evidence on regional variation is almost exclusively limited to the public sector, because it relies on Medicare data. Less is known about the corresponding patterns in the private sector. A venerable literature in economics has argued that private firms and their managers have stronger incentives to restrain costs and boost efficiency than their public counterparts.[2] In the health insurance context, Medicare does not face competition over premiums that might otherwise restrain its costs, and unlike private sector firms, Medicare does not have direct residual claimants whose standard of living improves with the efficiency of the enterprise.

To develop the implications of these incentive differences, this paper provides a theoretical and empirical analysis of how regional variation in health care differs across the public and the private sectors. We first examine conceptually how private efforts to control costs *within* a region, through selection of providers, might translate into differences in care *across* regions. In particular, our analysis implies that utilization controls within regions in the private sector should lead to lower regional variation in the private sector than in Medicare. However, the implications for variation in spending are less clear, because Medicare may also be able to better control prices through its greater monopsony power. If the private sector controls utilization while the public sector controls prices, the result is an ambiguous prediction for variation in spending.

We examine these implications empirically using individual-level data on patients with heart disease, comparing utilization and spending on patients who have private insurance with that on similar patients within Medicare. Data on the former come from a large database of private sector medical claims, and on the latter from the Medicare Current Beneficiary Survey. Both datasets include patient-level demographics and co-morbidities, which allow us to identify regional variation distinct from individual characteristics such as health. The focus on heart disease helps

2. Alchian and Demsetz (1972) showed a greater incentive for shirking and inefficiency in public enterprise, where managers' and employees' own standards of living are unaffected by poor performance. De Alessi (1974a, 1974b) observed that inefficient private firms disappear, whereas inefficient public firms can last for long periods. Spann (1977) argued that private firms typically produce similar goods and services at much lower cost than their public counterparts.

mitigate the confounding impact of regional differences in health status on our analysis.

Our main object of interest is the regional variation in utilization and spending across sectors that cannot be explained by variation in patient characteristics. Our data suggest greater variation in utilization in the public sector: our main analysis suggests that variation in the public sector is about 2.8 times as great for outpatient visits ($p < 0.01$) and 3.9 times as great for hospital days ($p = 0.09$) as in the private sector. There is some evidence of greater variation for the number of hospitalizations in the public sector, but this evidence is less robust. Prescription drug utilization serves as our "placebo" case of insurance that was privately provided in *both* samples during the period investigated. Significantly, and unlike other types of medical care, drug utilization exhibits less variation among Medicare patients. On the other hand, there is greater spending variation in the private sector, suggesting the potential importance of monopsony power in the public sector.

The paper proceeds as follows. Section I provides the conceptual analysis of how differing cost-control measures within a region might lead to differences in regional variation in utilization and spending. Section II reports our empirical analysis comparing regional variation in the public and the private sectors. Section III discusses how our findings relate to the existing literature on health care variation and the resulting policy implications. Section IV discusses some limitations of our analysis and presents several robustness tests. Section V concludes.

I. A Simple Analysis of Regional Variation in Utilization and Spending

This section presents a simple analysis of how private and public incentives interact to create different degrees of regional variation in health care utilization and spending between the public and the private sector.[3] A key assumption is that private insurers have stronger incentives to restrain costs and utilization than a public insurer such as Medicare. This assumption is based on the literature demonstrating that, unlike public enterprises,

3. This analysis is general enough to include several possible sources of regional differences, and in particular it allows for such differences to be efficient. However, differences in liability (Kessler and McClellan 1996, 2002a, 2002b, Baicker and Chandra 2007) or productivity (Chandra and Staiger 2007), for example, may imply differences in efficient levels of care.

private firms have to restrain costs in order to compete on price, and private firms' inefficiencies have direct impacts on the welfare of their owners and employees (Alchian and Demsetz 1972, De Alessi 1974a). For example, private payers may explicitly manage care and exert pressure on providers through utilization review and case management. They can also selectively contract with lower-cost providers, steer patients to preferred providers, and exclude inefficient doctors or hospitals. In addition, prior authorization of large expenditures is prevalent in the private health insurance sector, a practice that allocates major spending decisions to the payer rather than the provider. Finally, private payers can steer patients toward efficient care through benefits management—for example, by not covering certain services unless certain clinical criteria are met. In what follows, we use the shorthand of "utilization restrictions" (UR) to refer to all these practices.

We interpret UR as a limit on the provision of treatments whose costs exceed their benefits. This may still lead to regional variation in utilization, because there is substantial heterogeneity among apparently similar patients in the efficacy of different treatments. Excessive care for one patient may be cost-effective for another.

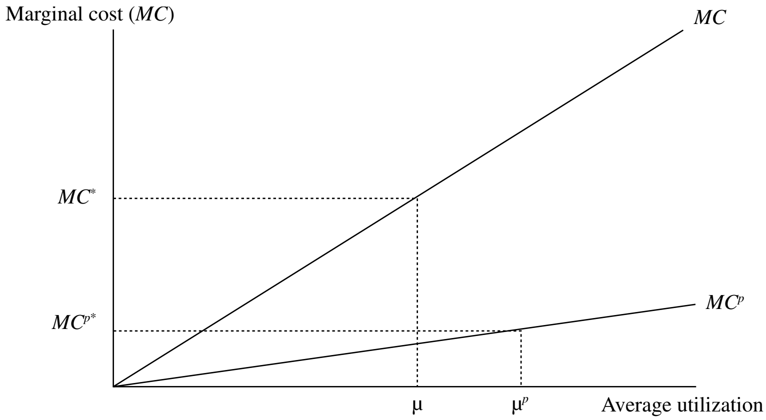
I.A. Causes of Sectoral Differences within Regions

We first consider the level of utilization in both the private and the public sectors. Define y^* as the efficient utilization level, that at which marginal benefit equals marginal cost. Following the earlier literature, we assume that private insurers have stronger incentives to limit utilization that rises above this level. They do this through UR, which we assume places an upper bound on utilization, $y_{UR} \geq y^*$, and perfectly eliminates inefficient utilization above that level.[4] The assumption of full efficiency is an analytical simplification; the positive predictions do not depend on it, and we do not emphasize the normative predictions.

Within any region there is a distribution of providers, who vary in the level of care they would provide to an identical patient. We characterize this distribution using the cumulative distribution function $F(y)$ for the random utilization variable Y. Private payers' UR procedures limit utilization and thereby truncate the support of the providers participating in their plans. This results in the private mean utilization level, $\mu = E(Y|Y \leq y_{UR})$.

4. Imperfect UR has qualitatively similar theoretical implications. The difference is one of degree rather than nature.

Figure 1. Variation in Spending Given Public Monopsony Power



Source: Authors' model described in the text.

This constrained private sector mean is thus lower than the unconstrained public sector mean, $\mu^p = E(Y)$.

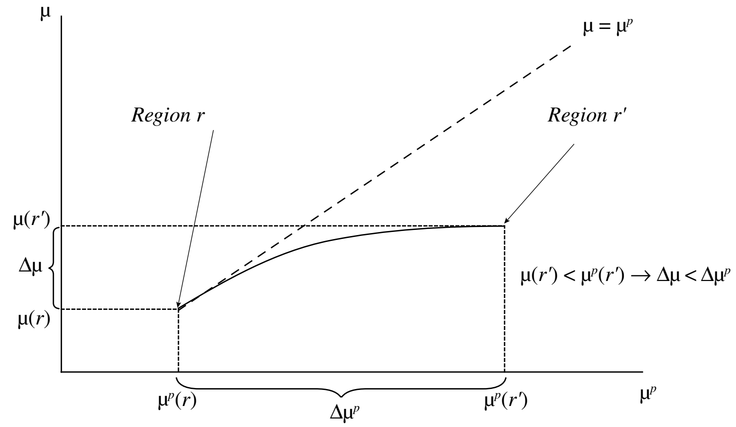
Now consider a pure increase in utilization, holding health status fixed. This can be represented as a rightward shift in the function $F(y)$. Assuming the efficient level of utilization remains fixed, the result is a greater difference in mean utilization across the two sectors, $\mu^p - \mu$. In other words, in regions with providers who have greater tendencies toward inefficiency, the difference in utilization between sectors will be larger.

The second key assumption is the presence of greater monopsony power in the public sector. The result is greater restraint of *prices,* as opposed to utilization, in the public sector. This affects the analysis of variation in spending, which combines the utilization effect and the price effect. If the government pays below-market prices through the exercise of either monopsony power or direct price regulation, the cost curves will differ across sectors. The result is depicted in figure 1. Average spending per patient in the private sector may exceed that in the public sector, if equilibrium marginal cost in the public sector, MC^{p*}, is less than equilibrium marginal cost in the private sector, MC^*.

I.B. Causes of Sectoral Differences across Regions

Next consider how mean utilization for each sector might vary across regions. Define the joint distribution $G(\mu^p, \mu)$ of mean utilization levels across regions. Specifically, suppose that the underlying distribution $F(y)$

Figure 2. Comparing Differences in Utilization across Sectors



Source: Authors' model described in the text.

differs across regions. Figure 2 illustrates how one might then characterize the relationship between changes in the public mean and the mean difference between sectors:

$$\frac{d(\mu^p - \mu)}{d\mu^p} = 1 - \frac{d\mu}{d\mu^p}.$$

For example, consider the case of normally distributed public sector utilization, $Y \sim N(\mu^p, \sigma^2)$. In this case, mean utilization in the private sector follows from the formula for the mean of a truncated normal random variable, $\mu = \mu^p + \sigma\lambda(\alpha)$, where $\lambda(\alpha) \equiv \frac{-\varphi(\alpha)}{\Phi(\alpha)}$ is the inverse Mills ratio and $\alpha = \frac{y_{UR} - \mu^p}{\sigma}$. This implies that the slope of the private mean as a function of the public mean is less than unity or, equivalently, that the between-sector difference rises with the public mean:[5]

$$0 \le \frac{d\mu}{d\mu^p} \le 1$$

$$\frac{d(\mu^p - \mu)}{d\mu^p} \ge 0.$$

5. We use the fact that the derivative of the inverse Mills ratio with respect to α is strictly between zero and 1, $\lambda'(\alpha) \in (0,1)$. (Sampford 1953).

When the public sector provides more care above the efficient level, this raises the between-sector difference. This in turn implies that the variance in the regional means in the public sector will exceed the variance in the regional means in the private sector: $V(\mu^p) > V(\mu)$.

This simple framework leads to several testable empirical predictions: Private provision should lead to lower mean utilization and less variance in mean utilization across regions, but not necessarily lower mean spending. In addition, the difference in utilization between sectors is likely to rise with the mean level of public utilization. Note that all these predictions hold patient health status constant.

II. Empirical Analysis of Regional Variation across Sectors

In this section we describe our empirical analysis of regional variation in the public and private sectors aimed at testing the implications discussed above.

II.A. Data and Empirical Specification

We compare regional variation between a sample of privately insured patients and a sample of Medicare patients. The private data come from a large database of health insurance claims. The data capture all health care claims, including prescription drugs and inpatient, emergency, and ambulatory services, by employees and retirees while they are enrolled in the health plans of 35 Fortune 500 firms. The analytical database integrates component datasets of medical claims, pharmacy claims, and enrollment records. This allows us to calculate spending and utilization for all services provided to the patients over our study period. The enrollment records allow us to identify basic demographics of the patients, including age, sex, and some information on income.[6] Importantly for our purposes, the data also include information on area of residence, coded by metropolitan statistical area (MSA) and 3-digit zip code. This allows us to analyze health care spending and utilization patterns at different levels of geographic aggregation.

Our Medicare sample is taken from the Medicare Current Beneficiary Survey (MCBS), which is administered to a nationally representative sample of aged, disabled, or institutionalized Medicare beneficiaries.

6. Our proxy for income is median household income at the 3-digit zip code level; this is taken from the 2000 Census.

Respondents, whether living in the community or residing in health care facilities, are interviewed up to 12 times over a 4-year period. Institutionalized respondents are interviewed by proxy. There is oversampling of the disabled under 65 years of age and of the oldest old (85 years of age or older). The MCBS uses a rotating panel design with limited periods of participation. Each fall a new panel is introduced with a target sample size of 12,000 respondents, and each summer a panel is retired. The MCBS data include detailed information on self-reported health status, health care use and expenditure, insurance coverage, and demographic characteristics. Additional Medicare claims data for beneficiaries enrolled in fee-for-service plans are also incorporated to provide more accurate information on health care use and expenditure. The MCBS data do not include actual claims data on prescription drugs; all information on prescription drug spending and utilization in the MCBS is self-reported. This leads to a known undercount of drug spending in the MCBS.[7]

Both datasets include information on medical claims that is used to compile utilization, spending, and baseline health information. That is, although the MCBS contains a survey component, all data on spending and utilization are compared with Medicare's administrative claims data (Eppig and Chulis 1997). However, since Medicare does not cover prescription drugs over our sample period, this validation procedure applies to medical care but not drugs. Finally, for both datasets we use information from 2000 to 2006. The one exception is prescription drug utilization and spending: to abstract from the complexities of Medicare Part D's introduction, we eliminate the 2006 data for these variables.

To mitigate differences in health status across sectors and regions, we condition inclusion in the sample on a diagnosis of ischemic heart disease (IHD).[8] We also use the diagnosis codes on medical claims to identify whether patients were treated for any of 30 different conditions in a

7. When estimating the cost of Medicare Part D (for example), the Congressional Budget Office scaled the reported MCBS prescription drug spending up by 33 percent for the noninstitutionalized population (Christensen and Wagner 2000).

8. Also called myocardial ischemia, IHD is characterized as reduced blood flow to the heart. In the private data we identify patients with IHD as those with at least one inpatient or outpatient claim with a primary diagnosis ICD-9 code of 410.xx, 411.xx, 412.xx, 413.xx, or 414.xx. In the public data we identify patients based on self-reports of ever diagnosed with heart disease. See the online data appendix (available on the *Brookings Papers* webpage at www.brookings.edu/economics/bpea.aspx, under "Conferences and Papers") for more information.

calendar year.[9] The claims-based measures of the number of diseases are available in both the MCBS and the private health insurance data.[10] This is important because unmeasured differences in severity across regions could lead to spurious positive correlation between sectors.

The primary geographic unit of analysis for our study is the MSA. An alternative candidate would be the hospital referral region (HRR), used by the Dartmouth Atlas. However, HRRs are not reported in either of our datasets, and the private sector data do not contain 5-digit zip codes, which are required to construct an individual's HRR. We restrict our sample to the 99 MSAs where we have the largest samples. MSAs are somewhat larger than HRRs, and this may compress the variation for both sectors in our data.

Our final sample contains 240,028 private patients and 24,800 public patients.[11] Since there are many fewer public patients, it is important to correct for the effects of sample size on our estimates. We derive and report these corrections in detail below.

Table 1 reports some summary statistics comparing demographic characteristics in the public and the private samples. As one would expect, the average age in the private sample is lower than in the sample of Medicare patients, most of whom are older than 65. The private sample contains a greater fraction of males, in part because it is influenced by current or past employment status. (The private sample contains both active workers and retirees receiving benefits from their current or past employers.) Average income is also higher in the private sample. The greater variance in income for the public sample is likely due to the fact that income is reported individually in the MCBS, but imputed at the local level in the private sample.

9. The specific conditions considered are essential hypertension, congestive heart failure, diabetes, asthma, hypercholesterolemia, ulcer, depression, chronic obstructive pulmonary disease, allergic rhinitis, migraine, arthritis, chronic sinusitis, anxiety disorder, cardiac disease, vascular disease, epilepsy, gastric acid disorder, glaucoma, gout, hyperlipidemia, irritable bowel syndrome, malignancy, psychotic illness, thyroid disorder, rheumatoid arthritis, tuberculosis, angina, human immunodeficiency virus infection, anemia, and stroke. Most co-morbidities are relatively uncommon, except for the ones involving heart disease (or risk factors for heart disease).

10. The MCBS also contains self-reports of a number of distinct health conditions, as well as the individual's self-reported general health status (coded 1 to 5, with 1 indicating poor and 5 indicating excellent). Our regression analysis relies on the claims-based, rather than self-reported, disease measures for both the public and the private samples. More details appear in our online data appendix.

11. See the footnotes to tables 1 and 2 for a few sample size issues specific to certain variables.

Table 1. Selected Patient Demographic and Health Characteristics[a]

	Private		*Public*	
Patient characteristic	*Mean*	*Standard deviation*	*Mean*	*Standard deviation*
Age	55.4	7.1	78.2	7.9
Percent male	65	48	46	50
Income (thousands of 2004 dollars)	42.8	10.8	28.5	46.8
Percent with heart disease in year	32	47	37	48
No. of adverse health conditions	1.4	1.4	2.9	2.4

Sources: Data on private patients come from a modified version of the Ingenix Touchstone product. Data on public patients come from the MCBS.

a. History of heart disease is self-reported in the public sample and identified using medical claims in the private sample. The private sample has 240,028 observations and the public sample 24,800 observations.

The table also compares the health of individuals in the two samples. Since both samples are limited to individuals with a history of heart disease, we include a variable indicating the fraction of individuals who are diagnosed with heart disease in a particular year. In all cases, the presence of disease is taken from claims rather than from self-reported data. The incidence of heart disease is similar in the two samples: 0.32 in the private sample and 0.37 in the public sample.

In addition, the table reports the average number of adverse health conditions (out of the total of 30, including heart disease) per patient. As with heart disease, the health conditions are determined using the ICD-9 diagnosis codes from medical claims in both the public and the private samples. Unsurprisingly, the elderly individuals in the public sample are much sicker on average, with 2.9 adverse health conditions in the year compared with 1.4 in the private sample.

As a matter of course, the public and the private samples are drawn from different populations. We include a number of controls and analyses designed to mitigate and test for the impact of these differences, but heterogeneity across samples remains a possibility. Later we discuss the sources of heterogeneity, the methods we have employed to address them, and their possible implications for the analysis.

II.B. Descriptive Statistics

Table 2 presents some descriptive statistics for health care spending and utilization in the public and the private samples aggregated over all regions and patient characteristics. We present not only the mean and the standard deviation but also the 25th-percentile, median, and 75th-percentile values. Our utilization measures (all measured as yearly averages per patient)

Table 2. Distributions of Spending and Utilization Measures[a]

Measure	*Sample*	*Mean*	*Standard deviation*	*25th percentile*	*Median*	*75th percentile*
Utilization (number per patient per year)						
Hospitalizations	Private	0.36	1.15	0	0	0
	Public	0.57	1.14	0	0	1
Hospital days	Private	1.23	7.13	0	0	0
	Public	2.93	8.59	0	0	1
Outpatient visits	Private	5.56	5.86	1	4	8
	Public[b]	8.59	11.05	1	5	12
Drug prescriptions[c]	Private	45.78	42.20	13	36	66
	Public	35.45	29.93	14	29	50
Spending (thousands of 2004 dollars)						
Total spending	Private	8.40	22.98	0.56	2.10	6.88
	Public	10.25	18.8	1.25	3.91	11.4
Inpatient spending	Private	4.02	18.36	0	0	0
	Public	4.94	13.21	0	0	4.65
Outpatient spending	Private	4.38	9.83	0.54	1.85	4.86
	Public	5.30	9.14	1.13	2.94	6.44
Prescription drug spending[c]	Private	2.80	5.78	0.53	1.67	3.42
	Public	1.92	2.05	0.58	1.39	2.63

Source: Authors' calculations.

a. Figures are yearly averages during 2000–06 (2000–05 for drug prescriptions) for patients with a history of heart disease, which is self-reported in the public sample and identified using medical claims in the private sample. Except where noted otherwise, the private sample has 240,028 observations and the public sample 24,800 observations.

b. Survey responses (used to cross-validate the claims data) were incomplete in 3,769 cases, so that the public sample has 21,031 observations.

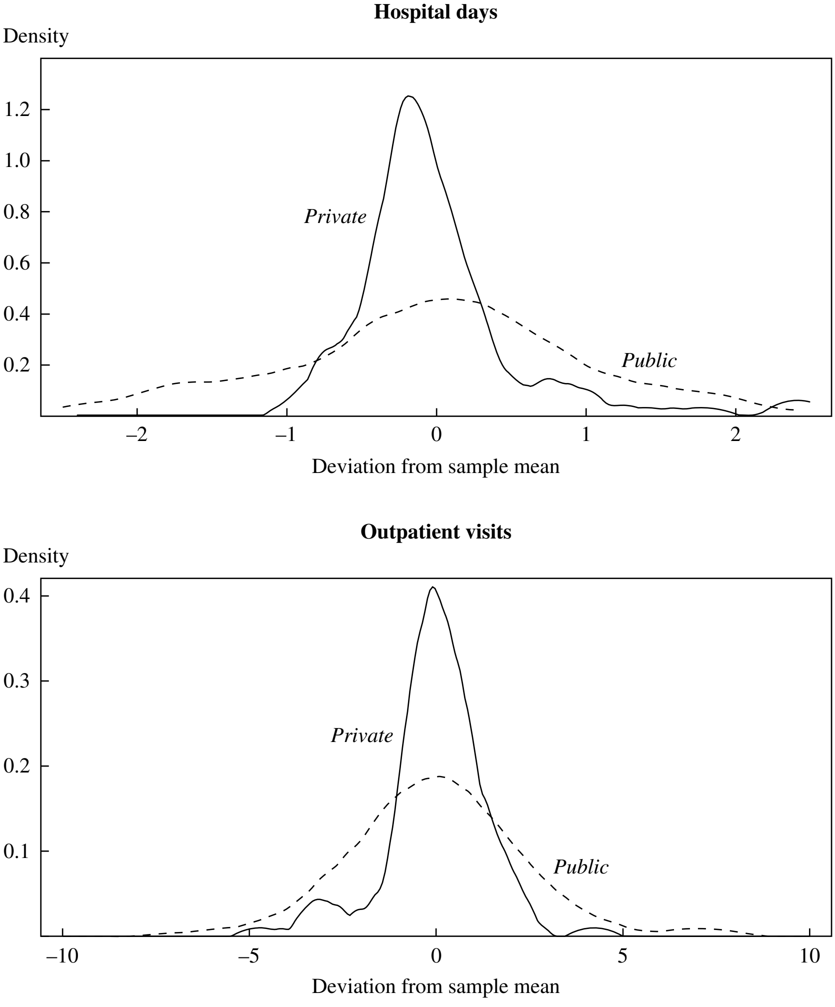
c. Because observations from 2006 are omitted, the private sample has 231,802 observations and the public sample 21,140 observations. Number of prescriptions is in terms of 30-day-equivalents.

include the number of hospitalizations, total hospital days across all hospitalizations, the number of outpatient visits, and the number of 30-day-equivalent prescriptions in both samples. For spending, we record total (inpatient plus outpatient), inpatient, and outpatient spending, as well as spending on prescription drugs.

Utilization, in terms of hospitalizations, hospital days, and outpatient visits, is lower for the private patients. Spending for this group also tends to be lower. Total medical spending for individuals in the private plans is $8,401 per year, compared with $10,245 for the Medicare patients—about a 20 percent difference. The exception to the pattern is prescription drugs, for which both utilization and spending are greater among private patients.

Figures 3 and 4 provide a broad sense of the variation present in our samples. Figure 3 reports for both samples the estimated kernel densities of MSA-level deviations from the mean for both hospital days and

Figure 3. Kernel Density Estimates of Regional Fixed Effects for Selected Utilization Measures[a]

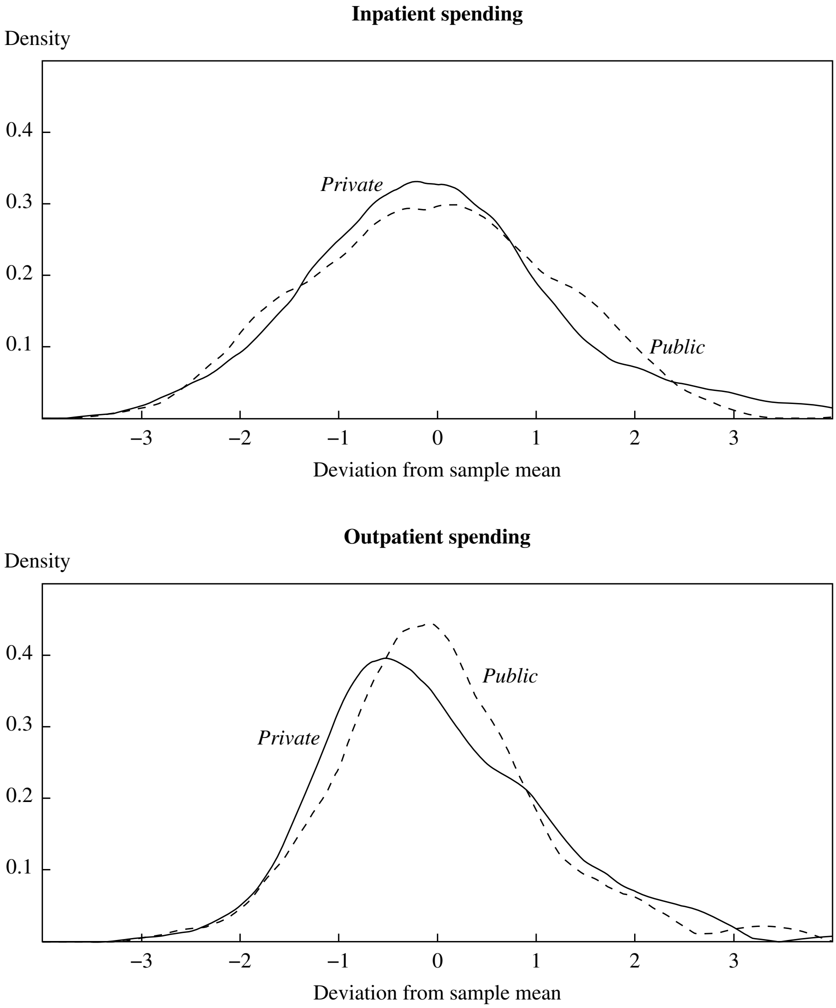


Source: Authors' calculations.

a. Estimated kernel densities of the deviation of mean hospital days and outpatient visits per patient per year across MSAs.

outpatient visits. Each data point underlying the kernel estimate is the difference between an MSA-level mean and the overall sample mean. For both variables, the distributions appear to be tighter for the private than for the public sample. However, these distributions are based on raw, unadjusted numbers that do not account for disease or other covariates.

Figure 4. Kernel Density Estimates of Regional Fixed Effects for Selected Spending Measures[a]

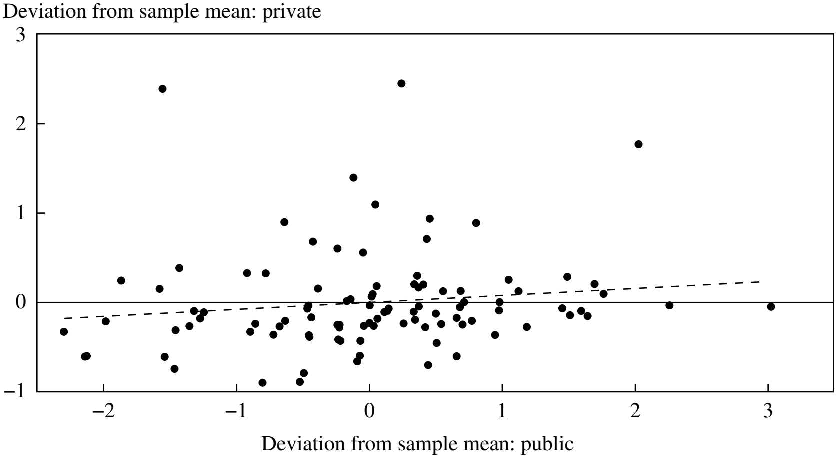


Source: Authors' calculations.

a. Estimated kernel densities of the deviation of mean inpatient and outpatient spending per patient per year across MSAs.

Figure 4 repeats this exercise for inpatient and outpatient spending. Here the findings are decidedly more mixed. For outpatient spending the distribution appears to be slightly tighter for the public sample. The figure for inpatient spending is harder to interpret visually, as the differences in the densities are small and asymmetric. In any event, the differences

Figure 5. Hospital Days: Relationship between Public and Private Deviations from the Mean[a]



Source: Authors' calculations.

a. Each observation pairs the deviation for a single MSA from mean hospital days per patient per year in the public sample with that for the same MSA for the private sample. The line represents the fitted ordinary least squares regression line.

observable visually between the spending and the utilization distributions suggest the possible importance of public sector price restraints, which would lower spending variation even with greater variation in utilization.

Finally, figure 5 plots the relationship between deviations from the MSA-level means for public and private hospital days. This is the empirical analogue to the theoretical relationship in figure 2. The figure suggests that mean private hospital days increase slightly with mean public hospital days, but much less than one for one. This is consistent with there being less regional variation in the private sector; we test this hypothesis more formally in the following analyses.

II.C. Framework for Estimating Regional Variation

We are particularly interested in the between-MSA variance in spending and utilization for the public and the private samples. We begin with the simplest possible approach that evaluates the variance between MSAs in the sample means. We then move to estimating the variance in regression-adjusted means, which we estimate from regressions that control for various factors that might also influence spending and utilization. In both cases we account for the relative bias that is created by the substantial differences in sample size across sectors: because the public samples are much smaller than the private samples, there is greater sampling variance in the public sector estimates and thus greater variation in the MSA-level means for Medicare patients. To estimate the true between-sector differences in regional variation, we estimate and remove the variability that is due to sample size differences alone.

Formally, the observed regional variation within a sector is due to the true variation and the sampling variance in estimating that variation. Denote by μ_r the true mean for region r and by $\hat{\mu}_r$ the corresponding sample estimate, whether unconditional or regression-adjusted. The sample mean is equal to the true mean plus sampling error, according to

$$\hat{\mu}_r = \mu_r + z_r.$$

The sampling error z_r has zero mean, and the covariance of the sampling error across regions is $E(z, z_s) = \sigma_{rs}$. Define $\bar{\mu} \equiv \frac{1}{R}\sum_{r=1}^{R}\mu_r$, the "grand mean" across regions. Similarly, define the corresponding sample analogue, $\bar{\hat{\mu}} \equiv \frac{1}{R}\sum_{r=1}^{R}\hat{\mu}_r$. Finally, define the average sampling error across regions, $\bar{z} \equiv \frac{1}{R}\sum_{r=1}^{R}z_r$. The object of interest is the degree of regional variation in the true MSA means, $RV \equiv \frac{1}{R}\sum_{r=1}^{R}\left(\mu_r - \bar{\mu}\right)^2$, which has the sample analogue $\widehat{RV} \equiv \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\mu}_r - \bar{\hat{\mu}}\right)^2$.

The observed variation $\widehat{RV}$ is a biased estimator of RV, as a result of sampling error in the estimates. Moreover, this bias is likely to be larger for our public sector estimates because of the smaller public sector sample size, which yields noisier estimates of public sector utilization. However, we can recover a consistent estimate of the bias and correct for it, according to

$$E\left(\widehat{RV}\right) = RV + \frac{1}{R}\sum_{r=1}^{R}E\left(z_r - \bar{z}\right)^2.$$

In the appendix we show how to estimate this bias from sample variances and covariances. Our formula works for both the case of unconditional sample means and the case of regression-adjusted means. In the simple case without covariance across regions or zero average error across regions, this expression simply states that the observed variation is the true variation plus the average squared standard error. More generally, the more precisely the sample means are estimated, the smaller is the bias correction.

In sum, the object of interest in our analysis is RV, which we estimate as $\widehat{RV} - \widehat{Bias}$ for both the public and the private sector. Using these estimates of regional variation, we report both the ratio of public to private variation and the difference between public and private variation. We construct standard errors around these by means of a bootstrap procedure, which samples individuals with replacement within each MSA, so that each bootstrap sample contains exactly the same number of individuals in each MSA as the original sample.[12] The bootstrap procedure reflects the nature of our sample design. We regard the set of MSAs as fixed but each sample within an MSA as a random sample of that MSA's population. Statistically, our set of MSAs approximates a population, but we have samples within each MSA.

Our regression-adjusted estimates employ a model with regional fixed effects that controls for disease severity and demographics.[13] For each sector s we estimate

$$Y_{irts} = \alpha_{0s} + \mathbf{X}_{its}\beta_s + \delta_{ts} + \delta_{rs} + \epsilon_{irts}.$$

Here Y_{irts} represents some measure of utilization or spending by patient i in region r at time t and in sector s. The vector $\mathbf{X}$ includes the following demographic characteristics for each patient: age, age squared, sex, income, income squared, age and age squared interacted with sex, as well as dummy variables for each of the adverse health conditions listed above. The terms δ_{ts} and δ_{rs} are sector-specific fixed effects for year and MSA, respectively. The sector-specific variance in the fixed effect δ_{rs} is the regression-adjusted analogue to the variance in the MSA-level sample means.

As a general matter, the covariates have relatively little predictive power within MSAs but a fair amount between MSAs. Across all specifications, for instance, the MSA means of the covariates explain about 50 to 70 percent of the between-MSA variation in utilization and spending, in the sense of R^2.

12. The alternative block-bootstrap that samples MSAs with replacement generates nearly identical inferences for statistical significance in our analysis, and so does a "flat" bootstrap.

13. A possible alternative is a random-effects model, but the Hausman test rejected this more efficient model in favor of the fixed-effects model in the majority of cases we analyzed.

Table 3. Regional Variation in Mean Utilization[a]

	Observed variation[b]		*Corrected variation*[c]		*Difference, public minus private*	*Ratio of public to private*
Utilization measure	*Private*	*Public*	*Private*	*Public*		
Unconditional means						
Hospitalizations	0.013	0.015	0.012	0.009	–0.003	0.728
					(0.003)	(0.204)
Hospital days	0.322	1.016	0.230	0.659	0.429**	2.870*
					(0.199)	(1.017)
Outpatient visits	1.736	5.154	1.676	4.585	2.909***	2.735***
					(0.502)	(0.323)
Drug prescriptions[d]	72.746	32.758	70.090	28.403	–41.687***	0.405***
					(3.896)	(0.043)
Regression-adjusted means[e]						
Hospitalizations	0.006	0.011	0.005	0.006	0.001	1.266
					(0.002)	(0.430)
Hospital days	0.169	0.610	0.080	0.313	0.233*	3.907*
					(0.124)	(1.684)
Outpatient visits	0.988	3.255	0.942	2.677	1.735***	2.841***
					(0.322)	(0.379)
Drug prescriptions[d]	29.190	25.856	27.086	21.131	–5.955*	0.780**
					(3.130)	(0.106)

Source: Authors' calculations.

a. Numbers in parentheses are standard errors on the difference between public and private variation or the ratio of public to private variation and are bootstrapped within MSAs, separately for public and private patients, with 200 bootstrap draws. For both sectors, then, the number of patients in each region in each bootstrapped sample is the same as the number of patients in the original sample. Asterisks indicate differences statistically significantly different from zero or ratios statistically significantly different from 1 at the ***1 percent, **5 percent, and *10 percent level.

b. Variance in the regional means or fixed effects of the utilization variables.

c. Unbiased measure of the true variance in the regional means or fixed effects corrected for sampling error, as described in the text and the appendix. All differences and ratios are based on these numbers.

d. 30-day-equivalents.

e. Estimates of regional fixed effects on each utilization variable from a regression that includes as other independent variables year fixed effects, quadratic specifications of patient age and income, patient sex, sex interacted with age, and dummy variables for 30 separate types of disease.

II.D. Regional Variance in Utilization and Spending

Table 3 reports the estimated regional variance in four utilization measures: number of hospitalizations, number of hospital days, number of outpatient visits, and number of prescription drugs (in terms of 30-day-equivalents). Again, prescription drug coverage is provided by the private sector in both populations throughout the sample period, and therefore we do not expect to see similar differences for prescription drugs as for the other measures.

The table shows overall between-MSA variation in the public and the private sectors. The observed variation (first two columns) is computed as the average MSA-level deviation from the overall mean. The top panel reports the variation based on unconditional means; in the bottom panel, both the overall mean and each MSA-level mean are regression-adjusted, as described above. The corrected variation (second two columns) is computed by subtracting the expected bias due to sampling error, as described above. The next column shows the absolute difference between the public and private variances, and the last column the ratio of the variances. Asterisks indicate statistically significant differences from zero for the differences, and from unity for the ratios.

Variation in hospital days is about three times, and variation in outpatient visits about two times, higher in the public sector. These differences are statistically significant at the 10 percent level or higher and appear regardless of whether we adjust for covariates (although the magnitudes differ somewhat). On the other hand, prescription drug utilization exhibits statistically less variation in the Medicare population; this is important because, again, even Medicare patients obtain their prescription drug insurance privately in our sample. Finally, there is no statistically significant difference in the variation for hospitalizations. It is likely that more statistical power is needed to pin down this variance, in one direction or the other. Overall, these results provide evidence suggesting higher variance in the public sector, but for a few of the outcomes our statistical tests lack the power to generate definitive results.

Table 4 reports the estimated regional variance in four spending measures: total spending, inpatient spending, outpatient spending, and prescription drug spending. The regression-adjusted estimates indicate that outpatient spending exhibits only about 35 percent as much variation in the public sector as in the private sector. Inpatient spending exhibits roughly equal variation in the two sectors. Finally, prescription drug spending varies less for Medicare patients. With that exception, these results are quite different from the utilization results and suggest that price restraints play a role in the public sector. In spite of greater variation in utilization, the public sector exhibits less variation in spending.

III. Comparisons with Existing Literature

Regional variation in spending and utilization in the public sector has been well documented in a literature that is almost 40 years old and well accepted by the academic community. In that sense, our contribution is to

Table 4. Regional Variation in Mean Spending[a]

Spending measure	*Observed variation*[b] *Private*	*Observed variation*[b] *Public*	*Corrected variation*[c] *Private*	*Corrected variation*[c] *Public*	*Difference, public minus private*	*Ratio of public to private*
Unconditional means						
Total medical spending	4.443	3.907	3.571	2.352	−1.219 (0.981)	0.659* (0.207)
Inpatient spending	1.634	1.587	1.109	0.842	−0.266 (0.416)	0.760 (0.247)
Outpatient spending	1.263	1.048	1.082	0.634	−0.447 (0.355)	0.586 (0.266)
Prescription drug spending	0.418	0.144	0.377	0.124	−0.253*** (0.049)	0.329*** (0.060)
Regression-adjusted means[d]						
Total medical spending	2.890	2.782	2.111	1.463	−0.647 (0.728)	0.693 (0.258)
Inpatient spending	1.186	1.357	0.698	0.695	−0.004 (0.324)	0.995 (0.298)
Outpatient spending	0.924	0.628	0.758	0.265	−0.494* (0.261)	0.349** (0.272)
Prescription drug spending	0.251	0.086	0.214	0.064	−0.150*** (0.043)	0.300*** (0.082)

Source: Authors' calculations.

a. Spending is measured in 2004 dollars. Numbers in parentheses are standard errors on the difference between public and private variation or the ratio of public to private variation and are bootstrapped within MSAs, separately for public and private patients, with 200 bootstrap draws. For both sectors, then, the number of patients in each region in each bootstrapped sample is the same as the number of patients in the original sample. Asterisks indicate differences statistically significantly different from zero or ratios statistically significantly different from 1 at the ***1 percent, **5 percent, and *10 percent level.

b. Variance in the regional means or fixed effects of the spending variables.

c. Unbiased measure of the true variance in the regional means or fixed effects corrected for sampling error, as detailed in the text and the appendix. All differences and ratios are based on these numbers.

d. Estimates of regional fixed effects on each spending variable from a regression that includes as other independent variables year fixed effects, quadratic specifications of patient age and income, patient sex, sex interacted with age, and dummy variables for 30 separate types of disease.

compare this with variation in the private sector, rather than to establish the existence of public sector variation.

Table 5 summarizes a few representative papers from this vast literature.[14] John Wennberg and Alan Gittelsohn (1973) provide an early example. Their study analyzed variation across "hospital service areas," a precursor to the HRRs typically analyzed in the modern Dartmouth Atlas

14. For useful summaries from both the economic and the clinical literatures on geographic variation in health care, see Wennberg and Cooper (1998), Phelps (2000), Fisher and others (2003a, 2003b), Chandra and Staiger (2007) and Sutherland, Fisher, and Skinner (2009).

Table 5. Key Findings on Variation in Regional Health Care Spending Using Medicare Data[a]

Study	*Geographic aggregation*	*Summary*	*Key findings*
Wennberg and Gittelsohn (1973)	Hospital service area	Studied geographic variation in utilization and spending in Vermont.	Found wide variations apparently due to differences in practice style rather than in population health. Hospital days in highest-use area were 1.5 times that in lowest.
Cutler and Sheiner (1999)	HRR	Calculated share of regional variation in Medicare spending attributable to regional differences in health and demographics.	Regional differences in demographics can explain about 70 percent of regional differences in Medicare spending, but significant differences remain unexplained.
Fisher and others (2003a)	HRR	Compared patients across regions holding other characteristics constant.	Patients in higher-spending regions received approximately 60 percent more care. The increased utilization mostly arose from more frequent physician visits.
Sirovich and others (2006)	MSA	Compared variation in intensity of treatment with physician perceptions of quality of care.	Medicare spending per capita in highest intensity quintile was 1.58 times that in lowest.
Chandra and Staiger (2007)	HRR	Specified a model of patient treatment choice with productivity spillovers and tested the model using treatment choices and health outcomes of heart attack patients.	Patterns of which patients benefit and which lose from intensive medical care are consistent with productivity spillover model.
Fowler and others (2008)	HRR	Used a patient survey to compare local variation in spending and utilization with patient perceptions of quality.	Regional differences in spending were not systematically related to differences in patient perceptions of care quality.

Wennberg and others (2008)	HRR	Summarized Dartmouth Atlas findings on geographic variation in Medicare spending and their implications.	Three states spent more than 20 percent above the national average of $46,412. Conversely, three states spent 25 percent below the national average or less. Interquartile ratio (75th percentile over 25th) is 1.26 for HRRs.
Rettenmaier and Saving (2009)	State	Studied how state rankings in medical spending per capita change when different definitions of spending are used, such as Medicare only or total spending by all payers.	Found a state-level correlation between Medicare spending and total spending of 0.21, and that variation in Medicare spending exceeds variation in private spending.
Sutherland, Fisher, and Skinner (2009)	HRR	Updated Dartmouth Atlas findings on geographic variation in Medicare spending and their implications.	Inpatient days per beneficiary in highest cost quintile were 1.50 times that in lowest; physician visits in highest cost quintile were 1.36 times that in lowest.
Chernew and others (2010)	HRR	Compared HRR-level variation in medical spending between Medicare and large firms.	Found substantial regional variation in spending, greater for large firms than for Medicare (coefficient of variation 0.21 v. 0.16). Correlation between private and public inpatient utilization was 0.59.
Gottlieb and others (2010)	HRR	Examined role of Medicare prices in driving geographic variations in health care.	Prices explain a small portion of variation in spending. The 80th percentile of price-adjusted Medicare Part B spending was 1.37 times the 20th percentile.

Sources: Literature cited.

a. HRR = hospital referral region; MSA = metropolitan statistical area.

of Health Care studies. The table also lists a couple of important studies that use states or MSAs. It is important to recognize this difference when comparing our MSA-level analysis with HRR-level analyses elsewhere, and it is important for future work to assess the potential implications of this difference.

The 2008 Dartmouth Atlas of Health Care reports that average spending on health care in the last 2 years of life (for deaths occurring from 2001 to 2005) ranged from a high of $59,379 in New Jersey to $32,523 in North Dakota (Wennberg and others 2008). This range, from 28 percent above to 30 percent below the national average, is similar to the range of quantity utilization reported across MSAs by MedPac: from 39 percent above the national average in Miami to 25 percent below in rural Hawaii (Medicare Payment Advisory Commission 2009).

These variations are not fully explained by factors such as age, insurance coverage, average income, and rates of illness or disease. David Cutler and Louise Sheiner (1999) investigate the extent to which variation in spending across HRRs can be explained by regional differences in illness, in the demand for health (for example, as measured by income and race), or in "exogenous differences in the structure of medical care markets" (for example, in the ratio of generalists to specialists). They find that regional demographics can explain about 70 percent of the variation in medical spending across regions, but the unexplained variation remains large. For example, when differences in demographics and the illness of the population are accounted for, bringing Medicare spending down to the level of the 10th-percentile region would reduce total spending by 15 percent.

Perhaps the existing study most closely related to ours is that of Michael Chernew and others (2010), who compare HRR-level variations in Medicare against those in a sample of large firms in the Thomson Reuters (Medstat) MarketScan Commercial Claims and Encounters Database. They estimate that the geographic variation in private sector spending is greater than that in Medicare spending (coefficient of variation of 0.21 versus 0.16). This is consistent with our findings for spending. They focus less on variation in utilization, although they do report a positive correlation between Medicare and non-Medicare inpatient days.

IV. Limitations of Our Analysis

There are several empirical questions that our data cannot address but that should be addressed in future work. The populations of privately insured and publicly insured patients differ, because the latter have often opted out of private health insurance options. The empirical implications of this are not clear a priori. Fee-for-service Medicare patients are likely to be sicker than their counterparts in private Medicare health maintenance organizations (HMOs), because HMOs attempt to select healthier patients (Morgan and others 1997). On the other hand, the privately insured nonelderly may also be healthier than the nonelderly overall, if private health insurers select against the sickest patients for similar reasons. The link between health insurance and employment in the nonelderly population adds further complexity, as those who are eligible for employment-based health insurance may be richer or healthier, or both, than their peers. Finally, the fact that our private sector data are based only on employees of large (Fortune 500) firms adds a further dimension of selection.

We ran several supplementary analyses to investigate some of these issues, but our data lack the power to reach definitive conclusions across the board. First, we narrowed the age range of our comparisons, to mitigate some of the differences in health status. We compared 60- to 64-year-olds in the commercially insured population with 66- to 70-year-olds in the fee-for-service Medicare population. As this restriction further reduces the sample, we limit our analysis to the 70 MSAs for which we have at least 25 observations in both samples.

Table 6 reports the result for the samples with the narrow age ranges. Generally, the point estimates based on these restricted age ranges are similar to those based on the full sample, but the precision of the estimates declines enough to eliminate statistical significance. The point estimates indicate that variation in the public sector is about 5.1, 3.4, and 1.2 times that in the private sector for hospital days, outpatient visits, and hospitalizations, respectively. As in the analysis based on the full sample, variation in prescription drug use is smaller in the public sector, about 53 percent as large as variation in the private sector.

Next we investigated the issue of selection based on employment by comparing our privately insured sample with Medicare patients who also have coverage from an employer. If an individual has such coverage, we know that he or she was employed and privately insured at one point. Roughly 35 percent of Medicare enrollees in our sample also have employer-provided coverage. They are slightly younger (averaging 77 years, compared with 79 years for those without such coverage), richer (average income is 58 percent higher), and more likely to be male (52 percent versus 40 percent) than the average Medicare enrollee. Having employer coverage is associated with very small differences in the fraction of total expenses paid for by Medicare: Medicare pays 39 percent

Table 6. Regional Variation in Regression-Adjusted Mean Utilization, Patients Aged 60 to 70[a]

Utilization measure	Observed variation[b]		Corrected variation[c]		Difference, public minus private	Ratio of public to private
	Private	Public	Private	Public		
Hospitalizations	0.015	0.035	0.011	0.014	0.003	1.244
					(0.010)	(0.921)
Hospital days	0.258	2.038	0.130	0.656	0.526	5.060
					(0.886)	(4.134)
Outpatient visits	1.460	7.207	1.335	4.524	3.190	3.390
					(2.316)	(1.695)
Prescriptions	52.869	47.198	46.741	24.765	−21.975**	0.530
					(11.090)	(0.209)

Source: Authors' calculations.

a. The private sample is restricted to patients aged 60 to 64 and the public sample to patients aged 66 to 70. Both samples are restricted to include only the 70 MSAs with at least 25 observations in both samples. The private sample has 67,414 observations and the public sample 3,568 observations. Numbers in parentheses are standard errors on the difference between public and private variation or the ratio of public to private variation and are bootstrapped within MSAs, and separately for public and private patients, with 200 bootstrap draws. For both sectors, then, the number of patients in each region in each bootstrapped sample is the same as the number of patients in the original sample. Asterisks indicate differences statistically significantly different from zero or ratios statistically significantly different from 1 at the ***1 percent, **5 percent, and *10 percent level.

b. Variance in the regional means or fixed effects of the utilization variables.

c. Unbiased measure of the true variance in the fixed effects corrected for sampling error, as detailed in the text and the appendix. All differences and ratios are based on these numbers.

of the expenses of those without employer coverage and 38 percent of those with such coverage. The lack of a disparity is due to the fact that once an elderly Medicare beneficiary retires, the employer-provided coverage becomes secondary to Medicare. In our data just 9 percent of individuals in the Medicare sample with employer coverage are working, so for the vast majority Medicare is the primary payer. It thus seems reasonable to assume that Medicare is the primary driver of resource allocation for these individuals. A number of MSAs are left with very small samples after this restriction, so we limit our analysis to the 77 MSAs where we have at least 50 observations in both samples.

These results are presented in table 7. Again, the point estimates are similar to those based on the full sample, but the precision of the estimates declines enough to eliminate much of the statistical significance. The point estimates indicate that variation in the public sector is about 4.1, 3.8, and 1.6 times that in the private sector for hospital days, outpatient visits, and hospitalizations, respectively. The greater variation in outpatient visits in the public sample is statistically significant at the 1 percent level. The other

Table 7. Regional Variation in Regression-Adjusted Mean Utilization, Patients with Some Private, Employer-Provided Coverage[a]

Utilization measure	*Observed variation*[b]		*Corrected variation*[c]		*Difference, public minus private*	*Ratio of public to private*
	Private	*Public*	*Private*	*Public*		
Hospitalizations	0.006	0.017	0.005	0.008	0.003	1.572
					(0.004)	(0.984)
Hospital days	0.144	0.709	0.049	0.198	0.149	4.058
					(0.240)	(3.861)
Outpatient visits	3.774	4.462	0.911	3.438	2.527***	3.774***
					(0.518)	(0.614)
Prescriptions	28.982	34.778	26.885	25.107	−1.778	0.934
					(6.032)	(0.213)

Source: Authors' calculations.

a. The public sample is restricted to patients who report at least some form of private, employer-provided insurance coverage. Both samples are restricted to include only the 77 MSAs that have at least 50 observations in both samples. The private sample has 202,202 observations and the public sample 8,416 observations. Numbers in parentheses are standard errors on the difference between public and private variation or the ratio of public to private variation and are bootstrapped within MSAs, and separately for public and private patients. For both sectors, then, the number of patients in each region in each bootstrapped sample is the same as the number of patients in the original sample. Asterisks indicate differences statistically significantly different from zero or ratios statistically significantly different from 1 at the ***1 percent, **5 percent, and *10 percent level.

b. Variance in the regional means or fixed effects of the utilization variables.

c. Unbiased measure of the true variance in the fixed effects corrected for sampling error, as detailed in the text and the appendix. All differences and ratios are based on these numbers.

differences are not significant at the 10 percent level. Variation in prescription drug use is slightly lower in the public sector, about 93 percent as large as in the private sector.

V. Concluding Remarks

It has long been recognized that public and private enterprises face different incentives to control costs. This paper has analyzed these differences in the health insurance context, along with their implications for variation in care. Public payers are likely able to restrain prices better than private payers but have weaker incentives to control costs through utilization controls. As a result, one might expect greater variation in utilization for the public sector, but the effects on total spending are ambiguous. Using samples of heart disease patients, we presented empirical evidence consistent with these implications.

Further research should focus more closely on the issue of whether and to what extent variations across sectors are the result of differences in the baseline health of the publicly and privately insured populations. Work is also needed to assess whether our basic findings can be generalized across other disease categories and geographical classifications. In addition, the analysis of health outcomes must be integrated into the analyses of utilization and of spending. As a related point, although we have focused on estimated variations, further research should be conducted into the sources of the mean differences in utilization and spending. Finally, and perhaps most important, research is needed to draw out the normative implications of variations in care both within and between sectors.

The normative implications of variation in care are not straightforward, in spite of the conventional wisdom that greater variation implies inefficiency. On the one hand, the literature has consistently found that areas exhibiting higher utilization of health care services do not exhibit demonstrably better outcomes for patients (Fisher and others 2003a). This has led many to conclude that these areas are practicing "flat-of-the-curve" medicine, where the marginal benefit approaches zero. However, Amitabh Chandra and Douglas Staiger (2007) demonstrate that productivity spillovers and specialization can explain regional variation in the utilization of intensive procedures, without resorting to inefficiency. Most notably, their model can reconcile the seemingly contradictory evidence that intensive treatments such as most surgery are often highly effective at the individual level, but that regions using these treatments more intensively do not have better average health outcomes. Chandra and Staiger observe that regions specializing in intensive treatment will find it optimal to provide that treatment to more patients; therefore, the marginal patient in such regions will be less suited to it than the marginal patient elsewhere. This mitigates the greater benefits of intensive treatment.

For this and other reasons, the efficiency implications of variation in care require a more careful analysis. Even in our simple framework, the normative impact of variation is unclear. For instance, if the private sector is pricing and producing efficiently, then the theory suggests that the public sector is engaging in inefficiently high utilization and inefficiently low pricing. On the other hand, if private sector prices are too high or if utilization is too low, then the effects of public insurance may actually represent second-best improvements to welfare. Evidently, it is important to investigate the baseline efficiency properties of the private health insurance market and to characterize how these are affected by the presence of publicly financed health insurance.

Regardless of the conclusions, normative analysis of this issue will likely generate a number of important policy implications. Many have noted that

Medicare has lower administrative costs than the private sector. This is often interpreted as part of the value generated by centralized insurance. This is a typical finding when one is comparing a centralized with a decentralized model, but it could also be explained by the cost of administering utilization controls in the private sector. If so, any efficiency benefits of utilization controls would need to be weighed against these administrative costs. The benefits generated by administrative costs are often neglected in the policy debate, as are related issues such as the deadweight costs of the tax revenue required to fund public enterprise, the efficiency gains of marketing activities by private firms, and higher rates of fraud in the public Medicare system. The last of these is directly related to lax utilization controls. A fuller analysis of the costs and benefits of public versus private health insurance is needed.

The relative merits of public enterprise have a number of policy implications. The first concerns the appropriate size of Medicare Advantage, which operates through publicly provided premium subsidies to private HMOs. Medicare Advantage plans are not directly comparable to private payers, because they compete on quality rather than price, as long as there is no price competition through competitive bidding for plan members. Thus differences in incentives for utilization control operate through the need to enhance quality, subject to available premium resources, or result from residual claims on profits. Future research needs to investigate more carefully the differences and similarities in cost-control measures from this type of coverage and their effects on regional variations and efficiency.

The second implication regards the timely issue of comparative effectiveness research (CER), which has been offered as a means of raising health care quality and reducing costs. The rationale for CER is to generate better evidence, and to disseminate it to patients, payers, and providers, about what works and does not work in health care. Indeed, a common motivation for the use of CER is to reduce cost inefficiencies due to regional differences in care. Awareness of CER has been heightened recently by its significant public subsidization through the American Recovery and Reinvestment Act of 2009.[15] An overriding question raised by our analysis is whether regional variation in care occurs because of a lack of information or a lack of incentives for utilization control in the public sector.

15. The explicit use of comparative effectiveness assessments is much more common outside the United States, particularly in the European Union. However, this is a relatively recent trend: no European countries formally required economic assessments for pricing and reimbursement decisions as of 1993, but a majority had such a policy either in place or in development by 1999 (Drummond and others 1993, 1999).

Health economists have not yet paid sufficient attention to the differences in incentives across the public and the private sectors or to the corresponding implications for health care variation. The regional variations documented in the Dartmouth Atlas of Health Care have led several prominent researchers to conclude that high-use regions ought to model themselves after their low-use peers (Fisher, Bynum, and Skinner 2009). Our study suggests the importance of research focusing on another, different question: whether or not public sector health insurers ought to model themselves after their peers in the private sector.

APPENDIX

Sampling Error in Estimation of Regional Variation

In the text we outlined our approach for obtaining a consistent estimate of regional variation, defined as

$$RV \equiv \frac{1}{R}\sum_{r=1}^{R}\left(\mu_r - \bar{\mu}\right)^2.$$

In this appendix we show how we solve for the bias in the sample analogue, $\widehat{RV} \equiv \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\mu}_r - \bar{\hat{\mu}}\right)^2$, and estimate it consistently using the variance-covariance matrix of the estimates. Recall the definitions from the text: μ_r is the true population fixed-effect parameter for region r, $\hat{\mu}_r$ is the corresponding sample estimate, and z_r is a mean-zero sampling error with covariance across regions $E(z_r z_s) = \sigma_{rs}$. The sample estimate is the true value plus sampling error,

$$\hat{\mu}_r = \mu_r + z_r.$$

Define $\bar{\mu} \equiv \frac{1}{R}\sum_{r=1}^{R}\mu_r$, the mean regional fixed effect across regions; $\bar{\hat{\mu}} \equiv \frac{1}{R}\sum_{r=1}^{R}\hat{\mu}_r$, its sample analogue; and $\bar{z} \equiv \frac{1}{R}\sum_{r=1}^{R} z_r$, the average sampling error across regions.

Using the definitions above, we can write

$$E\left(\widehat{RV}\right) \equiv \frac{1}{R}\sum_{r=1}^{R}E\left(\hat{\mu}_r - \bar{\hat{\mu}}\right)^2 = \frac{1}{R}\sum_{r=1}^{R}E\left(\mu_r + z_r - \bar{\mu} - \bar{z}\right)^2,$$

where we rely on the fact that we are dealing with regional fixed effects, rather than random effects, to move the expectations operator inside the summation. Expanding the right-hand side of the expression results in

$$E\left(\widehat{RV}\right) = \frac{1}{R}\sum_{r=1}^{R}\left[E(\mu_r - \bar{\mu})^2 - 2E(\mu_r - \bar{\mu})(z_r - \bar{z}) + E(z_r - \bar{z})^2\right].$$

Since μ_r and $\bar{\mu}$ are both scalars, this simplifies to

$$E\left(\widehat{RV}\right) = \frac{1}{R}\sum_{r=1}^{R}\left[(\mu_r - \bar{\mu})^2 - 2(\mu_r - \bar{\mu})E(z_r - \bar{z}) + E(z_r - \bar{z})^2\right].$$

The distributional assumptions on z imply that $E(z_r) = E(\bar{z}) = 0$. Therefore, we can write

$$E\left(\widehat{RV}\right) = RV + \frac{1}{R}\sum_{r=1}^{R}E(z_r - \bar{z})^2.$$

To characterize the bias, note that

$$E(z_r - \bar{z})^2 = E(z_r^2) - 2E\left(z_r \frac{1}{R}\sum_{s=1}^{R} z_s\right) + E\left(\frac{1}{R}\sum_{s=1}^{R} z_s\right)^2,$$

which we can write in terms of the variance and covariance parameters as

$$E(z_r - \bar{z})^2 = \sigma_{rr} - \frac{2}{R}\sum_{s=1}^{R}\sigma_{rs} + \frac{1}{R^2}\sum_{s=1}^{R}\sum_{t=1}^{R}\sigma_{st}.$$

The bias due to sampling variance is equal to the above expression, averaged across all regions. A consistent estimate of the bias can be calculated by summing up and taking the appropriate averages of estimated variances of and covariances between the estimated regional fixed effects. The more precisely the regional fixed effects are estimated, the smaller is the bias correction.

ACKNOWLEDGMENTS We thank the editors, David Cutler, Mark McClellan, and seminar participants at the University of Chicago, RAND, and the Brookings Institution for comments. Financial support from the George Stigler Center for the Economy and the State at the University of Chicago, the Leonard Schaeffer Center for Health Economics and Policy at the University of Southern California, and Pfizer Inc. is gratefully acknowledged.

References

Alchian, Armen A., and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *American Economic Review* 62, no. 5: 777–95.

Baicker, Katherine, Elliott S. Fisher, and Amitabh Chandra. 2007. "Malpractice Liability and the Practice of Medicine in the Medicare Program." *Health Affairs* 26, no. 3: 841–52.

Chandra, Amitabh, and Douglas O. Staiger. 2007. "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks." *Journal of Political Economy* 115, no. 1: 103–40.

Chernew, Michael E., and others. 2010. "Geographic Correlation between Large-Firm Commercial Spending and Medicare Spending." *American Journal of Managed Care* 16, no. 2: 131–38.

Christensen, Sandra, and Judith Wagner. 2000. "The Costs of a Medicare Prescription Drug Benefit." *Health Affairs (Millwood)* 19, no. 2: 212–18.

Cutler, David M., and Louise Sheiner. 1999. "The Geography of Medicare." *American Economic Review* 89, no. 2: 228–233.

De Alessi, Louis. 1974a. "An Economic Analysis of Government Ownership and Regulation: Theory and the Evidence from the Electric Power Industry." *Public Choice* 19: 1–42.

———. 1974b. "Managerial Tenure under Private and Government Ownership in the Electric Power Industry." *Journal of Political Economy* 82, no. 3: 645–53.

Drummond, Michael, and others. 1993. "Economic Evaluation of Pharmaceuticals: A European Perspective." *Pharmacoeconomics* 4, no. 3: 173–86.

———. 1999. "Current Trends in the Use of Pharmacoeconomics and Outcomes Research in Europe." *Value in Health* 2, no. 5: 323–32.

Eppig, Franklin J., and George S. Chulis. 1997. "Matching MCBS (Medicare Current Beneficiary Survey) and Medicare Data: The Best of Both Worlds." *Health Care Financing Review* 18, no. 3: 211–29.

Fisher, Elliott S., Julie P. Bynum, and Jonthan S. Skinner. 2009. "Slowing the Growth of Health Care Costs—Lessons from Regional Variation." *New England Journal of Medicine* 360, no. 9: 849–52.

Fisher, E. S., and others. 2003a. "The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care." *Annals of Internal Medicine* 138, no. 4: 273–87.

———. 2003b. "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care." *Annals of Internal Medicine* 138, no. 4: 288–99.

Fowler, Floyd J., and others. 2008. "Relationship between Regional per Capita Medicare Expenditures and Patient Perceptions of Quality of Care." *JAMA* 299, no. 20: 2406–12.

Gottlieb, Daniel J., and others. 2010. "Prices Don't Drive Regional Medicare Spending Variations." *Health Affairs* 29, no. 3: 537–43.

Kessler, Daniel, and Mark McClellan. 1996. "Do Doctors Practice Defensive Medicine?" *Quarterly Journal of Economics* 111, no. 2: 353–90.

———. 2002a. "Malpractice Law and Health Care Reform: Optimal Liability Policy in an Era of Managed Care." *Journal of Public Economics* 84, no. 2: 175–97.

———. 2002b. "How Liability Law Affects Medical Productivity." *Journal of Health Economics* 21, no. 6: 931–55.

Medicare Payment Advisory Commission. 2009. "Measuring Regional Variation in Service Use." Washington.

Morgan, Robert O., and others. 1997. "The Medicare-HMO Revolving Door—The Healthy Go In and the Sick Go Out." *New England Journal of Medicine* 337, no. 3: 169–75.

Phelps, Charles E. 2000. "Information Diffusion and Best Practice Adoption." In *Handbook of Health Economics,* vol. 1A, edited by Anthony J. Culyer and Joseph P. Newhouse. Amsterdam: Elsevier Science.

Rettenmaier, Andrew J., and Thomas R. Saving. 2009. "Perspectives on the Geographic Variation in Health Care Spending." Special Publication. Washington: National Center for Policy Analysis (June).

Sampford, M. R. 1953. "Some Inequalities on Mill's Ratio and Related Functions." *Annals of Mathematical Statistics* 24, no. 1: 130–32.

Sirovich, Brenda E., and others. 2006. "Regional Variations in Health Care Intensity and Physician Perceptions of Quality of Care." *Annals of Internal Medicine* 144, no. 9: 641–49.

Spann, Robert M. 1977. "The Macroeconomics of Unbalanced Growth and the Expanding Public Sector: Some Simple Tests of a Model of Government Growth." *Journal of Public Economics* 8, no. 3: 397–404.

Sutherland, Jason M., Elliott S. Fisher, and Jonathan S. Skinner. 2009. "Getting Past Denial—The High Cost of Health Care in the United States." *New England Journal of Medicine* 361, no. 13: 1227–30.

Wennberg, John, and Alan Gittelsohn. 1973. "Small Area Variations in Health Care Delivery." *Science* 182, no. 4117: 1102–08.

Wennberg, John E., and Megan McAndrew Cooper, eds. 1998. *Dartmouth Atlas of Health Care, 1998.* Chicago: Dartmouth Medical School and American Hospital Association.

Wennberg, John E., and others. 2008. *Tracking the Care of Patients with Severe Chronic Illness: The Dartmouth Atlas of Health Care 2008.* Lebanon, N.H.: Dartmouth Institute for Health Policy and Clinical Practice.

Comment and Discussion

COMMENT BY

DAVID M. CUTLER The issue of geographic variation in medical care spending played an important role in the recent health care reform debate. Supporters of reform cited the wide variation in spending across U.S. regions as evidence that care is inefficiently provided. According to a widely cited analysis by Elliott Fisher and others (2003), "If the United States as a whole could safely achieve spending levels comparable to those of the lowest-spending regions, annual savings of up to 30% of Medicare expenditures could be achieved." One-third of the nation's medical spending amounts to over $700 billion annually—a huge savings.

Few dispute that health care is characterized by significant inefficiency, but how best to limit that inefficiency is a subject of great debate. Broadly speaking, there are two approaches. The first is to use consumer demand to limit wasteful care. In this approach, informed, incentivized consumers would shop for efficient care just as they do for other products. Private insurers responding to this demand would then squeeze out excessive spending. The main barriers to this favorable competition, proponents argue, are the tax subsidy favoring the purchase of more generous insurance by employers in the private sector, and the lack of good stewardship of Medicare and Medicaid in the public sector. In the latter category, the biggest issue is the absence of a profit incentive to manage care for the bulk of Medicare and Medicaid enrollees.

The second approach focuses on the supply side. Here the key issue is how physicians and hospitals are paid for treating people. Doctors are reimbursed more for providing more care, but not for providing higher-quality care or limiting the need for care in the first place. Thus surgeons are paid for performing operations, but primary care physicians are not paid for the management involved in preventing surgery. Care will exceed the optimum because quantity is favored over quality.

These two theories have very different implications for health care reform. In the demand-side view, competition among insurers is the ideal, and private markets are preferred to public programs. Thus Medicare would do better to encourage entry by private plans than to focus on improving the fee-for-service system. In the supply-side view, the key is changing the financial incentives for providers. This is best accomplished by reforming the payment system in the traditional Medicare program, and working to spread that change throughout the health care system. Thus encouraging more private plans would not be a helpful step for Medicare, and might be harmful.

This paper by Tomas Philipson and his colleagues is directed at this debate. The idea behind the paper is to examine whether private insurers do a better job at limiting health care utilization than Medicare does. Evidence that private plans are better at rationing care would support the demand-side view of reform. The test that the authors propose is to look at the variability of care across a sample of metropolitan statistical areas. If private plans are better at eliminating excessive care, they should have lower spending in metropolitan areas where care is otherwise overprovided—defined in the paper as areas where Medicare fee-for-service costs are higher. Analytically, this would show up as lower variance of health care utilization across areas in the private sector.

Philipson and his coauthors test this hypothesis using data on patients with ischemic heart disease, a condition that bridges the over- and under-65 populations. They reach two conclusions. First, they show that the variation in two measures of utilization—the number of hospital days and outpatient visits—is indeed greater in Medicare than in private insurance. For both measures, the standard deviation across areas in Medicare is three times that across areas in private insurance. Interestingly, the average number of hospital stays per patient is not more variable in Medicare; rather, it is the average length of stay (hospital days) that is more variable. Private insurers seem better at getting people out of the hospital sooner.

In contrast, however, spending is much less variable across areas than is utilization. Areas with low private sector utilization have higher private sector prices than areas with higher utilization. This is not true in Medicare, where prices are set nationally.

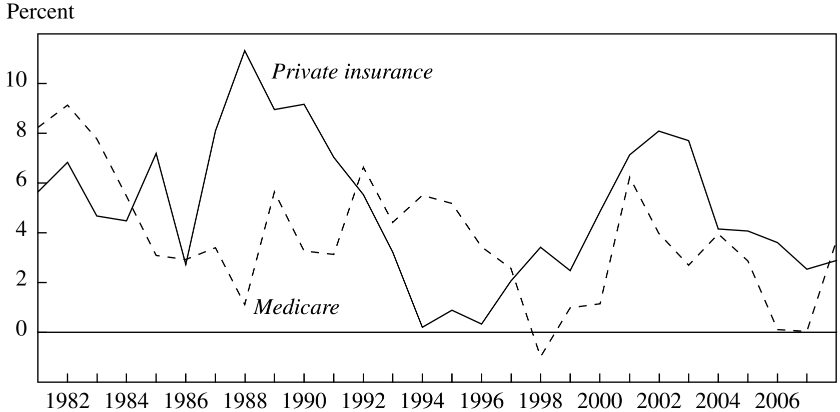
Whether one concludes from these findings that the public or the private sector is more efficient depends on whether one puts more weight on utilization or on spending. Since greater utilization has a real resource cost, whereas higher prices paid for given resources are just a transfer (if one excludes the deadweight loss of raising tax revenue to support the public programs), one might imagine that lower utilization at higher prices is preferable to higher utilization at lower prices. This logic favors the private sector. On the other hand, the additional care that people receive in Medicare is not necessarily wasteful. If the added care in the public sector has high marginal value, more care received at lower prices would be preferred to less care at higher prices, and Medicare would be judged superior.

A priori, it is not clear whether the additional care provided in high-utilization areas is worth the cost. If health care obeyed standard market theory, the care that well-informed patients facing price incentives decide not to purchase (either when the care is offered, or earlier, when they choose a less generous insurance plan) would be the care that is least valued. But patients are not always sure about what care they need, and providers have imperfect incentives to fully inform them. Thus one natural extension of the paper would be to compare health *outcomes* in the public and private sector, to see whether patients get care of real value. The data that Philipson and his coauthors employ do not include ideal measures of outcomes, but some outcome assessment is possible: for example, one could examine the rate at which patients in the sample are later admitted to a hospital with cardiac complications.

There is also an obvious alternative explanation for the greater geographic variance of utilization in Medicare than in private insurance: health status may vary more in the Medicare population than in the privately insured population. To control for this variation, Philipson and coauthors perform a variety of risk adjustments. Having already limited the sample to people with ischemic heart disease, they further control for other conditions that those patients may have. But these adjustments are imperfect, as any risk adjustment is.

The reason to suspect that greater variation in health status in the Medicare population is important is that the reduction in variation in the private sector occurs at both the top and the bottom end of the utilization distribution. The theory that the authors put forward is that private insurers should ration care more when underlying utilization is higher. Thus private sector utilization should be disproportionately lower in high-use areas. That is the point of figure 2 in the paper. In contrast, however, their figure 3 shows a uniform reduction in the standard deviation of care in the private sector. That is, high-utilization areas are less far above the mean in private insurance, but low-utilization areas are less far below the mean.

Figure 1. Growth in Real Spending per Enrollee in Private Health Insurance and Medicare, 1981–2008



Source: Centers for Medicare and Medicaid Services, National Health Expenditure Accounts.

The reason why care in low-utilization areas would be greater in the private sector is not clear. One might imagine that providers are skimping on the Medicare population in low-utilization areas, and that private insurers encourage more care than providers would otherwise prescribe. But the analysis by Fisher and his coauthors does not suggest significant underuse of care in the low-spending Medicare areas, nor do patients report themselves less satisfied in those areas. Given the approximately proportional reduction in variation in the private sector compared with Medicare, it seems that a theory of variation in underlying health status garners at least as much support as a theory of efficient rationing.

This conclusion is buttressed by a consideration of aggregate differences between Medicare and private insurance. An analysis of aggregate spending data does not suggest any superiority of private insurers in limiting cost increases over time. My figure 1 shows annual data on growth of real Medicare spending per beneficiary and private insurance premiums per enrollee. In each case the spending measure is limited to a common set of benefits, including hospital inpatient and outpatient care and physician services. The two series generally mirror each other. Cost growth was high in the 1980s, low in the 1990s, and resurgent in the 2000s. Overall, cost increases were greater in the private sector (averaging 4.9 percent per year) than in Medicare (3.7 percent).

A few anomalies in the pattern are of interest. Private insurance premiums rose more rapidly than Medicare spending in the 1980s, when prospective payment was introduced in the public sector. The managed care era of the 1990s saw moderately slower cost growth in the private sector than in Medicare. Managed care used the type of utilization restrictions that Philipson and his coauthors argue are effective.

But a good share of these utilization controls were reversed in the 2000s, when a backlash against managed care caused insurers to change their practices. Requirements for prior authorization were substantially weakened, and financial incentives for providers to perform less care were loosened. Further, providers responded to the presence of managed care by merging, and this led to higher prices in the private sector, where prices are flexible. The result was a decade of much higher private sector cost increases than Medicare cost increases.

The data that Philipson and coauthors analyze are from the 2000s. Figure 1 shows the peculiarity of the case that they make. They argue for the superiority of private utilization controls using data from a time when private sector controls on utilization were falling and costs were rising rapidly. It may be that people were wrong to reject the managed care strictures of the 1990s, and that a return to those practices would be beneficial. But few people argue that the version of managed care that prevailed in the 2000s is a sustainable model.

In the recent health care reform debate, the conclusion that public and private insurance were relatively similar in cost growth was more commonly accepted than the argument that private insurers were superior in reducing excess utilization. Added to this was the sense that private insurers engaged in significant and costly risk selection, wasting money and making insurance difficult to obtain. Thus, regulating private insurers to limit risk selection is a significant element of the recently enacted Patient Protection and Affordable Care Act (PPACA), whereas opening up new markets in Medicare for private insurers is not. Indeed, the act reduces Medicare payments to private insurers. In its place are a series of pilot programs and demonstration projects aimed at reforming the incentives in the existing fee-for-service program.

Whether the strategy underlying the PPACA is the right one or not will be determined in the next few years. If reform fails, it may be because private insurers were strangled too tightly. If so, this paper may point the way to a future reform. But for now we are likely to see tighter restrictions on private insurance before any significant new utilization controls.

REFERENCE FOR THE CUTLER COMMENT

Fisher, Elliott S., David E. Wennberg, Thérèse A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Étoile L. Pinder. 2003. "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care." *Annals of Internal Medicine* 138: 288–98.

GENERAL DISCUSSION Christopher Sims raised the question of how secondary private insurance fits into the classification of public versus private. Many Medicare beneficiaries have supplementary coverage, and whether an individual has it or not could affect his or her utilization choices. If there is regional variation in the share or beneficiaries using Medicare supplements, it could be affecting the paper's results.

Daniel Sacks asked whether the privately covered individuals in the sample were truly representative of people with private coverage generally. If the sample is drawn from Fortune 500 companies only, their health insurance plans may be better than the average private plan and have better, or at least different, cost management. He also wondered what incentives the executives of both public and private insurance plans actually face and whether they are as different as the paper assumes.

David Romer observed that the theory laid out in the paper makes a clear prediction of where the expected change in the distribution of fixed effects ought to appear: The incentives facing the private sector should lead to compression of the upper tail of the distribution in that sample but not the lower tail, yet no such pattern is visible in the data.

BROOKINGS

www.brookings.edu

ISBN 978-0-8157-0513-0