Dynamic Modeling and Econometrics in Economics and Finance 20

Marco Gallegati
Willi Semmler *Editors*

Wavelet Applications in Economics and Finance

Springer

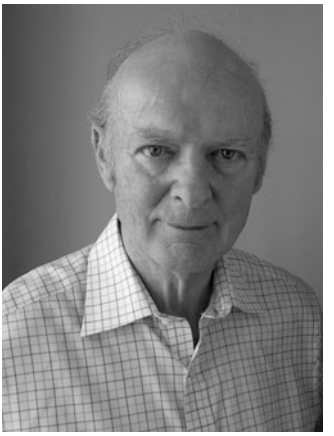
Dynamic Modeling and Econometrics in Economics and Finance

Volume 20

Editors
Stefan Mittnik
University of Munich
Munich, Germany

Willi Semmler
Bielefeld University
Bielefeld, Germany
and
New School for Social Research
New York, USA

For further volumes:
http://www.springer.com/series/5859



James B. Ramsey

Marco Gallegati • Willi Semmler
Editors

Wavelet Applications in Economics and Finance

Springer

Editors
Marco Gallegati
Faculty of Economics "G.Fuà"
Polytechnic University of Marche
Ancona
Italy

Willi Semmler
New School for Social Research
The New School University
New York
USA

ISSN 1566-0419 Dynamic Modeling and Econometrics in Economics and Finance
ISBN 978-3-319-07060-5 ISBN 978-3-319-07061-2 (eBook)
DOI 10.1007/978-3-319-07061-2
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014945649

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

Mater semper certa est, pater numquam ("The mother is always certain, the father is always uncertain") is a Roman-law principle which has the power of *praesumptio iuris et de iure*. This is certainly true for biology, but not for *wavelets* in economics which have a true father: James Ramsey.

The most useful property of wavelets is its ability to decompose a signal into its time scale components. Economics, like many other complex systems, include variables simultaneously interacting on different time scales so that relationships between variables can occur at different horizons. Hence, for example, we can find a stable relationship between durable consumption and income. And the literature is soaring: from money–income relationship to Phillips curve, from financial market fluctuations to forecasting. But this feature threatens to undermine the very foundations of the Walrasian construction. If variables move differently at different time scales (stock market prices in nanoseconds, wages in weeks, and investments in months), then also a linear system can produce chaotic effects and market self-regulation is lost. If validated, wavelet research becomes a silver bullet.

James is also an excellent sailor (in 2003 he sailed across the Atlantic to keep his boat from North America to Turkey), and his boat braves the streams with "nonchalance": by the way if you are able to manage wavelets, you are also ready for waves.

Ancona, Italy
March 2, 2014

Mauro Gallegati

Preface

James Bernard Ramsey received his B.A. in Mathematics and Economics from the University of British Columbia in 1963, and his M.A. and Ph.D. in Economics from the University of Wisconsin, Madison in 1968 with the thesis "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis". After being Assistant and Associate Professor at the Department of Economics of Michigan State University, he became Professor and Chair of Economics and Social Statistics at the University of Birmingham, England, from 1971 to 1973. He went back to the US as Full Professor at Michigan State University until 1976 and finally moved to New York University as Professor of Economics and Chair of the Economics Department between 1978 and 1987, where he remained for 37 years until his retirement in 2013. Fellow of the American Statistical Association, Visiting Fellow at the School of Mathematics (Institute for Advanced Study) at Princeton in 1992–1993, and ex-president of the Society for Nonlinear Dynamics and Econometrics, James Ramsey was also a jury member of the Econometric Game 2009. He has published 7 books and more than 60 articles on nonlinear dynamics, stochastic processes, time series, and wavelet analysis with special emphasis on the analysis of economic and financial data.

This book intends to honor James B. Ramsey and his contribution to economics on occasion of his recent retirement from academic activities at the NYU Department of Economics. This festschrift, as it is called in the German tradition, intends to honor an exceptional scholar whose fundamental contributions have influenced a wide range of disciplines, from statistics to econometrics and economics, and whose lifelong ideas have inspired more than a generation of researchers and students.

He is widely acclaimed for his pioneering work in the early part of his career on the general specifications test for the linear regression model, Ramsey's RESET test, which is part of any econometric software now. He is also well known for his contributions to the theory and empirics of chaotic and nonlinear dynamical systems. A significant part of his work has also been devoted to the development of genuine new ways of processing data, as for instance the application of functional data analysis or the use of wavelets in terms of nonparametric analysis.

Each year the Society for Nonlinear Dynamics and Econometrics, at its Annual Conference, awards two James Ramsey prizes for top graduate papers in econometrics. This year there will also be a set of special sessions dedicated to his research. One of these sessions will be devoted to wavelet analysis, an area where James work has had a great outstanding impact in the last twenty years. James Ramsey and his coauthors have provided early applications of wavelets in economics and finance by making use of discrete wavelet transform (DWT) in decomposing economic and financial data. These works paved the way to the application of wavelet analysis for empirical economics. The articles in this book are comprised of contributions by colleagues, former students, and researchers covering a wide range of wavelet applications in economics and finance and are linked to or inspired by the work of James Ramsey.

We have been working with James continuously over the last 10 years and have always been impressed by his competence, motivation, and enthusiasm. Our collaboration with James was extraordinarily productive and an inspiration to all of us. Working together we developed a true friendship strengthened by virtue of the pleasant meetings held periodically at James office on the 7th floor of the NYU Department of Economics, which became an important space for discussing ongoing as well as new and exciting research projects. As one of his students has recently written, rating James' Statistics class: "He is too smart to be teaching!" Sometimes our impression was that he could also have been too smart for us as coauthor. This book is a way to thank him for the privilege we have had to met and work with him.

Ancona, Italy — Marco Gallegati
New York, NY — Willi Semmler
March 2014

Introduction

Although widely used in many other disciplines like geophysics, engineering (sub-band coding), physics (normalization groups), mathematics (C-Z operators), signal analysis, and statistics (time series and threshold analysis), wavelets still remain largely unfamiliar to students of economics and finance. Nonetheless, in the past decade considerable progress has been made, especially in finance and one might say that wavelets are the "wave of the future". The early empirical results show that the separation by time scale decomposition analysis can be of great benefit for a deeper understanding of economic relationships that operate simultaneously at several time scales. The "short and the long run" can now be formally explored and studied.

The existence of time scales, or "planning horizons", is an essential aspect of economic analysis. Consider, for example, traders operating in the market for securities: some, the fundamentalists, may have a very long view and trade looking at market fundamentals and concentrate their attention on "long run variables" and average over short run fluctuations. Others, the chartists, may operate with a time horizon of only weeks, days, or even hours. What fundamentalists deem to be variables, the chartists deem constants. Another example is the distinction between short run adaptations to changes in market conditions; e.g., merely altering the length of the working day, and long run changes in which the firm makes strategic decisions and installs new equipment or introduces new technology.

A corollary of this assumption is that different planning horizons are likely to affect the structure of the relationships themselves, so that they might vary over different time horizons or hold at certain time scales, but not at others. Economic relationship might also show negative relationship over some time horizon, but a positive one over others. These different time scales of variation in the data may be expected to match the economic relationships more precisely than a single time scale using aggregated data. Hence, a more realistic assumption should be to separate out different time scales of variation in the data and analyze the relationships among variables at each scale level, not at the aggregate level. Although the concepts of the "short-run" and of the "long-run" are central for modeling economic and financial

decisions, variations in those relationships across time scales are seldom discussed nor empirically studied in economics and finance.

The theoretical analysis of time, or "space series" split early on into the "continuous wavelet transform", CWT, and into "discrete wavelet transform". DWT. The latter is often more useful for applying to regular time series analysis with observations at discrete intervals. Wavelets provide a multi-resolution decomposition analysis of the data and can produce a synthesis of the economic relationships that is parameter preserving. The output of wavelet transforms enables one to decompose the data in ways that are potentially revealing relationships that are not visible using standard methods on "scale aggregated" data. Given their ability to isolate the bounds on the temporary frequency content of a process as a function of time, it is a great advantage of those transforms to be able to rely only on local stationarity that is induced by the system, although Gabor transforms provide a similar service for Fourier series and integrals.

The key lesson in synthesizing the wavelet transforms is to facilitate and develop the theoretical insight into the interdependence of economic and financial variables. New tools are most likely to generate new ways of looking at the data and new insights into the operation of the finance–real interaction.

The 11 articles collected in this volume, all strictly refereed, represent original up-to-date research papers that reflect some of the latest developments in the area of wavelet applications for economics and finance.

In the first chapter James provides a personal retrospective of a decade's research that highlights the links between CWT, DWT wavelets and the more classical Fourier transforms and series. After stressing the importance of analyzing the various basis spaces, the exposition evaluates the alternative bases available to wavelet researchers and stresses the comparative advantage of wavelets relative to the alternatives considered. The appropriate choice of class of function, e.g., Haar, Morlet, Daubchies, etc., with rescaling and translation provide appropriate bases in the synthesis to yield parsimonious approximations to the original time or space series.

The remaining papers examine a wide variety of applications in economics and finance that reveal more complex relationships in economic and financial time series and help to shed light on various puzzles that emerged in the literature since long; on business cycles, traded assets, foreign exchange rates, credit markets, forecasting, and labor market research. Take, for example, the latter. Most economists agree that productivity increases welfare, but whether productivity also increases employment is still controversial. As economists have shown using data from the EU and the USA, productivity may rise, but employment may be de-linked from productivity increases. Recent work has shown that the analysis of the relationship between productivity and employment is one that can only properly be analyzed after decomposition by time scale. The variation in the short run is considerably different from the variation in the long run. In the chapter "Does Productivity Affect Unemployment? A Time-Frequency Analysis for the US", Marco Gallegati, Mauro Gallegati, James B. Ramsey, and Willi Semmler, applying parametric and

nonparametric approaches to US post-war data, conclude that productivity creates unemployment in the short and medium term, but employment in the long run.

The chapters "The Great Moderation Under the Microscope: Decomposition of Macroeconomic Cycles in US and UK Aggregate Demand" and "Nonlinear Dynamics and Wavelets for Business Cycle Analysis" contain articles using wavelets for business cycles analysis. In the paper by P.M. Crowley and A. Hughes Hallett the Great Moderation is analyzed employing both static and dynamic wavelet analysis using quarterly data for both the USA and the UK. Breaking the GDP components down into their frequency components they find that the "great moderation" shows up only at certain frequencies, and not in all components of real GDP. The article by P.M. Addo, M. Billio, and D. Guégan applies a signal modality analysis to detect the presence of determinism and nonlinearity in the US Industrial Production Index time series by using a complex Morlet wavelet.

The chapters "Measuring the Impact Intradaily Events Have on the Persistent Nature of Volatility" and "Wavelet Analysis and the Forward Premium Anomaly" deal with foreign exchange rates. In their paper M.J. Jensen and B. Whitcher measure the effect of intradaily events on the foreign exchange rates level of volatility and its well-documented long-memory behavior. Volatility exhibits the strong persistence of a long-memory process except for the brief period after a market surprise or unanticipated economic news announcement. M. Kiermeier studies the forward premium anomaly using the MODWT and estimate the relationship between forward and corresponding spot rates on foreign exchange markets on a scale-by-scale basis. The results show that the unbiasedness hypothesis cannot be rejected if the data is reconstructed using medium-term and long-term components.

Two papers analyzing the influence of several key traded assets on macroeconomics and portfolio behavior are included in the chapters "Oil Shocks and the Euro as an Optimum Currency Area" and "Wavelet-Based Correlation Analysis of the Key Traded Assets". L. Aguiar-Conraria, T.M. Rodrigues, and M.J. Joana Soares study the macroeconomic reaction of Euro countries to oil shocks after the adoption of the common currency. For some countries, e.g., Portugal, Ireland, and Belgium, the effects of an oil shock have become more asymmetric over the past decades. J. Baruník, E. Kočenda, and L. Vácha, in their paper, provide evidence for different dependence between gold, oil, and stocks at various investment horizons. Using wavelet-based correlation analysis they find a radical change in correlations after the 2007–2008 in terms of time-frequency behavior.

A surprising implication of the development of forecasting techniques to real and financial economic variables is the recognition that the results are strongly dependent on the analysis of scale. Only in the simplest of circumstances will forecasts based on traditional time series aggregates accurately reflect what is revealed by the time scale decomposition of the time series. The chapter "Forecasting via Wavelet Denoising: The Random Signal Case" by J. Bruzda presents a wavelet-based method of signal estimation for forecasting purposes based on wavelet shrinkage combined with the MODWT. The comparison of the random signal estimation with analogous methods relying on wavelet thresholding suggests that the proposed approach may be useful especially for short-term forecasting. Finally, the chapters

"Short and Long Term Growth Effects of Financial Crises" and "Measuring Risk Aversion Across Countries from the Consumption-CAPM: A Spectral Approach" contain two articles using the spectral approach. F.N.G. Andersson and P. Karpestam investigate to what extent financial crises can explain low growth rates in developing countries. Distinguishing between different sources of crises and separating short- and long-term growth effects of financial crises, they show that financial crises have reduced growth and that the policy decisions have caused them to be worsened and prolonged. In their paper E. Panopolou and S. Kalyvitis adopt a spectral approach to estimate the values of risk aversion over the frequency domain. Their findings suggest that at lower frequencies risk aversion falls substantially across countries, thus yielding in many cases reasonable values of the implied coefficient of risk aversion.

Contents

Contributors

Peter Martey Addo Université Paris 1 Panthéon-Sorbonne, Paris, France

Luís Aguiar-Conraria NIPE and Economics Department, University of Minho, Braga, Portugal

Fredrik N.G. Andersson Department of Economics, Lund University, Lund, Sweden

Jozef Baruník Institute of Economic Studies, Charles University, Prague, Czech Republic

Monica Billio Department of Economics, Università CaFoscari of Venice, Italy

Joanna Bruzda Department of Logistics Faculty of Economic Sciences and Management, Nicolaus Copernicus University, Torun, Poland

Patrick M. Crowley Economics Group, College of Business, Texas A&M University, Corpus Christi, TX, USA

Marco Gallegati Department of Economics and Social Sciences, Università Politecnica delle Marche, Ancona, Italy

Mauro Gallegati Department of Economics and Social Sciences, Università Politecnica delle Marche, Ancona, Italy

Dominique Guégan Université Paris 1 Panthéon-Sorbonne, Paris, France

Mark J. Jensen Federal Reserve Bank of Atlanta, Atlanta, GA, USA

Sarantis Kalyvitis Department of International and European Economic Studies, Athens University of Economics and Business, Athens, Greece

Peter Karpestam Department of Economics, Lund University, Lund, Sweden

Michaela M. Kiermeier Fachbereich Wirtschaft, Hochschule Darmstadt, Dieburg, Germany

Evžen Kočenda CERGE-EI, Charles University and the Czech Academy of Sciences, Prague, Czech Republic

Ekaterini Panopoulou Department of Statistics and Insurance Science, University of Piraeus, Athens, Greece and University of Kent, UK

James B. Ramsey Department of Economics, New York University, New York, NY, USA

Teresa Maria Rodrigues Economics Department, University of Minho, Braga, Portugal

Willi Semmler Department of Economics, New School for Social Research, New York, NY, USA

Maria Joana Soares NIPE and Department of Mathematics and Applications, University of Minho, Braga, Portugal

Lukáš Vácha Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague, Czech Republic

Brandon Whitcher Pfizer Worldwide Research & Development, Cambridge, MA, USA

Functional Representation, Approximation, Bases and Wavelets

James B. Ramsey

Abstract After stressing the importance of analyzing the various basis spaces, the exposition evaluates the alternative bases available to wavelet researchers. The next step is to demonstrate the impact of choice of basis for the representation or projection of the regressand. The similarity of formulating a basis is explored across a variety of alternative representations. This development is followed by a very brief overview of some articles using wavelet tools. The comparative advantage of wavelets relative to the alternatives considered is stressed.

1 Introduction

The paper begins with a review of the main features of wavelet analysis which are contrasted with other analytical procedures, mainly Fourier, splines, and linear regression analysis. A review of Crowley (2007), Percival and Walden (2000), Bruce and Gao (1996), the excellent review by Gençay et al. (2002), or the Palgrave entry for Wavelets by Ramsey (2010) before proceeding would be beneficial to the neophyte wavelet researcher.

The next section contains a non-rigorous development of the theory of wavelets and contains discussions of wavelet theory in contrast to the theory of Fourier series and splines. The third section discusses succinctly the practical use of wavelets and compares alternative bases; the last section concludes.

Before proceeding the reader should note that all the approximating systems are characterized by the functions that provide the basis vectors, e.g. $\mathrm{Sin}(k\omega_t)$, $\mathrm{Cos}(k\omega_t)$ for Fourier series, or "t" for the monomials, e^{λ_i} for the exponentials, etc.

J.B. Ramsey (✉)
Department of Economics, New York University, New York, NY, USA
e-mail: james.ramsey@nyu.edu

M. Gallegati and W. Semmler (eds.), *Wavelet Applications in Economics and Finance*,
Dynamic Modeling and Econometrics in Economics and Finance 20,
DOI 10.1007/978-3-319-07061-2_1,

For a regular regression framework, the basis is the standard Euclidean space, E_N. For the Fourier projections we have the frequency scaled sine and cosine functions that produce a basis of infinite power, high resolution in the frequency domain, but no resolution in the time domain; e.g.

$$Re^{i2\pi ft} \quad \text{or alternatively expressed:}$$
$$1, sin(k\omega t), cos(k\omega t) \quad k = 1, 2, 3, \ldots$$

are highly differential, but are not suitable for analyzing signals with discrete changes and discontinuities.

The basis functions for splines are polynomials that are also differential and are defined over a grid determined by the knots; various choices for the differentiability at the knots determine the flexibility and smoothness of the spline approximation and the degree of curvature between knots.

Obviously, the analysis of any signal involves choosing both the approximating function and the appropriate basis vectors generated from the chosen function.

The concepts of "projection" and analysis of a function are distinguished; for the former one considers the optimal manner in which an N dimensional basis space can be projected onto a K dimensional subspace.

For a given level of approximation one seeks the smallest K for the transformed basis. Alternatively, a given function can be approximated by a series expansion, which implies that one is assuming that the function lies in a space defined in turn by a given class of functions, usually defined to be a Hilbert space. Projection and representation of a function are distinguished.

2 Functional Representation and Basis Spaces

2.1 An Overview of Bases in Regression Analysis

Relationships between economic variables are characterized by three universal components. Either the variable is a functional defined by an economic equation as a function of itself lagged to its past, i.e. is autoregressive; or is a function of time, i.e. is a "time series"; or it is a projection onto the space spanned by a set of functions, labeled, "regressors" each of which in turn may be autoregressive, or a vector of "time series". The projection of the regressand on the space spanned by the regressors provides a relationship between the variables, which is invariant to permutations of the indexing of the variables:

$$Y = X\beta + u$$
$$Y_{perm} = X_{perm}\beta + u_{perm}$$

where Y is the regressand, Y_{perm} the permuted values of Y, X_{perm}, represents a conformable permutation of the rows of X, and u_{perm}, a conformable permutation to Y_{perm}. However, if the formulation of the model involves an "ordering" of the variables over space, or over time, the model is then not invariant to permutation of the index of the ordering. It is known, but seldom recognized as a limitation of the projection approach, that least squares approximations are invariant to any permutation of the ordering. Consequently, the projection approach omits the information within the ordering in the space spanned by the residuals, which is, of course, the null space. Another distinguishing characteristic is that added to the functional development of the variable known as the "regressand" is an unobserved random variable, "u", which may be represented by a solitary pulse, or may have a more involved stochastic structure. In the former case, the regressand vector is contained in the space spanned by the regressors, where as in the latter case the regressand is projected onto the space spanned by the designated regressors.

The usual practice is to represent regressors and the regressand in terms of the standard Euclidean N dimensional space; i.e. the i_{th} component of the basis vector is "1", the remaining entries are zero; in this formulation, we can interpret the observed terms, x_i, y_i, $i = 1, 2, \ldots k$; as N dimensional vectors relative to the linear basis space, E_N.

The key question the analyst needs to resolve is to derive an appropriate procedure for determining reasonable values for the unknown parameters and coefficients of the system; i.e. estimation of coefficients and forecasting of declared regressands. Finally, if the postulated relationship is presumed to vary over space or time, special care will be needed to incorporate those changes in the relationship over time or over the sample space.

Consider as a first example, a simple non-linear differentiable function of a single variable x, $f(x|\theta)$, which can be approximated by a Taylor's series expansion about the point a_1 in powers of x:

$$y = f(x|\theta) = f(a_1|\theta) + f^1(a_1|\theta)(x - a_1) + \frac{f^2(a_1|\theta)(x - a_1)^2}{2!} + \frac{f^3(a_1|\theta)(x - a_1)^3}{3!} + \frac{R(\xi)}{4!} \tag{1}$$

for some ξ value. This equation approximately represents the variation of y in terms of powers of x. Care must be taken in that the derived relationship is not exact, as the required value for ξ in the remainder term will vary for different values for a_1, x, and the highest derivative used in the expansion. Under the assumption that $R(\xi)$ is approximately zero, the parameters θ given the coefficient a_1 can be estimated by least squares using N independent drawings on the regressand's error term. Assuming the regressors are observed error free; one has:

$$\min_{\theta} \{\Sigma_1^N (y_i - f(x_i|\theta))^2\} \tag{2}$$

A single observation, i, on this simple system is:

$$y_i \ \{x_i, x_i^2 \ x_i^3\} \tag{3}$$

$$i = 1, 2, 3 \ldots . N. \tag{4}$$

This model is easily extended to differential functions which are themselves functions of multivariate regressors. The key aspect of the above formulation is that the estimators are obtained by a projection onto the space spanned by the regressors. Other, perhaps more suitable spaces, can be used instead. The optimal choice for a basis, as we shall see, is one that reduces significantly the required number of coefficients to represent the function y_i with respect to the chosen basis space. Different choices for the basis will yield different parameterizations; the research analyst is interested in minimizing the number of coefficients, actually the dimension of the supporting basis space.

2.2 Mononomial Basis

An alternative, ancient, procedure is provided by the monomials:

$$\{1, t^1, t^2, t^3, t^4, t^5, \ldots\} \tag{5}$$

that is, we consider the projection of a vector y on the space spanned by the monomials, $t^0, t^1, \ldots, t^k$, or as became popular as a calculation saving device, one considers the projection of y on the orthogonal components of the sequence in Eq. (5), see Kendall and Stuart (1961).

These first two procedures indicate that the underlying concept was that insight would be gained if the projections yielded approximations that could be specified in terms of very few estimated coefficients. Further very little structure was imposed on the model, either in terms of the statistical properties of the model or in terms of the restrictions implied by the underlying theory.

Two other simple basis spaces are the exponential

$$\{e^{\lambda_1 t}, \ e^{\lambda_2 t}, \ e^{\lambda_3 t} \ldots . . e^{\lambda_k t}\} \tag{6}$$

and the power base:

$$\{t^{\lambda_1}, \ t^{\lambda_2}, \ t^{\lambda_3}, \ \ldots . . . . t^{\lambda_k}\}. \tag{7}$$

The former is most useful in modeling differential equations, the latter in modeling difference equations.

2.3 Spline Bases

A versatile basis class is defined by the spline functions. A standard definition of a version of the spline basis, the B-spline, $S_B(t)$, is:

$$S_B(t) = \sum_{k=1}^{m+L-1} c_k B_{k,(t,\tau)} \tag{8}$$

where $S_B(t)$ is the spline approximation, c_k, are the coefficients of the projection, $B_{k,(t,\tau)}$ is the B-spline function at position k, with knot structure, τ. The vector τ designates the number of knots, L, and their position which defines the subintervals that are modeled in terms of polynomials of degree m. At each knot the polynomials are constrained to be equal in value for polynomials of degree 1, agreement for the first derivative for polynomials of degree 2, etc. Consequently, adjacent spline polynomials line up smoothly.

B-Splines are one of the most flexible basis systems, so that it can easily fit locally complex functions. An important use of splines is to interpolate over the grid created by the knots in order to generate a differential function, or more generally, a differential surface. Smoothing is a local phenomenon.

2.4 Fourier Bases

The next procedure in terms of longevity of use is Fourier analysis. The basis for the space spanned by Fourier coefficients is given by:

$$1, \sin(k\omega t), \cos(k\omega t), \tag{9}$$

$$k = 1, 2, 3, \ldots$$

$$i.e. \exp(i\omega_k) \tag{10}$$

where ω is the fundamental frequency. The approximating sequences are given most simply by:

$$y = f(t) \cong \sum_{k=1}^{K} c_k \phi_k \tag{11}$$

where the sequence c_k specifies the coefficients chosen to minimize the squared errors between the observed sequence and the known functions shown in Eq. (11), ϕ_k is the basis function as used in Eq. (9), and the coefficients are given by

$$c_k = \int f(t)\phi_k(t)\, dt \tag{12}$$

The implied relationships between the basis function, ϕ, the basis space given by ϕ_k, $k = 1, 2, 3, ..$, and representation of the function $f(t)$, are given in abstract form in Eq. (12), in order to emphasize the similarities between the various basis spaces.

We note two important aspects of this equation. We gain in understanding if the number of coefficients are few in number; i.e. k is "small". We gain if the function "f" is restricted to functions of a class that can be described in terms of the superposition of the basis functions, e.g. trigonometric functions and their derivatives for Fourier analysis. The fit for functions that are continuous, but not every where differential, can only be approximated using many basis functions. The equations generating the basis functions, ϕ_k, based on the fundamental frequency, ω, are re-scaled versions of that fundamental frequency. The concept of re-scaling a "fundamental" function to provide a basis will occur in many guises.

Fourier series are useful in fitting global variation, but respond to local variation only at very high frequencies thereby substantially increasing the required number of Fourier coefficients to achieve a given level of approximation For example, consider fitting a Fourier basis to a "Box function", any reasonable degree of fit will require very many terms at high frequency at the points of discontinuity (see Bloomfield 1976; Korner 1988).

Economy of coefficients can be obtained for local fitting by using windows, that is, instead of

$$\hat{h}(\omega) = \frac{1}{2\pi} \sum_{-\infty}^{\infty} \hat{R}(s) \cos(s\omega)$$

where $\hat{R}(s)$ is the sample covariance at lag "s", we consider

$$\hat{h}(\omega) = \frac{1}{2\pi} \sum_{-M}^{M} \lambda(s) \hat{R}(s) \cos(s\omega) \tag{13}$$

where $\lambda(s)$ is the "window function" which has maximum effect at $s = 0, +/- 2k\pi$, $k = 1, 2, 3, \ldots$. Distant correlations are smoothed, the oscillations of local events are enhanced (see Bloomfield 1976).

A more precise formulation is provided by stating that for the function "f" defined for a mapping from the real line modulo 2π to R, the Fourier coefficients of "f" are given by:

$$\begin{aligned} \hat{f}(r) &= (2\pi)^{-1} \int_0^{2\pi} f(t) \exp(-irt)\, dt \\ &= (2\pi)^{-1} \int_T f(t) \exp(-irt)\, dt \end{aligned} \tag{14}$$

For simple functions we have the approximation (Korner 1988):

$$S_n(f,t) = \sum_{-n}^{n} \hat{f}(r) \exp irt \longrightarrow f(t) \quad (15)$$
$$as\ n \longrightarrow \infty$$

$T \longrightarrow C$ is continuous everywhere and has a continuous bounded derivative except at a finite number of points, then $S_n(f,\ ) \longrightarrow f$ uniformly (Korner 1988). The problem we have to face is the behavior of the function at points of discontinuity and to be aware of the difficulties imposed by even a finite number of discontinuities. For example, consider:

$$h(x) = x, \quad -\pi < x < \pi \quad (16)$$
$$h(\pi) = 0$$
$$\hat{h}(0) = 0$$

As pointed out by Korner (1988), the difficulty is due to the confusion between "the limit of the graphs and the graph of the limit of the sum". This insight was presented by Gibbs and illustrated practically by Michelson; that is $S_n(h,t) \longrightarrow h(t)$ pointwise; that is the blips move towards the discontinuity but pointwise convergence of f_n to f does not imply that the graph of f_n starts to look like f for large N shown in (16). The important point to remember is that the difference is bounded from below in this instance by:

$$\frac{2}{\pi} \int_0^{\pi} \frac{\sin x}{x}\, dx > 1.17 \quad (17)$$

The main lesson here for the econometrician is that observed data may well contain apparently continuous functions that are not only sampled at discrete intervals, but that may in fact contain significant discontinuities. Indeed, one may well face the problem of estimating a continuous function that is nowhere differential,the so called "Weierstrass functions" (see, for example, Korner 1988).

It is useful to note that, whether we are examining wavelets (to be defined below), or sinusoids or Gabor functions, we are in fact approximating $f(t)$ by "atoms".[1] We seek to obtain the best M atoms for a given $f(t)$ out of a dictionary of P atoms. There are three standard methods for choosing the M atoms in this over sampled situation. The first is "matching pursuit" in which the M atoms are chosen one at a time; this procedure is referred to as greedy and sub-optimal (see Bruce and Gao 1996). An alternative method is the best basis algorithm which begins with a

[1] A collection of atoms is a "dictionary".

dictionary of bases. The third method, which will be discussed in the next section, is known as basis pursuit, where the dictionary is still over complete. The synthesis of $f(t)$ in terms of $\phi_i(t)$ is under-determined.

This brief discussion indicates that the essential objective is to choose a good basis. A good basis depends upon the resolution of two characteristics; *linear independence* and *completeness*. Independence ensures uniqueness of representation and completeness ensures that any $f(t)$ within a given class of functions can be represented in terms of the basis vectors. Adding vectors will destroy independence, removing vectors will destroy completeness. Every vector v or function $v(t)$ can be represented uniquely as:

$$v = \sum b_i v_i \qquad (18)$$

$$or$$

$$v(t) = \sum b_i v_i(t)$$

provided the coefficients b_i satisfy:

$$A||v||^2 \leq \sum |b_i|^2 \leq B||v||^2 \, with \, A > 0. \qquad (19)$$

This is the defining property of a Riesz basis (see, for example, Strang and Nguyen 1996).

If $0 < A < B$ and Eq. (19) holds and the basis generating functions are defined within a Hilbert space, then we have defined a frame and A, B are the frame bounds. If A equals B the bounds are said to be tight; if further the bounds are unity, i.e. $A = B = 1$, one has a orthonormal basis for the transformation. For example, consider a frame within a Hilbert space, C, given by: $e_1 = (0, 1)$, $e_2 = (-\frac{\sqrt{3}}{2} - \frac{1}{2})$, $e_3 = (-\frac{\sqrt{3}}{2}, \frac{1}{2})$. For any v in the Hilbert space we have:

$$\sum_{j=1}^{3} \left|\langle v, e_j \rangle\right|^2 = \frac{3}{2} \|v\|^2 \qquad (20)$$

where the redundancy ratio is $3/2$, i.e. three vectors in a two dimensional space (Daubechies 1992).

2.5 Wavelet Bases

Much of the usefulness of wavelet analysis has to do with its flexibility in handling a variety of nonstationary signals. Indeed, as wavelets are constructed over finite intervals of time and are not necessarily homogeneous over time, they are localized in time and scale. The projection of the analizable signal onto the wavelet function

by time scale and translation produces an orthonormal transformation matrix, W, such that the wavelet coefficients, w, are represented by:

$$w = Wx \tag{21}$$

where x is the analizable signal (see Eq. (21)). While theoretically this is a very useful relationship which clarifies the link between wavelet coefficients and the original data, it is decidedly not useful in reducing the complexity of the relationships and does not provide a suitable mechanism for evaluating the coefficients (Bruce and Gao 1996).

The experienced Waveletor knows also to consider the shape of the basis generating function and its properties at zero scale. This concern is an often missed aspect of wavelet analysis. Wavelet analysis, unlike Fourier analysis, can consider a wide array of generating functions. For example, if the function being examined is a linearly weighted sum of Gaussian functions, or of the second derivatives of Gaussian functions, then efficient results will be obtained by choosing the Gaussian function, or the second derivative of the Gaussian function in the latter case. This is a relatively under utilized aspect of wavelet analysis, which will be discussed more fully later.

Further any moderately experienced "Waveletor" knows to choose his wavelet generating function so as to maximize the "number of zero moments", to ascertain the number of continuous derivatives (as a measure of smoothness), and to worry about the symmetry of the underlying filters although one may consider models for which asymmetry in the wavelet generating function is appropriate. While many times the choice of wavelet generating function makes little or no difference there are times when such considerations are important for the analysis in hand. For example, the inappropriate use of the Haar function for resolving continuous smooth functions, or using smooth functions to represent samples of discontinuous paths. Wavelets provide a vast array of alternative wavelet generating functions, e.g. Gaussian, Gaussian first derivative, Mexican hat, the Daubechies series, the Mallat series, and so on. The key to the importance of the differences lies in choosing the appropriate degree and nature of the oscillation within the supports of the wavelet function. With the Gaussian, first, and second derivatives as exceptions, the generating functions are usually derived from applying a pair of filters to the data using subsampled data (Percival and Walden 2000).

I have previously stated that at each scale the essential operation is one of differencing using weighted sums; the alternative rescaleable wavelet functions provide an appropriate basis for such differences. Compare for example:

$$\begin{aligned} Haar \;:\; (h_0,\, h_1) &= (-\frac{1}{2},\, \frac{1}{2}) \\ Daubchies(D4) &= (h_0,\, h_1,\, h_2,\, h_3) \\ &= (\frac{1-\sqrt{3}}{4\sqrt{2}},\, \frac{-3+\sqrt{3}}{4\sqrt{2}},\, \frac{3+\sqrt{3}}{4\sqrt{2}},\, \frac{-1-\sqrt{3}}{4\sqrt{2}}) \end{aligned} \tag{22}$$

The Haar transform is of width two, the Daubchies ($D4$) is of width 4. The Haar wavelet generates a sequence of paired differences at varying scales 2^j. In comparison, the Daubechies transform provides a "nonlinear differencing" over sets of four scaled elements, at scales 2^j.

Alternatively, wavelets can be generated by the conjunction of high and low pass filters, termed "filter banks" by Strang and Nguyen (1996) to produce pairs of functions $\Psi(t)$, $\Phi(t)$ that with rescaling yield a basis for the analysis of a function f_t. Unlike the Fourier transform, which uses the sum of certain basis functions (sines and cosines) to represent a given function and may be seen as a decomposition on a frequency-by-frequency basis, the wavelet transform utilizes some elementary functions (father Φ and mother wavelets Ψ) that, being well-localized in both time and scale, provide a decomposition on a "scale-by-scale" basis as well as on a frequency basis. The inner product Φ with respect to f is essentially a low pass filter that produces a moving average; indeed we recognize the filter as a linear time-invariant operator. The corresponding wavelet filter is a high pass filter that produces moving differences (Strang and Nguyen 1996). Separately, the low pass and high pass filters are not invertible, but together they separate the signal into frequency bands, or octaves. Corresponding to the low pass filter there is a continuous time scaling function $\phi(t)$. Corresponding to the high pass filter is a wavelet $w(t)$.

For any set of filters that satisfy the following conditions

$$\sum_{l=0}^{L-1} h_l = 0 \tag{23}$$

$$\sum_{l=0}^{L-1} h_l^2 = 0 \tag{24}$$

$$\sum_{l=0}^{L-1} h_l h_{l+2n} = 0 \tag{25}$$

defines a wavelet function and so is both necessary and sufficient for the analysis of a function "f". However, this requirement is insufficient for defining the synthesis for a function "f". To achieve synthesis, one must add the constraint that:

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\psi(\hat{\omega})|^2}{\omega} d\omega < \infty \tag{26}$$

see Chui (1992).

This gives wavelets a distinct advantage over a purely frequency domain analysis. Because Fourier analysis presumes that any sample is an independent drawing, Fourier analysis requires "covariance stationarity", whereas wavelet analysis may analyze both stationary and long term non-stationary signals. This approach provides a convenient way to represent complex signals. Expressed differently, spectral decomposition methods perform a global analysis, whereas wavelet methods act

locally in both frequency and time. Fourier analysis can relax local non-stationarity by windowing the time series as was indicated above. The problem with this approach is that the efficacy of this approach depends critically on making the right choice of window and, more importantly, presuming its constancy over time.

Any pair of linear filters that meets the following criteria can represent a wavelet transformation (Percival and Walden 2000). Equation (23) gives the necessary conditions for an operator to be a wavelet: h_l denotes the high pass filter, and the corresponding low pass filter is given by:

$$\begin{aligned} g_l &= (-1)^{l+1} h_{L-l-1} \\ &\quad or \\ h_l &= (-1)^l g_{L-l-1} \end{aligned} \tag{27}$$

Equation (27) indicates that the filter bank depends on both the lowpass and high pass filters. Recall the high pass filter for the Daubechies $D(4)$, see Eq. (22), the corresponding low pass filter is:

$$g_0 = -h_3,\ g_1 = h_1,\ g_2 = -h_1,\ g_3 = h_0 \tag{28}$$

For wavelet analysis however, as we have observed, there are two basic wavelet functions, father and mother wavelets, $\phi(t)$ and $\psi(t)$. The former integrates to 1 and reconstructs the smooth part of the signal (low frequency), while the latter integrates to 0 and can capture all deviations from the trend The mother wavelets, as said above, play a role similar to sines and cosines in the Fourier decomposition. They are compressed or dilated, in the time domain, to generate cycles to fit actual data. The approximating wavelet functions $\phi_{J,k}(t)$ and $\psi_{J,k}(t)$ are generated from father and mother wavelets through scaling and translation as follows:

$$\phi_{J,k}(t) = 2^{-\frac{J}{2}} \phi\left(\frac{t - 2^J k}{2^J}\right) \tag{29}$$

and

$$\psi_{J,k}(t) = 2^{-\frac{J}{2}} \psi\left(\frac{t - 2^J k}{2^J}\right) \tag{30}$$

where j indexes the scale, so that 2^j is a measure of the scale, or width, of the functions (scale or dilation factor), and k indexes the translation, so that $2^j k$ is the translation parameter.

Given a signal $f(t)$, the wavelet series coefficients, representing the projections of the time series onto the basis generated by the chosen family of wavelets, are given by the following integrals:

$$\begin{aligned} d_{j,k} &= \int \psi_{j,k}(t) f(t)\, dt \\ s_{J,k} &= \int \phi_{J,k}(t) f(t)\, dt \end{aligned} \tag{31}$$

where $j = 1, 2, \ldots, J$ is the number of scales and the coefficients d_{jk} and s_{Jk} are the *wavelet transform coefficients* representing, respectively, the projection onto mother and father wavelets. In particular, the detail coefficients $d_{Jk}, \ldots., d_{2k}, d_{1k}$ represent progressively finer scale deviations from the smooth behavior (thus capturing the higher frequency oscillations), while the smooth coefficients s_{Jk} correspond to the smooth behavior of the data at the coarse scale 2^J (thus capturing the low frequency oscillations).

Finally, given these wavelet coefficients, from the functions

$$S_{J,k} = \sum_k s_{J,k}\phi_{J,k}(t) \text{ and } D_{j,k} = \sum_k d_{J,k}\psi_{J,k}(t) \tag{32}$$

we may obtain what are called the *smooth* signal, $S_{J,k}$, and the *detail* signals, $D_{j,k}$, respectively. The sequence of terms $S_J, D_J, .., D_j, \ldots, D_1$ for $j = 1, 2, \ldots, J$ represents a set of signal components that provide representations of the original signal $f(t)$ at different scales and at an increasingly finer resolution level.

It is very useful to view the use of wavelets in "regression analysis" in greater generality than as a simple exercise in "least squares fitting". As indicated above the use of wavelets involves the properties of the implicit filters used in the construction of the wavelet function. Such an approach to determining the properties of wavelet analysis provides for a structured, but highly flexible, system that is characterized by a "scarce transformation matrix"; that is, most coefficients in the transformed space are zero. Indeed, the source of the benefit from creating a spanning set of basis vectors, both for Fourier analysis and wavelets, is the reduction in degrees of freedom from N, in the given Euclidean space, to K in the transformed space, where K is very much smaller than N; simple linear regression models illustrate the same situation and perform a similar transformation.

The argument so far, has compared wavelets to splines and to Fourier series or integrals. A discussion of the differences is required. Splines are easily dealt with in that the approximations implied by the spline procedure is to interpolate smoothly a sequence of observations from a smooth differential signal. The analysis is strictly local, even though most spline algorithms average over the whole sample space. The fit is almost entirely determined by the observed data points, so that, little structure is imposed on the process. What structure is predetermined is generated by the position of the knots.

Fourier series, or Fourier integrals, are strictly global over time or space, notwithstanding the use of windows to obtain useful local estimates of the coefficients. Wavelets, however, can provide a mixture of local and global characteristics of the signal, and are easily modified to incorporate restrictions of the signal over time or space. Wavelets generalize Fourier integrals and series in that each frequency band, or octave, groups together, frequencies separated by the supports at each scale. A research analyst can incorporate the equivalent of a windowed analysis of Fourier integrals and incorporate time scale variations as in Ramsey and Zhang (1996, 1997). Further, as illustrated by cosine wave packets (Bruce and Gao 1996), and the

wide choice for low and high pass filters (Strang and Nguyen 1996), considerable detail can be captured, or suppressed, and basic oscillations can be incorporated using band pass filters to generate oscillatory wavelets.

3 Some Examples of the Use of Wavelets

While it is well recognized that wavelets have not been as widely used in Economics as in other disciplines, I hope to show that there is great scope for remedying the situation. The main issue involves the gain in insight to be stimulated by using wavelets; quite literally, the use of wavelets encourages researchers to generalize their conception of the problem at hand.

3.1 Foreign Exchange and Waveform Dictionaries

A very general approach using time frequency atoms is especially useful in analyzing financial markets. Consider the equation $g_\gamma(t)$ where $\gamma = (s, u, \xi)$:

$$g_\gamma(t) = \frac{1}{\sqrt{s}} g(\frac{t-u}{s}) e^{i\xi t} \tag{33}$$

We impose the conditions $||g|| = 1$ where $||g||$ is L^2 and $g(0) \neq 0$. For any scale parameter s, frequency modulation ξ and translation parameter u: the factor $1/\sqrt{s}$ normalizes the norm of $g(t)$ to 1; $g(t)$ is centered at the abscissa u and its energy is concentrated in the neighborhood of u, size is proportional to s; the Fourier transform is centered at the frequency ξ and its energy is concentrated in the neighborhood of ξ and size is proportional to $1/s$. Matching pursuit was used to determine the values of the coefficients; i.e. the procedure picks the coefficients with the greatest contribution to the variation in the function being analyzed. Raw tick by tick data on three foreign exchange rates were obtained from October 1, 1992 to September 30, 1993 (see Ramsey and Zhang 1997). The waveform analysis indicates that there is efficiency of structure, but only at the lowest frequencies equivalent to periods of 2 h with little power. There are some low frequencies that wax and wane in intensity. Most of the energy of the system seems to be in the localized energy frequency bursts.

The frequency bursts provide insights into market behavior. One can view the dominant market reaction to news as a sequence of short bursts of intense activity that are represented by narrow bands of high frequencies. For example, only the first one hundred structures provide a good fit to the data at all but the highest frequencies. Nevertheless the isolated bursts are themselves unpredictable.

The potential for the observable frequencies to wax and wane militates against use of the Fourier approach. Further, the series most likely is a sequence of observations on a continuous, but nowhere differential process. Further analysis is needed to consider the optimal basis generating function.

3.2 Instrumental Variables and "Errors in the Variables"

To begin the discussion on the "errors in variables" problem one notes that the approaches are as unstructured as they have always been; that is, we endeavor to search for a strong instrumental variable, but have no ability to recognize one even if considered. Further, it is as difficult to recognize a weak instrument that if used would yield worse results. I have labeled this approach "solution by assumption" since one has in fact no idea if a putative variable is, or is not, a useful instrumental variable.

Wavelets can resolve the issue: see Ramsey et al. (2010), Gençay and Gradojevic (2011), Gallegati and Ramsey (2012) for an extensive discussion of this critical problem. The task is simple: use wavelets to decompose the observed series into a "noise" component and a structural component, possibly refined by thresholding the coefficient estimates (Ramsey et al. 2010). The benefits from recognizing the insights to be gained from this approach are only belatedly coming to be realized. Suppose all the variables in a system of equations can be factored into a structural component, itself decomposable into a growth term, an oscillation term and into a noise term, e.g.

$$\begin{aligned} y_i &= y_i^* + \varepsilon_i \\ x_i &= x^* + \eta_i \\ z_i &= z_i^* + \omega_i \end{aligned}$$

where the starred terms are structural and the terms $\varepsilon_i, \eta_i, \omega_i$ are random variables, either modeled as simple pulses or have a far more complex stochastic structure, including having distributions that are functions of the structural terms. If we wish to study the structure of the relationships between the variables, we can easily do so (see Silverman 2000; Johnstone 2000). In particular, we can query the covariance between the random error terms, select suitable instrumental variables, solve the simultaneous equation problem, and deal effectively with persistent series.

Using some simulation exercises Ramsey et al. (2010) demonstrated how the structural components revealed by the wavelet analysis yield nearly ideal instrumental variables for variables observed with error and for co-endogenous variables in simultaneous equation models. Indeed, the comparison of the outcomes with current standard procedures indicates that as the nonparametric approximation to the structural component improves, so does the convergence of the near structural estimates.

While I have posed the situation in terms of linear regression, the benefits of this approach are far greater for non-linear relationships. The analysis of Donoho and Johnstone (1995) indicates that asymptotic convergence will yield acceptable results and convergence is swift.

3.3 Structural Breaks and Outlier Detection

Most economic and financial time series evolve in a nonlinear fashion over time, are non-stationary and their frequency characteristics are often time-dependent, that is, the importance of the various frequency components is unlikely to remain stable over time. Since these processes exhibit quite complicated patterns like abrupt changes, jumps, outliers and volatility clustering, a locally adaptive filter like the wavelet transform is particularly well suited for evaluation of such models.

An example of the potential role to be played by wavelets is provided by the detection and location of outliers and structural breaks. Indeed, wavelets can provide a deeper understanding of structural breaks with respect to standard classical analysis given their ability to identify the scale as well as the time period at which the inhomogeneity occurs. Specifically, based on two main properties of the discrete wavelet transform (DWT), i.e. the energy preservation and approximate decorrelation properties, a wavelet-based test for homogeneity of variance (see Whitcher 1998; Whitcher et al. 2002) can be used for detecting and localizing regime shifts and discontinuous changes in the variance.

Similarly, structural changes in economic relationships can be usefully detected by the presence of shifts in their phase relationship. Indeed, although a standard assumption in economics is that the delay between variables is fixed, Ramsey and Lampart (1998a,b) have shown that the phase relationship (and thus the lead/lag relationship) may well be scale dependent and vary continuously over time. Therefore examining "scale-by-scale" overlaid graphs between pairs of variables can provide interesting insights into the nature of the relationship between these variables and their evolution over time (Ramsey 2002). A recent example of this approach is provided in Gallegati and Ramsey (2013) where the analysis of such variations in phase is proven to be useful for detecting and interpreting structural changes in the form of smooth changes in the q-relationship proposed by Tobin. To consider an extreme example, suppose that the economy was composed entirely of discrete jumps, the only suitable wavelet would be based on the Haar function. Less restrictive is the assumption, that the process is continuous, except for a finite number of discontinuities. The analysis can proceed in two stages; first isolate the discontinuities using Haar wavelets, next analyze the remaining data using an appropriate contiguous wavelet generating function.

Finally, wavelets provide a natural way to seek for outliers in that wavelets allow for local distributions at all scales and outliers are at the very least a "local" phenomenon [see for a very brief introduction (Wei et al. 2006; Greenblatt 1996)]. The idea of thresholding (Bruce and Gao 1996; Nason 2008), is that the noise

component is highly irregular, but with a modest amplitude of variation, which is dominated by the variation of the structural component. Naively, outliers are observations drawn from a different distribution; intuitively one tends to consider observations for which the modulus squared is very large relative to the modulus of the remainder of the time series, or cross-sectional data. But outliers may be generated in far more subtle ways and not necessarily reveal themselves in terms of a single large modulus, but in terms of a temporary shift in the stochastic structure of the error terms. In these cases "thresholding, in particular soft thresholding" (Bruce and Gao 1996; Nason 2008), will prove to be very useful especially in separating the coefficient values of "structural components" from noise contamination.

3.4 Time Scale Relationships

The separation of aggregate data into different time scale components by wavelets can provide considerable insights into the analysis of economic relationships between variables. Indeed, economics is an example of a discipline in which time scale matters. Consider, for example, traders operating in the market for securities: some, the fundamentalists, may have a very long view and trade looking at firm or market fundamentals; some others, the chartists, may operate with a time horizon of weeks or days. A corollary of this assumption is that different planning horizons are likely to affect the structure of the relationships themselves, so that such relationships might vary over different time horizons or hold at several time scales, but not at others.

Although the concepts of the "short-run" and of the "long-run" are central for modeling economic and financial decisions, variations in the relationship across time scales are seldom discussed in economics and finance. We should begin by recognizing that for each variable postulated by the underlying theory we admit the possibility that:

$$y_s = g_s(y_{j,s}\, x_{i,s}),$$

where y_s is the dependent variable at scale s, $g_s(.)$ are arbitrary functions specified by the theory, which might differ across scales, $y_{j,s}$ represents the codependent variables at scale s, and $x_{i,s}$ represents exogenous variables x_i at scale s; that is, the relationships between economic variables may well be scale dependent.

Following Ramsey and Lampart (1998a,b) many authors have confirmed that allowing for different time scales of variation in the data can provide a fruitful understanding of the complex dynamics of economic relationships among variables with non-stationary or transient component variables. For example, relationships that are veiled when estimated at the aggregate level, may be consistently revealed after allowing for a decomposition of the variables into different time scales. In general, the results indicate that by using wavelet analysis it is possible to uncover relationships that are at best puzzling using standard regression methods

and that ignoring time and frequency dependence between variables when analyzing relationships in economics and finance can lead to erroneous conclusions.

3.5 Comments on Forecasting

The standard concerns about forecasting carry over to the use of wavelets, but, as might have been anticipated, wavelets incorporate a degree of refinement and flexibility not available using conventional methods (see, for example, Diebold 1998). With wavelets, one can choose the scale at which the forecast is to be made, treating each scale level as a separate series for forecasting purposes. Secondly, one should note that at any given point in time the "forecast" will depend on the scales at which one wishes to evaluate the forecast; for example, at all scales for a point in time, t_0, or for a subset of scales at time t_0. Further, one might well choose to consider, at a given minimum scale whether to forecast a range, given the chosen minimum scale, or to forecast a point estimate at time t_0.

These comments indicate a potentially fruitful line of research and indicates that the idea of "forecasting" is more subtle than has been recognized so far. Forecasts need to be expressed conditional on the relevant scales, and the usual forecasts are special cases of a general procedure. Indeed, one concern that is ignored in the conventional approach is to recognize across scales the composition of the variance involved in term of the variances at each scale level. For examples, see Gallegati et al. (2013), Yousefi et al. (2005), Greenblatt (1996). Linking forecasts to the underlying scale indicates an important development in the understanding of the information generated by wavelets. There is not a single forecast at time t_{0+h} made at time t_0, but a forecast at each relevant time scale.

3.6 Some Miscellaneous Examples

Fan and Gencay (2010) have explored the gain in efficiency in discovering unit roots and applying tests for cointegration using wavelet procedures. Further, using MODWT multi-resolution techniques the authors demonstrate a significant gain in power against near unit root processes. In addition, the wavelet approach leads to a novel interpretation of Von Neumann variance ratio tests.

Gallegati et al. (2009, 2011) reviewed the literature on the "wage Phillips curve" using U.S. data. The most significant result of the multiscale analysis is the long run one to one relationship between wage and price inflation and the close relationship between nominal changes and unemployment rate at business cycle scales. Over all, the paper suggests that allowing for different time scales of variation in the data can provide a richer understanding of the complex dynamics of economic relationships between variables. Relationships that are puzzling when tested using standard methods can be consistently estimated and structural properties revealed

using timescale analysis. The authors note with some humor that Phillips himself can be considered as the first user of wavelets in Economics!

One of the most cogent rationalizations for the use of wavelets and timescale analysis is that different agents operate at different timescales. In particular, one might examine the behavior of central banks to elucidate their objectives in the short and long run. This is done in Aguiar-Conraria and Soares (2008) in assessing the relationship between central bank decision-making and government decision-making. The authors confirm that the macro relationships have changed and evolved over time.

In Rua and Nunes (2009), Rua (2010) interesting results are obtained which concentrate on the role of wavelets in the analysis of the co-movement between international stock markets. In addition, the authors generalize the notion of co-movement across both time and frequency. In Samia et al. (2009), a wavelet approach is taken in assessing values for VaR's and compared favorably to the conventional ARMA-GARCH processes.

4 Conclusions and Recommendations

The functional representation of regression functions projected onto basis spaces was elucidated. The first step began with standard Euclidean N space and demonstrated a relationship to Taylor's series approximations, monomials, exponential and power bases. Fourier series were used to illustrate the relationship to wavelet analysis in that both versions included a concept of rescaling a fundamental function to provide a basis. Spline bases were also defined and related to wavelets. In the discussion and development of wavelets a number of aspects not normally considered were discussed and the concept of atoms was introduced. One can characterize the research analyst's objective as seeking to obtain the best M atoms for a given $f(t)$ out of a dictionary of P atoms. The overall objective is to choose a good basis which depends upon the resolution of two characteristics; linear independence and completeness. Independence guarantees uniqueness of representation and completeness ensures that any $f(t)$ is represented. Adding vectors will destroy independence, removing vectors will destroy completeness. The generality of wavelet analysis is enhanced by the choices available of functional forms to suit specific characteristics of the vector space in which the function resides; for example Haar, Gaussian, Gaussian first derivative, Mexican Hat and so on. In addition, further generalization of the approximation provided by wavelets is illustrated in terms of the Waveform dictionary which uses a triplet of parameters to represent translation, scaling, and is centered around a fundamental frequency $e^{i\xi t}$.

While the discussion above has demonstrated the wide usefulness of the wavelet approach, one might speculate that many more insights are liable to occur as the implications of this unique space are explored. Not enough attention has yet been expended on the wide variation in the formation of wavelet forms and their application in practical problems. In short, attention may well be concentrated in the

future on capturing variation within the function's supports and thereby providing alternative determinations of very short run behavior. The implied flexibility of wavelets provides deconvolution of very short run phenomena as well as the medium run and long run phenomena.

The paper also contains brief reviews of a variety of applications of wavelets to economic examples which are of considerable importance to economists interested in accurate evaluations of policy variables. A wide variety of data sources have been examined, including both macroeconomic and financial data (Bloomfield 1976).

In these models the problem of errors in the variables is critical, but wavelets provide the key to resolving the issue. Some papers examine data for structural breaks and outliers. Comments on forecasting were presented. These thoughts indicate that forecasting is more subtle than is currently believed in that forecasts require to be calculated conditional on the scales involved in the forecast. Some forecasts might well involve only a particular subset of the time scales included in the entire system.

References

Aguiar-Conraria L, Soares MJ (2008) Using wavelets to decompose the time-frequency effects of monetary policy. Phys A 387:2863–2878

Bloomfield R (1976) Fourier analysis of time series: an introduction. Wiley, New York

Bruce A, Gao H (1996) Applied wavelet analysis with S-Plus. Springer, New York

Chui CK (1992) An introduction to wavelets. Academic Press, San Diego

Crowley P (2007) A guide to wavelets for economists. J Econ Surv 21:207–267

Daubechies I (1992) Ten lectures on wavelets. Society for Industrial and Applied Mathematics, Philadephia

Diebold FX (1998) Elements of forecasting. South-Western, Cincinnati

Donoho DL, Johnstone IM (1995) Adapting to unknown smoothness via wavelet shrinkage. J Am Stat Assoc 90:1200–1224

Fan Y, Gencay R (2010) Unit root tests with wavelets. Econ Theory 26:1305–1331

Gallegati M, Gallegati M, Ramsey JB, Semmler W (2009) The US wage Phillips curve over different time horizons. Giornale degli Economisti e Annali di Economia 68:113–148

Gallegati M, Gallegati M, Ramsey JB, Semmler W (2011) The US wage Phillips curve across frequencies and over time. Oxf Bull Econ Stat 73:489–508

Gallegati M, Ramsey JB (2012) Errors-in-variables and the wavelet multiresolution approximation approach: a Monte Carlo study. In: Badi H, Baltagi BH, Hill RC, Newey WK, White HL (eds) Essays in honor of Jerry Hausman. Advances in econometrics, Vol 29. Emerald Group Publishing Limited, Bingley, pp 149–171

Gallegati M, Ramsey JB (2013) Structural change and phase variation: a re-examination of the q-model using wavelet exploratory analysis. Struct Change Econ Dyn 25:60–73

Gallegati M, Ramsey JB, Semmler W (2013) Time scale analysis of interest rate spreads and output using wavelets. Axioms 2:182–207

Gençay R, Selçuk S, Whitcher BJ (2002) An introduction to wavelets and other filtering methods in finance and economics. San Diego Academic Press, San Diego

Gençay R, Gradojevic G (2011) Errors-in-variables estimation with wavelets. J Stat Comput Simul 81:1545–1564

Greenblatt SA (1996) Wavelets in econometrics: an application to outlier testing. In Gilli M (ed) Computational economic systems: models, methods, and econometrics. Advances in computational economics. Kluwer Academic Publishers, Dordrecht, pp 139–160

Johnstone IM (2000) Wavelets and the theory of non-parametric function estimation. In: Silverman BW, Vassilicos JC (eds) Wavelets: the key to intermittent information. Oxford University Press, Oxford, pp 89–110

Kendall MG, Stuart A (1961) The advanced theory of statistics, vol 2. Griffin and Co., London

Korner TW (1988) Fourier analysis. Cambridge University Press, Cambridge

Nason GP (2008) Wavelet methods in statistics with R. Springer, New York

Percival DB, Walden AT (2000) Wavelet methods for time series analysis. Cambridge University Press, Cambridge

Ramsey JB, Zhang Z (1996) The application of waveform dictionaries to stock market index data. In Kravtsov YA, Kadtke J (eds) Predictability of complex dynamical systems. Springer, New York, pp 189–208

Ramsey JB, Zhang Z (1997) The analysis of foreign exchange data using waveform dictionaries. J Empir Financ 4:341–372

Ramsey JB, Lampart C (1998a) The decomposition of economic relationship by time scale using wavelets: money and income. Macroecon Dyn Econ 2:49–71

Ramsey JB, Lampart C (1998b) The decomposition of economic relationship by time scale using wavelets: expenditure and income. Stud Nonlinear Dyn Econ 3:23–42

Ramsey JB (2002) Wavelets in economics and finance: past and future. Stud Nonlinear Dyn Econ 6:1–29.

Ramsey JB (2010) Wavelets. In: Durlauf SN, Blume LE (eds) The new Palgrave dictionary of economics. Palgrave Macmillan, Basingstoke, pp 391–398

Ramsey JB, Gallegati Marco, Gallegati Mauro, Semmler W (2010) Instrumental variables and wavelet decomposition. Econ Model 27:1498–1513

Rua A, Nunes L (2009) International comovement of stock market returns: a wavelet analysis. J Empir Financ 16:632–639

Rua A (2010) Measuring comovement in the time-frequency space. J Macroecon 32:685–691

Samia M, Dalenda M, Saoussen A (2009) Accuracy and conservatism of V aR models: a wavelet decomposed VaR approach versus standard ARMA–GARCH method. Int J Econ Financ 1:174–184

Silverman BW (2000) Wavelets in statistics: beyond the standard assumptions. In: Silverman BW, Vassilicos JC (ed) Wavelets: the key to intermittent information. Oxford University Press, Oxford, pp 71–88

Strang G, Nguyen T (1996) Wavelets and filter banks. Wellesley-Cambridge Press, Wellesley

Yousefi S, Weinreich I, Reinarz D (2005) Wavelet-based prediction of oil prices. Chaos Solitons Fractals 25:265–275

Wei Z, Bo L, Zhang XT, Xiong X, Kou Y (2006) Application of the wavelet based multi-fractal for outlier detection in financial high frequency time series data. In: IEEE International Conference on Engineering of Intelligent Systems, pp 420–425

Whitcher BJ (1998) Assessing nonstationary time series using wavelets. PhD Thesis, University of Washington

Whitcher BJ, Byers SD, Guttorp P, Percival DB (2002) Testing for homogeneity of variance in time series: long memory. Wavelets and the Nile river. Water Resour Res 38:1054–1070

Part I
Macroeconomics

Does Productivity Affect Unemployment? A Time-Frequency Analysis for the US

Marco Gallegati, Mauro Gallegati, James B. Ramsey, and Willi Semmler

Abstract The effect of increased productivity on unemployment has long been disputed both theoretically and empirically. Although economists mostly agree on the long run positive effects of labor productivity, there is still much disagreement over the issue as to whether productivity growth is good or bad for employment in the short run. Does productivity growth increase or reduce unemployment? This paper try to answer this question by using the property of wavelet analysis to decompose economic time series into their time scale components, each associated to a specific frequency range. We decompose the relevant US time series data in different time scale components and consider co-movements of productivity and unemployment over different time horizons. In a nutshell, we conclude that, according to US post-war data, productivity creates unemployment in the short and medium terms, but employment in the long run.

1 Introduction

Productivity growth is recognized as a major force to increase the overall performance of the economy, as measured for example by the growth of output, real wages, and cost reduction, and a major source of the observed increases in the standard of

M. Gallegati (✉) • M. Gallegati
DISES and SIEC, Polytechnic University of Marche, Ancona, Italy
e-mail: marco.gallegati@univpm.it; mauro.gallegati@univpm.it

J.B. Ramsey
Department of Economics, New York University, New York, NY, USA
e-mail: james.ramsey@nyu.edu

W. Semmler
Department of Economics, New School for Social Research, New York, NY, USA
e-mail: semmlerw@newschool.edu

M. Gallegati and W. Semmler (eds.), *Wavelet Applications in Economics and Finance*, Dynamic Modeling and Econometrics in Economics and Finance 20,
DOI 10.1007/978-3-319-07061-2_2,

living (Landes 1969). Economists in the past, from Ricardo to Schumpeter to Hicks, have explored the phenomenon of whether new technology and productivity in fact increase unemployment. The relationship between productivity and employment is also very important in the theoretical approach followed by the mainstream models: Real Business Cycle (RBC) and DSGE. In particular, RBC theorists have postulated technology shocks as the main driving force of business cycles. In RBC models technology shocks, either to output and employment (measured as hours worked) are predicted to be positively correlated.[1] This claim has been made the focus of numerous econometric studies.[2] Employing the Blanchard and Quah (1989) methodology Gali (1999), Gali and Rabanal (2005), Francis and Ramey (2005) and Basu et al. (2006) find a negative correlation between employment and productivity growth, once the technology shocks have been purified taking out demand shocks affecting output.

Although economists mostly agree on the long run positive effects of labor productivity, significant disagreements arise over the issue as to whether productivity growth is good or bad for employment in the short run. Empirical results have been mixed (e.g. in Muscatelli and Tirelli 2001, where the relationship between productivity growth and unemployment is negative for several G7 countries and not significant for others) and postulate a possible trade-off between employment and productivity growth (Gordon 1997). Such empirical findings have been also complicated by the contrasting evidence emerging during the 1990s between the US and Europe as to the relationship between (un)employment and productivity growth. Whereas the increase in productivity growth in the US in the second half of the 1990s is associated with low and falling unemployment (Staiger et al. 2001), in Europe the opposite tendency was visible. Productivity growth appears to have increased unemployment.

The labor market provides an example of a market where the strategies used by the agents involved, firms and workers (through unions), can differ by time scale. Thus, the "true" economic relationships among variables can be found at the disaggregated (scale) level rather than at the usual aggregate level. As a matter of fact, aggregate data can be considered the result of a time scale aggregation procedure over all time scales and aggregate estimates a mixture of relationships across time scales, with the consequence that the effect of each regressor tends to be mitigated by this averaging over all time scales.[3] Blanchard et al. (1995) were the first ones to hint at such a research agenda. They stressed that it may be useful to distinguish between the short, medium and long-run effects of productivity growth, as the effects of productivity growth on unemployment may show different

[1] See the volume by Cooley (1995), and see also Backus and Kehoe (1992) among the others.

[2] For details of the evaluations, see Gong and Semmler (2006, ch.6).

[3] For example in Gallegati et al. (2011) where wavelet analysis is applied to the wage Phillips curve for the US.

co-movements depending on the time scales.[4] Similar thoughts are also reported in Solow (2000) with respect to the different ability of alternative theoretical macroeconomic frameworks to explain the behavior of an economy at the aggregate level in relation to their specific time frames[5] and, more recently, the idea that time scales can be relevant in this context has also been expressed by Landmann (2004).[6]

Following these insights, studies are now emerging arguing that researchers need to disentangle the short and long-term effects of changes in productivity growth for unemployment. For example, Tripier (2006), studying the co-movement of productivity and hours worked at different frequency components through spectral analysis, finds that co-movements between productivity and unemployment are negative in the short and long run, but positive over the business cycle.[7] This paper is related to the above mentioned literature by focussing on the relationship of unemployment and productivity growth at different frequency ranges. Indeed, wavelets with respect to other filtering methods are able to decompose macroeconomic time series, and data in general, into several components, each with a resolution matched to its scale. After the first applications of wavelet analysis in economics and finance provided by Ramsey and his co-authors (1995; 1996; 1998a; 1998b), the number of wavelet applications in economics has been rapidly growing in the last few years as a result of the interesting opportunities provided by wavelets in order to study economic relationships at different time scales.[8]

The objective of this paper is to provide evidence on the nature of the time scale relationship between labor productivity growth and the unemployment rate using wavelet analysis, so as to provide a new challenging theoretical framework, new empirical results as well as policy implications. First, we perform wavelet-based exploratory analysis by applying the continuous wavelet transform (CWT) since tools such as wavelet power, coherency and phase can reveal interesting features

[4]Most of the attention of economic researchers who work on productivity has been devoted to measurement issues and to resolve the problem of data consistency, as there are many different approaches to the measurement of productivity linked to the choice of data, notably the combination of employment, hours worked and GDP (see for example the OECD Productivity Manual, 2001).

[5]"At short term scales, I think, something sort of Keynesian is a good approximation, and surely better than anything straight neoclassical. At very long scales, the interesting questions are best studied in a neoclassical framework.... At the 5–10 years time scale, we have to piece things together as best as we can, and look for an hybrid model that will do the job" (Solow 2000, p. 156).

[6]"The nature of the mechanism that link [unemployment and productivity growth] changes with the time frame adopted" because one needs "to distinguish between an analysis of the forces shaping long-term equilibrium paths of output, employment and productivity on the one hand and the forces causing temporary deviations from these equilibrium paths on the other hand" (Landmann 2004, p. 35).

[7]Qualitative similar results are also provided using time domain techniques separating long-run trends from short run phenomena.

[8]For example Gençay et al. (2005), Gençay et al. (2010), Kim and In (2005), Fernandez (2005), Crowley and Mayes (2008), Gallegati (2008), Ramsey et al. (2010), Gallegati et al. (2011).

about the structure of a process as well as information about the time-frequency dependencies between two time series. Hence, after decomposing both variables into their time-scale components using to the maximum overlap discrete wavelet transform (MODWT), we analyze the relationship between labor productivity and unemployment at the different time scales using parametric and nonparametric approaches. The results indicate that in the medium-run, at business cycle frequency, there is a positive relationship of productivity and unemployment, whereas in the long-run we can observe a negative co-movement, that is productivity creates employment.

The paper proceeds as follows. In Sect. 2 a wavelet-based exploratory analysis is performed by applying several CWT tools to labor productivity growth and the unemployment rate. In Sect. 3, we analyze the "scale-by-scale" relationships between productivity growth and unemployment by means of parametric and nonparametric approaches. Section 4 provides interpretation of results according to alternative labor market theories and Sect. 5 concludes the paper.

2 Continuous Wavelet Transforms

The essential characteristics of wavelets are best illustrated through the development of the continuous wavelet transform (CWT).[9] We seek functions $\psi(u)$ such that:

$$\int \psi(u)\, du = 0 \tag{1}$$

$$\int \psi(u)^2\, du = 1 \tag{2}$$

The cosine function is a "large wave" because its square does not converge to 1, even though its integral is zero; a wavelet, a "small wave" obeys both constraints. An example would be the Haar wavelet function:

$$\psi^H(u) = \begin{cases} -\frac{1}{\sqrt{2}} & -1 < u < 0 \\ \frac{1}{\sqrt{2}} & 0 < u < 1 \\ 0 & otherwise \end{cases} \tag{3}$$

Such a function provides information about the variation of a function, $f(t)$, by examining the differences over time of partial sums. As will be illustrated below

[9] Wavelets, their generation, and their potential use are discussed in intuitive terms in Ramsey (2010), while Gençay et al. (2001) generate an excellent development of wavelet analysis and provide many interesting economic examples. Percival and Walden (2000) provide a more technical exposition with many examples of the use of wavelets in a variety of fields, but not in economics.

general classes of wavelet functions compare the differences of weighted averages of the function $f(t)$. Consider a signal, $x(u)$ and the corresponding "average":

$$\frac{1}{b-a}\int_a^b x(u)\,du = \alpha(a,b) \tag{4}$$

Let us choose the convention that we assess the value of the "average" at the center of the interval and let $\lambda \equiv b-a$ represent the scale of the partial sums. We have the expression:

$$\begin{aligned} A(\lambda,t) &\equiv \alpha(t-\lambda/2,\ t+\lambda/2) \\ &= \frac{1}{\lambda}\int_{t-\lambda/2}^{t+\lambda/2} x(u)\,du \end{aligned} \tag{5}$$

A(λ,t) is the average value of the signal centered at "t" with scale λ. But what is of more use is to examine the differences at different values for λ and at different values for "t". We define:

$$\begin{aligned} D(\lambda,t) &= A(\lambda,t+\lambda/2) - A(\lambda,t-\lambda/2) \\ &= \frac{1}{\lambda}\int_t^{t+\lambda} x(u)\,du - \frac{1}{\lambda}\int_{t-\lambda}^{t} x(u)\,du \end{aligned} \tag{6}$$

This is the basis for the continuous wavelet transform, CWT, as defined by the Haar wavelet function. For an arbitrary wavelet function, $W(\lambda\ t)$, the wavelet transform, ψ is:

$$\begin{aligned} W(\lambda,\ t) &= \int_{-\infty}^{\infty} \psi_{\lambda,\,t}(u)x(u)\,du \\ \psi_{\lambda,\,t}(u) &\equiv \frac{1}{\sqrt{\lambda}}\left(\frac{u-t}{\lambda}\right) \end{aligned} \tag{7}$$

where λ is a scaling or dilation factor that controls the length of the wavelet and t a location parameter that indicates where the wavelet is centered (see Percival and Walden 2000).

2.1 Wavelet Power Spectrum

Let $W_x(\lambda,t)$ be the continuous wavelet transform of a signal $x(.)$, $|W_x|^2$ represents the wavelet power and can be interpreted as the energy density of the signal in the time-frequency plane. Among the several types of wavelet families available, that is Morlet, Mexican hat, Haar, Daubechies, etc., the Morlet wavelet is the most widely

used because of its optimal joint time frequency concentration. The Morlet wavelet is a complex wavelet that produces complex transforms and thus can provide us with information on both amplitude and phase. It is defined as

$$\psi_\eta(t) = \pi^{-\frac{1}{4}} e^{i\omega_0\eta} - e^{-\frac{\eta^2}{2}}. \tag{8}$$

where $-1/4$ is a normalization term, $\eta = t/\lambda$ is the dimensionless time parameter, t is the time parameter, λ is the scale of the wavelet. The Morlet coefficient ω_0 governs the balance between time and frequency resolution. We use the value $\omega_0 = 6$ since this particular choice provides a good balance between time and frequency localization (see Grinsted et al. 2004) and also simplifies the interpretation of the wavelet analysis because the wavelet scale, λ, is inversely related to the frequency, $f \approx 1/\lambda$.

Plots of the wavelet power spectrum provide evidence of potentially interesting structures, like dominant scales of variation in the data or "characteristic scales" according to the definition of Keim and Percival (2010).[10] Since estimated wavelet power spectra are biased in favor of large scales, the bias rectification proposed by Liu et al. (2007) is applied, where the wavelet power spectrum is divided by the scale coefficient so that it becomes physically consistent and unbiased. Specifically, the adjusted wavelet power spectrum is obtained by dividing the power at each point in the spectrum by the corresponding scale based on the energy definition (the transform coefficient squared is divided by the scale it associates). This allows for a comparison of the spectral peaks across scales.

Time is recorded on the horizontal axis and the vertical axis gives us the periods and the corresponding scales of the wavelet transform. Reading across the graph at a given value for the wavelet scaling, one sees how the power of the projection varies across the time domain at a given scale. Reading down the graph at a given point in time, one sees how the power varies with the scaling of the wavelet (see Ramsey et al. 1995). A black contour line testing the wavelet power 5 % significance level against the null hypothesis that the data generating process is generated by a stationary process is displayed,[11] as is the cone of influence represented by a shaded area corresponding to the region affected by edge effects.[12]

[10]The CWT has been computed using the MatLab package developed by Grinsted et al. (2004). MatLab programs for performing the bias-rectified wavelet power spectrum and partial wavelet coherence are provided by Ng and Kwok at http://www.cityu.edu.hk/gcacic/wavelet.

[11]The statistical significance of the results obtained through wavelet power analysis was first assessed by Torrence and Compo (1998) by deriving the empirical (chi-squared) distribution for the local wavelet power spectrum of a white or red noise signal using Monte Carlo simulation analysis.

[12]As with other types of transforms, the CWT applied to a finite length time series inevitably suffers from border distortions; this is due to the fact that the values of the transform at the beginning and the end of the time series are always incorrectly computed, in the sense that they involve missing values of the series which are then artificially prescribed; the most common choices are zero padding extension of the time series by zeros or periodization. Since the effective support of

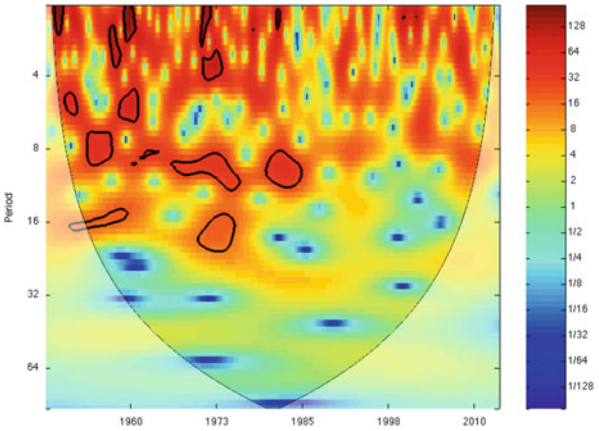


Fig. 1 Rectified wavelet power spectrum plots for labor productivity growth. Note: contours and a cone of influence are added for significance. A *black contour line* testing the wavelet power 5 % significance level against a *white* noise null is displayed as is the cone of influence, represented by a *shaded area* corresponding to the region affected by edge effects

In Figs. 1 and 2 we report estimated wavelet spectra for labor productivity growth and the unemployment rate, respectively.[13] The comparison between the power spectra of the two variables reveals important differences as to their characteristic features. In the case of labor productivity growth there is evidence of highly localized patterns at lower scales, with high power regions concentrated in the first part of the sample (until late eighties). By contrast, for the unemployment rate significant power regions are evident at scales corresponding to business cycle frequencies throughout the sample.

Although useful for revealing potentially interesting features in the data like "characteristic scales", the wavelet power spectrum is not the best tool to deal with the time-frequency dependencies between two time-series. Indeed, even if two variables share similar high power regions, one cannot infer that their comovements look alike.

---

the wavelet at scale λ is proportional to λ, these edge effects also increase with λ. The region in which the transform suffers from these edge effects is called the cone of influence. In this area of the time-frequency plane the results are unreliable and have to be interpreted carefully (see Percival and Walden 2000).

[13]We use quarterly data for the US between 1948:1 and 2013:4 from the Bureau of Labor Statistics. Labor productivity is defined as output per hour of all persons in the Nonfarm Business Sector, Index 1992 = 100, and transformed into its growth rate as $400 * ln(x_t / x_{t-1})$. Unemployment rate is defined as percent Civilian Unemployment Rate.

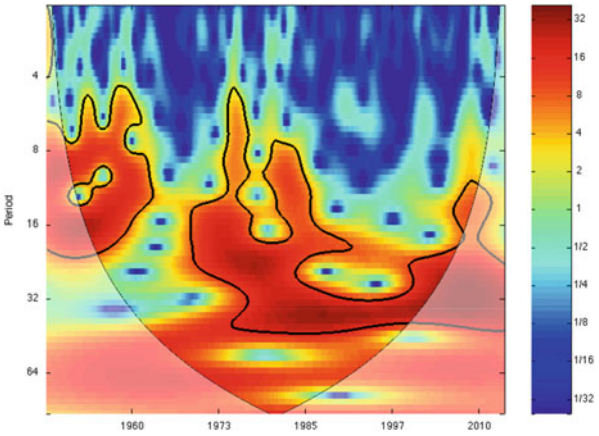


Fig. 2 Rectified wavelet power spectrum for the unemployment rate. Note: see Table 1

2.2 Wavelet Coherence

In order to detect and quantify relationships between variables suitable wavelet tools are the cross-wavelet power, wavelet coherence and wavelet phase difference. Let W_x and W_y be the continuous wavelet transform of the signals $x(.)$ and $y(.)$, the cross-wavelet power of the two series is given by $|W_{xy}|=|W_x W_y|$ and depicts the local covariance of the two time series at each scale and frequency (see Hudgins et al. 1993). The wavelet coherence is defined as the modulus of the wavelet cross spectrum normalized to the single wavelet spectra and is especially useful in highlighting the time and frequency intervals where two phenomena have strong interactions. It can be considered as the local correlation between two time series in time frequency space. The statistical significance level of the wavelet coherence is estimated using Monte Carlo methods. The 5 % significance level against the null hypothesis of red noise is shown as a thick black contour. The cone of influence is marked by a black thin line: again, values outside the cone of influence should be interpreted very carefully, as they result from a significant contribution of zero padding at the beginning and the end of the time series.

Complex-valued wavelets like Morlet wavelet have the ability to provide the phase information, that is a local measure of the phase delay between two time series as a function of both time and frequency. The phase information is coded by the arrow orientation. Following the trigonometric convention the direction of arrows shows the relative phasing of the two time series and can be interpreted as indicating a lead/lag relationship: right (left) arrow means that the two variables are in phase (anti-phase). If the arrows points to the right and up, it means the unemployment rate is lagging. If they points to the right and down, unemployment rate is leading. If the arrows are to the left and up, it means unemployment rate is leading and if they are to the left and down, unemployment rate is lagging. The relative phase

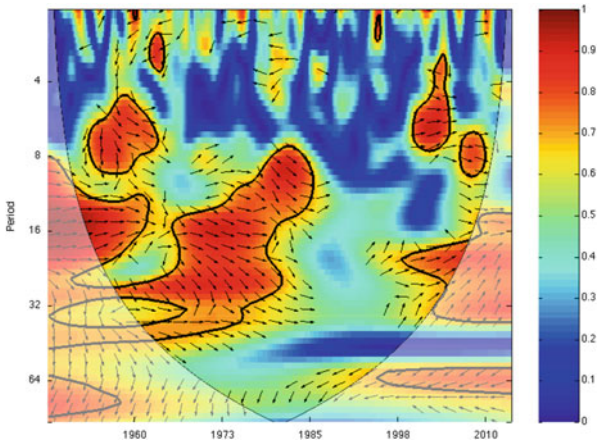


Fig. 3 Wavelet coherence between the unemployment rate and productivity growth. The color code for power ranges from *blue* (low coherence) to *red* (high coherence). A pointwise significance test is performed against an almost process independent background spectrum. 95 % confidence intervals for the null hypothesis that coherency is zero are plotted as contours in *black* in the figure. The cone of influence is marked by *black lines* (Color figure online)

information is graphically displayed on the same figure with wavelet coherence by plotting such arrows inside and close to regions characterized by high coherence, so that the coherence and the phase relationship are shown simultaneously.

In Fig. 3 regions of strong coherence between productivity and unemployment are evident at business cycle scales, i.e. at scales corresponding to periods between 2 and 8-years, except for the mid 1980s–mid 1990s period where no relationship is evident at any scale. The analysis of the phase difference reveals an interesting difference in the phase relationship of the two variables. If at scales corresponding to business cycle frequencies the two series are generally in phase, the low frequency region of the wavelet coherence reveals the presence of an anti-phase relationship between productivity and unemployment.

3 Discrete Wavelet Transform

So far we have considered only continuously labeled decompositions. Nonetheless there are several difficulties with the CWT. First, it is computationally impossible to analyze a signal using all wavelet coefficients. Second, as noted, $W(\lambda, t)$ is a function of two parameters and as such contains a high amount of redundant information. As a consequence, although the CWT provides a useful tool for analyzing how the different periodic components of a time series evolve over time, both individually (wavelet power spectrum) and jointly (wavelet coherence and phase-difference), in practice a discrete analogs of this transform is developed.

We therefore move to the discussion of the discrete wavelet transform (DWT), since the DWT, and in particular the MODWT, a variant of the DWT, is largely predominant in economic applications.[14]

The DWT is based on similar concepts as the CWT, but is more parsimonious in its use of data. In order to implement the discrete wavelet transform on sampled signals we need to discretize the transform over scale and over time through the dilation and location parameters. Indeed, the key difference between the CWT and the DWT lies in the fact that the DWT uses only a limited number of translated and dilated versions of the mother wavelet to decompose the original signal. The idea is to select t and λ so that the information contained in the signal can be summarized in a minimum number of wavelet coefficients. The discretized transform is known as the discrete wavelet transform, DWT.

The discretization of the continuous time-frequency decomposition creates a discrete version of the wavelet power spectrum in which the entire time-frequency plane is partitioned with rectangular cells of varying dimensions but constant area, called Heisenberg cells (e.g. in Fig. 4).[15] Higher frequencies can be well localized in time, but the uncertainty in frequency localization increases as the frequency increases, which is reflected as taller, thinner cells with increase in frequency. Consequently, the frequency axis is partitioned finely only near low frequencies. The implication of this is that the larger-scale features of the signal get well resolved in the frequency domain, but there is a large uncertainty associated with their location. On the other hand, the small-scale features, such as sharp discontinuities, get well resolved in the time domain, even if there is a large uncertainty associated with their frequency content. This trade-off is an inherent limitation due to the Heisenberg's uncertainty principle that states that the resolution in time and frequency cannot be arbitrarily small because their product is lower bounded. Therefore, owing to the uncertainty principle, an increased resolution in the time domain for the time localization of high-frequency components comes at a cost of an increased uncertainty in the frequency localization, that is one can only trade time resolution for frequency resolution, or vice versa.

The general formulation for a continuous wavelet transform can be restricted to the definition of the "discrete wavelet transform", the properties of which can be summarized by the equation:

$$\psi_{j,k}(t) = 2^{-j/2}\psi\left(\frac{t - 2^j k}{2^j}\right) \tag{9}$$

[14]The number of the papers applying the DWT is far greater than those using the CWT. As a matter of fact, the preference for DWT in economic applications can be explained by the ability of the DWT to facilitate a more direct comparison with standard econometric tools than is permitted by the CWT, e.g. time scales regression analysis, homogeneity test for variance, nonparametric analysis.

[15]Their dimensions change according to their scale: the windows stretch for large values of λ to measure the low frequency movements and compress for small values of λ to measure the high frequency movements.

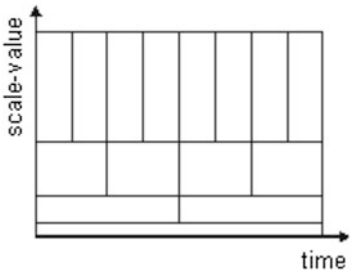


Fig. 4 DWT time-scale partition

which is known as the "mother wavelet". This function represents a sequence of rescaleable functions at a scale of $\lambda = 2^j, j = 1, 2, \ldots J$, and with time index $k, k = 1, 2, 3, \ldots N/2^j$. The wavelet transform coefficient of the projection of the observed function $f(t)$ for $i = 1, 2, 3, \ldots N, N = 2^J$ on the wavelet $\psi_{j,k}(t)$ is given by:

$$d_{j,k} \approx \int \psi_{j,k}(t) f(t)\, dt,$$
$$j = 1, 2, ..J \tag{10}$$

For a complete reconstruction of a signal f(t), one requires a scaling function, $\phi(.)$, that represents the smoothest components of the signal. While the wavelet coefficients represent weighted "differences" at each scale, the scaling coefficients represent averaging at each scale. One defines the scaling function, also know as the "father wavelet", by:

$$\phi_{J,k}(t) = 2^{-J/2} \phi \left(\frac{t - 2^J k}{2^J} \right) \tag{11}$$

And the scaling function coefficients vector is given by:

$$s_{J,k} \approx \int \phi_{J,k}(t) f(t)\, dt, \tag{12}$$

By construction, we have an orthonormal set of basis functions, whose detailed properties depend on the choices made for the functions, $\phi(.)$ and $\psi(.)$, see for example the references cited above as well as Daubechies (1992) and Silverman (1999). At each scale, the entire real line is approximated by a sequence of "non-overlapping" wavelets. The deconstruction of the function $f(t)$ is therefore:

$$f(t) \approx \sum_k s_{J,k} \phi_{J,k}(t) + \sum_k d_{J,k} \psi_{J,k}(t) +$$
$$\sum_k d_{J-1,k} \psi_{J-1,k}(t) + \ldots + \sum_k d_{1,k} \psi_{1,k}(t) \tag{13}$$

The above equation is an example of the Discrete Wavelet Transform, DWT based on an arbitrary wavelet function, $\phi(.)$. Using economic variables, the degree of relative error is approximately on the order of 10^{-13} in many cases, so that one can reasonably claim that the wavelet decomposition is very good. While it would appear that wavelets involve large numbers of coefficients, it is also true that the number of coefficients greater than zero is very small; the arrays are said to be "sparse". In the literature quite complicated functions are approximated to a high level of accuracy with a surprisingly small number of coefficients. As a corollary to this general statement, other scholars have noted the extent to which the distribution of coefficients under the null hypothesis of zero effect, rapidly approaches the Gaussian distribution.

Further, the approximation can be re-written in terms of collections of coefficients at given scales. Define;

$$\begin{aligned} S_J &= \sum_k s_{J,k}\phi_{J,k}(t) \\ D_J &= \sum_k d_{J,k}\psi_{J,k}(t) \\ D_{J-1} &= \sum_k d_{J-1,k}\psi_{J-1,k}(t) \\ &\cdots\cdots \\ D_1 &= \sum_k d_{1,k}\psi_{1,k}(t) \end{aligned} \tag{14}$$

Thus, the approximating equation can be restated in terms of coefficient crystals as:

$$f(t) \approx S_J + D_J + D_{J-1} + \ldots D_2 + D_1 \tag{15}$$

where S_J contains the "smooth component" of the signal, and the D_j, $j = 1, 2, ..J$, the detail signal components at ever increasing levels of detail. S_J provides the large scale road map, D_1 shows the pot holes. The previous equation indicates what is termed the multiresolution decomposition, MRD.

3.1 Time Scale Decomposition Analysis

The orthonormal discrete wavelet transform (DWT), even if widely applied to time series analysis in many disciplines, has two main drawbacks: (1) the dyadic length requirement (i.e. a sample size divisible by 2^J), and (2) the wavelet and scaling coefficients are not shift invariant. Because of the practical limitations of DWT wavelet analysis is generally performed by applying the *maximal overlap discrete*

wavelet transform (MODWT), a non-orthogonal variant of the classical discrete wavelet transform (DWT) that, unlike the DWT, is translation invariant, as shifts in the signal do not change the pattern of coefficients, can be applied to data sets of length not divisible by 2^J and provides at each scale a number of coefficients equal to the length of the original series.

For our analysis we select the Daubechies least asymmetric (LA) wavelet filter of length $L = 8$ based on eight non-zero coefficients (Daubechies 1992), with reflecting boundary conditions, and apply the MODWT up to a level $J = 5$ that produces one vector of smooth coefficients s_5, representing the underlying smooth behavior of the data at the coarse scale, and five vectors of details coefficients d_5, d_4, d_3, d_2, d_1, representing progressively finer scale deviations from the smooth behavior. Through the synthesis, or reconstruction, operation we can reassemble the original signal from the wavelet and scaling coefficients using the inverse stationary wavelet transform.[16] Specifically, with $J = 5$ we reconstruct five wavelet details vectors D_5, D_4, D_3, D_2, D_1 and one wavelet smooth vector, S_5, each associated with a particular time scale 2^{J-1}. In particular, since we use quarterly data the first detail level D_1 captures oscillations between 2 and 4 quarters, while details D_2, D_3, D_4 and D_5 capture oscillations with a period of 1–2, 2–4, 4–8 and 8–16 years, respectively.[17]

The smooth and detail components obtained from the reconstruction process take the form of non-periodic oscillating waves representing the long-term trend and the deviations from it at an increasing level of detail. According to Ramsey (2002) the visual inspection between pairs of variables provides an excellent exploratory tool for discovering time varying delays or phase variations between variables. Indeed, by examining the phase relationship in a bivariate context we can obtain useful insights on the timing (lagging, synchro or leading) of the linkage between variables as well as on the existence of a fixed or changing relationship.[18]

In Fig. 5 we plot the smooth and detail components, i.e. S_5, D_5, D_4 and D_3, as a sequence of pairs where the unemployment rate (dotted lines) is plotted against labor productivity growth (solid lines). The visual inspection of the long-run components indicate an anti-phase relationship between variables, with productivity growth slightly leading the unemployment rate.[19] The pattern displayed by the top right

[16]Since the J components obtained by the application of MODWT are not orthogonal, they do not sum up to the original variable.

[17]Detail levels D_1 and D_2, represent the very short-run dynamics of a signal (and contains most of the noise of the signal), levels D_3 and D_4 roughly correspond to the standard business cycle time period (Stock and Watson 1999), while the medium-run component is associated to level D_5. Finally, the smooth component S_5 captures oscillations with a period longer than 16 years corresponding to the low-frequency components of a signal.

[18]Although a standard assumption in economics is that the delay between variables is fixed, the phase relationship may well be scale dependent and vary continuously over time (e.g. in Ramsey and Lampart 1998a,b; Gallegati and Ramsey 2013).

[19]This leading behavior is consistent with the findings reported in the previous section using wavelet coherence.

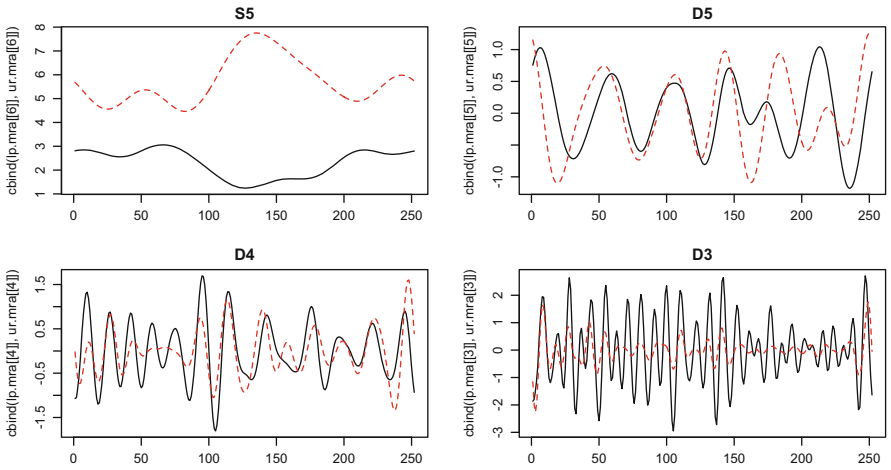


Fig. 5 Phase shift relationships of smooth and detail components for unemployment (*dotted lines*) and productivity (*solid lines*)

panel in Fig. 5 reveals that the two components are mostly in phase at the D_5 scale level, with unemployment slightly leading productivity growth. Nonetheless, the plot also shows that the two series at this level have been moving into antiphase at the beginning of the 1990s, as a consequence of a shift in the phase relationship, and then have been moving in-phase again in the last part of the sample. At the D_4 scale level unemployment and productivity are in-phase throughout the sample with the exception of the 1960s. Finally, at the lower scale levels productivity growth and unemployment rate components show very different amplitude fluctuations. This pattern suggests how a well known feature of aggregate productivity growth quarterly data, that is its very high volatility, can be ascribed to high frequency components.

3.2 Parametric Analysis

Wavelets provide a unique tool for the analysis of economic relationships on a scale-by-scale basis. Time scale regression analysis allows the researcher to examine the relationship between variables at each j scale where the variation in both variables has been restricted to the indicated specific scale. In order to perform a time scale regression analysis first we need to partition each variable into a set of different components by using the discrete wavelet transform (DWT), such that each component corresponds to a particular range of frequencies, and then run regression analysis on a scale-by-scale basis (e.g. Ramsey and Lampart 1998a,b; Kim and

Table 1 Aggregate and time scale regression analysis (1948:1–2013:4)—OLS

$ur_t = \alpha + \beta \Delta lp_t + \epsilon_t$							
	Aggregate	S_5	D_5	D_4	D_3	D_2	D_1
α_j	**5.7694**	**9.8850**	−5.13e−09	7.41e−10	−3.46e−09	−5.29e−10	2.26e−10
	(0.2248)	(0.2773)	(0.0616)	(0.0530)	(0.0259)	(0.0079)	(0.0018)
β_j	0.0257	**−1.8614**	**0.6862**	**0.5217**	**0.1902**	**0.0285**	**−0.0092**
	(0.0316)	(0.1166)	(0.1496)	(0.0557)	(0.0217)	(0.0101)	(0.0022)
$\overline{R}^2$	0.0027	0.8077	0.2375	0.4247	0.3777	0.0450	0.0735
S.E.	1.6646	0.5359	0.4606	0.4197	0.2671	0.1586	0.0704

Note: *HAC* standard errors in parenthesis, *S.E.* is the regression standard error
Regressors significant at 5 % in bold

In 2005; Gallegati et al. 2011).[20] Therefore, after decomposing the regression variables into their different time scale components using the MOWDT we estimate a sequence of least squares regressions using

$$ur[S_J]_t = \alpha_J + \beta_J lp[S_J]_t + \epsilon_t \tag{16}$$

and

$$ur[D_j]_t = \alpha_j + \beta_j lp[D_j]_t + \epsilon_t \tag{17}$$

where $ur[S_J]_t$, and $lp[S_J]_t$ represent the components of the variables at the longest scale, and $ur[D_j]_t$, and $lp[D_j]_t$ represent the components of the variables at each scale j, with $j = 1, 2, \ldots, J$.

In Table 1 we present the results from least squares estimates at the aggregate and individual scale levels. First of all, we notice that although at the aggregate level the relationship between productivity and unemployment is not significant, the "scale-by-scale" regressions reveal a positive significant relationship at almost each scale level and that the effects of productivity on unemployment rate differ widely across scales in terms of sign and estimated size effect. Specifically, if at scales D_1 and D_2 the estimated size effect of productivity growth on unemployment is negligible, at business cycles and medium run scales, i.e. from D_3 to D_5, the size and significance of the estimated coefficients indicate a positive relationship that is higher for the D_4 scale level. Finally, long run trends are negatively related. A 1 % fall in the long-run productivity growth rate increases the unemployment rate by 1.86 %.[21]

[20]Thus, we test for frequency dependence of the regression parameter by using timescale regression analysis since the approaches used to detect and model frequency dependence such as spectral regression approaches (Hannan 1963; Engle 1974, 1978) present several shortcomings because of their use of the Fourier transformation. For examples of the use of this procedure in economics, see Ramsey and Lampart (1998a,b), Gallegati et al. (2011).

[21]This estimated magnitude of the impact of growth on unemployment is in line with those obtained in previous studies. For example, Pissarides and Vallanti (2007) a panel of OECD countries estimate that a 1 % decline in the growth rate leads to a 1.3–1.5 % increase in unemployment.

This finding is not new. A negative link between unemployment and productivity growth at low frequencies is also documented in Staiger et al. (2001), Ball and Moffitt (2002), where the trending behavior of productivity growth is called for in the explanation of low and falling inflation combined with low unemployment experienced by the US during the second half of the 1990s, as well as in Muscatelli and Tirelli (2001) for several G7 countries. Similar results have been also obtained in Tripier (2006), Chen et al. (2007) using different methods. In the first by using measures of co-movements in the frequency domain it is shown that co-movements between variables differ strongly according to the frequency, that is negative in the short and long run, but positive over the business cycle. In the latter, the authors, disaggregating data into their short and long-term components and using two different econometric methods (Maximum Likelihoof and structural VAR), find that productivity growth affects unemployment positively in the short run and negatively in the long run.[22]

In sum, when we consider different time frames we find that the effects of productivity growth on unemployment are frequency-dependent: in the long run an increase in productivity releases forces that stimulate innovation and growth in the economy and thus determine a reduction of unemployment, whereas at intermediate and business cycle time scales productivity gains cause unemployment to increase.

3.3 Nonparametric Analysis

In this section we apply a methodology that allows us to explore the robustness of the issues related to the relationship between productivity growth and unemployment without making any a priory explicit or implicit assumption about the form of the relationship: nonparametric regression analysis. Indeed, nonparametric regressions can capture the shape of a relationship without us prejudging the issue, as they estimate the regression function $f(.)$ linking the dependent to the independent variables directly.[23]

There are several approaches available to estimate nonparametric regression models,[24] and most of these methods assume that the nonlinear functions of the independent variables to be estimated by the procedures are *smooth* continuous functions. One such model is the locally weighted polynomial regression pioneered

[22]Recently, a negative long-run relationship between productivity growth and unemployment has also been obtained by Schreiber (2009) using a co-breaking approach and Miyamoto and Takahashi (2011) using band-pass filtering.

[23]The traditional nonlinear regression model introduce nonlinear functions of dependent variables using a limited range of transformed variables to the model (quadratic terms, cubic terms or piecewise constant function). An example of a methodology testing for nonlinearity without imposing any a priory assumption about the shape of the relationship is the smooth transition regression used in Eliasson (2001).

[24]See Fox (2000a,b) for a discussion on nonparametric regression methods.

by Cleveland (1979). This procedure fits the model $y = f(x_1, \ldots, x_k) + \epsilon$ nonparametrically, that is without assuming a parametric form for $f(x_1, \ldots, x_k)$. The low-degree polynomial, generally first or second degree (that is, either locally linear or locally quadratic), is fit using weighted least squares, so that the data points are weighted by a smooth function whose weights decrease as the distance from the center of the window increases. The value of the regression function is obtained by evaluating the local polynomial at each particular value of the independent variable, x_i. A fixed proportion of the data is included in each given local neighborhood, called the *span* of the local regression smoother (or the smoothing parameter)[25] and the fitted values are then connected in a nonparametric regression curve.

In Fig. 6 we report the scatter plots of the productivity growth-unemployment relationship at the different scale levels, from S_5 (top left panel) to D_1 (top right panel). In each panel of Fig. 6 a solid line drawn by connecting the points of the fitted values for each function against its regressor is superimposed on each scatter plot. The smooth plots represented by the solid lines depict the loess fit using a smoothing parameter value of $2/3$.[26] These lines can be used to reveal the shape of the estimated relationship between the dependent (unemployment rate) and the independent variable (labor productivity).

The loess fits shown on the plots in Fig. 6 support the conclusions obtained from the parametric results reported in Table 1. In particular, the shape of the nonparametric fitted regression function suggests a negative long-run relationship between labor productivity and unemployment. By contrast, a positive relationship is evident at lower wavelet scales, especially at the frequency band corresponding to periods of 2–8 years. To summarize, we find that unemployment is positively associated with productivity in the short and medium term, but negatively in the long term.

4 Interpretation

Notwithstanding the question of how productivity growth affects unemployment has received much attention in the recent literature, the theoretical approach is far from being uniform (e.g. in search and matching theories of the labor market in Pissarides (1990), Aghion and Howitt (1994), Mortensen and Pissarides (1998)). Theoretical predictions of the impact of productivity growth on unemployment depend on the quantitative importance of the "capitalization" and "creative destruction" effects

[25]The smoothing parameter controls the flexibility of the loess regression function: large values of produce the smoothest functions that wiggle the least in response to fluctuations in the data, the smaller the smoothing parameter is, the closer the regression function will conform to the data

[26]We use different smoothing parameters, but our main findings do not show excess sensitivity to the choice of the span in the loess function within what appear to be reasonable ranges of smoothness (i.e. between 0.4 and 0.8).

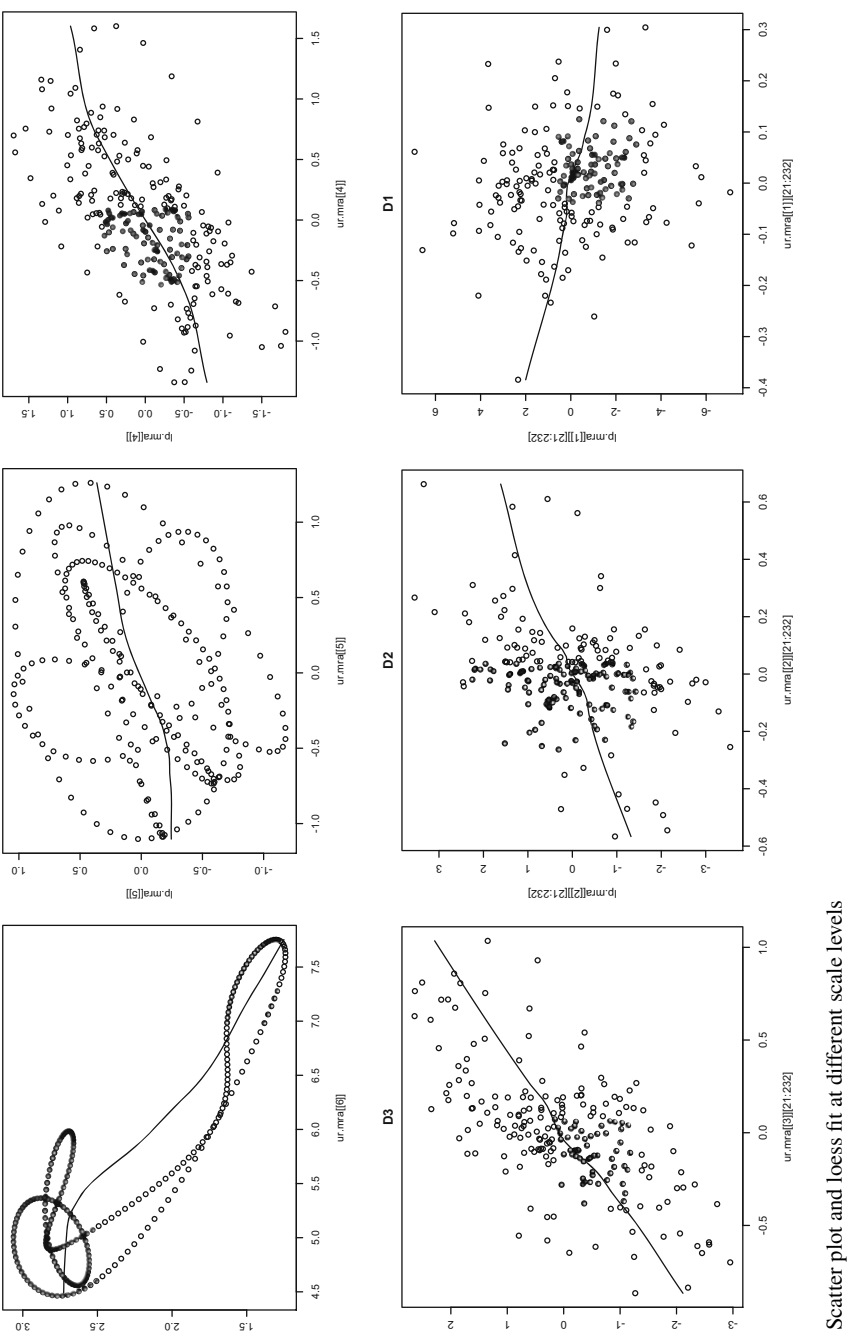


Fig. 6 Scatter plot and loess fit at different scale levels

which, in turn, reflect the extent to which the two forms of technical change discussed in this literature, that is embodied and disembodied technology,[27] are embodied in production factors.

In the model with disembodied technological progress it is suggested that higher productivity growth reduces the long run unemployment rate through the so called "capitalization effect" (Pissarides 1990, 2000). By contrast, in the model with embodied technological progress, faster technical change increases long run unemployment through a "creative destruction effect" (Aghion and Howitt 1994, 1998; Postel-Vinay 2002). Inconsistency between these findings is resolved in Mortensen and Pissarides (1998) by building up a matching model with embodied technical progress in which both types of effects, that is "capitalization" and "creative destruction", can be obtained depending on "whether new technology can be introduced into ongoing jobs, or it needs to be embodied in new job creation" (Pissarides and Vallanti 2007). As a result, whether the overall impact of productivity growth on unemployment is positive or negative is assumed to depend upon the relative strength of the "capitalization" and "creative destruction" effects.

What effect is likely to prevail is a question that can be addressed by considering the different time horizon of "capitalization" and "creative destruction" effects, and their associated effects on job creation and jobs destruction, respectively. The time horizon of job creation can be radically different from that of job destruction. Indeed, firm's time horizon when creating jobs can be very long, and definitely much longer than firm's horizon when destructing jobs, since job creation involves computing the expected present discounted value of future profits from new jobs. As a consequence, we can expect that the relevance of the capitalization effect as to the creative destruction effect (and the net effect of productivity growth on employment) could be different across different time horizons since the latter effect induces more job destruction and less job creation than the first one. In particular, we should observe a positive relationship between productivity growth and unemployment if the creative destruction effect dominates over the capitalization effect, and conversely a negative relationship if the capitalization effect dominates.

The empirical evidence provided using wavelet analysis hints that the "creative destruction" effect dominates over the "capitalization" effect at short- to medium term scales, whereas the "capitalization" effect dominates at the longest scale. In this way we can interpret the negative long-run connection between productivity growth and unemployment as consistent with models where technological progress is purely disembodied (see Pissarides and Vallanti 2007) or the positive "capitalization effect"

[27]Embodied technical change is embedded in (new) capital goods or jobs and can benefit only jobs that explicitly invest in new technology. By contrast, disembodied technical change is not tied to any factor of production and can benefit all existing jobs. According to the "capitalization" effect an increase in growth raises the capitalized value of those returns obtained from creating jobs, thereby reducing the equilibrium rate of unemployment by increasing the job-finding rate. The second effect is the creative destruction, according to which an increase in growth raises the equilibrium level of unemployment both directly, by raising the job-separation rate, and indirectly, by discouraging the creation of job vacancies.

of disembodied technological progress dominates the "creative destruction" effect of embodied technology.[28] On the other hand, the positive impact of productivity growth on unemployment at intermediate scales support the "creative destruction" hypothesis of several labour market models in which the "capitalization" effect is too weak to reverse a "creative destruction" effect.[29]

To summarize, we argue that co-movements of productivity and unemployment at short-term scales can be markedly different from those at the longest scale. In particular, our results indicate that this "opposite" relationship displayed by unemployment and productivity growth at different time frames can be determined by the relative strength of the "capitalization" and "creative destruction" effects. All in all, what emerges is a more complex picture of the relationship in which the two effects have different strengths at different time horizons and the aggregate effect is simply the interaction of the relative strength of the two effect at different time horizons.

Furthermore, these results have other relevant economic implications. First of all, as regards the Okun's (1962) law, the US employment seems to be decoupled from economic growth, the so-called "jobless growth". In the US there is a slowly recovering unemployment rate, though the annual growth rates of productivity are higher than in Europe. Due to high productivity growth rates, in the US one can observe some aspect of jobless growth. This might be a short run phenomenon. In the long term this could be turned into a negative relationship of productivity and unemployment.

Finally, as to the controversial hypothesis of the RBC models that employment is rising with positive productivity shocks, the critics (such as Basu et al. 2006) are presumably correct to state a nonsignificant relationship between technology shocks and employment, or even a negative relationship of those variables. So the RBC postulate of a positive relationship between productivity and employment seems to be incorrect in the short and medium run, but in the long run, when productivity growth makes the firms and the country more competitive, the increase in productivity may cause employment to rise.

5 Conclusion

The effect of productivity increases on unemployment is controversial. Economic theories have postulated strong comovements of productivity shocks and employment. Yet, in the 1990s, Europe was seen to suffer from higher growth rates of

[28]These long run effects maybe also based on the sluggishness of real wage adjustments as suggested by models where wage setting depends on backward looking reservation wages (Blanchard and Katz 1999). Results compatible with this evidence are reported in Gallegati et al. (2011) where wages do not adjust fully to productivity changes in the long-run.

[29]Higher productivity growth is often accompanied by structural change wherein "old jobs" are replaced by "new ones" since technology could enhance the demands for new products.

productivity that did not show up in the labor market as higher employment. US is now viewed as suffering from jobless growth, so that the question is whether the low reaction of employment to increases in productivity is a short or long run phenomenon. The issue is therefore how does productivity affect unemployment at different time horizons. Such relationships, and, in particular, the medium and long-run relationships between productivity growth and unemployment are generally analyzed in the empirical literature looking at average aggregate data, generally decades, because from a time series perspectives the rate of growth of labor productivity is a very volatile series whose implications in terms of the movements of the other supply-side variables are difficult to interpret, particularly in the short-run and medium run.[30]

The key to the empirical results obtained in the past is to examine the empirical relationships on a "scale-by-scale" basis. This is because the result is an empirical issue and the outcome depends at each scale on the elasticity of response of demand to price, new products, and/or re-engineered products to the new technology. The results in the short and intermediate run indicate that a reduction of employment is plausible, especially if the elasticity of response of demand to price reductions is unsubstantial. But the opposite seems to be the case for the long run. However, even though the sign of the relationship between employment and productivity may well stay constant over long periods of time, one would expect there to be large differences in the relative magnitudes of the net response over time caused by different market and technology conditions.

In this paper, we use wavelets to analyze the productivity-unemployment relationship over different time frames and demonstrate the usefulness of wavelet analysis in disentangling the short, medium and long run effects of changes in productivity growth for unemployment. In a nutshell, we find a strong negative long run relationship between labor productivity and unemployment, but also a positive significant relationship at lower scales, especially at scales corresponding to business cycle frequency bands. In the medium run, new technology is likely to be labor reducing, and thus adding to unemployment,[31] as was visible in Europe during the 1990s. In the long run, however, new technology replacing labor (process innovation) increase productivity and makes firms and the economy more competitive and may reduce unemployment.[32] Finally, our results suggest some relevant implications concerning the interpretation of search-matching models of unemployment, Okun's law, the RBC hypothesis of a positive co-movement of productivity shocks and employment, and the US employment prospects.

When Thomas More (Utopia 1516) was asserting: sheep are eating men, he was, in the short run, right. Due to agricultural innovations, profits in the primary sector

[30]Indeed, the relationship between productivity and the unemployment rate may appear weaker when we reduce the time period used for aggregating data (see Steindel and Stiroh 2001).

[31]A statement like this goes back to David Ricardo who has pointed out that if machinery is substituted for labor unemployment is likely to increase.

[32]This point is made clear in a simple text book illustration by Blanchard (2005).

were rising, less labor force was employed in agriculture and more lands were devoted to pastureland. People had to "invent" new jobs, i.e. people were stimulated into creating new products that the new technology made possible.

Acknowledgements The paper has been presented at the Workshop on "Frequency domain research in macroeconomics and finance", held at the Bank of Finland, Helsinki, 20–21 October 2011. We thank all participants for valuable comments and suggestions, particularly Jouko Vilmunen and Patrick Crowley.

References

Aghion P, Howitt P (1994) Growth and unemployment. Rev Econ Stud 61:477–94
Aghion P, Howitt P (1998) Endogenous growth theory. MIT Press, Cambridge
Backus DK, Kehoe PJ (1992) International evidence on the historical evidence of business cycles. Am Econ Rev 82:864–888
Ball L, Moffitt R (2002) Productivity growth and the Phillips curve. In: Krueger AB, Solow R (ed) The roaring nineties: Can full employment be sustained? Russell Sage Foundation, New York, pp 61–90
Basu S, Fernald JG, Kimball MS (2006) Are technology improvement contractionary? Am Econ Rev 96:1418–1448
Blanchard OJ (2005) Macroeconomics, 4th edn. Prentice Hall, New Jersey
Blanchard OJ, Quah D (1989) The dynamic effects of aggregate demand and supply disturbances. Am Econ Rev 79:655–673
Blanchard OJ, Katz L (1999) Wage dynamics: reconciling theory and evidence. NBER Working Paper, No. 6924
Blanchard OJ, Solow R, Wilson BA (1995) Productivity and unemployment. MIT Press, unpublished
Chen P, Rezai A, Semmler W (2007) Productivity and Unemployment in the Short and Long Run. SCEPA Working Paper, 2007–8
Cleveland WS (1979) Robust Locally-Weighted Regression and Scatterplot Smoothing. J Am Stat Assoc 74:829–836
Cooley TF (1995) Frontiers of business cycle research. Princeton University Press, Princeton
Crowley PM, Mayes DG (2008) How fused is the euro area core? An evaluation of growth cycle co-movement and synchronization using wavelet analysis. J Bus Cycle Measur Anal 4:76–114
Daubechies I (1992) Ten lectures on wavelets. In: CBSM-NSF regional conference series in applied mathematics. SIAM, Philadelphia
Eliasson AC (2001) Detecting equilibrium correction with smoothly time-varying strength. Stud Nonlinear Dyn Econ 5:Article 2
Engle RF (1974) Band spectrum regression. Int Econ Rev 15:1–11
Engle RF (1978) Testing price equations for stability across spectral frequency bands. Econometrica 46:869–881
Fernandez VP (2005) The international CAPM and a wavelet-based decomposition of value at risk. Stud Nonlinear Dyn Econ 9(4):4
Fox J (2000a) Nonparametric simple regression: smoothing scatterplots. Sage, Thousands Oaks
Fox J (2000b) Multiple and Generalized Nonparametric Regression. Sage, Thousands Oaks CA.
Francis N, Ramey VA (2005) Is the technology-driven real business cycle hypothesis dead? Shocks and aggregate fluctuations revisited. J Monet Econ 52:1379–1399
Gali J (1999) Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations? Am Econ Rev 89:249–271

Gali J, Rabanal P (2005) Technology shocks and aggregate fluctuations: How well does the RBC model fit postwar U.S. data? IMF Working Papers 04/234
Gallegati M (2008) Wavelet analysis of stock returns and aggregate economic activity. Comput Stat Data Anal 52:3061–3074
Gallegati M, Ramsey JB (2013) Structural change and phase variation: A re-examination of the q-model using wavelet exploratory analysis. Struct Change Econ Dyn 25:60–73
Gallegati M, Gallegati M, Ramsey JB, Semmler W (2011) The US wage Phillips curve across frequencies and over time. Oxf Bull Econ Stat 73:489–508
Gençay R, Selçuk F, Whitcher B (2001) An Introduction to Wavelets and Other Filtering Methods in Finance and Economics. San Diego Academic Press, San Diego
Gençay R, Selçuk F, Whitcher B (2005) Multiscale systematic risk. J Int Money Financ 24:55–70
Gençay R, Gradojevic N, Selçuk F, Whitcher B (2010) Asymmetry of information flow between volatilities across time scales. Quant Financ 10:895–915
Gordon RJ (1997) Is there a trade-off between unemployment and productivity growth? In Snower D, de la Dehesa G (ed) Unemployment policy: government options for the labor market. Cambridge University Press, Cambridge, pp 433–463
Gong G, Semmler W (2006) Stochastic dynamic macroeconomics: theory and empirical evidence. Oxford University Press, New York
Grinsted A, Moore JC, Jevrejeva S (2004) Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Processes Geophys 11:561–566
Hannan EJ (1963) Regression for time series with errors of measurement. Biometrika 50:293–302
Hudgins L, Friehe CA, Mayer ME (1993) Wavelet transforms and atmospheric turbulence. Phys Rev Lett 71:3279–3282
Keim MJ, Percival DB (2010) Assessing Characteristic Scales Using Wavelets. arXiv:1007.4169
Kim S, In FH (2005) The relationship between stock returns and inflation: new evidence from wavelet analysis. J Empir Financ 12:435–444
Landes DS (1969) The unbound Prometheus: technological change and industrial development in Western Europe from 1750 to the present. Cambridge University Press, London
Landmann O (2004) Employment, productivity and output growth. In: World Employment Report 2004 International Labour Organization, Geneva
Liu Y, Liang XS, Weisberg RH (2007) Rectification of the bias in the wavelet power spectrum. J Atmos Oceanic Technol 24:2093–2102
Miyamoto H, Takahashi Y (2011) Productivity growth, on-the-job search, and unemployment. Economics & Management Series 2011–06, IUJ Research Institute
Mortensen DT, Pissarides C (1998) Technological progress, job creation and job destruction. Rev Econ Dyn 1:733–753
Muscatelli VA, Tirelli P (2001) Unemployment and growth: some empirical evidence from structural time series models. Appl Econ 33:1083–1088
OECD (2001) Measuring productivity OECD manual. OECD, Paris
Okun A (1962) Potential GNP: Its measurement and significance. In: Proceedings of the business and economic statistics section, American Statistical Association
Percival DB, Walden AT (2000) Wavelet methods for time series analysis. Cambridge University Press, Cambridge
Pissarides C (1990) Equilibrium unemployment theory. Blackwell, Oxford
Pissarides C (2000) Equilibrium unemployment theory, 2nd edn. MIT Press, Cambridge
Pissarides CA, Vallanti G (2007) The impact of TFP growth on steady-state unemployment. Int Econ Rev 48:607–640
Postel-Vinay F (2002) The dynamic of technological unemployment. Int Econ Rev 43:737–60
Ramsey JB (2002) Wavelets in economics and finance: Past and future. Stud Nonlinear Dyn Econ 6:1–29.
Ramsey JB (2010) Wavelets. In: Durlauf SN, Blume LE (ed) The new Palgrave dictionary of economics. Palgrave Macmillan, Basingstoke, pp 391–398
Ramsey JB, Zhang Z (1995) The analysis of foreign exchange data using waveform dictionaries. J Empir Financ 4:341–372

Ramsey JB, Zhang Z (1996) The application of waveform dictionaries to stock market index data. In: Kravtsov YA, Kadtke J (ed) Predictability of complex dynamical systems. Springer, New York, pp 189–208

Ramsey JB, Lampart C (1998a) The decomposition of economic relationship by time scale using wavelets: money and income. Macroecon Dyn Econ 2:49–71

Ramsey JB, Lampart C (1998b) The decomposition of economic relationship by time scale using wavelets: expenditure and income. Stud Nonlinear Dyn Econ 3:23–42

Ramsey JB, Uskinov D, Zaslavsky GM (1995) An analysis of U.S. stock price behavior using wavelets. Fractals 3:377–389

Ramsey JB, Gallegati M, Gallegati M, Semmler W (2010) Instrumental variables and wavelet decomposition. Econ Model 27:1498–1513

Schreiber S (2009) Explaining shifts in the unemployment rate with productivity slowdowns and accelerations: a co-breaking approach. Kiel Working Papers 1505, Kiel Institute for the World Economy

Silverman B (1999) Wavelets in statistics: beyond the standard assumptions. Phil Trans R Soc Lond A 357:2459–2473

Solow RM (2000) Towards a macroeconomics of the medium run. J Econ Perspect 14:151–158

Staiger D, Stock JH, Watson MW (2001) Prices, wages and the U.S. NAIRU in the 1990s. NBER Working Papers no. 8320

Steindel C, Stiroh KJ (2001) Productivity: What Is It, and Why Do We Care About It? Federal Reserve Bank of New York Working Paper

Stock JH, Watson MW (1999) Business cycle fluctuations in US macroeconomic time series. In: Taylor JB, Woodford M (ed) Handbook of macroeconomics. North-Holland, Amsterdam

Torrence C, Compo GP (1998) A practical guide to wavelet analysis. Bull Am Meteorol Soc 79:61–78

Tripier F (2006) Sticky prices, fair wages, and the co-movements of unemployment and labor productivity growth. J Econ Dyn Control 30:2749–2774

The Great Moderation Under the Microscope: Decomposition of Macroeconomic Cycles in US and UK Aggregate Demand

Patrick M. Crowley and Andrew Hughes Hallett

Abstract In this paper the relationship between the growth of real GDP components is explored in the frequency domain using both static and dynamic wavelet analysis. This analysis is carried out separately for both the US and the UK using quarterly data, and the results are found to be substantially different in the two countries. One of the key findings in this research is that the "great moderation" shows up only at certain frequencies, and not in all components of real GDP. We use these results to explain why the incidence of the great moderation has been so patchy across GDP components, countries and time periods. This also explains why it has been so hard to detect periods of moderation (or otherwise) reliably in the aggregate data. We argue it cannot be done without breaking the GDP components down into their frequency components across time and these results show why: the predictions of traditional real business cycle theory often appear not to be upheld in the data.

1 Introduction

The frequency domain offers economists a different perspective for analysis of economic data from time series approaches. The number of contributions in economics that use frequency domain analysis is woefully small, and yet a number of important advances have been made in frequency domain methods which have not yet filtered fully into the economics literature. In particular, there are now methods available which permit simultaneous analysis of economic time series in both the

P.M. Crowley (✉)
Economics Group, College of Business, Texas A&M University, Corpus Christi, TX, USA
e-mail: patrick.crowley@tamucc.edu

A. Hughes Hallett
School of Public Policy, George Mason University, VA, USA

M. Gallegati and W. Semmler (eds.), *Wavelet Applications in Economics and Finance*, Dynamic Modeling and Econometrics in Economics and Finance 20,
DOI 10.1007/978-3-319-07061-2_3,

frequency and time domains, and these techniques are typically referred to as "time-frequency analysis" While time-frequency techniques are still not yet part of the standard toolbox for analysis of time series in economics, these techniques are standard in other disciplines such as engineering, acoustics, neurological sciences, physics, geology and environmental sciences.

The contribution contained in this paper uses discrete wavelet analysis to analyse fluctuations in the components of US and UK growth,[1] and to look at the interactions between the components of US or UK growth over time at different frequencies.

2 Some Basic Correlations

We begin by noting the correlations for the US between the growth rates of the main components of aggregate demand in real GDP, namely personal consumption expenditures, private investment, government expenditures (both current and capital) and export of goods and services. The data is chained real quarterly data from 1948 to end 2012, and growth rates are calculated as year-over-year changes in the logged values of each component.

Using a basic Fisher correlation test (- reported in the table with * referring to significance at the 10 % level, ** = 5 % level and *** = 1 % level) for a null hypothesis of zero correlation, only the correlations between C, I and G are significant. Unsurprisingly, the highest correlation between annual changes in components of US growth is between consumption and investment. Government expenditures appear to be negatively related to both consumption and investment as might be expected due to counter-cyclical fiscal policy. However, although neither of these correlations are that high, both outstrip the contemporaneous correlation of exports with consumption or investment (Table 1).

We now compare these initial correlations with those for the UK in Table 2. The data is from the UK National Statistics Office and is quarterly chained real quarterly data from 1955 to the third quarter of 2012.

Once again not all the reported correlations are significant—those between C and I and X are, but none of the correlations with G are significant. Again, the largest correlation is between consumption and investment; but the size of the correlation is lower than for the US. This time the correlation between C and G is positive if small, indicating a weak pro-cyclical (near a-cyclical) use of government spending, whereas that between G and I is negative. The correlations between X and C and I are all positive, with quite high correlation between X and I in particular. For the UK the correlation between X and G is small, insignificant and negative (Table 3).

We now repeat the same exercise, but for the 1987–2007 period, which corresponds to the period referred to as the "great moderation", and we do this first for the US:

[1] An analysis of fluctuations in real GNP itself has already been undertaken in Crowley (2010).

Table 1 Correlation of US GDP components

	C	I	G	X
C	1	0.629***	−0.123**	0.073
I		1	−0.220***	0.104*
G			1	0.000
X				1

Table 2 Correlation of UK GDP components

	C	I	G	X
C	1	0.589***	0.092	0.149**
I		1	−0.167**	0.309***
G			1	−0.090
X				1

Table 3 Correlation of US GDP components: 1997–2007

	C	I	G	X
C	1	0.664***	−0.050	0.160
I		1	−0.371**	0.551***
G			1	−0.617***
X				1

Table 4 Correlation of UK GDP components: 1997–2007

	C	I	G	X
C	1	−0.158	0.148	−0.024
I		1	−0.376**	0.056
G			1	−0.057
X				1

Apart from the high correlations of C and I, these results are surprising. Correlations between X and all other variables are much higher in this subperiod than for the entire period, with a high correlation between I and X, and a large and significant negative correlation between X and G.

In this table for the UK, we also get markedly different results, but in this case the results are even more surprising. The correlations between C and I, G and X are now not significant, which is a markedly different result from the correlations for the whole dataset. The correlation between I and G is negative and is now higher and significant (Table 4).

Taken together this set of four tables of correlations implies that there is a shifting contemporaneous relationship between the components of GDP for both the US and the UK, which likely has different underlying dynamics for each of the two countries concerned. It also highlights the different dynamics that were in play during the great moderation. The reason why this is the case is unclear, but it clearly merits further investigation. To start with, not only do these simple statistics show that many of these correlations are not significant, but they also ignore two important considerations: (1) that lead or lag relationships may exist between components of GDP which may change our interpretation of the facts (for example: two perfectly correlated variables that are out of phase by half a cycle will show correlation of −1); and (2) that (possibly variable) cycle relationships might be significant between

the constituent components of GDP, which are only weakly related at other non-business cycle lengths.

Clearly simple correlation coefficients are not going to reveal the size or causal direction of these relationships, and more appropriate frequency domain tools are required to explore if any "hidden" relationships exist. Two obvious examples will make the point: (a) two perfectly correlated data series out of phase will yield contemporaneous correlations close to zero or negative; contrast the correlations between C and G which should be in phase, but negative if there is any smoothing, with correlations between C and I which are likely to be out of phase but positively correlated if they are driven by a common cycle. And (b) how strong should we expect the C, I correlations to be? Since C will be subject to short business cycles, and I to business and longer investment cycles, there is likely to be some (positive) correlation—but not that strong, unless I's cycle length is a multiple of that for C. Perhaps this is the reason why we observe a negative and insignificant relationship between C and I in the "great moderation" subperiod for the UK.

To address these considerations we use discrete wavelet analysis to analyze the relationship between the components of GDP, in both the US and UK economies.

3 Rationale and Data

3.1 Rationale

The rationale for looking at cyclical interactions between the major components of output is twofold:

(a) there are obviously some interactions between the components that occur through the business cycles—notably between consumption and investment through inventories and government policies, and between consumption and exports through the international transmission of business cycles. These interactions have important policy implications; and
(b) the real business cycle literature focused on these interactions as justification for technology "shocks" driving fluctuations in the economy and hence the business cycle. A deeper understanding of the interaction between the GDP components may better inform model-building in terms of modelling the transmission of fluctuations or shocks between spending units in the macro-economy.

The latter concern is particularly relevant here. In King et al. (1988) it was first noted that real business cycle models do not reproduce the same variability in the components of output, notably investment and consumption, and much effort has been expended in this literature to attempt to construct models that exhibit the same degree of co-movement in investment and consumption over time (see Christiano and Fitzgerald 1998; Rebelo 2005). One solution to this has been explored in

research which allows for investment-specific technology shocks[2]—as noted in research using New Keynesian Dynamic Stochastic General Equilibrium (DSGE) models by Furlanetto and Seneca (2010), a positive consumption response can be obtained in a standard DSGE model with nominal rigidities when preferences are non-separable in consumption and hours. It suggests that both real business cycle and New Keynesian models have difficulty in generating the empirically observed movements in consumption and investment, and it is here that this research might shed some light on the interaction at different frequencies between consumption, investment and exports.

3.2 Data

The data used is quarterly chain-weighted quarterly real GDP data and it's major components. The US data was sourced from the Bureau of Economic Analysis for 1947Q1 to 2012Q4 (giving 260 datapoints), and the data was transformed by logging the source data and then taking annual differences. The UK data was sourced from the National Statistics Office for 1954Q1 to 2012Q3 (giving 233 datapoints) and is transformed in the same manner. Figure 1 plots the data for the US while Fig. 2 does the same for the UK.[3]

It should be noted that in the recent downturn government spending is still rising, while all the other components of aggregate demand have clearly been falling.

3.3 Discrete Wavelet Analysis

Discrete wavelet analysis uses wavelet filters to extract cycles at different frequencies from the data. It uses a given discrete function which is passed through the series and "convolved"[4] with the data to yield a coefficient, otherwise known as a "crystal". In the basic approach (the discrete wavelet transform or DWT) these data points or crystals will be increasingly sparse for lower frequency (long) cycles if the

[2]These are shocks from new investment which contains new technology rather than investment that either replaces depreciated equipment or just adds to the stock of existing capital without upgrading the technology.

[3]Note that the vertical axes are scaled differently for each component.

[4]In mathematics and, in particular, functional analysis, convolution is a mathematical operation on two functions f and g, producing a third function that is typically viewed as a modified version of one of the original functions. Convolution is similar to cross-correlation. It has applications that include statistics, computer vision, image and signal processing, electrical engineering, and differential equations.

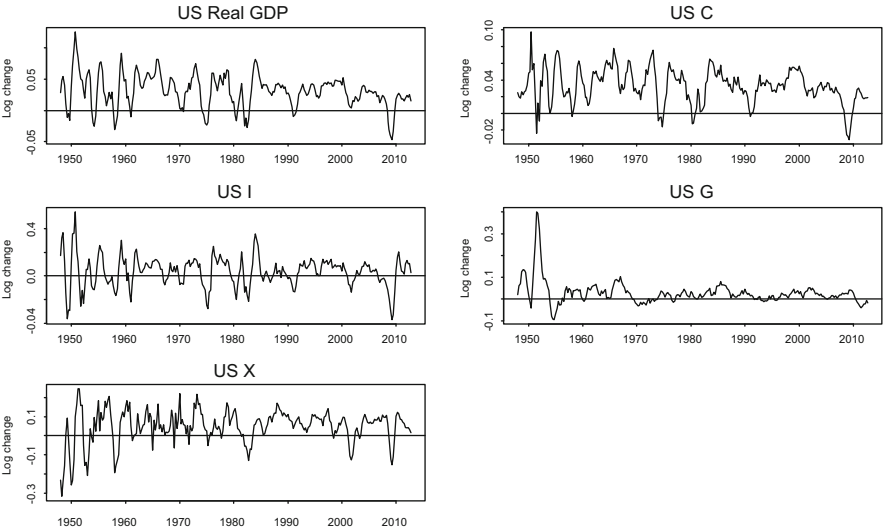


Fig. 1 US GDP and components: log annual change

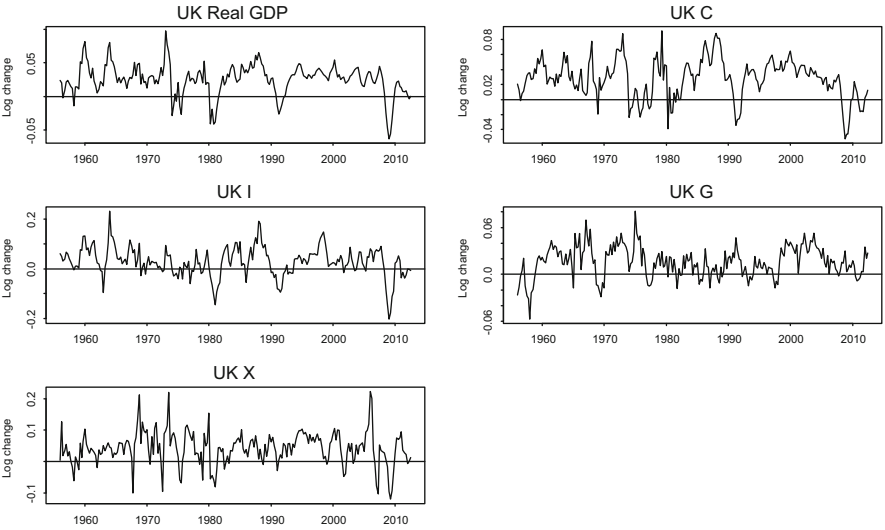


Fig. 2 UK GDP and components: log annual change

wavelet function is applied to the series over consecutive data spans.[5] So another way of obtaining crystals corresponding to all data points in each frequency range is to pass the wavelet function down the series by data observation,[6] rather than moving the whole wavelet function down the series to cover a completely new data span. This is the basis of the maximal overlap discrete wavelet transform (MODWT), and is the technique used here.

As shown in Bruce and Gao (1996), the wavelet coefficients can be approximated by the integrals for father and mother wavelets as:

$$s_{J,k} \approx \int x(t)\phi_{J,k}(t)\,dt \tag{1}$$

$$d_{j,k} \approx \int x(t)\psi_{j,k}(t)\,dt \tag{2}$$

respectively, where $j = 1, 2, \ldots J$ such that J is the maximum scale sustainable with the data to hand, then a multiresolution representation of the signal $x(t)$ is can be given by:

$$x(t) = \sum_k s_{J,k}\phi_{J,k}(t) + \sum_k d_{J,k}\psi_{J,k}(t) + \sum_k d_{J-1,k}\psi_{J-1,k}(t) + \ldots + \sum_k d_{1,k}\psi_{1,k}(t) \tag{3}$$

where the basis functions $\phi_{J,k}(t)$ and $\psi_{J,k}(t)$ are assumed to be orthogonal. The multiresolution decomposition (MRD) of the variable or signal $x(t)$ is then defined by the set of "crystals" or coefficients:

$$\{s_J, d_J, d_{J-1}, \ldots d_1\} \tag{4}$$

The interpretation of the MRD using the DWT is of interest as it relates to the frequency at which activity in the time series occurs.[7] For example with a quarterly time series Table 5 shows the frequencies captured by each scale crystal.

Note that as quarterly data is used in the present study, to capture the conventional business cycle length scale, crystals need to be obtained for five scales. This requires at least 64 observations. But to properly resolve at the lowest frequency it would help to have 128 observations, and as we have at least 214 observations for all 8 series this

[5] But given that we seek the same resolution of cycles at different frequencies, this is still the most efficient way to estimate the crystals.

[6] Given the previous footnote, it is obvious that by doing this, it will lead to "redundancy" as the wavelet coefficients have already been combined with most of the same datapoints.

[7] One of the issues with spectral time-frequency analysis is the Heisenberg uncertainty principle, which states that the more certainty that is attached to the measurement of one dimension (- frequency, for example), the less certainty can be attached to the other dimension (- here the time location).

Table 5 Frequency interpretation of MRD scale levels

Scale crystals	Quarterly frequency resolution
d1	2–4 = 6 m–1 yr
d2	4–8 = 1–2 yrs
d3	8–16 = 2–4 yrs
d4	16–32 = 4–8 yrs
d5	32–64 = 8–16 yrs
d6	64–128 = 16–32 yrs
d7	etc.

is easily accomplished. Hence we can use six crystals here even though resolution for the d6 crystal is not high. It should be noted that if conventional business cycles are usually assumed to range from 12 quarters (3 years) to 32 quarters (8 years), then crystal d4 together with the d3 crystal should contain the business cycle.

The variance decomposition for all series considered in this paper is calculated using:

$$E_j^d = \frac{1}{E^d} \sum_{k=1}^{\frac{n}{2^j}} d_{j,k}^2 \tag{5}$$

where $E^d = \sum_j E_j^d$. represents the energy or variance in the detail crystals E_j^d.

Although extremely popular due to its intuitive approach, the DWT suffers from two drawbacks: dyadic length requirements for the series to be transformed and the fact that the DWT is non-shift invariant (- so if datapoints from the beginning of the series are put aside, the lower frequencies will yield different crystals with completely different values). In order to address these two drawbacks, as noted above, we use the maximal-overlap DWT (MODWT)[8] in this study. The MODWT was originally introduced by Shensa (1992) and a phase-corrected version was added and found superior to other methods of frequency decomposition[9] by Walden and Cristan (1998). The MODWT gives up the orthogonality property of the DWT to gain other features, given in Percival and Mofjeld (1997), such as the ability to handle any sample size regardless of whether the series is dyadic or not, increased resolution at coarser scales as the MODWT oversamples the data, translation-

[8] As Percival and Walden (2000) note, the MODWT is also commonly referred to by various other names in the wavelet literature such as non-decimated DWT, time-invariant DWT, undecimated DWT, translation-invariant DWT and stationary DWT. The term "maximal overlap" comes from its relationship with the literature on the Allan variance (the variation of time-keeping by atomic clocks)—see Greenhall (1991).

[9] The MODWT was found superior to both the cosine packet transform and the short-time Fourier transform.

invariance, and more asymptotically efficient wavelet variance estimator than the DWT.

Both Gençay et al. (2001) and Percival and Walden (2000) give a thorough and accessible description of the MODWT using matrix algebra. Crowley (2007) also provides an "intuitive" introduction to wavelets, written specifically for economists, and references the (limited) contributions made by economists using discrete wavelet analysis.[10] The first real usage of wavelet analysis in economics was pioneered by James Ramsey (Lampart and Ramsey 1998), and the first application of wavelets to economic growth (in the form of industrial production) was by Gallegati and Gallegati (2007) and in the form of GDP in a working paper by Crowley and Lee (2005) and then more recently in a published article by Yogo (2008). There are now a few articles that have been published in macroeconomics using wavelet methods in economics, most notably Crowley (2010), Aguiar-Conraria and Soares (2011), Aguiar-Conraria et al. (2012), and Gallegati et al. (2011).

4 Maximal Overlap Discrete Wavelet Transform Results

In this section and the next we review the output from the MODWT for both US and UK real GDP and their aggregate demand components. We first review the US results.

4.1 US Results

The plots for the US in Fig. 3 show the phase-adjusted crystals for each of the frequency bands contained in the detail crystals (or frequency-resolved series) d1–d6, plus the smoothed trend residual from the series (often referred to as the "smooth"), d6, which is obtained after extracting the fluctuations corresponding to the detail crystals. The most obvious observation is that the "great moderation" clearly appears in the data from 1983 through to around 2007; but most noticeably in the d1, d2 and d3 crystals (i.e. for cycles between 6 months and 4 years periodicity), and less obviously in the 4–8 year cycle (d4 crystal) and not at all in the 8–16 year cycle (d5 crystal). There also appears to be the possibility of a longer 30-year cycle in the data, which appears here in s6, the smooth.[11] Note that these observations could not be made using a traditional time series analysis approach: the "great moderation" for all its appeal at the time, appears not to have been a systematic or permanent phenomenon. It is also noteworthy that in the current recovery cycles

[10]These can also be accessed online at: http://faculty.tamucc.edu/pcrowley/Research/frequency_domain_economics.html.

[11]This also appears in GNP data as shown in Crowley (2010).

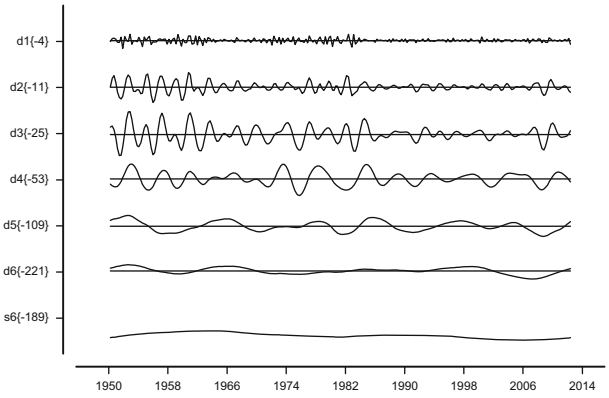


Fig. 3 MODWT decomposition of log change in US GDP

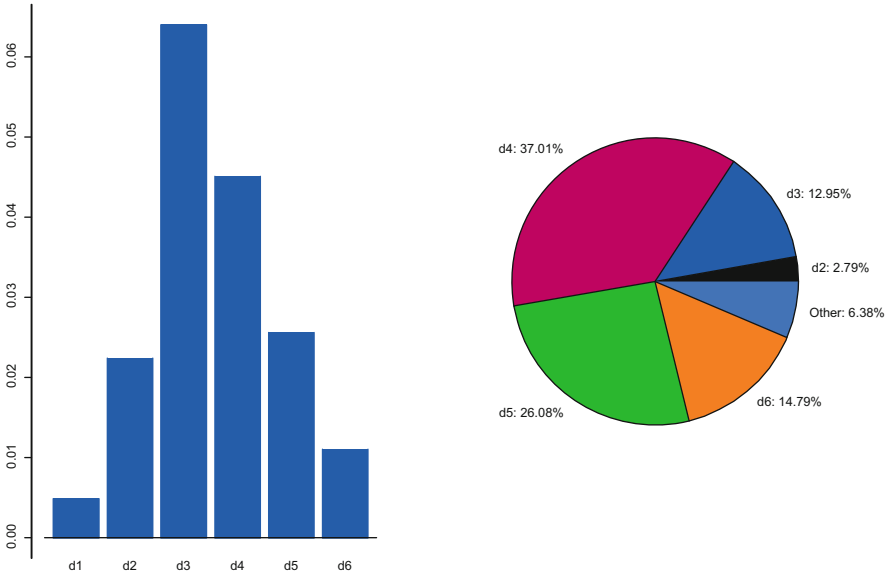


Fig. 4 Variance decomposition by scale for US GDP

at different frequencies are not necessarily concordant—d2 and d4 are falling while d3 and d5 are rising. This interaction of cyclical activity likely gives rise to the uneven pace of US economic growth.

Figure 4 shows the variance decomposition by crystal over the entire data span. Clearly the strongest cycle is contained in crystal d3 (representing 2–4 year cycles), with d4 (representing 4–8 year cycles) following close behind; then d2 (1–2 years) and d5 (8–16 year cycles) contain roughly the same amount of energy. As noted before though, the amount of volatility in any given crystal can change over time.

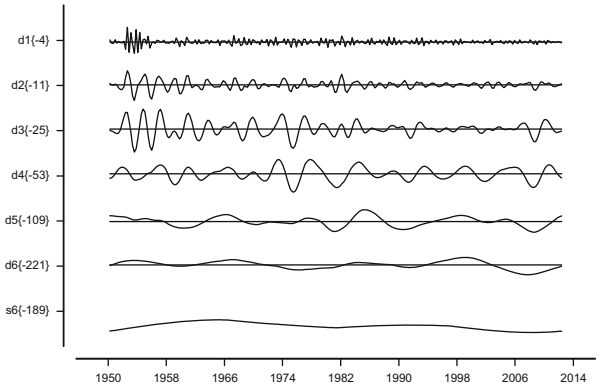


Fig. 5 MODWT decomposition of log change in US consumption

So during the "great moderation", crystals d4 and d5 (4–16 year cycles) appear to dominate fluctuations in growth, but not necessarily during other periods. Hence the great moderation in fact appears to have been a phenomenon in which volatility was shifted from short and business cycle lengths, to the longer cycles (up to 16 years in length). This would certainly explain the observation that recessions or economic slowdowns now appear to take place every 10–15 years, but the periods between are more stable than they used to be.

As might be expected, the MODWT plot in Fig. 5 for consumption expenditures shows relatively similar cyclical patterns to overall GDP, with a clear fall in volatility after 1983 in crystal d3 (2–4 year cycles) but less so for d1, d2 or d4. This is also reflected in the variance decomposition plot in Fig. 6 where there is now more volatility in longer cycles, relative to the shorter cycles, reflecting the success of consumption smoothing over time. As with the moderation in GDP volatility, this fall in volatility after 1983 clearly shows the smoothing power of the strict monetary controls introduced by the Volker regime at the Fed. The more recent movements in US consumption are interesting, with shorter cycles up to an 8 year frequency showing a downturn in consumption, but all longer cycles showing an upturn.

Figure 7 shows the MODWT plot for US private investment, and it is clearly apparent that the "great moderation" for investment spending took place after around 1987, that is later than in consumption; and again this was mostly confined to fluctuations in d2 and d3 crystals, but does not appear in d4 and d5, and hardly at all in d1. In terms of overall energy, the variance decomposition plots in Fig. 8 show that most energy lies in crystal d3 (2–4 year cycles), with both d2 (1–2 year cycles) and d4 (4–8 year cycles) also containing some cyclical activity. In d2, this mostly occurred towards the beginning of the time series, whereas in d4 this appears to have been more consistent through time and likely relates to the business cycle. This finding clearly highlights the rich dynamics at play within the components of output. It shows that the great moderation started at different points within the components of GDP, and this observation would be missed if using only total GDP to measure

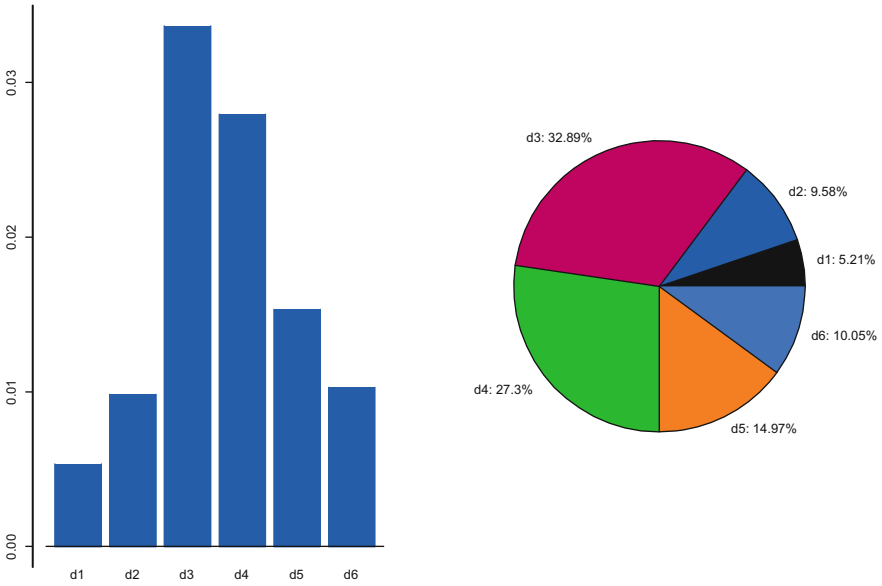


Fig. 6 Variance decomposition by scale for US consumption

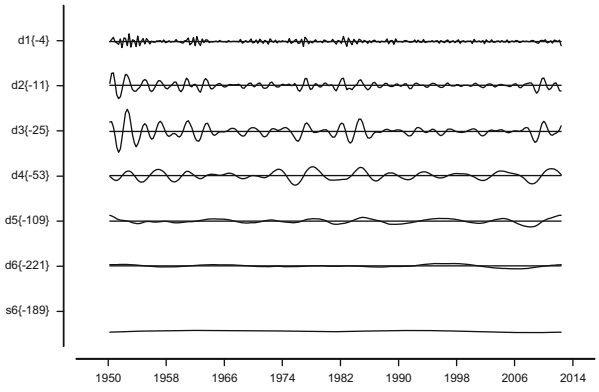


Fig. 7 MODWT decomposition of log change in US private investment

the onset of lower volatility in output. When looking at more recent trends, d1, d2 and d4 are turning downwards, and d3, d5 and d6 are all turning up. Also the d5 (at the business cycle, appears to be becoming more volatile through time.

Government expenditures, since they contain both automatic stabilizers and for more severe recessions, discretionary spending programs should display some cyclical activity at business cycle frequencies. However Fig. 9 shows that apart from the very beginning of the series, there is relatively little cyclicality in this series and where there is, it clearly lies at around the business cycle in crystals d3, d4 and

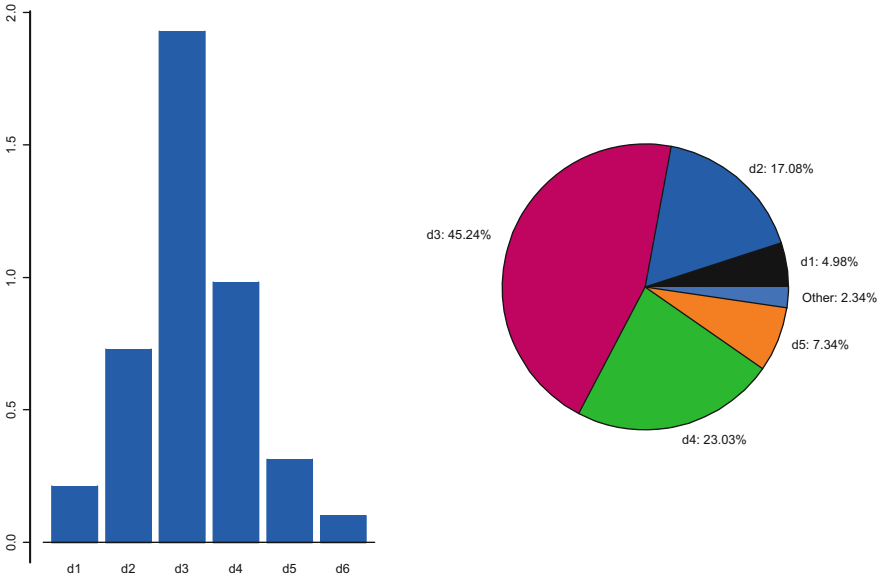


Fig. 8 Variance decomposition by scale of US private investment

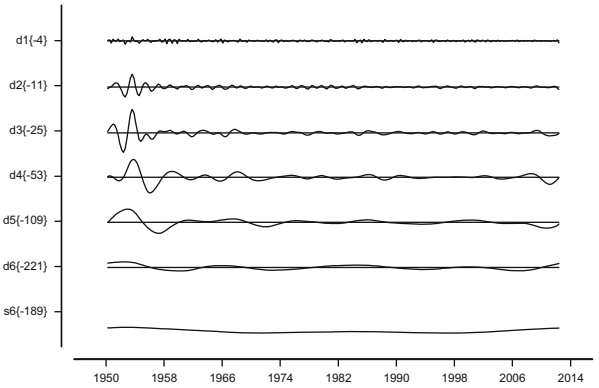


Fig. 9 MODWT decomposition of log change in US G

d5 (2–16 year cycles). Compared to the other components of GDP the volatility in the crystals of government spending is extremely weak, signifying the relatively minor movements in government expenditures compared to private sector activity. Interestingly also there is virtually no energy at short term horizons (6 month to 1 year cycles), and activity in other crystals dies down to only small fluctuations after the mid-1970s, indicating that discretionary fiscal policies had largely been abandoned as an instrument of demand management at that point.

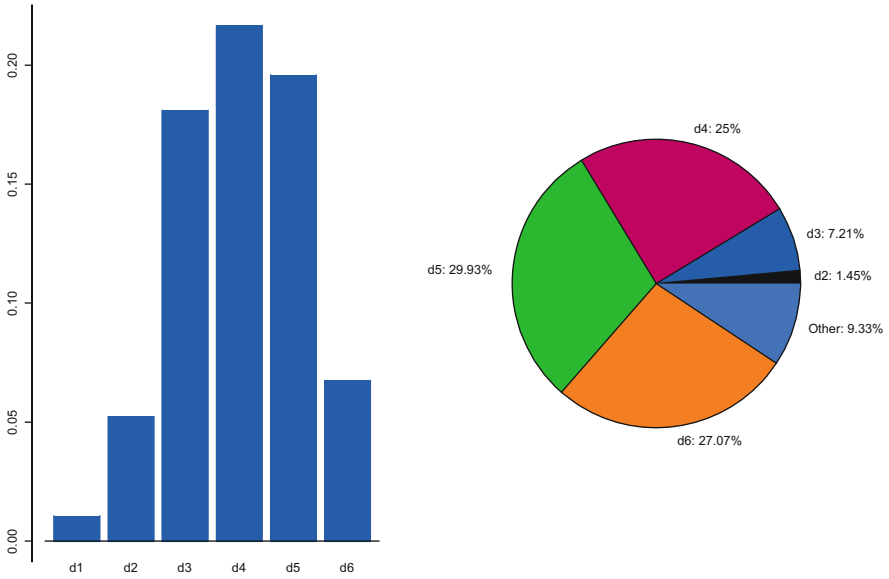


Fig. 10 Variance decomposition of US log G by scale

These results are also to be seen in the variance decomposition by scale which is shown in Fig. 10. Here crystals d4 and d5 have the highest variance. These results also help answer an old debate on whether fiscal policies have been anti-cyclical (stabilizing) or pro-cyclical (destabilizing). In the US, there is little cyclical movement in government spending at any frequency after 1960 which suggests it has largely been a-cyclical in practice. That means the US did not succeed in stabilizing her economy through fiscal policy (or possibly hasn't tried), but she hasn't made it worse either, as some claim. It is also worth noting that Table 1 and the text which follows indicate that G has been a better or more effective shock absorber (stabilizer) than the export markets.

The MODWT plot shown in Fig. 11 for exports is rather surprising. It shows a clear reduction in volatility for crystal d1 from around the mid-1970s with a reduction in volatility in crystal d2 in roughly 1983, followed by reductions in volatility in d3 and d4 in the late 1980s. Surprisingly, volatility then picks up again for crystals d2, d3, d4 and d5 in the late 1990s and continues into the 2000s. This is not matched in the d1 crystal, which shows hardly any short-term movements in exports. Figure 12 shows that most of the energy in the series resides in crystals d3 and d4, with cycle frequencies between 2 and 8 years, corresponding to the business cycle.

These last results require some explanation, but offer an important insight into the vexed issue of whether the exchange rate acts as a shock absorber, or equilibriating device, to offset various external or internal imbalances; or whether it is an additional source of uncertainty in itself. Most business leaders and many

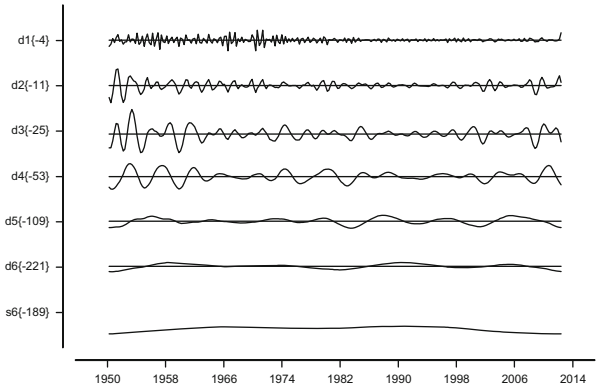


Fig. 11 MODWT decomposition of log change in US X

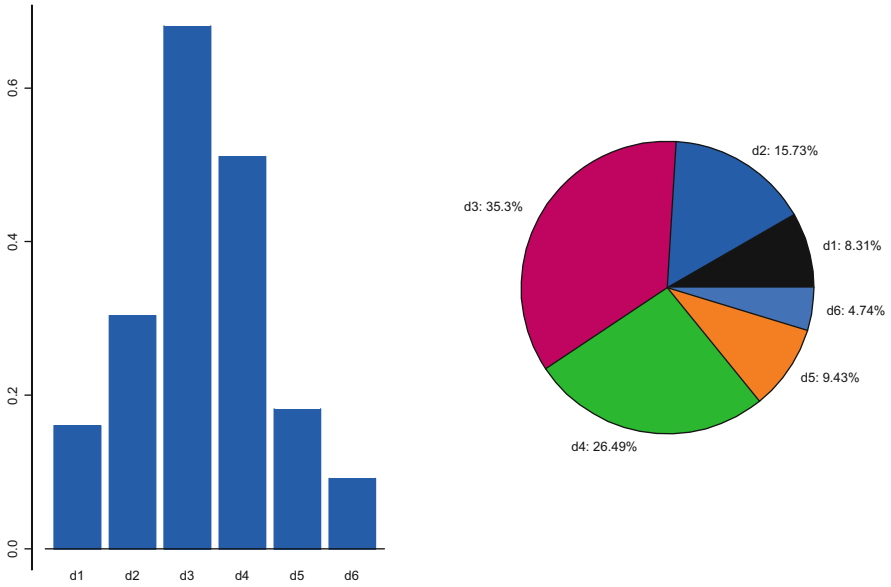


Fig. 12 Variance decomposition of US log X by scale

economists claim that it is primarily a source of uncertainty, whereas it is easy to demonstrate that, in a flexible economy, price volatility is beneficial because it enables us to take advantage of windfall profits/switch to cheaper imports if relative prices or the exchange rate rise. Similarly price volatility allows us to reduce or retrench output and imports when those prices or the exchange rate fall.

In the case at hand, the fall in the volatility of US exports coincides with the start of the dollar's floating exchange rate regime in 1971. That fall in volatility is mostly in short cycles to start, but then spreads to the US business cycle frequencies

later on. That shows the more market sensitive monetary policies of the 1980s were used to stabilize the economy; but that this in turn affected the exchange rate, converting it into a shock absorber and stabilizing exports at the same time. In effect, the exchange rate becomes volatile and exports stable at those frequencies. But the pattern changes in the mid-1990s. At that point the volatility in US exports begins to pick up at business cycle frequencies. The explanation is that by the late-1990s and into the 2000s, US monetary policy had become more activist in pursuit of low and stable inflation, *de facto* inflation targeting.

The result of course was a more stable exchange rate in this period, and hence rising export volatility, as can be seen in our decomposed cyclical data—except at short cycles, reflecting the somewhat greater short run monetary policy activity. To the extent that an inflation targeting regime depends on interest rates as a policy instrument, it should lead to increased volatility in investment and consumption in the same period. And it does—as can be seen in Figs. 5 and 7, principally at cycle lengths d3 and d4, though the increases appear to be fairly small (not surprisingly since other factors are also involved, and neither variable experiences volatility increases back to the pre-1985 levels). The increases in the export volatility are, by contrast, rather larger as we might expect—but again not up to the pre-1985 levels, which suggests that business cycles have become more synchronized across countries than they were.

These results help resolve the controversy: the US exchange rate has acted as a shock absorber more than a source of uncertainty. The US, being a relatively flexible economy, has adjusted as required to remove or balance off external or internal imbalances against each other. When it moved to targeting inflation, some export volatility returned but with a persistent trade deficit since the easiest way to keep inflation low is to let the exchange rate appreciate. The second conclusion is therefore that what often passes for exchange rate uncertainty is in fact fluctuations in the variables that underlie the exchange rate, not random shocks in the exchange itself.

To summarize, it is clear that the "great moderation", although discernable in GDP growth data for the US, is more apparent at various frequencies and in various components of GDP than in others. Nor does it represent some kind of long term paradigm shift. The timing and dynamics that lead to the "great moderation" do not translate directly back to the components of GDP growth. Consumption and investment appear to be the sources for the "great moderation", with consumption volatility moderation occurring in the early 1980s and investment volatility moderation occurring in the later part of the 1980s. Changes in government expenditure and export expenditures do not appear to be major sources of the origins of lower volatility in real GDP growth. Lower volatility is also therefore not a result of government stabilisation policies. Instead monetary policy, with effects on the exchange rate, must be the culprit because the residuals (s6) and short term shocks (d1) play little or no role in these moderations after the mid-1950s. These are all features that cannot be detected from aggregate data on output, or with traditional time series analysis.

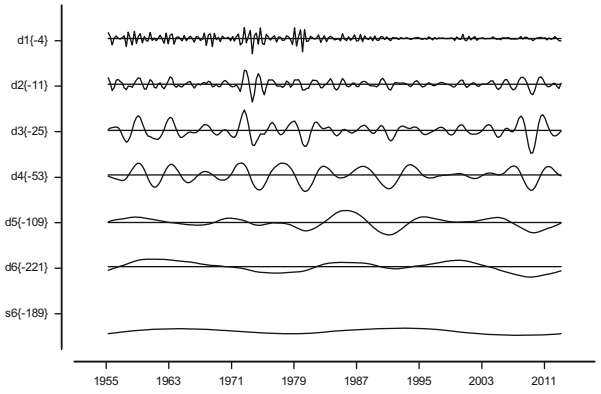


Fig. 13 MODWT decomposition of log change in UK GDP growth

4.2 UK Results

The same exercise is now repeated for the UK. In Fig. 13 we observe the same patterns for UK GDP as in US GDP, with crystals d1, d2 and d3 exhibiting lower volatility after the era of the miners and other strikes in 1984–1985 and after the Thatcher policies took hold, but with d4 exhibiting slightly lower volatility and d5 and d6 hardly changing. The longer residual cycle is once again weak, and has a periodicity of approximately 35 years. Figure 14 once again shows that most of the variance resides in d3, d4 and d5 (2–16 year periodicities), with d4 (4–8 year cycles) containing most energy. However, compared with the US, the volatility is more evenly spread across cycles. It is also evident that d1–d3 show the great moderation like the US, while d4 and d5 actually get *less* stable in the moderation period. This again suggests a mechanism that shifts short run instability to long term instability. Most recently, the double-dip downturn in the UK can be seen quite clearly in d1–d3, where as d5 and d6 (cycles over 8 years in length) point to a longer term recovery.

In Figs. 15 and 16 the MODWT and the variance decomposition by scale are shown for UK consumption growth. In Fig. 15 the "great moderation" is evident from 1983 in d1, but doesn't occur until roughly 1991 in d2 and d3, and not until 1995 in d4. In terms of volatility, d4 and d5 clearly have most energy and, although d4 has been less volatile until the recent downturn, d5 has not. Apparently a new cycle appears to also have emerged since the mid-1970s in the d6 crystal with roughly a 16 year periodicity. There appears to be little cyclicality beyond this frequency. Once again the volatility is spread across a wider range of frequencies compared to the US. As with the GDP data, these results show a much richer and more complex set of dynamics than could be captured by traditional real business cycle models.

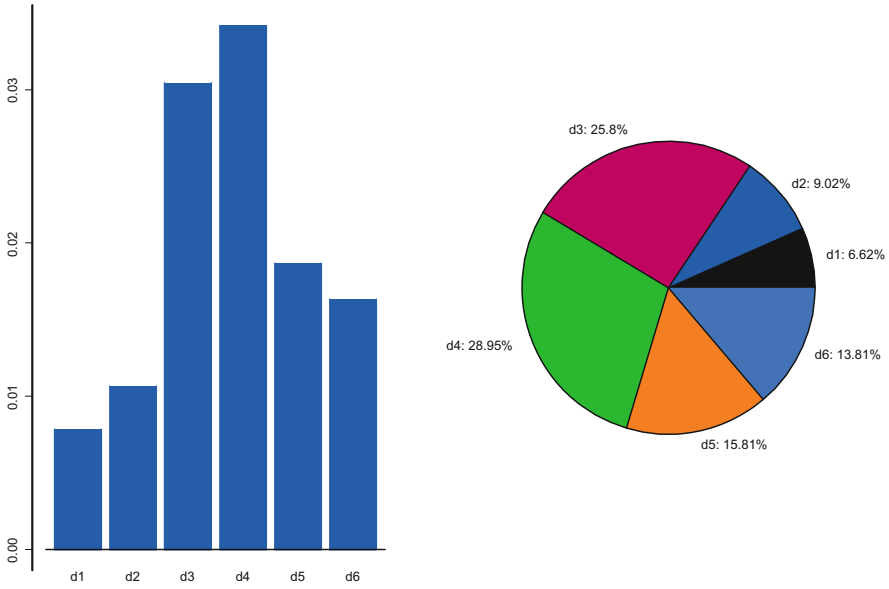


Fig. 14 Variance decomposition of UK GDP by scale

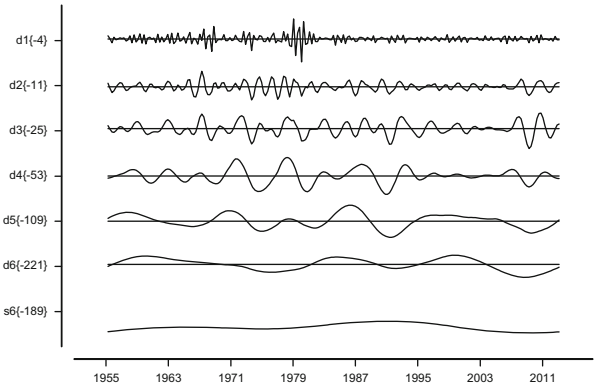


Fig. 15 MODWT decomposition of log change in UK consumption

In Fig. 17 the change in UK private investment expenditures are decomposed using the MODWT, and here much more cyclicality is detected than with the US, with only one of the crystals, d1, exhibiting any real lowering in volatility, and then only after 1990. This is a surprising result (given that the great moderation effect is hardly evident in the data), and definitely does not match that obtained for the US. In Fig. 18 most of the volatility lies in crystals d3 and d4, with clearly a recent increase in volatility in d5, perhaps reflecting a lengthening of the business cycle. Compared with the US though there is more volatility in longer and shorter cycles.

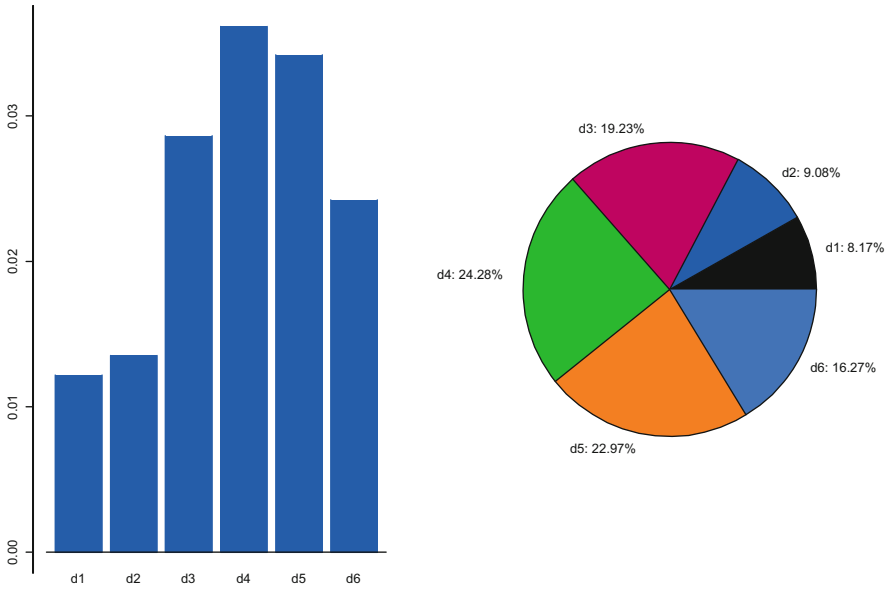


Fig. 16 Variance decomposition of UK C by scale

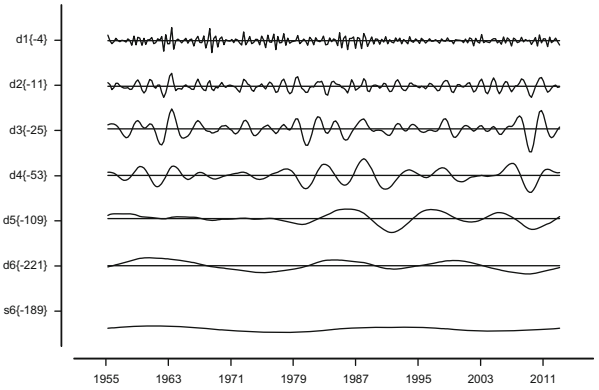


Fig. 17 MODWT decomposition of log change in UK investment

A reasonable question is, why does the UK show more volatility in investment spending than the US—especially at frequencies shorter and longer than business cycles? It will be observed from Fig. 17 that this higher volatility is mostly in the boom years of the mid-1980s and late 1990s, and is largely restricted to d2–d5. In addition, because this extra volatility does not show up (proportionately) in the other components of UK GDP, nor is there any excess volatility in the residuals or short cycles while the investment itself is less well coordinated/correlated with C and G but better coordinated with UK exports, we can conclude that the extra

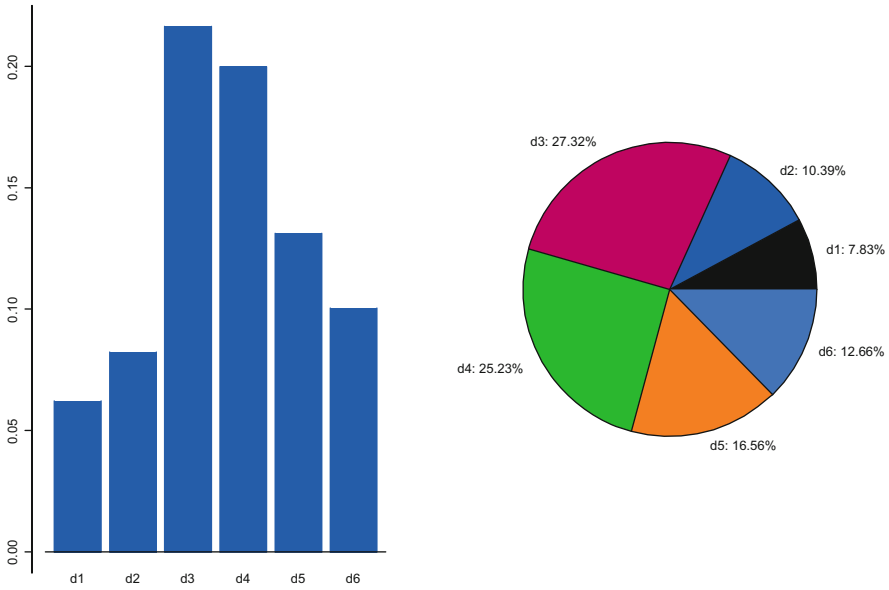


Fig. 18 Variance decomposition of UK I by scale

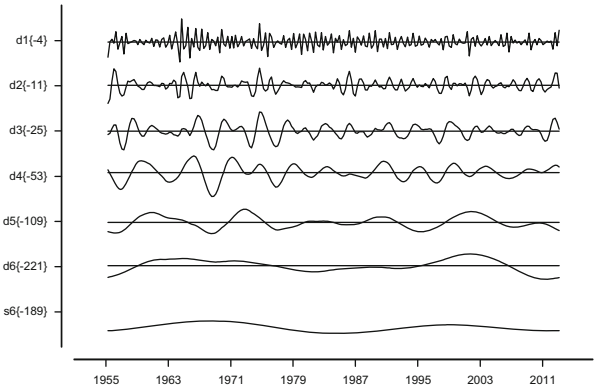


Fig. 19 MODWT decomposition of log change in UK government spending

investment volatility is due to the UK's successful record of attracting FDI in those boom periods.

With the log annual change in UK government expenditures, there is also much more volatility than with the US measure, as Fig. 19 shows. Here there appears to have been a dampening of volatility in d1 beginning only in the mid-1990s. And while for d2 very little change has occurred, for d3 and d4 a dampening of volatility appears to have taken place around 1982, a time when monetary policy moved away from monetarism and fiscal policy started to be more closely managed.

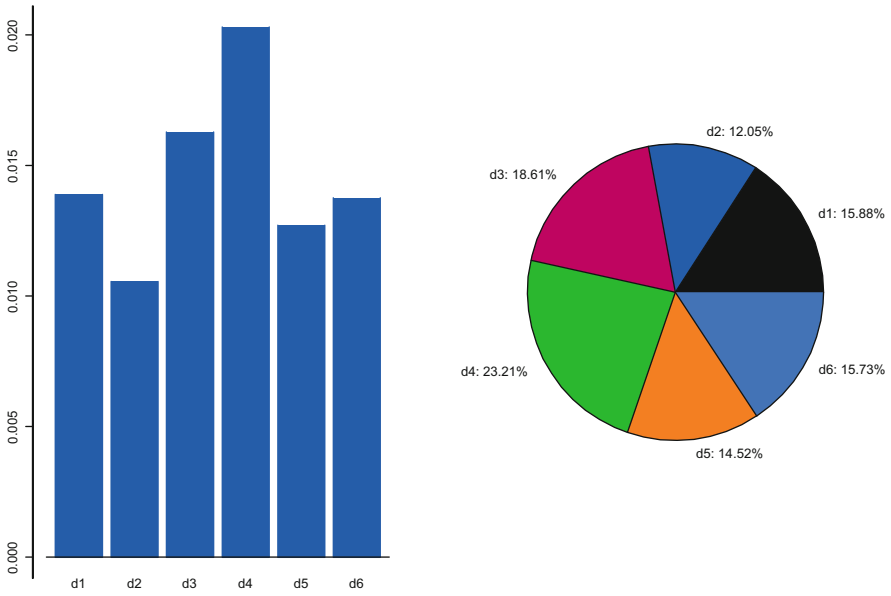


Fig. 20 Variance decomposition of UK government spending by scale

There seem to be cycles operating at lower frequencies as well, with a very irregular cycle captured by the d5 crystal and rather strange semi-cyclical movements in the d6 crystal, which almost certainly means that the UK moderation has been achieved by policy actions not by a smoother operating economy. The implication therefore is that the better and smoother performance of the UK economy in the Thatcher and Blair years was held together by policy actions, rather than by favourable market and institutional reforms that promote smoother running markets. The contrast with the US post-1970 for any cycle is instructive. Further, there are no obvious breaks in behaviour (except possibly d5 and d6 after the 1970s). It is also noteworthy that recent fiscal austerity measures in the UK can be seen at all frequencies except those for 4–8 year frequencies.

Figure 20 shows that most energy resides in the d4 crystal, but what is surprising here is that a significant amount of movement is found in d1, which contains cycles of 6 months to 1 year duration. Here the volatility is fairly evenly distributed across different cycles with noise less important than business cycles. Hence automatic stabilisers must have been at work. There is also no obvious shift in weight from short to long run, so it is difficult to see a distinction between discretionary policy vs automatic stabilizers.[12]

[12]Separating automatic from discretionary fiscal policies in a cyclical environment is not an easy matter. Bernoth et al. (2013) review different methods, and show how it can be done by combining real time and ex-post data.

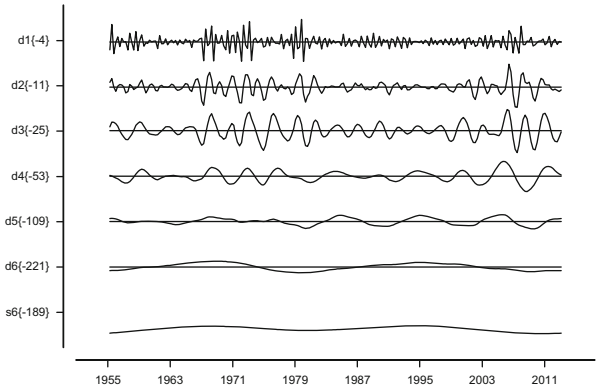


Fig. 21 MODWT decomposition of log change in UK exports

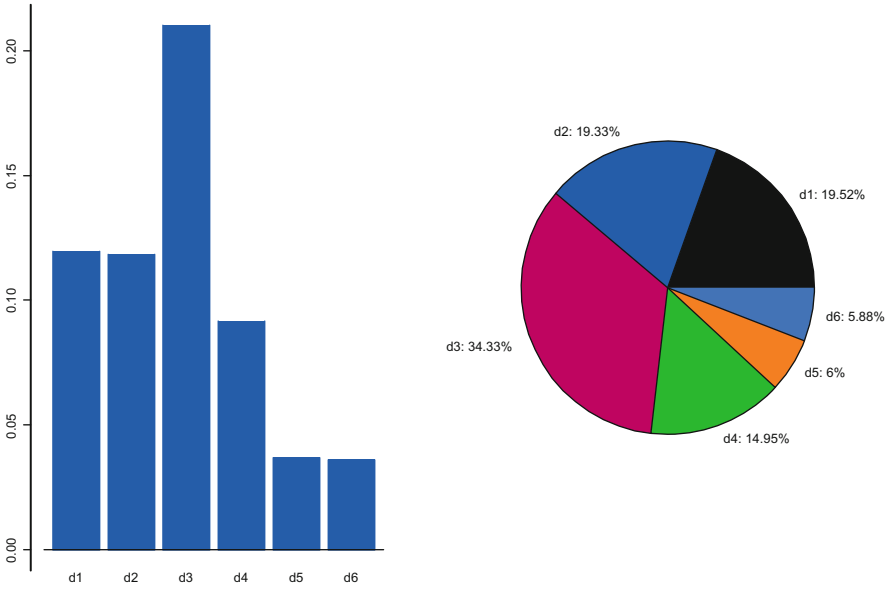


Fig. 22 Variance decomposition of UK X by scale

In Figs. 21 and 22 the MODWT decomposition of expenditures on UK exports is plotted together with the variance decomposition by scale crystal. Here, perhaps surprisingly, there are two episodes of high volatility in export expenditures, presumably in this instance mostly related to the fortunes of the British currency. After the infamous 1967 devaluation of the pound by the Wilson government, this clearly led to greater volatility in export growth, which then continued with the collapse of the Bretton Woods system in 1973.

Much smaller fluctuations are observed in d1–d3 (and to some extent in d4) after 1983. But by 2005 the volatility in export expenditures had clearly returned. At that point d5 appears to suggest that a regular 10 year cycle has emerged and d6 suggests a weak cycle at roughly a 27 year periodicity. So what is notable here is the moderation in short-run cycles (noise) and post 1980 (up to d3), a moderation that was lost again by 2004. The explanation for this result is the same as in the US. During the 1980s the UK became a convinced exchange rate floater, which meant the exchange rate became the shock absorber that lowered the volatility of exports (at least in the shorter cycles). But, after 1994 and the unsuccessful EMS period, she had adopted explicit inflation targets which led to smoother monetary policies and a (mostly) smoother exchange rate path, and with it higher export volatilities, once the new monetary regime had settled down. However these results also show that there is no case for saying that fixing the exchange rate stabilizes the economy, at least for the UK. Significant regime changes, like fixing the exchange rate in 1990–1992, or the EMS crisis which seriously unfixed them again, do not destabilize the economy or exports. Those events just do not show up in the data. Figure 21 shows that higher frequency cycles (with periodicity less than 4 years) dominate the variance decomposition in this case—once again, a quite different result from that observed in the US—suggesting that the great moderation was transitory in the UK, and largely took the form of shifting short-term volatility (d3 and lower) to longer term cycles (d4 and higher). These changes are clearly seen in Fig. 21.

Once again, the conclusion is that the exchange rate has acted as a shock absorber rather than as a source of uncertainty—albeit a little less successfully than in the US because the UK economy is less flexible and is less well stabilized. Figure 22 shows that d3 is the most important cycle in UK exports, but has a variance of only 0.2 (or 30 % of total variance in exports), compared to 0.65 or 34 % for the US.

5 Conclusions

In this paper we have used wavelet transformations to decompose the separate parts of domestic expenditures which make up real GDP for the US and the UK into their component cycles. The first finding is that decomposing the components of real GDP growth separately into different cycles reveals characteristics of the cycles in growth, and the relationships between them, that cannot be seen in an analysis of the aggregate data for real GDP alone. That is because the cycles of the various components offset each other to a degree, leading to a loss of information at the aggregate level. The second finding is that although the "great moderation" is found in most of the data, it is not consistent across different frequency cycle lengths, appearing only in cycles generally shorter than or equal to the business cycle. This is an important finding, as it demonstrates (as we now know) that the so-called "great moderation" was not as significant as economists had thought, given that the business cycle still was evident, and that longer cycles did not abate in strength at all. The "great moderation" in fact appears to have been more a case of shifting

short term volatility to long cycle volatility, than moderating volatility as such. This means that changes in volatility, like the "great moderation", will be very difficult to detect with any certainty without a full frequency decomposition of the components of GDP.

In terms of the comparison between the US and the UK, the volatility in components at the specified frequency ranges in discrete wavelet analysis is markedly different. The analysis shows that there is much more volatility in GDP components at very short and longer frequencies for the UK than there is in the US. This is particularly the case for government expenditures, where activist fiscal policies have clearly had a much greater impact than in the US. There has also been a tendency for volatility to have been shifted from shorter cycles to longer cycles, more in the UK than the US. This we put down to the changes in monetary regimes, and hence exchange rate arrangements, which focussed first on stabilisation and then on explicit or implicit inflation targeting. Fiscal policies, by contrast, have largely been acylical or ineffective for stabilisation in the US; but pro-cyclical and destabilising in the UK.

Acknowledgements This research was completed while Crowley was visiting the School of Public Policy at George Mason University in Fairfax, VA, USA during the fall of 2009. Dean Kingsley Haynes should be thanked for hosting Crowley at George Mason University in 2009 and Texas A&M University - Corpus Christi is acknowledged for providing faculty development leave funding.

References

Aguiar-Conraria L, Soares M (2011) Business cycle synchronization and the euro: a wavelet analysis. J Macroecon 33(3):477–489

Aguiar-Conraria L, Martins M, Soares M (2012) The yield curve and the macro-economy across time and frequencies. J Econ Dyn Control 36(12):1950–1970

Bernoth K, Hughes-Hallett A, Lewis J (2013) The cyclicality of automatic and discretionary fiscal policy: what can real time data tell us? Macroecon Dyn 5: 1–23

Bruce A, Gao HY (1996) Applied wavelet analysis with S-PLUS. Springer, New York

Christiano L, Fitzgerald T (1998) The business cycle: it's still a puzzle. Fed Reserve Bank Chic Econ Perspect 22:56–83

Crowley P (2007) A guide to wavelets for economists. J Econ Surv 21(2):207–267

Crowley P (2010) Long cycles in growth: explorations using new frequency domain techniques with US data. Bank of Finland Discussion Paper 6/2010, Helsinki

Crowley P, Lee J (2005) Decomposing the co-movement of the business cycle: a time-frequency analysis of growth cycles in the euro area. Bank of Finland Discussion Paper 12/05, Helsinki

Furlanetto F, Seneca M (2010) Investment-specific technology shocks and consumption. Norges Bank Discussion Paper 30-2010, Oslo

Gallegati M, Gallegati M (2007) Wavelet variance analysis of output in G-7 countries. Stud Nonlinear Dyn Econ 11(3):1435–1455

Gallegati M, Gallegati M, Ramsey J, Semmler W (2011) The US wage phillips curve across frequencies and over time. Oxf Bull Econ Stat 73:489–508

Gençay R, Selçuk F, Whicher B (2001) An introduction to wavelets and other filtering methods in finance and economics. Academic, San Diego

Greenhall C (1991) Recipes for degrees of freedom of frequency stability estimators. IEEE Trans Instrum Meas 40:994–999
King R, Plosser C, Rebelo S (1988) Production, growth and business cycles i: the basic neoclassical model. J Monet Econ 21:195–232
Lampart C, Ramsey J (1998) Decomposition of economic relationships by timescale using wavelets. Macroecon Dyn 2(1):49–71
Percival D, Mofjeld H (1997) Analysis of subtidal coastal sea level flusctuations using wavelets. J Am Stat Assoc 92:868–80
Percival D, Walden A (2000) Wavelet methods for time series analysis. Cambridge University Press, Cambridge
Rebelo S (2005) Real business cycle models: past, present and future. Scand J Econ 107(2):217–238
Shensa M (1992) The discrete wavelet transform: wedding the à trous and mallat algorithms. IEEE Trans Signal Process 40:2464–2482
Walden A, Cristan C (1998) The phase-corrected undecimated discrete wavelet packet transform and its application to interpreting the timing of events. Proc R Soc Lond Math Phys Eng Sci 454(1976):2243–2266
Yogo M (2008) Measuring business cycles: a wavelet analysis of economic time series. Econ Lett 100(2):208–212

Nonlinear Dynamics and Wavelets for Business Cycle Analysis

Peter Martey Addo, Monica Billio, and Dominique Guégan

Abstract We provide a signal modality analysis to characterize and detect nonlinearity schemes in the US Industrial Production Index time series. The analysis is achieved by using the recently proposed "delay vector variance" (DVV) method, which examines local predictability of a signal in the phase space to detect the presence of determinism and nonlinearity in a time series. Optimal embedding parameters used in the DVV analysis are obtained via a differential entropy based method using Fourier and wavelet-based surrogates. A complex Morlet wavelet is employed to detect and characterize the US business cycle. A comprehensive analysis of the feasibility of this approach is provided. Our results coincide with the business cycles peaks and troughs dates published by the National Bureau of Economic Research (NBER).

P.M. Addo (✉)
European Doctorate in Economics–Erasmus Mundus (EDEEM), Université Paris 1 - Panthéon-Sorbonne, MSE-CES UMR8174, 106-113 Boulevard de l'hopital, 75013 Paris, France

Department of Economics, Università Ca'Foscari of Venice, Venice, Italy

Université Paris 1- Panthéon-Sorbonne, Paris, France
e-mail: peter.addo@univ-paris1.fr

M. Billio
Department of Economics, Università Ca'Foscari of Venice, Venice, Italy

D. Guégan
Université Paris I—Panthéon Sorbonne, Paris, France

M. Gallegati and W. Semmler (eds.), *Wavelet Applications in Economics and Finance*, Dynamic Modeling and Econometrics in Economics and Finance 20,
DOI 10.1007/978-3-319-07061-2_4,

1 Introduction

In general, performing a nonlinearity analysis in a modeling or signal processing context can lead to a significant improvement of the quality of the results, since it facilitates the selection of appropriate processing methods, suggested by the data itself. In real-world applications of economic time series analysis, the process underlying the generated signal, which is the time series, are a priori unknown. These signals usually contain both linear and nonlinear, as well as deterministic and stochastic components, yet it is a common practice to model such processes using suboptimal, but mathematically tractable models. In the field of biomedical signal processing, e.g., the analysis of heart rate variability, electrocardiogram, hand tremor, and electroencephalogram, the presence or absence of nonlinearity often conveys information concerning the health condition of a subject (for an overview Hegger et al. 1999). In some modern machine learning and signal processing applications, especially biomedical and environmental ones, the information about the linear, nonlinear, deterministic or stochastic nature of a signal conveys important information about the underlying signal generation mechanism. In the analysis of economic indicators, the presence of nonlinearity in the data provides information about both structural and behavioral changes that can occur in the economy across time. In particular, nonlinearity in an economic indicator conveys information on possible existence of different *states of the world* or regimes in the economy. There has been an increasing concerns on the forecasting performance of some nonlinear models in modeling economic variables. Nonlinear models often provide superior in-sample fit, but rather poor out-of-sample forecast performance (Stock and Watson 1999). In cases were the nonlinearity is spurious or relevant for only a small part of the observations, the use of nonlinear models will lead to forecast failure (Terasvirta 2011). It is, therefore, essential to investigate the intrinsic dynamical properties of economic time series in terms of its deterministic/stochastic and nonlinear/linear components reveals important information that otherwise remains not clear in using conventional linear methods of time series analysis.

Since the early work by Burns and Mitchell (1946), many attempts have been made to measure and forecast business cycles. Many business cycle indicators present asymmetric features that have long been recognized in economics (Mitchell 1927; Keynes 1936). Putting it simply, there are sharp retractions during downturns in the economy as opposed to gradual upswings during recoveries (Kontolemis 1997; Sichel 1993; Ashley and Patterson 1936; Brock and Sayers 1988). Asymmetry has been recognized as a nonlinear phenomenon in several recent studies investigating various economic time series. Nonlinear models are therefore required to capture the features of the data generating mechanism of inherently asymmetric realizations of some of the macroeconomic business cycle series, since linear models are incapable of generating such behavior (Granger and Terasvirta 1993; Terasvirta and Anderson 1992; Terasvirta 1994; Dias 2003). Many nonlinear models are only identified when the alternative hypothesis holds (the model is genuinely nonlinear) but not when the null hypothesis is valid. Since the parameters of an

unidentified model cannot be estimated consistently, testing linearity before fitting any of these models is an unavoidable step in nonlinear modeling (Luukkonen et al. 1988; van Dijk et al. 2002). However, most linearity testing against nonlinearity in literature, usually require a specification of a stationary nonlinear model under the alternative hypothesis. This makes such tests restrictive to the dynamics of the specified nonlinear model. There is therefore a need to use procedures that tests linearity against any form of nonlinearity in the economic time series.

Several methods for detecting nonlinear nature of a signal have been proposed over the past few years. The classic ones include the "deterministic versus stochastic" (DVS) plots (Weigend and Casdagli 1994), the Correlation Exponent and "δ-ϵ" method (Kaplan 1994). For our purpose, it is desirable to have a method which is straightforward to visualize, and which facilitates the analysis of predictability, which is a core notion in online learning. In this paper, we adopt to the recently proposed phase space based "delay vector variance" (DVV) method (Gautama et al. 2004a), for signal characterization, which is more suitable for signal processing application because it examines the nonlinear and deterministic signal behavior at the same time. This method has been used for understanding the dynamics of exchange rates (Addo et al. 2012), detecting nonlinearity in financial markets (Addo et al. 2013a), qualitative assessment of machine learning algorithms, analysis of functional magnetic resonance imaging (fMRI) data, as well as analyzing nonlinear structures in brain electrical activity and heart rate variability (HRV) (Gautama et al. 2004b). Optimal embedding parameters will be obtained using a differential entropy based method proposed in Gautama et al. (2003), which allows for simultaneous determination of both the embedding dimension and time lag needed for the DVV analysis. Surrogate generation used in this study will be based on both the Iterative Amplitude Adjusted Fourier Transform (iAAFT) (Schreiber and Schmitz 1996, 2000) and a recently refined iAAFT with a wavelet-based approach, denoted WiAAFT (Keylock 2006).

Wavelet analysis has successfully been applied in a great variety of applications like signal filtering and denoising, data compression, imagine processing and also pattern recognition. The application of wavelet transform analysis in science and engineering really began to take off at the beginning of the 1990s, with a rapid growth in the numbers of researchers turning their attention to wavelet analysis during that decade (Ramsey and Lampart 1998; Guttorp et al. 2000; Jensen 1999; Ramsey 1999). The wavelet transforms has the ability to perform local analysis of a time series revealing how the different periodic components of the time series change over time. Wavelets are able to locate precisely in time regime shifts, discontinuities, and isolated shocks to a dynamical system. A wavelet approach has the ability to deal with non-stationarity of stochastic innovations that are inevitably involved with economic and financial time series (Ramsey 1999). The maximum overlap discrete wavelet transform (MODWT) has commonly been used by some economists (Guttorp et al. 2000; Gallegati and Gallegati 2007; Gallegati 2008, among others). The MODWT can be seen as a kind of compromise between the discrete wavelet transform (DWT) and the continuous wavelet transform (CWT); it

is a redundant transform, because while it is efficient with the frequency parameters it is not selective with the time parameters. The CWT, unlike the DWT, gives us a large freedom in selecting our wavelets and yields outputs that makes it much easier to interpret. The continuous wavelet transform has emerged as the most favored tool by researchers as it does not contain the cross terms inherent in the Wigner-Ville transform and Choi-Williams distribution (CWD) methods while possessing frequency-dependent windowing which allows for arbitrarily high resolution of the high frequency signal components. Moreover, the time invariance property of the CWT implies that the wavelet transform of a time-delayed version of a signal is a time-delayed version of its wavelet transform. This serves as an important property in terms of pattern recognition. In other words, the identification of the business cycle turning points for a subset of the entire time series does not change through time for a given time series. From an economic point of view, this ensures an effective dating chronology since it avoids revisions through time. This nice property of the CWT is not readily obtained in the case of DWT and MODWT (Addison 2005; Walden and Percival 2000).

Dating is an *ex post* exercise, and in this respect accuracy is an important criterion since dating is useful for economic decision-making (Addo et al. 2013b). Governments and central banks are usually very sensitive to indicators showing signs of deterioration in growth to allow them to adjust their policies sufficiently in advance, avoiding more deterioration or a recession (Anas et al. 2008). As such, it will be interesting to choose a wavelet that will provide a better interpretation of the results from an economic point of view and also enhance accurate detection of the dates. In this respect, the choice of the wavelet is important. We are concerned with information about cycles and as such complex wavelets serves as a necessary and better choice. We need complex numbers to gather information about the phase, which, in turn, tells us the position in the cycle of the time-series as a function of frequency and the associated magnitude in this position. This will enable extraction of information about the economy-wide fluctuations in production that occur around a long-term growth trend. Thus, detecting and studying periods of relatively rapid economic growth (an expansion), and periods of relative stagnation or decline (recession) in the economy. There are many continuous wavelets to choose from; however, by far the most popular are the Mexican hat wavelet and the Morlet wavelet. In this work, we employ a complex Morlet wavelet which satisfies these requirements and has optimal joint time-frequency concentration, meaning that it is the wavelet which provides the best possible compromise in these two dimensions.

In this paper, we provide a novel methodology for business cycle modeling which encompasses different existing methods successfully applied in physics and engineering. In particular, we first study the structure of a chosen economic indicator[1] via a phase-space representation using the differential entropy based

[1]The economic indicator used in the work is the US Industrial Production Index. We remark that our methodology can easily be applied to any known business cycle indicator.

method with both iAAFT and wavelet-based surrogates. We then use the DVV method to detect the nonlinear nature of the economic indicator using the values of the optimal embedding parameters obtained via the differential entropy method with surrogates. Finally, we perform wavelet analysis with a complex Morlet wavelet using a continuous wavelet transform to discover patterns or hidden information that cannot be captured with traditional methods of business cycle analysis such as spectral analysis. Our results is consistent with business cycle dates published by the National Bureau of Economic Research (NBER). We are able to detect these business cycle dates and study these fluctuations in the economy over frequency and time. This serves as an important finding in terms of forecasting and pattern recognition.

The paper is organized as follows: The concept of wavelet analysis and our choice of analyzing wavelet is presented in Sect. 2.1. Surrogate generation methodology and differential entropy based method for determining optimal embedding parameters of the phase-space representation of time series are then presented in Sects. 2.2 and 2.3 respectively. Lastly, we present, in Sect. 2.4, an overview of the "delay vector variance" method with illustrations. In Sect. 3, we present a comprehensive analysis of the feasibility of this approach to analyze the US Business cycle. Section 4 concludes.

2 Background: Wavelet Analysis and "Delay Vector Variance" Method

In this section, we present an overview of different existing methods successfully applied in physics and engineering. In particular, we show the usefulness of these methods over other methods and then explain how we merged these methods to business cycle modeling. Our methodology encompasses wavelet analysis, surrogate generation methods, differential entropy method for determining the optimal embedding parameters in phase-space, and the DVV method.

2.1 Wavelet and Wavelet Analysis

It is a time-frequency signal analysis method which offers simultaneous interpretation of the signal in both time and frequency allowing local, transient or intermittent components to be elucidated. These components are often not clear due to the averaging inherent within spectral only methods like the fast Fourier transform (FFT).

A wavelet is a function, ψ, which has a small concentrated burst of finite energy in the time domain and exhibits some oscillation in time. This function must be in the space of measurable functions that are absolutely and squared-integrable, i.e.

$\psi \in \mathcal{L}^1(\mathbb{R}) \cap \mathcal{L}^2(\mathbb{R})$, to ensure that the Fourier transform of ψ is well-defined and ψ is a finite energy signal. A single wavelet function generates a family of wavelets by dilating (stretching and contracting) and translating (moving along the time axis) itself over a continuum of dilation and translation values. If ψ is a wavelet analyzing function then the set $\{\tau_t D_s \psi\}$ of all the dilated (by $s \neq 0$) and translated (by t) versions of ψ is that wavelet family generated by ψ. Dilation in time by contracting values of scale ($s > 1$) corresponds to stretching dilation in the frequency domain.

The basic concept in wavelet transforms is the projection of data onto a basis of wavelet functions in order to separate large-scale and fine-scale information (Bruce et al. 2002). Thus, the signal is decomposed into a series of shifted and scaled versions of a mother wavelet function to make possible the analysis of the signal at different scales and resolutions. For reconstruction of a signal, it is necessary that ψ be such that $\{\tau_t D_s \psi\}$ span a large enough space of interest.

- Thus, every signal f of interest should be representable as a linear combination of dilated and translated versions of ψ.
- Knowing all the inner products $\{\langle f, \tau_t D_s \psi \rangle\}$, the signal should be recoverable.

The wavelet ψ is assumed to satisfy the admissibility condition,

$$C_{adm,\psi} = \int_{\mathbb{R}} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty, \tag{1}$$

where $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(\xi)$, $\hat{\psi}(\omega) = \int_{\mathbb{R}} \psi(\xi) e^{-i\omega\xi} d\xi$. The admissibility condition (1) implies $\hat{\psi}(0) = \int_{\mathbb{R}} \psi(\xi) d\xi = 0$. For s restricted to $\mathbb{R}^+$, the condition (1) becomes

$$C_{adm_+,\psi} = \int_0^{\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < \infty. \tag{2}$$

This means that the wavelet has no zero-frequency component. The value of the admissibility constant, $C_{adm,\psi}$ or $C_{adm_+,\psi}$ depends on the chosen wavelet. This property allows for an effective localization in both time and frequency, contrary to the Fourier transform, which decomposes the signal in terms of sines and cosines, i.e. infinite duration waves.

There are essentially two distinct classes of wavelet transforms: the continuous wavelet transform and the discrete wavelet transform. We refer the reader to Addison (2005), Walden and Percival (2000) for a review on wavelet transforms. In this work, we employ a complex wavelet via a continuous wavelet transform (CWT) in order to separate the phase and amplitude information, because the phase information will be useful in detecting and explaining the cycles in the data. We provide in Appendix "Continuous Wavelet Transform (CWT)" an overview of CWT and its relevance to our work.

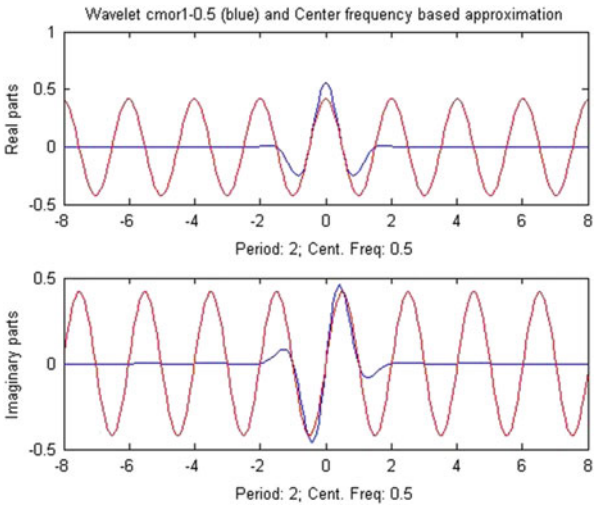


Fig. 1 Complex Morlet wavelet with $f_b = 1$ and $f_c = 0.5$

2.1.1 Choice of Wavelet

The *Morlet wavelet* is the most popular complex wavelet used in practice. A complex Morlet wavelet (Teolis 1998) is defined by

$$\psi(\xi) = \frac{1}{\sqrt{\pi f_b}} e^{i 2\pi f_c \xi - \frac{\xi^2}{f_b}} \tag{3}$$

depending on two parameters: f_b and f_c, which corresponds to a bandwidth parameter and a wavelet center frequency respectively. The Fourier transform of ψ is $\hat{\psi}(\zeta) = e^{-\pi^2 f_b (\zeta - f_c)^2}$, which is well-defined since $\psi \in \mathcal{L}^1(\mathbb{R})$. It can easily be shown that the Morlet wavelet (3) is a modulated Gaussian function and involutive, i.e. $\psi = \tilde{\hat{\psi}}$. The Fourier transform $\hat{\psi}$ has a maximum value of 1 which occurs at f_c, since $\|\psi\|_1 := \int |\psi| = 1$. This wavelet has an optimal joint time-frequency concentration since it has an exponential decay in both time and frequency domains, meaning that it is the wavelet which provides the best possible compromise in these two dimensions. In addition, it is infinitely regular, complex-valued and yields an exactly reconstruction of the signal after the decomposition via CWT.

In this work, the wavelet that best detects the US business is the complex Morlet wavelet with $f_b = 1$ and $f_c = 0.5$. In this case, the Morlet wavelet becomes

$$\psi(\xi) = \frac{1}{\sqrt{\pi}} e^{i\pi\xi - \xi^2}, \tag{4}$$

which we will often refer to as Morlet wavelet. The nature of our choice of wavelet function and the associated center frequency is displayed in Fig. 1. It illustrates the

oscillating nature of the wavelet with short duration of the time support. In other words, the wavelet is bounded, centered around the origin, and have time support (respectively frequency support).

2.2 Surrogate Data Method

Surrogate time series, or "surrogate" for short, is non-parametric randomized linear version of the original data which preserves the linear properties of the original data. For identification of nonlinear/linear behavior in a given time series, the null hypothesis that the original data conform to a linear Gaussian stochastic process is formulated. An established method for generating constrained surrogates conforming to the properties of a linear Gaussian process is the Iterative Amplitude Adjusted Fourier Transform (iAAFT), which has become quite popular (Teolis 1998; Schreiber and Schmitz 1996, 2000; Kugiumtzis 1999). This type of surrogate time series retains the signal distribution and amplitude spectrum of the original time series, and takes into account a possibly nonlinear and static observation function due to the measurement process. The method uses a fixed point iteration algorithm for achieving this, for the details of which we refer to Schreiber and Schmitz (1996, 2000).

Wavelet-based surrogate generation is a fairly new method of constructing surrogate for hypothesis testing of nonlinearity which applies a wavelet decomposition of the time series. The main difference between Fourier transform and wavelet transform is that the former is only localized in frequency, whereas the latter is localized both in time and frequency. The idea of a wavelet representation is an orthogonal decomposition across a hierarchy of temporal and spatial scales by a set of wavelet and scaling functions.

The iAAFT-method has recently been refined using a wavelet-based approach, denoted by WiAAFT (Keylock 2006), that provides for *constrained realizations* of surrogate data that resembles the original data while preserving the local mean and variance as well as the power spectrum and distribution of the original except for randomizing the nonlinear properties of the signal. The WiAAFT-procedure follows the iAAFT-algorithm but uses the Maximal Overlap Discrete Wavelet Transform (MODWT) where the iAAFT-procedure is applied to each set of wavelet detail coefficients $D_j(n)$ over the dyadic scales 2^{j-1} for $j = 1, \cdots, J$, i.e., each set of $D_j(n)$ is considered as a time series of its own. The main difference between iAAFT and wiAAFT algorithms is that the former is designed to produce constrained, linear realizations of a process that can be compared with the original time series on some measure, while the later algorithm restricts the possible class of realizations to those that retain some aspect of the local mean and variance of the original time series (Keylock 2008).

Statistical analysis by the concept of *surrogate data* tests for a difference between a *test statistic* computed for the original and linearized versions of the data, i.e.,

an ensemble of realizations of the null hypothesis linear dynamics. For statistical testing of the null hypothesis of linearity, we follow Theiler et al. (1992) by using a non-parametric rank-order test. The degree of difference between the original and surrogate data is given by the ranked position of the data asymmetry with respect to the surrogates. For a right-tailed test, we generate at least $N_s = \frac{1}{\alpha} - 1$ surrogates, where α is the level of significance and N_s denotes the number of surrogates. The rank-threshold (or critical value) for right-tailed rank-order test is given by $(1-\alpha)(N_s+1)$. The null of linearity is rejected as soon as the rank-order statistic is greater than the rank-threshold. To achieve a minimal significance requirement of 95 % ($\alpha = 0.05$), we need at least 19 surrogates time series for right-tailed tests. Increasing the number of surrogates can increase the discrimination power (Schreiber and Schmitz 1996, 2000; Theiler et al. 1992). The concept of *surrogate data* will be incorporated into the *Delay Vector Variance* method *(below)* to examine the dynamics of an underlying economic indicator.

2.3 Optimal Embedding Parameters

In the context of signal processing, an established method for visualizing an attractor of an underlying nonlinear dynamical signal is by means of time delay embedding (Hegger et al. 1999). By time-delay embedding, the original time series $\{x_k\}$ is represented in the so-called "phase space" by a set of delay vectors (DVs) of a given embedding dimension, m, and time lag, τ : $x(k) = [x_{k-\tau}, \cdots, x_{k-m\tau}]$. Gautama et al. (2003) proposed a differential entropy based method for determining the optimal embedding parameters of a signal. The main advantage of this method is that a single measure is simultaneously used for optimizing both the embedding dimension and time lag. We provide below an overview of the procedure:

The "Entropy Ratio" is defined as

$$R_{ent}(m,\tau) = I(m,\tau) + \frac{m \ln N}{N}, \tag{5}$$

where N is the number of delay vectors, which is kept constant for all values of m and τ under consideration,

$$I(m,\tau) = \frac{H(x,m,\tau)}{\langle H(x_{s,i},m,\tau)\rangle_i} \tag{6}$$

where x is the signal, $x_{s,i}$ $i = 1, \cdots, T_s$ surrogates of the signal x, $\langle\cdot\rangle_i$ denotes the average over i, $H(x,m,\tau)$ denotes the differential entropies estimated for time delay embedded versions of a time series, x, which an inverse measure of the *structure* in the phase space. Gautama et al. (2003) proposed to use the Kozachenko-Leonenko (K-L) estimate (Leonenko and Kozachenko 1987) of the differential entropy given by

$$H(x) = \sum_{j=1}^{T} \ln(T\rho_j) + \ln 2 + C_E \tag{7}$$

where T is the number of samples in the data set, ρ_j is the Euclidean distance of the j-th delay vector to its nearest neighbor, and $C_E(\approx 0.5772)$ is the Euler constant. This ratio criterion requires a time series to display a clear structure in the phase space. Thus, for time series with no clear structure, the method will not yield a clear minimum, and a different approach needs to be adopted, possibly one that does not rely on a phase space representation. When this method is applied directly to a time series exhibiting strong serial correlations, it yields embedding parameters which have a preference for $\tau_{opt} = 1$. In order to ensure robustness of this method to the dimensionality and serial correlations of a time series, Gautama et al. (2003) suggested to use the iAAFT method for surrogate generation since it retains within the surrogate both signal distribution and approximately the autocorrelation structure of the original signal. In this Paper, we opt to use wavelet-based surrogate generation method, WiAAFT by in Keylock (2006), for reasons already discussed in the previous section.

2.4 *"Delay Vector Variance" Method*

The characterization of signal nonlinearities, which emerged in physics in the mid-1990s, have been successfully applied in predicting survival in heart failure cases and also adopted in practical engineering applications (Ho et al. 1997; Chambers and Mandic 2001). The "delay vector variance" (DVV) method (Gautama et al. 2004a) is a recently proposed phase space based method for signal characterization. It is more suitable for signal processing application because it examines the deterministic[2] nature of a time series and when combined with the concept of surrogate data, provides as additional account of the nonlinear behavior of the time series. The DVV-analysis is based on the calculation of the *target variance*, σ^{*2}, which is an inverse measure of the predictability of a time series. The algorithm is summarized below:

- For an optimal embedding dimension m and time lag τ, generate delay vector (DV): $x(k) = [x_{k-\tau}, \cdots, x_{k-m\tau}]$ and corresponding target x_k
- The mean μ_d and standard deviation, σ_d, are computed over all pairwise distances between DVs, $\|x(i) - x(j)\|$ for $i \neq j$.
- The sets Ω_k are generated such that $\Omega_k = \{x(i) | \|x(k) - x(i)\| \leq \varrho_d\}$, i.e., sets which consist of all DVs that lie closer to $x(k)$ than a certain distance ϱ_d, taken

[2]This means that the underlying process that generate the data can theoretically be described precisely by a set of linear or nonlinear equations. Thus, the component of a time series that can be predicted from a number of previous samples (Wold 1938).

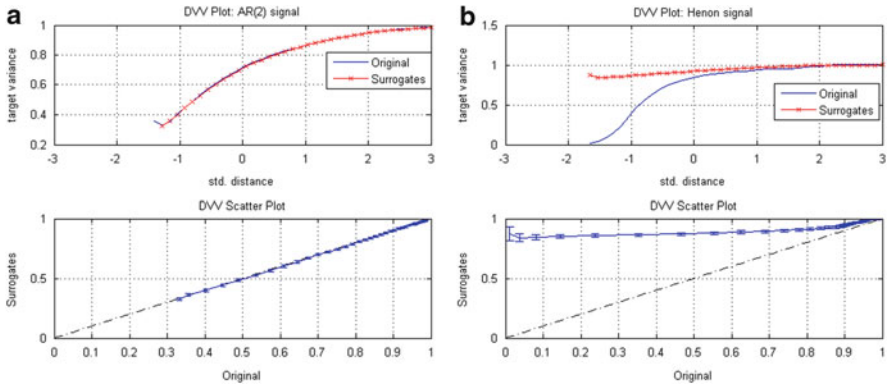


Fig. 2 Nonlinear and deterministic nature of signals. The *first row* of (**a**) and (**b**) are DVV plots for a linear benchmark signal: AR(2) signal and a nonlinear benchmark signal: Henon signal, where the *red line* with *crosses* denotes the DVV plot for the average of 25 WiAAFT-based surrogates while the *blue line* denotes that for the original signal. The *second row* of (**a**) and (**b**) denote the DVV scatter diagrams for those two signals, where *error bars* denote the standard deviation of the target variances of surrogates. (**a**) AR(2) signal. (**b**) Henon signal (Color figure online)

from the interval $[\min\{0, \mu_d - n_d\sigma_d\}; \mu_d + n_d\sigma_d]$, e.g., uniformly spaced, where n_d is a parameter controlling the span over which to perform the DVV analysis.

- For every set Ω_k, the variance of the corresponding targets, σ_k^2, is computed. The average over all sets Ω_k, normalized by the variance of the time series, σ_x^2, yields the *target variance*, σ^{*2}:

$$\sigma^{*2}(\varrho_d) = \frac{\frac{1}{N}\sum_{k=1}^{N}\sigma_k^2(\varrho_d)}{\sigma_x^2} \tag{8}$$

where N denotes the total number of sets $\Omega_k(\varrho_d)$

Graphical representation of DVV-analysis is obtained by plotting $\sigma^{*2}(\varrho_d)$ as function of the standardized distance, ϱ_d. The minimum *target variance*, $\sigma_{min}^{*2} = \min_{\varrho_d}[\sigma^{*2}(\varrho_d)]$, which corresponds to the lowest point of the curve, is a measure for the amount of noise which is present in the time series. Thus, σ_{min}^{*2} is inversely related to prevalence of the deterministic component over the stochastic one, lowest σ_{min}^{*2} indicating a strong deterministic component. At the extreme right, the DVV plots smoothly converge to unity, as illustrated in Fig. 2a,b. The reason behind this is that for maximum spans, all DVs belong to the same set, and the variance of the targets is equal to the variance of the time series.

The analysis addressing the linear or nonlinear nature of the original time series is examined by performing DVV analysis on both the original and a set of WiAAFT surrogate time series. Due to the standardization of the distance axis, these plots can be conveniently combined within a scatter diagram, where the horizontal axis corresponds to the DVV plot of the original time series, and the vertical to that of

the surrogate time series. If the surrogate time series yield DVV plots similar to that of the original time series, as illustrated by the first row of Fig. 2a, the DVV scatter diagram coincides with the bisector line, and the original time series is judged to be linear, as shown in second row of Fig. 2a. If not, as illustrated by first row of Fig. 2b, the DVV scatter diagram will deviate from the bisector line and the original time series is judged to be nonlinear, as depicted in the second row of Fig. 2b. Statistical testing of the null of linearity using a non-parametric rank-order test (Theiler et al. 1992) is performed to enhance robust conclusion of results obtained via the DVV-analysis. We refer the reader to Appendix "DVV Plots of Simulated Processes" for more on DVV analysis of some simulated process.

We provide below a summary of our methodology which can be characterized in two stages:

1. Stage One: Detection of Nonlinearity in the underlying time series.
   (a) We study the structure of the economic indicator via a phase-space representation using the differential entropy-based method with both iAAFT and wiAAFT surrogates. Embedding parameters that yields lower entropy ratio is selected for the DVV analysis in next step. The main advantage of this differential entropy-based method is that a single measure is simultaneously used to obtain the embedding dimension, m, and time lag, τ.
   (b) In order to detect the nonlinear behavior in the underlying time series, we use the DVV method discussed in Sect. 2.4. We are able to generate delay vectors necessary for the DVV analyzes using the (m, τ) obtained in the step (a). Unlike classical nonlinearity testing procedures, this non-parametric method is essentially data-driven and carry no a priori assumptions about the intrinsic properties or mathematical structure of the underlying time series. In particular, this method provides a straightforward visualization and interpretation of results. With this approach, we are able to obtain important information on the underlying economic indicator, which is essential in choosing the appropriate class of models suggested by the data itself. It is noteworthy that this procedure does not need the underlying time series to be stationary. Statistical testing of the null of linearity using a non-parametric rank-order test (Theiler et al. 1992) is performed to enhance robust conclusion of results obtained via the DVV-analysis.
2. Stage Two: Detection and explaining the business cycle.

   The next stage of the methodology deals with the problem of discovering pattern or hidden information that cannot be captured with traditional methods of business cycle analysis such as spectral analysis, which in only localized in frequency. In this work, we perform wavelet analysis using a complex-valued wavelet via a continuous wavelet transform in order to separate the phase and amplitude information. The phase information will be useful in explaining the economy-wide fluctuations in production that occur around a long-term growth trend. Information on the magnitude of such cycles across time will be obtained from the amplitude information.

3 Data Analysis

It is now well-known that the United States and all other modern industrial economies experience significant swings in economic activity. In this section, we perform analysis to characterize and detect nonlinear schemes for the US business cycle considering the monthly industrial production. The Industrial Production index is a business cycle indicator that has been widely used in business cycle analysis (see Artis et al. 2002, 2003; Anas et al. 2008; Billio et al. 2012a; Addo et al. 2013b). Firstly, we characterize the *nature* of the time series using the DVV method with both iAAFT and WiAAFT surrogates and then employ complex Morlet wavelet to discover the cycles or hidden information in the data. In particular, we show that this new methodology permits to study the dynamics of the underlying economic indicator without a priori assumptions on the statistical properties and also allows for the detection of recessions periods. In addition, we attempt to establish a comparison between the late-2000s financial crisis and the Great Depression of the 1930s.

The monthly US Industrial Production Index (IPI) time series[3] spanning over the period January, 1919 to July, 2012 is considered for the data analysis. Figure 3a is the plot of the monthly IPI series for the period: 1919:01–2012:07, implying 1123 observations, where the shaded regions corresponds to NBER[4] published dates for US recessions from 1920. Figure 3b is the plot of the IPI spectrum which can be interpreted as a presence of long memory dynamics in the data.

We now give a comprehensive analysis of the IPI in level. To begin with, we opted for the differential-entropy based method (Gautama et al. 2003) to determine the optimal embedding parameters, i.e., the embedding dimension, m, and the time lag, τ, for the DVV method with both the iAAFT surrogates and WiAAFT surrogates. We consider two approaches for estimating (m, τ). In the first case, the optimal embedding parameters are estimated using wiAAFT surrogates are $m = 3$ and $\tau = 1$ with an *entropy ratio*, $R_{ent}(m, \tau) = 0.7923$, indicated as an open circle in the diagram with a clear structure in Fig. 4b. This result indicates the presence of time correlations, in the time series, implying a higher degree of structure, thus, a lower amount of disorder. The second case is by using the iAAFT surrogates of which the estimated values of the optimal embedding parameters are $m = 4$ and $\tau = 7$ with *entropy ratio*, $R_{ent}(m, \tau) = 0.7271$, which is less than that obtained via wiAAFT surrogates. In selecting the embedding parameters to generate delayed vectors needed to perform the DVV analysis, we choose the estimates with lower *entropy ratio* implying a higher degree of structure. In this case, $m = 4$ and $\tau = 7$ is used to generate the delayed vectors needed to perform the DVV analysis.

[3]The data can be downloaded from Federal Reserve Bank of St. Louis: http://research.stlouisfed.org/fred2/.

[4]National Bureau of Economic Research: http://www.nber.org/cycles.html.

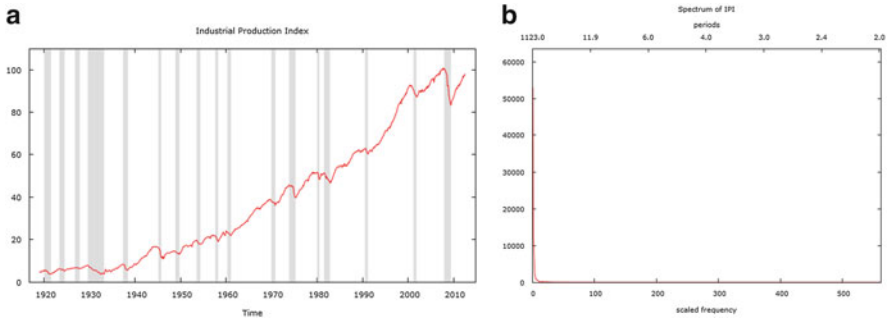


Fig. 3 US Industrial Production Index (IPI) time series. (**a**) is the plot of the monthly IPI series for the period: 1919:01–2012:07 ($n = 1123$), where the *shaded regions* corresponds to the US recessions from 1920 published by NBER. (**b**) is the plot of the Spectrum of IPI. (**a**) Industrial Production Index with the shaded areas indicating the US recessions. (**b**) The Spectrum of the Industrial Production Index (IPI) which can be interpreted as a presence of long memory dynamics

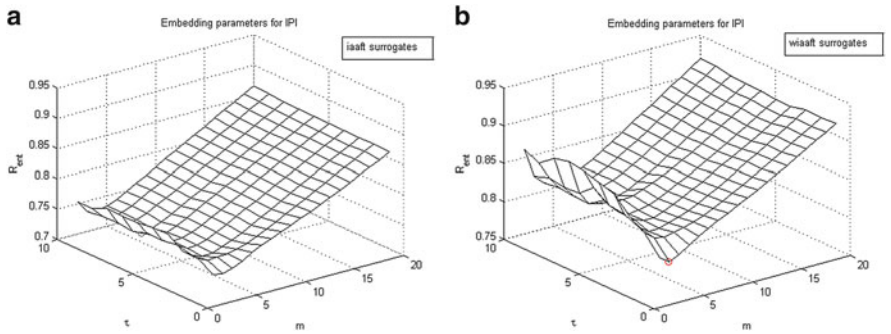


Fig. 4 The optimal embedding parameters obtained via the Differential-Entropy based method using the two types of surrogates are indicated as an *open circle* in the diagrams with a clear structure. We obtain a lower entropy $R_{ent}(m, \tau)$ with iAAFT surrogates which corresponds to a higher degree of structure. The values will be used is creating delay vectors needed for the DVV analysis. (**a**) Differential-Entropy based method with iAAFT surrogates. The optimal embedding values are $m = 4$ and $\tau = 7$ with *entropy ratio*, $R_{ent}(m, \tau) = 0.7271$. (**b**) Differential-Entropy based method with wiAAFT surrogates. The optimal embedding values are $m = 3$ and $\tau = 1$ with $R_{ent}(m, \tau) = 0.7923$

Based on the optimal embedding parameters $m = 4$ and $\tau = 7$, we generate delay vectors necessary for the DVV Analysis. The results from the DVV analysis, in Fig. 5, with iAAFT surrogates performed on the IPI indicates a clear deviation from the bisector on the DVV scatter diagram. The DVV plot also shows that the process is neither strictly deterministic or strictly stochastic. Thus, the original time series, IPI, exhibits nonlinear dynamics since the iAAFT surrogates are linear realizations of the original (Schreiber and Schmitz 1996, 2000). Statistical testing of the null of linearity using the non-parametric rank-order test, Table 1, indicates that the IPI is nonlinear. Thus, the DVV analysis suggests that the time series

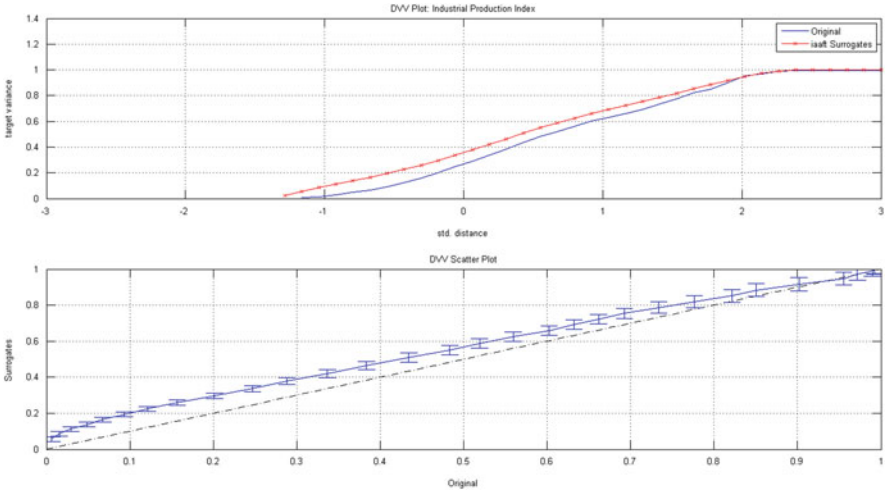


Fig. 5 This is the DVV analysis with iAAFT surrogates performed on the IPI using the embedding parameters obtained via the differential entropy-based method. We clearly observe a deviation from the bisector on the DVV scatter diagram. The DVV plot also indicates that the process is neither strictly deterministic or strictly stochastic. Thus, the original time series, IPI, exhibits nonlinear dynamics since the surrogates are linear realizations of the original (Schreiber and Schmitz 1996, 2000)

Table 1 Results of the non-parametric rank-order test. The null of linearity is rejected as soon as the *Rank-Order* is greater than *Rank-Threshold*. The code H takes the value 0 or 1, where $H = 0$ corresponds to failure of rejecting the null of linearity and $H = 1$ the rejection of linearity for nonlinearity. The number of iAAFT surrogates considered for the DVV-analysis is 25, which is greater than the minimum requirement of 19 surrogates for testing at $\alpha = 0.05$ level of significance

Data	Code, H	Rank-order	Rank-threshold	Decision
IPI	1	26	24.7	Nonlinear dynamics

under consideration, IPI, behaves more of a nonlinearity with neither a strictly deterministic or strictly component. The nonlinearity in the data could be due to both structural and behavioral changes that can occur in the economy across time. In otherwords, the nonlinearity may be as a result of existence of different *states of the world* or regimes in the economy. Many business cycle indicators present asymmetric features that have long been recognized in economics (Mitchell 1927; Keynes 1936). Putting it simply, there are sharp retractions during downturns in the economy as opposed to gradual upswings during recoveries (Kontolemis 1997; Sichel 1993; Ashley and Patterson 1936; Brock and Sayers 1988). Asymmetry has been recognised as a nonlinear phenomenon in several recent studies investigating various economic time series. Nonlinear models are therefore required to capture the features of the data generating mechanism of inherently asymmetric realizations of some of the macroeconomic business cycle series, since linear models are incapable of generating such behaviour (Granger and Terasvirta 1993; Terasvirta and

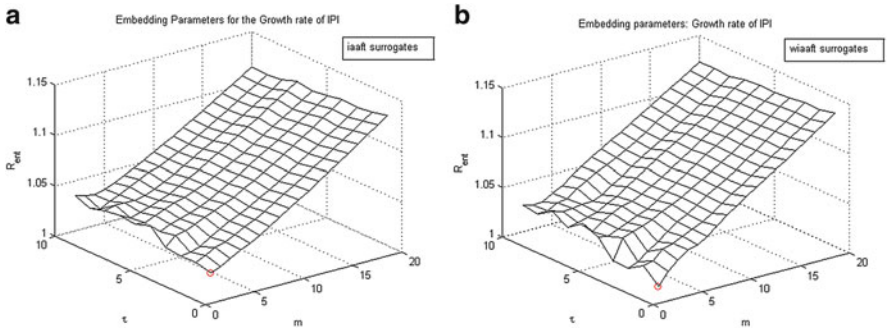


Fig. 6 The optimal embedding parameters obtained via the Differential-Entropy based method using the two types of surrogates are indicated as an *open circle* in the diagrams with a clear structure. The values of the embedding parameters are $m = 2$ and $\tau = 1$ in both diagrams. This result $\tau = 1$ indicates the presence of time correlations, in the growth rate of IPI, implying a higher degree of structure, thus, a lower amount of disorder. (**a**) Differential-Entropy based method with iAAFT surrogates on the growth rate of the IPI. (**b**) Differential-Entropy based method with wiAAFT surrogates on the growth rate of the IPI

Anderson 1992; Terasvirta 1994; Dias 2003). Thus, some possible class of nonlinear models such as Markov switching models, smooth transition autoregressive (STAR) models, threshold autoregressive models, could capture such nonlinear behavior (Kim and Nelson 1999; Granger and Terasvirta 1993; Franses and van Dijk 2000; Addo et al. 2014; Billio et al. 2012b).

On understanding the dynamics of the growth rate (suppose we denote the IPI as X_t, then the growth rate of IPI defined as $Y_t = log(X_t) - log(X_{t-1})$) of IPI to study the business cycle, we obtain the same estimates of embedding parameters $m = 2$ and $\tau = 1$ using the differential entropy-based method with both iAAFT surrogates and WiAAFT surrogates (Fig. 6). Using the values of the embedding parameters, we are able to generate the phase space representation as displayed in Fig. 7 and perform the DVV analysis in Fig. 8. The purpose of studying the growth rate of IPI is not to ensure stationarity but to enable a better comparison of IPI dynamics over time. Business cycles are usually measured by considering the growth rate of industrial production index or the growth rate of real gross domestic product.

The DVV analysis, in Fig. 8 and the statistical testing of the null of linearity using the non-parametric rank-order test, Table 2, suggests that the time series under consideration behaves more of a nonlinear stochastic process than a deterministic one.

In the following step, we perform CWT on the IPI growth rate using wavelets of the form in Eq. (3) at different bandwidths f_b and center frequency f_c. In detecting the recession dates, the wavelet analysis was first performed for the period 1919:02–1940:01 using the US recession dates published by NBER as benchmark. The Morlet wavelet that captures the recession dates in this sample is then chosen as the wavelet to be used for the whole sample period. The Morlet wavelet that best detect cycles and hidden information in the data for the period 1919:02–1940:01, is given in Eq. (4). The colormap used in the coefficient plots and scalogram plot ranges from

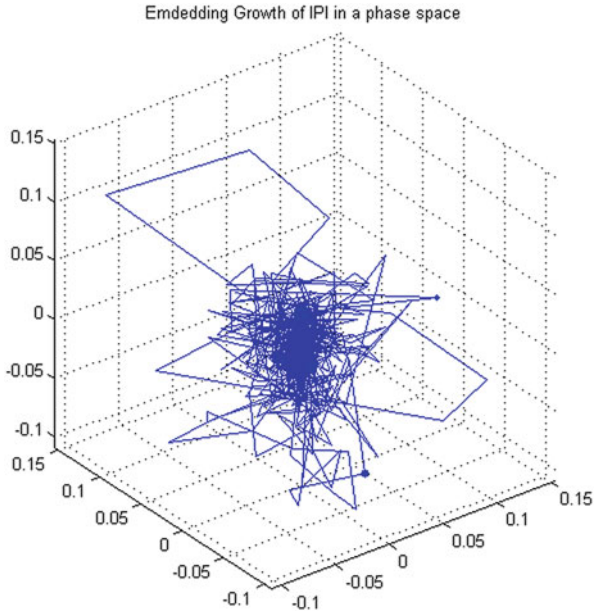


Fig. 7 Phase Space reconstruction using the embedding parameters $m = 2$ and $\tau = 1$. This represents the embedding of the underlying time series, growth rate of IPI, in phase space. The attractor is clearly visualized

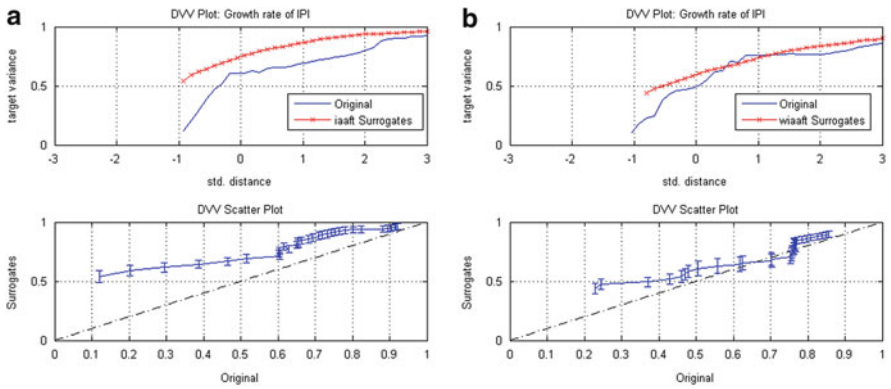


Fig. 8 The DVV analysis on Growth rate of IPI indicates that it is characterize by nonlinear dynamics. (**a**) DVV with iAAFT surrogates. (**b**) DVV with wiAAFT surrogates

blue to red, and passes through the colors cyan, yellow, and orange. The blue *root-like* structures on the phase-angle plot of the Coefficient plots, in Fig. 9, corresponds to recession periods of the economy. These are the periods where the economy-wide fluctuations in production are below the long-term growth trend.

Table 2 Results of the non-parametric rank-order test. The null of linearity is rejected as soon as the *Rank-Order* is greater than *Rank-Threshold*. The code H takes the value 0 or 1, where $H = 0$ corresponds to failure of rejecting the null of linearity and $H = 1$ the rejection of linearity for nonlinearity. The number of surrogates considered for the DVV-analysis is 25, which is greater than the minimum requirement of 19 surrogates for testing at $\alpha = 0.05$ level of significance

Data	Surrogates	Code, H	Rank-order	Rank-threshold	Decision
Growth rate of IPI	wiAAFT	1	25	24.7	Nonlinear dynamics
Growth rate of IPI	iAAFT	1	26	24.7	Nonlinear dynamics

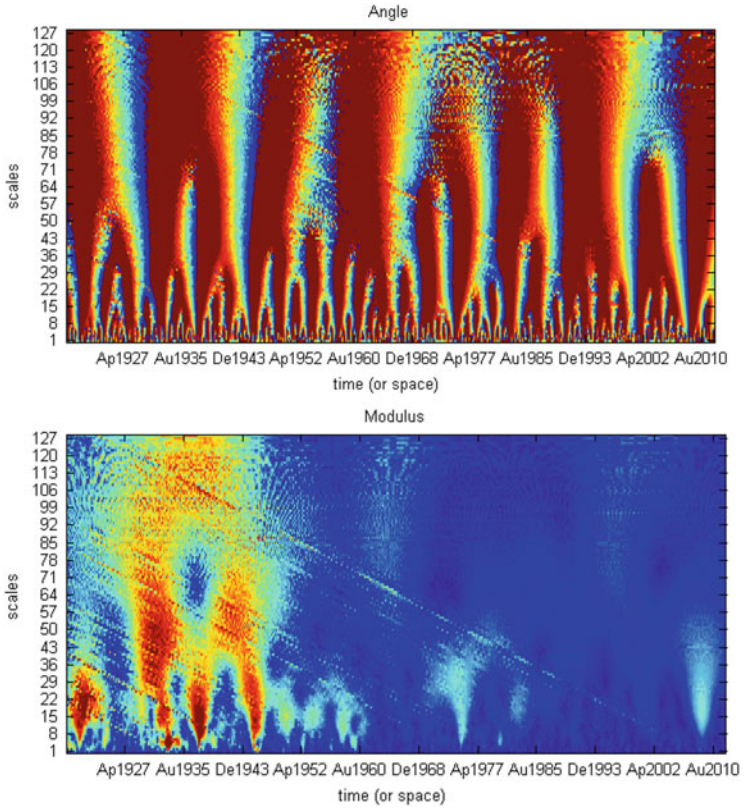


Fig. 9 Coefficients plots obtained from the CWT using complex Morlet wavelet on the growth rate of IPI: *First row* represents the phase (angle) plot and *second row* is the corresponding Modulus plot. The colormap ranges from *blue* to *red*, and passes through the colors *cyan*, *yellow*, and *orange*. The *blue regions* on the Angle Coefficient plot corresponds to periods of relative stagnation in the economy from 1920. Thus, we consider only such structures with a minimum of 6 months as recession in the economy. The corresponding amplitudes can be read from the Modulus plot (Color figure online)

Table 3 Business cycle peaks and troughs in the United States, 1920–2009. The peak and trough dates, in the format YYYY:MM, represent the start and end of "episodes" of some sort. (http://www.nber.org/cycles.html)

Peak	Trough	Time index
1920:01	1921:07	13–31
1923:05	1924:07	53–67
1926:10	1927:11	94–107
1929:08	1933:03	128–171
1937:05	1938:06	221–235
1945:02	1945:10	314–322
1948:11	1949:10	359–370
1953:07	1954:05	415–425
1957:08	1958:04	464–472
1960:04	1961:02	496–506
1969:12	1970:11	613–623
1973:11	1975:04	659–676
1980:01	1980:07	733–739
1981:07	1982:11	751–767
1990:07	1991:03	859–867
2001:03	2001:11	987–995
2007:12	2009:06	1068–1086

The detection of the recession dates are represented by blue *root-like* structures on the angle coefficient plot in Fig. 9. We consider only such structures with a minimum of 6 months[5] as recession in the economy. The corresponding magnitude of these cycles can be read from the modulus plot of the coefficient plot in Fig. 9. The Wall Street Crash of 1929, followed by the Great Depression of the 1930s—the largest and most important economic depression in the twentieth century—are well captured on the phase-angle coefficient plot in the Fig. 9 for time period (128–235), reported on Table 3. The three recessions between 1973 and 1982: the oil crisis—oil prices soared, causing the stock market crash are shown on the blue *root-like* structures of the phase-angle coefficient plot in the Fig. 9 for the time periods (659–676), (733–739), (751–767). Furthermore, the bursting of dot-com bubble—speculations concerning Internet companies crashed is also detected for the time periods (987–995). The wavelet energy at time periods are displayed on the scalogram in Fig. 10 and the pseudo-frequency corresponding to scales are displayed in Fig. 11. This interesting finding provide support for the use of wavelet methodology in business cycle modeling.

In order to compare the late-2000s financial crisis with the Great Depression of the 1930s, we perform the wavelet analysis on the growth rate of the IPI. The IPI growth rate dynamics are well captured by the phase-angle coefficient plot in Fig. 9, where the blue *root-like* structures corresponds to periods of relative stagnation in the economy from 1920. The amplitudes associated with these economic fluctuations can be read from the modulus plot in Fig. 9. Looking at

[5]This is a known censuring period accepted in Business Cycle literature and National Bureau of Economic Research: http://www.nber.org/cycles.html.

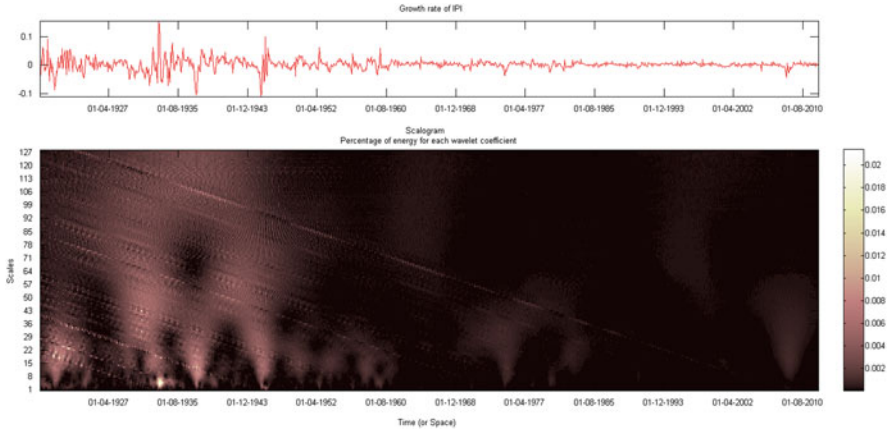


Fig. 10 The IPI growth rate and associated Scalogram from the CWT. The bar with colour scales on the *left-hand side* of the scalogram plot indicates the percentage of energy for each wavelet coefficient. Higher energy levels can be clearly observed for the Great Depression of the 1930s compared to the period of late-2000s financial crisis, also known as the Global Financial Crisis (Color figure online)

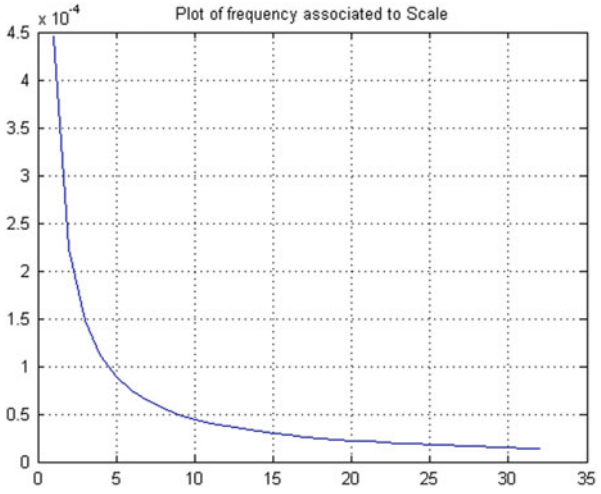


Fig. 11 The pseudo-frequency associated to scale, in Hertz (Hz). The horizontal axis represents the scales and the vertical axis corresponds to the frequency associated to a scale

the modulus plot in Fig. 9 and scalograms in Figs. 10 and 12, we clearly observe higher amplitude and energy levels on the interval (0.008–0.016 %) corresponding to the Great depression of the 1930s compared to the late-2000s financial crisis with energy levels below 0.004 %. These results, based on the data set we used,

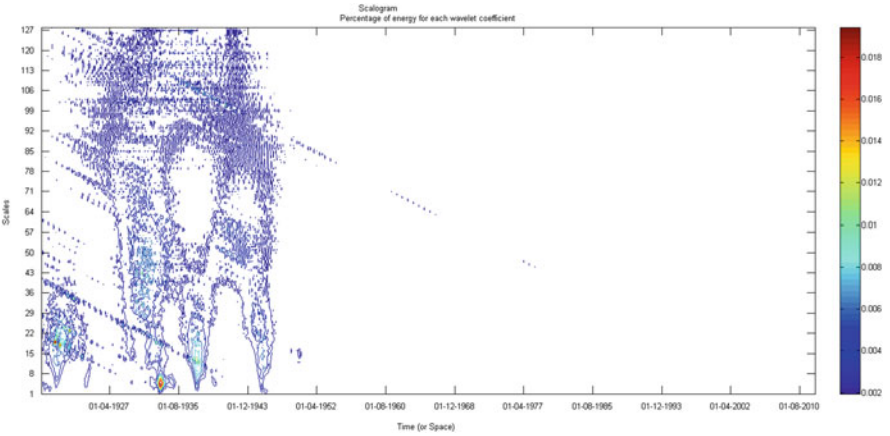


Fig. 12 A contour representation of Scalogram, Fig. 10, associated with the US IPI growth rate

suggests that the intensity of the late-2000s financial crisis, also known as the Global Financial Crisis (GFC) is at the moment not so high as compared to the Great Depression of the 1930s.

4 Conclusion

In this paper, we have proposed a methodology in business cycle modeling which encompasses different existing methods successfully applied in physics and engineering. Our proposed procedure allows to first study the dynamics of the underlying economic indicator using non-parametric methods which are essentially data-driven and carry no a priori assumptions on the statistical properties, such as possible non-stationarity, or mathematical structure of the time series. We have provided a comprehensive analysis of the feasibility of our approach as essential in selecting the appropriate class of models suggested by the data itself. Finally, we have demonstrated the usefulness of wavelets in discovering patterns or hidden information that vary in nature across time which cannot be captured using traditional methods of business cycle analysis such as the correlation analysis and spectral analysis.

Acknowledgements The authors are grateful to the Editors and anonymous referees for their careful revision, valuable suggestions, and comments that have improved this paper. We thank the conference participants of ISCEF 2012, CFE–ERCIM 2012, COMPSTAT 2012 and the participants at the Econometrics Internal Seminar at Center for Operations Research and Econometrics (CORE) for their participation and interest. We also would like to thank Sébastien Van Bellegem, Luc Bauwens, Christian Hafner, Timo Terasvirta and Yukai Kevin Yang for their remarks and questions. In this paper, we made use of the algorithms in wavelet toolbox in Matlab and DVV toolbox avail-

able from www.commsp.ee.ic.ac.uk/ mandic/dvv.htm. The wavelet-based (wiAAFT) surrogates algorithm used in this paper may be downloaded from http://www.chriskeylock.net/page2.aspx. The first author acknowledges financial support under Erasmus Mundus fellowship.

Appendix 1: Wavelet Analysis

Notation and Operators

Consider the mapping

$$\Gamma : \mathcal{L}^2(\mathbb{R}) \longrightarrow \mathcal{L}^2(\mathbb{R})$$

Let $f \in \mathcal{L}^2(\mathbb{R})$, $\alpha, \beta \in \mathbb{R}$ and $s \in \mathbb{R}^+$ where $\mathbb{R}^+ := \{t \in \mathbb{R} : t > 0\}$. Unless otherwise stated, the complex conjugate of $z \in \mathbb{C}$ is denoted $\bar{z}$ and the magnitude of z is denoted $|z|$. The symbol i will represent the square root of -1, i.e., $i^2 = -1$. We present in Table 4 some notations and operators that will be often referred to in this manuscript.

Continuous Wavelet Transform (CWT)

The continuous wavelet transform (CWT) differs from the more traditional short time Fourier transform (STFT) by allowing arbitrarily high localization in time of high frequency signal features. The CWT permits for the isolation of the high frequency features due to it's variable window width related to the scale of observation. In particular, the CWT is not limited to using sinusoidal analyzing functions but allows for a large selection of localized waveforms that can be employed as long as they satisfy predefined mathematical criteria (described below).

Let $\mathcal{H}$ be a Hilbert space, the CWT may be described as a mapping parameterized by a function ψ

$$C_\psi : \mathcal{H} \longrightarrow C_\psi(\mathcal{H}). \tag{9}$$

The CWT of a one-dimensional function $f \in \mathcal{L}^2(\mathbb{R})$ is given by

$$\begin{aligned} C_\psi : \mathcal{L}^2(\mathbb{R}) &\longrightarrow C_\psi(\mathcal{L}^2(\mathbb{R})) \\ f &\mapsto \langle f, \tau_t D_s \psi \rangle_{\mathcal{L}^2(\mathbb{R})} \end{aligned} \tag{10}$$

where $\tau_t D_s \psi$ is a dilated (by s) and translated (by t) version of ψ given as

$$(\tau_t D_s \psi)(\xi) = \frac{1}{|s|^{\frac{1}{2}}} \psi\Big(\frac{\xi - t}{s}\Big) \tag{11}$$

Table 4 Notations and operators

Operator, Γ	Notation, Γf	Output	Inverse, $\Gamma * f = \Gamma^{-1} f$	Fourier transform, $(\Gamma f)\hat{}$
Dilation	$(D_s f)(t)$	$\frac{1}{s^{1/2}} f(\frac{t}{s})$	$D_{s^{-1}} f$	$D_{s^{-1}} \hat{f}$
Involution	$\tilde{f}$	$\bar{f}(-t)$	$\tilde{f}$	$\bar{\hat{f}}$
Translation	$(\tau_\alpha f)(t)$	$f(t-\alpha)$	$\tau_{-\alpha} f$	$e_{-a} \hat{f}$
Modulation	$(e_\alpha f)(t)$	$e^{i2\pi\alpha t} f(t)$	$e_{-\alpha} f$	$\tau_\alpha \hat{f}$
Reflection	$(Rf)(t)$	$f(-t)$	Rf	$R\hat{f}$

Thus, the CWT of one-dimensional signal f is a two-dimensional function of the real variables time t, and scale $s \neq 0$. For a given ψ, the CWT may be thought of in terms of the representation of a signal with respect to the wavelet family generated by ψ, that is, all it's translated and dilated versions. The CWT may be written as

$$(C_\psi f)(t,s) := \langle f, \tau_t D_s \psi \rangle \tag{12}$$

For each point (t,s) in the time-scale plan, the wavelet transform assigns a (complex) numerical value to a signal f which describes how much f like a translated by t and scaled by s version of ψ.

The CWT of a signal f is defined as

$$(C_\psi f)(t,s) = \frac{1}{|s|^{\frac{1}{2}}} \int_{\mathbb{R}} f(\xi) \bar{\psi}\big(\frac{\xi - t}{s}\big) d\xi \tag{13}$$

where $\bar{\psi}(\xi)$ is the complex conjugate of the analyzing wavelet function $\psi(\xi)$. Given that ψ is chosen with enough time-frequency localization,[6] the CWT gives a gives a picture of the time-frequency characteristics of the function f over the whole time-scale plane $\mathbb{R} \times (\mathbb{R} \backslash \{0\})$. When $C_{adm,\psi} < \infty$, it is possible to find the inverse continuous transformation via the relation known as *Calderón's reproducing identity*,

$$f(\xi) = \frac{1}{C_{adm,\psi}} \int_{\mathbb{R}^2} \langle f, \tau_t D_s \psi \rangle \tau_t D_s \psi(\xi) \frac{1}{s^2} ds\, dt. \tag{14}$$

and if s restricted in $\mathbb{R}^+$, then the *Calderón's reproducing identity* takes the form

$$f(\xi) = \frac{1}{C_{adm_+,\psi}} \int_{-\infty}^{\infty} \int_0^{\infty} \langle f, \tau_t D_s \psi \rangle \tau_t D_s \psi(\xi) \frac{1}{s^2} ds\, dt. \tag{15}$$

[6]The time-frequency concentrated functions, denoted $TF(\mathbb{R})$, is a space of complex-valued finite energy functions defined on the real line that decay faster than $\frac{1}{t}$ simultaneously in the time and frequency domains. This is defined explicitly as $TF(\mathbb{R}) := \{\varphi \in \mathcal{L}^2(\mathbb{R} : |\varphi(t)| < \eta(1+|t|)^{-(1+\varepsilon)} \quad and \quad |\hat{\varphi}(\gamma)| < \eta(1+|\gamma|)^{-(1+\varepsilon)} for \quad \eta < \infty, \varepsilon > 0\}$.

Let α and β be arbitrary real numbers and f, f_1, and f_2 be arbitrary functions in $\mathcal{L}^2(\mathbb{R})$ The CWT, C_ψ, with respect to ψ satisfies the following conditions:

1. Linearity
   - $(C_\psi(\alpha f_1 + \beta f_2))(t,s) = \alpha(C_\psi f_1)(t,s) + \beta(C_\psi f_2)(t,s)$
2. Time Invariance
   - $(C_\psi(\tau_\beta f))(t,s) = (C_\psi f)(t-\beta, s)$
3. Dilation
   - $(C_\psi(D_\alpha f))(t,s) = (C_\psi f)(\alpha t, \alpha^{-1}s)$
4. Negative Scales
   - $C_\psi f(t,-s) = (C_\psi Rf)(-t,s)$

The time invariance property of the CWT implies that the wavelet transform of a time-delayed version of a signal is a time-delayed version of its wavelet transform. This serves as an important property in terms of pattern recognition. This nice property is not readily obtained in the case of Discrete wavelet transforms (Addison 2005; Walden and Percival 2000).

The contribution to the signal energy at the specific scale s and location t is given by

$$\mathcal{E}(t,s) = |C_\psi|^2 \tag{16}$$

which is a two-dimensional wavelet energy density function known as the scalogram. The wavelet transform C_ψ corresponding to a complex wavelet is also complex valued. The transform can be separated into two categories:

- Real part $\mathcal{R}\{C_\psi\}$ and Imaginary part $\mathcal{I}\{C_\psi\}$
- Modulus (or Amplitude), $|C_\psi|$ and phase (or phase-angle), $\Phi(t,s)$,

which can be obtained using the relation:

$$C_\psi = |C_\psi| e^{i\,\Phi(t,s)} \quad \textit{and} \quad \Phi(t,s) = \arctan\left(\frac{\mathcal{I}\{C_\psi\}}{\mathcal{R}\{C_\psi\}}\right). \tag{17}$$

Maximal Overlap Discrete Wavelet Transform

The Maximal Overlap Discrete Wavelet Transform (MODWT), also related to notions of "cycle spinning" and "wavelet frames", is with the basic idea of downsampling values removed from discrete wavelet transform. The MODWT unlike the conventional discrete wavelet transform (DWT), is non-orthogonal and highly redundant, and is defined naturally for all sample sizes, N (Walden and Percival 2000). Given an integer J such that $2^J < N$, where N is the number

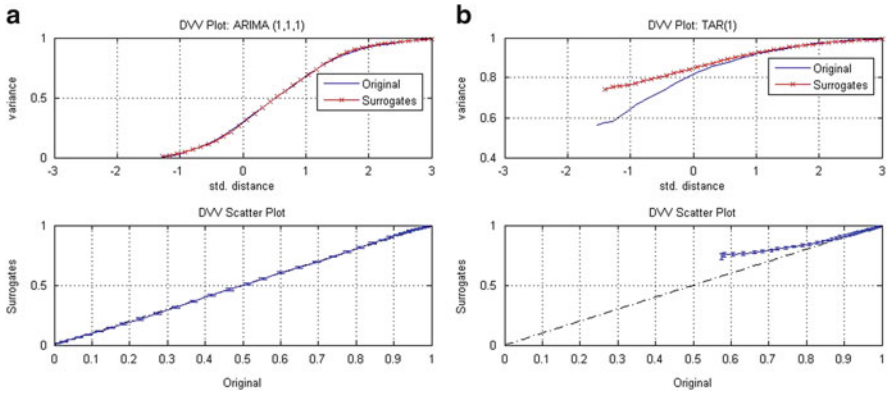


Fig. 13 DVV analysis on ARIMA and Threshold Autoregressive signals. (**a**) DVV analysis on ARIMA(1,1,1) signal. (**b**) DVV analysis on TAR(1) signal

of data points, the original time series represented by the vector $X(n)$, where $n = 1, 2, \cdots, N$, can be decomposed on a hierarchy of time scales by details, $D_j(n)$, and a smooth part, $S_J(n)$, that shift along with X:

$$X(n) = S_J(n) + \sum_{j=1}^{J} D_j(n) \tag{18}$$

with $S_j(n)$ generated by the recursive relationship

$$S_{j-1}(n) = S_J(n) + D_J(n). \tag{19}$$

The MODWT details $D_j(n)$ represent changes on a scale of $\tau = 2^{j-1}$, while the $S_j(n)$ represents the smooth or approximation wavelet averages on a scale of $\tau_J = 2^{J-1}$. Gallegati and Gallegati (2007) employed this wavelet transform to investigate the issue of moderation of volatility in G-7 economies and also to detect the importance of the various explanations of the moderation.

Appendix 2: DVV Plots of Simulated Processes

We provide the structure of the DVV analysis on some simulated processes such as: a Threshold autoregressive process (TAR), linear autoregressive integrated moving average (ARIMA) signal, a Generalized autoregressive conditional heteroskadastic process (GARCH), and a Bilinear process (Figs. 13 and 14).

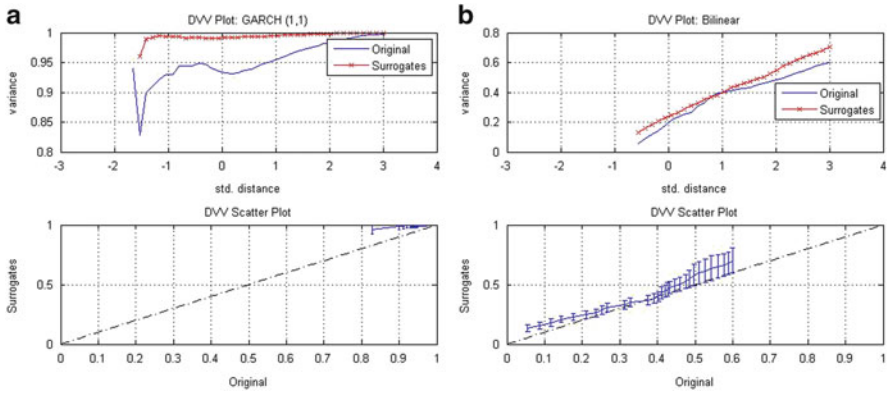


Fig. 14 DVV analysis on GARCH and Bilinear signals. (**a**) DVV analysis on GARCH(1,1). (**b**) DVV analysis on Bilinear signal

References

Addison PS (2005) Wavelet transforms and ecg : a review. Physiol Meas 26:155–199

Addo PM, Billio M, Guégan D (2012) Understanding exchange rate dynamics. In: Colubi A, Fokianos K, Kontoghiorghes EJ (eds) Proceedings of the 20th international conference on computational statistics, pp 1–14

Addo PM, Billio M, Guégan D (2013a) Nonlinear dynamics and recurrence plots for detecting financial crisis. North Am J Econ Financ. http://dx.doi.org/10.1016/j.najef.2013.02.014

Addo PM, Billio M, Guégan D (2013b) Turning point chronology for the eurozone: a distance plot approach. J Bus Cycle Meas Anal (forthcoming)

Addo PM, Billio M, Guégan D (2014) The univariate mt-star model and a new linearity and unit root test procedure. Comput Stat Data Anal. http://dx.doi.org/10.1016/j.csda.2013.12.009

Anas J, Billio M, Ferrara L, Mazzi GL (2008) A system for dating and detecting turning points in the euro area. Manchester Schools 76(5):549–577.

Artis MJ, Marcellino M, Proietti T (2002) Dating the euro area business cycle. CEPR Discussion Papers No 3696 and EUI Working Paper, ECO 2002/24

Artis MJ, Marcellino M, Proietti T (2003) Dating the euro area business cycle. CEPR Discussion Papers (3696)

Ashley RA, Patterson DM (1936) Linear versus nonlinear macroeconomics: A statistical test. Int Econ Rev 30:685–704

Billio M, Casarin R, Ravazzolo F, van Dijk HK (2012a) Combination schemes for turning point predictions. Q Rev Econ Financ 52:402–412

Billio M, Ferrara L, Guégan D, Mazzi GL (2013) Evaluation of regime-switching models for real-time business cycle analysis of the euro area. J. Forecast 32:577–586. doi:10.1002/for.2260

Brock WA, Sayers CL (1988) Is the business cycle characterised by deterministic chaos. J Monet Econ 22:71–90

Bruce L, Koger C, Li J (2002) Dimentionality reduction of hyperspectral data using discrete wavelet transform feature extraction. IEEE Trans Geosci Remote Sens 40:2331–2338

Burns AF, Mitchell WC (1946) Measuring business cycles. NBER

Chambers D, Mandic J (2001) Recurrent neural networks for prediction: learning algorithms architecture and stability. Wiley, Chichester

Dias FC (2003) Nonlinearities over the business cycle: An application of the smooth transition autoregressive model to characterize gdp dynamics for the euro-area and portugal. Bank of Portugal, Working Paper 9-03

Franses PH, van Dijk D (2000) Non-linear time series models in empirical finance. Cambridge University Press, Cambridge

Gallegati M (2008) Wavelet analysis of stock returns and aggregate economic activity. Comput Stat Data Anal 52:3061–3074

Gallegati M, Gallegati M (2007) Wavelet variance analysis of output in g-7 countries. Stud Nonlinear Dyn Econ 11(3):6

Gautama T, Mandic DP, Hulle MMV (2003) A differential entropy based method for determining the optimal embedding parameters of a signal. In: Proceedings of ICASSP 2003, Hong Kong IV, pp 29–32

Gautama T, Mandic DP, Hulle MMV (2004a) The delay vector variance method for detecting determinism and nonlinearity in time series. Phys D 190(3–4):167–176

Gautama T, Mandic DP, Hulle MMV (2004b) A novel method for determining the nature of time series. IEEE Trans Biomed Eng 51:728–736

Granger CWJ, Terasvirta T (1993) Modelling nonlinear economic relationship. Oxford University Press, Oxford

Guttorp B, Whitcher P, Percival D (2000) Wavelet analysis of covariance with application to atmospheric time series. J Geophys Res 105:14941–14962

Hegger R, Kantz H, Schreiber T (1999) Practical implementation of nonlinear time series methods: The tisean package. Chaos 9:413–435

Ho A, Moody K, Peng G, Mietus C, Larson J, Levy M, Goldberger D (1997) Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics. Circulation 96:842–848

Jensen M (1999) An approximate wavelet mle of short and long memory parameters. Stud Nonlinear Dyn Econ 3:239–253

Kaplan D (1994) Exceptional events as evidence for determinism. Phys D 73(1):38–48

Keylock CJ (2006) Constrained surrogate time series with preservation of the mean and variance structure. Phys Rev E 73:036707

Keylock CJ (2008) Improved preservation of autocorrelative structure in surrogate data using an initial wavelet step. Nonlinear Processes Geophys 15:435–444

Keynes JM (1936) The general theory of employment, interest and money. Macmillan, London

Kim CJ, Nelson CR (1999) State-space models with regime-switching: classical and Gibbs-sampling approaches with applicationss. MIT Press, Cambridge

Kontolemis ZG (1997) Does growth vary over the business cycle? some evidence from the g7 countries. Economica 64:441–460

Kugiumtzis D (1999) Test your surrogate data before you test for nonlinearity. Phys Rev E 60:2808–2816

Leonenko NN, Kozachenko LF (1987) Sample estimate of the entropy of a random vector. Probl Inf Transm 23:95–101

Luukkonen R, Saikkonen P, Terasvirta T (1988) Testing linearity against smooth transition autoregressive models. Biometrica 75:491–499

Mitchell WC (1927) Business cycles. the problem and its setting. National Bureau of Economic Research, New York

Ramsey J (1999) The contributions of wavelets to the analysis of economic and financial data. Philos Trans R Soc Lond A 357:2593–2606

Ramsey J, Lampart C (1998) The decomposition of economic relationships by time scale using wavelets: expenditure and income. Stud Nonlinear Dyn Econ 3:23–42

Schreiber T, Schmitz A (1996) Improved surrogate data for nonlinearity tests. Phys Rev Lett 77:635–638.

Schreiber T, Schmitz A (2000) Surrogate time series. Phys D 142:346–382

Sichel DE (1993) Business cycle asymmetry: a deeper look. Econ Inquiry 31:224–236
Stock MW, Watson JH (1999) Forecasting inflation. J Monet Econ 44(2):293–335
Teolis A (1998) Computational signal processing with wavelets. Birkhauser, Basel
Terasvirta T (1994) Specification, estimation and evaluation of smooth transition autoregressive models. J Am Stat Assoc 89:208–218
Terasvirta T (September 2011) Modelling and forecasting nonlinear economic time series, wGEM workshop, European Central Bank
Terasvirta T, Anderson HM (1992) Characterizing nonlinearities in business cycles using smooth transition autoregressive models. J Appl Econ 7:119–136
Theiler JD, Eubank J, Longtin S, Galdrikian A, Farmer B (1992) Testing for nonlinearity in time series: the method of surrogate data. Phys A 58:77–94
van Dijk D, Terasvirta T, Franses PH (2002) Smooth transition autoregressive models -a survey of recent developments. Econ Rev 21:1–47
Walden DB, Percival AT (2000) Wavelet methods for time series analysis. Cambridge University Press, Cambridge
Weigend MC, Casdagli AS (1994) Exploring the continuum between deterministic and stochastic modelling, in time series prediction: Forecasting the future and understanding the past. Addison-Wesley, Reading, pp 347–367
Wold HOA (1938) A study in the analysis of stationary time series. Almquist and Wiksell, Uppsala

Part II
Volatility and Asset Prices

Measuring the Impact Intradaily Events Have on the Persistent Nature of Volatility

Mark J. Jensen and Brandon Whitcher

Abstract In this chapter we measure the effect a scheduled event, like the opening or closing of a regional foreign exchange market, or a unscheduled act, such as a market crash, a political upheaval, or a surprise news announcement, has on the foreign exchange rate's level of volatility and its well documented long-memory behavior. Volatility in the foreign exchange rate is modeled as a non-stationary, long-memory, stochastic volatility process whose fractional differencing parameter is allowed to vary over time. This non-stationary model of volatility reveals that long-memory is not a spurious property associated with infrequent structural changes, but is a integral part of the volatility process. Over most of the sample period, volatility exhibits the strong persistence of a long-memory process. It is only after a market surprise or unanticipated economic news announcement that volatility briefly sheds its strong persistence.

1 Introduction

The current class of long-memory, stochastic volatility models assume the conditional variance is a stationary long-memory process with a fractional differencing parameter that is constant and does not change its value over time. These models of volatility assume that a stationary environment exists in the financial world and ignore the regular intradaily market micro-structure style events and unexpected market crashes, surprises, and political disruptions that are prevalent in the data

M.J. Jensen (✉)
Federal Reserve Bank of Atlanta, 1000 Peachtree Street, Atlanta, GA 3030–4470, USA
e-mail: Mark.Jensen@atl.frb.org

B. Whitcher
Pfizer Worldwide Research & Development, 620 Main Street, Cambridge, MA 02139, USA
e-mail: brandon.whitcher@pfizer.com

M. Gallegati and W. Semmler (eds.), *Wavelet Applications in Economics and Finance*, Dynamic Modeling and Econometrics in Economics and Finance 20,
DOI 10.1007/978-3-319-07061-2_5,

(see Andersen and Bollerslev 1998).[1] In this paper, we combine high-frequency intradaily behavior with low-frequency interdaily long-memory dynamics and model volatility as a time-varying, long-memory, stochastic volatility model where the fractional differencing parameter of Breidt et al. (1998) and Harvey (2002) long-memory, stochastic volatility model are allowed to vary over time.

We define the model:

$$y_t = \sigma(t) \exp\{h_t/2\}\epsilon_t, \tag{1}$$

$$(1-B)^{d(t)} h_t = \sigma_\eta(t)\eta_t, \qquad t = 1, \ldots, T, \tag{2}$$

where at time t the mean corrected return from holding a financial instrument is y_t and h_t is the unobservable log-volatility, as a time-varying, long-memory, stochastic volatility (TVLMSV) model. At every t, $|d(t)| < 1/2$, and ϵ_t and η_t are uncorrelated Gaussian white noise processes. The fractional differencing operator, $(1-B)^{d(t)}$, where B is the lag operator, $x_{t-s} = B^s x_t$, is defined by the binomial expansion:

$$(1-B)^{d(t)} = \sum_{l=0}^{\infty} \frac{\Gamma(l-d(t))}{\Gamma(l+1)\Gamma(-d(t))} B^l,$$

where $\Gamma(\cdot)$ is the gamma function. The parameter $\sigma_\eta(t)$ is the standard deviation of the log-volatility and $\sigma(t)$ is the modal instantaneous volatility.

The TVLMSV is a non-stationary model belonging to the Dahlhaus (1997) class of locally stationary processes. In the TVLMSV model, the fractional differencing parameter, $d(t)$, the modal instants volatility, $\sigma(t)$, and the volatility of volatility, $\sigma_\eta(t)$, are smooth functions that change value over time. This time-varying property allows the TVLMSV model to produce responses to volatility shocks that are not only persistent in the classical long-memory sense of a slow decaying autocorrelation function, but depending on when a shock takes place, the level of persistence associated with the shock can vary. For example, the time-varying fractional differencing parameter enables the TVLMSV to model levels of persistence unique to volatility over the operating hours of a regional exchange, or to its dynamics over an entire 24 hour trading day. The TVLMSV is also flexibly enough that by setting $d(t) = d$, $\sigma(t) = \sigma$, and $\sigma_\eta(t) = \sigma_\eta$, for all t, it becomes the stationary, long-memory, stochastic volatility model. Hence, the time-varying differencing parameter of the TVLMSV model equips us with the means of determining whether the long-memory found in volatility is structural ($d(t) > 0$, over a wide range of values for t) or just a spurious artifact of unaccounted regime changes or shocks ($d(t) = 0$, over a wide range of values of t) as suggest by Russell et al. (2008), Jensen and Liu (2006), Diebold and Atusushi (2001), Lamoureux and Lastrapes (1990), and Lastrapes (1989).

[1] Departure from the assumption of stationarity can be found in Stăricică and Granger (2005).

We estimate the latent volatility's time-varying fractional differencing parameter by projecting log-squared returns into the time-scale domain of the wavelet basis. Where Fourier analysis is well designed to study stationary processes with basis functions that are localized in the frequency domain, the wavelet is constructed to analyze non-stationary, time-varying processes. With basis functions that are localized in both time and frequency space, wavelets are well equipped to detect long-memory behavior like the Fourier's spectral density. But with the additional time dimension of its basis function, the wavelet is also able to locate time dependent behavior like structural breaks and discontinuities in its equivalent spectral density measure. The wavelet accomplishes this with basis functions that are small and tight on the time domain for high frequencies, while having basis functions with large and long time support for low frequencies. This inverse relationship between the wavelet basis's frequency and time support allows the wavelet to synthesis the long-memory behavior found in the TVLMSV power spectrum, without giving up the TVLMSV model's time-varying behavior.

Applying our wavelet estimator of $d(t)$ to a years worth of five-minute, log-squared returns of the Deutsche mark-US dollar exchange rate, we find the values of $d(t)$ are significantly positive over a large portion of the year. In response to economics news and market surprise, the value of $d(t)$ quickly declines in value, becoming either slightly negative or insignificant from zero. This decline, however, is very short lived and last for approximately a day after the event causing the drop. The drop in the value of the time-varying fractional differencing parameter suggests that volatility is generally strongly persistent, but when the foreign exchange market receives new public information the volatility remains persistent but for a short period of time in a negative way (anti-persistent).

The intraday average of the estimated $d(t)$ at each five-minute interval over the entire year reveals a diurnal pattern in volatility's degree of long-memory. On average $d(t)$ is at its largest value as the Asian markets open. It then slowly declines over the course of the Asian market's trading day, continuing to drop in value as the Asian markets close and the European markets open. The intradaily average of $d(t)$ declines until it reaches its lowest value of the trading day exactly at the same moment the London market closes. The time-varying differencing parameter then begins to slowly rise through the operating hours of the North American markets. Through the course of a 24 hour trading day, the only market operations affecting the value of $d(t)$ on average is the opening of the Tokyo exchange and the closing of the markets in London.

The chapter proceeds as follows. In Sect. 2 we define the non-stationary class of locally stationary processes and in Sect. 3 present a stochastic volatility model where latent volatility follows a locally stationary, long-memory process. We show in Sect. 4 how the discrete wavelet transform of log-squared returns produces a log-linear relationship between the wavelet coefficients local variance and the scaling parameter that is equal to the time-varying, long-memory parameter. We then apply in Sect. 5 the estimator to a years worth of five-minute Deutsche mark-US dollar foreign exchange rate data.

2 Locally Stationary Processes

Dahlhaus (1997) defines the locally stationary class of non-stationary processes as a triangular array, $X_{t,T}$, $t = 1, \ldots, T$, that adheres to the following definition.

Definition 1. Define the class of non-stationary processes as **locally stationary processes** if the triangular array, $X_{t,T}$, $t = 1, \ldots, T$, with transfer function A^0, and drift μ, has the spectral representation:

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int e^{i\omega t} A^0_{t,T}(\omega)\, dZ(\omega) \tag{3}$$

where:

A1. the complex stochastic process $Z(\omega)$ is defined on $[-\pi, \pi]$ with $\overline{Z(\omega)} = Z(-\omega)$, $E[Z(\omega)] = 0$ and:

$$E[dZ(\omega)dZ(\omega')] = \zeta(\omega + \omega')\, d\omega\, d\omega'$$

where $\zeta(\omega) = \sum_{j=-\infty}^{\infty} \delta(\omega + 2\pi j)$ is the period 2π extension of the Dirac delta function, $\delta(\omega)$, where $\int f(\omega)\delta(\omega)\, d\omega = f(0)$, for all functions f continuous at 0.

A2. there exists a complex-valued 2π-periodic function, $A(u, \omega) : [0, 1] \times [-\pi, \pi] \to \mathscr{C}$, with $A(u, \omega) = \overline{A(u, -\omega)}$, that is uniformly Lipschitz continuous in both arguments with index $\alpha > 1/2$, and a constant K such that:

$$\sup_{t,\omega} \left| A^0_{t,T}(\omega) - A(t/T, \omega) \right| \leq KT^{-1} \tag{4}$$

for all T.[2]

A3. the drift $\mu(u)$ is continuous in $u \in [0, 1]$.

Except for the transfer function, $A^0_{t,T}$, having the time, t, and number of observations, T, subscripts, Eq. (3) is the same as the spectral representation of a stationary process (see Brockwell and Davis (1991), Theorem 4.8.2). Like a stationary process, Assumption A.1 ensures that $Z(\omega)$ is a continuous stochastic process with orthogonal increments, such as a Brownian motion process. The only difference between the spectral representation of a stationary process and the representation of the locally stationary process is in Assumption A.2. Assumption A.2 precisely quantifies the smoothness of the time-varying transfer function, $A(t/T, \omega)$, over not only the frequency domain, but also over the time domain. The class of uniformly Lipschitz functions includes transfer functions that are not only smooth

[2] Our notation is such that u will always represent a time point in the rescaled time domain $[0, 1]$; i.e., $u = t/T$.

differentiable functions in time and frequency, but this function class also contains non-differentiable transfer functions that have bounded variation.

Under Assumption A.2, Dahlhaus (1996) shows that the locally stationary process, $X_{t,T}$, has a well defined time-varying spectrum. In a manner analogous to the spectral density measuring a stationary process's energy at different frequencies, the time-varying spectral density of $X_{t,T}$ equals $f(u,\omega) \equiv |A(u,\omega)|^2$. Combining the triangular process, $X_{t,T}$, with the smoothness conditions of $A(u,\omega)$, Dahlhaus (1996, Theorem 2.2) overcomes the redundant nature of the time-frequency representation of a non-stationary process. Rather than there being a number of spectral representations of $X_{t,T}$, Dahlhaus shows that if there exists a spectral representation of the form found in Definition 1 with a "smooth" $A(u,\omega)$ as quantified by Assumption A.2, then $f(u,\omega)$ is unique. In other words, there is not just one spectral representation associated with $X_{t,T}$. However, the representation found in Definition 1 is the only one where the transfer function is "smooth" and the $f(u,\omega)$ is unique.

In Definition 1, as $T \to \infty$, more and more of $X_{t,T}$ behavior as determined by $A(u_o,\omega)$, where $u \in [u_o - \epsilon, u_o + \epsilon]$, will be observed. This definition of asymptotic theory is similar to those found in nonparametric regression estimation. Since future observations of $X_{t,T}$ tell us nothing about the behavior of a non-stationary process at an earlier t, in our setting $T \to \infty$ has the interpretation of measuring the series over the same time period but at a higher sampling rate. Phillips (1987) referred to this form of asymptotic as continuous record asymptotic since in the limit a continuous record of observations is obtained. In the context of Phillips (1987), the locally-stationary process $X_{t,T}$ can be regarded as a triangular array of a dually index random variable, $\{\{X_{nt}\}_{t=1}^{T_n}\}_{n=1}^{\infty}$ where as $n \to \infty$, $T_n \to \infty$ and the length between observations, k_n, approaches zero such that $k_n T_n = N$ so that the time interval $[0, N]$ may be considered fixed. Given that the existing asymptotic results for locally stationary processes are mostly due to Dahlhaus (1996, 1997), we choose to follow his notation and use the triangular array, $X_{t,T}$, to describe a locally stationary series.

2.1 *Locally Stationary Long-Memory Stochastic Volatility*

Applying the asymptotics found in Definition 1 to the TVLMSV model of Eqs. (1)–(2), we define the triangular array:

$$y_{t,T} = \sigma(t/T)\exp\{h_{t,T}/2\}\epsilon_t \tag{5}$$

$$(1-B)^{d(t/T)}h_{t,T} = \sigma_\eta(t/T)\eta_t \tag{6}$$

where $t = 1,\ldots,T$, $|d(u)| < 1/2$, and $d(u)$, $\sigma(u)$, and $\sigma_\eta(u)$ are continuous on $\mathbb{R}$ with $d(u) = d(0)$, $\sigma(u) = \sigma(0)$, $\sigma_\eta(u) = \sigma_\eta(0)$ for $u < 0$, and $d(u) = d(1)$, $\sigma(u) = \sigma(1)$, $\sigma_\eta(u) = \sigma_\eta(1)$ for $u > 1$, and differentiable for $u \in (0,1)$ with bounded derivatives.

Like a stationary stochastic volatility model, a key feature of $y_{t,T}$ is that it can be made to be linear in the latent volatility, $h_{t,T}$ by applying the log-squared transformation to the returns:

$$y^*_{t,T} \equiv \log y^2_{t,T} = \log \sigma^2(t/T) + h_{t,T} + \log \epsilon^2_t.$$

Because ϵ_t is Gaussian white noise, $\log \epsilon^2_t$ will be distributed as a log-$\chi^2_{(1)}$ with mean, -1.2704, and variance, 4.93. By adding and subtracting this mean from $y^*_{t,T}$ we obtain:

$$y^*_{t,T} = \mu(t/T) + h_{t,T} + z_t$$

where $\mu(t/T) = \log \sigma^2(t/T) - 1.2704$ and $z_t = \log \epsilon^2_t + 1.2704$.

We now state in theorem that $y^*_{t,T}$ is a locally stationary process in the sense of Definition 1 and as such refer to Eqs. (5)–(6) as a locally-stationary, long-memory, stochastic volatility model.[3]

Theorem 1. *Let $y_{t,T}$ be the triangular array defined by Eqs. (5)–(6), where ϵ_t and η_t are uncorrelated Gaussian white noise processes, $|d(u)| < 1/2$, and $d(u)$, $\sigma(u)$, and $\sigma_\eta(u)$ are continuous on $\mathbb{R}$ with $d(u) = d(0)$, $\sigma(u) = \sigma(0)$, $\sigma_\eta(u) = \sigma_\eta(0)$ for $u < 0$, and $d(u) = d(1)$, $\sigma(u) = \sigma(1)$, $\sigma_\eta(u) = \sigma_\eta(1)$ for $u > 1$, and differentiable for $u \in (0, 1)$ with bounded derivatives. Then $y^*_{t,T}$ is the locally stationary process:*

$$y^*_{t,T} = \mu(t/T) + \int e^{i\omega t} A^o_{t,T}(\omega) dZ(\omega) + \sqrt{4.93/2\pi} \int e^{i\omega t} dZ_z(\omega),$$

with the smooth transfer function:

$$A(u, \omega) = \frac{\sigma_\eta(u)}{\sqrt{2\pi}} (1 - e^{-i\omega})^{-d(u)},$$

and time-varying spectral density:

$$f(u, \omega) = \frac{\sigma^2_\eta(u)}{2\pi} |1 - e^{-i\omega}|^{-2d(u)} + 4.93/2\pi, \tag{7}$$

where $\mu(t/T) = \log \sigma^2(t/T) - 1.2704$, $z_t = \sqrt{4.93/2\pi} \int e^{i\omega t}\, dZ_z(\omega)$, where Z_z is a complex stochastic process found in Assumption A.1.

From Theorem 1, $y^*_{t,T}$ is not only a random process, it is also a locally stationary process with time-varying spectrum, $f(u, \omega)$. The only difference between $f(u, \omega)$ and the spectrum from a stationary, long-memory stochastic volatility model is that the fractional differencing, the modal instantaneous volatility, and the variance of

[3] All proofs are rendered to Appendix 1.

volatility parameters are allowed to change their value over time. In the TVLMSV model the risk of holding a financial asset is time-varying in the context that volatility follows a stationary stochastic process, but now the risk is not time homogeneous. In other words, the level of persistence associated with a shock to the conditional variance now depends on when the shock takes place. Suppose that at some point in time $d(t)$ is close to $1/2$. Shocks to volatility during this time period will be more persistent than if the same size shock occurred during a time period when $d(t)$ is closer to zero.

3 Maximal Overlap Discrete Wavelet Transform

To estimate $d(t)$, we introduce the discrete wavelet transform (DWT). While wavelets have been making a substantial impact in a broad array of disciplines from statistics to computer imagery, the fields of economics and finance have just skimmed the surface of its usefulness (see, for example, Gallegati et al. 2014; Rua and Nunes 2009; Crowley 2007; Hong and Kao 2004; Gençay et al. 2001, 2005; Hong and Lee 2001; Jensen 2004, 2000, 1999a,b; Ramsey 1999; Ramsey and Lampart 1998a,b). In contrast to the well-localized frequency basis functions of Fourier analysis and its spectral representation of stationary processes, wavelet analysis is designed around well-localized basis functions in both time and frequency. Being localized in time the wavelet is ideally suited for locally stationary processes. For example, wavelets capture the short, intraday volatility patterns found in foreign exchange rate data with basis functions that have small time support; whereas the long-run, interday dynamics of the foreign exchange data are captured by those wavelet's basis functions with time supports that are large. The intuition of the frequency properties of wavelets runs in the opposite direction. When short-run (long-run) properties of a time series are analyzed, the support of the wavelet basis function's Fourier transform is on an octave of high (low) frequencies.

In this paper we use a modified version of the DWT called the maximal overlap discrete wavelet transform (MODWT).[4] Both the DWT and the MODWT draw on multiresolution analysis to decompose a time series into lower and lower levels of resolutions. In wavelet terminology, these different levels of resolution are referred to as the wavelet scale.[5] In terms of multiresolution analysis, the wavelet transform decomposes a time series into weighted moving average values ("smooths") and the information required to reconstruct the signal ("details") from the averages. At each scale the MODWT coefficients constitute a time series describing the original series at coarser and coarser levels of resolution, not in a time aggregate manner, but in a manner where information that is being lost as the original series is aggregated over

[4] See Percival and Walden (2000) for an introduction to the MODWT.

[5] See Mallat (1989) for the seminal article on wavelets as presented from a multiresolution analysis point of view.

longer and longer time intervals. For example, if one observes five minute return data and aggregates these returns into ten minute returns the "details" at time scale five minutes would equal the information needed to construct the five minute returns from the ten minute returns.

Unlike the DWT, the MODWT is not an orthogonal basis. Instead, it is redundant and uses an approximate zero-phase filter to produce an over-determined representation of a time series. An advantage of this redundancy is that the "details" at each time scale and the "smooth" have the same number of observations as the original time series. This enables the "details" and the "smooth" to be aligned in time with the original series so that the impact of an event can be analyzed over different time scales.

To facilitate the introduction of our estimator of $d(t)$, consider applying the MODWT to the locally stationary process defined in Definition 1, $X_{t,T}$. Letting $j = 1, \ldots, J$, be the scale parameters, where $J \leq \log_2 T$ is the longest time interval over which the original time series is aggregated, and $\{\tilde{h}_{j,l} \,|\, l = 0, \ldots, L_j\}$ be the level-j, real-valued, MODWT wavelet filters, where $L_1 = L < T$ is an even, positive, integer and $L_j = (2^j - 1)(L - 1) + 1$. The level-$j$ MODWT coefficients of $X_{t,T}$ are obtained from the linear, circular filter[6]:

$$\tilde{W}_{j,t,T} = \sum_{l=0}^{L_j - 1} \tilde{h}_{j,l} X_{t-l \bmod T,T} \quad t = 1, \ldots, T, \tag{8}$$

where the wavelet filter $\{\tilde{h}_{j,l}\}$ satisfies:

A.4.

$$\sum_{l=0}^{L_j-1} \tilde{h}_{j,l} = 0, \; \sum_{l=0}^{L_j-1} \tilde{h}_{j,l}^2 = 2^{-j},$$

$$\sum_{l=0}^{L_j-1-2n} \tilde{h}_{j,l}\tilde{h}_{j,l+2n} = \begin{cases} 2^{-j}, & n = 0 \\ 0, & n = 1, 2, \ldots, (L_j - 2)/2. \end{cases}$$

Notice from the MODWT coefficients, $\tilde{W}_{j,t,T}$, $t \geq L_j$, (those that do not involve the circularity assumption), that the above filter is compactly supported on the time interval $[t - L_j + 1, t]$. From the above definition of L_j this time support increases as the scale j increases. In Appendix 2 we show that just the opposite is the case in the frequency domain representation of the MODWT; i.e., as j increases the frequency support of the transfer function of $\{\tilde{h}_{j,l}\}$ shrinks and covers a lower octave of frequencies. This inverse relationship between time and frequency

[6]The third equation in Assumption A.4 guarantees the orthogonality of the filters to double shifts and the first two conditions ensure that the wavelet has at least one vanishing moment and normalizes to one, respectively.

domain properties where a large (small) time support is associated with low (high) frequencies is one of the wavelets many strengths.

Now define the MODWT scaling filters as the quadrature mirror filters of the MODWT wavelet filters:

$$\tilde{g}_{j,l} \equiv (-1)^{l+1}\tilde{h}_{j,L_j-1-l}, \qquad l = 0, \ldots, L_j - 1. \tag{9}$$

From *A.4* it follows that the MODWT scaling filter satisfies the conditions:

$$\sum_{l=0}^{L_j-1} \tilde{g}_{j,l} = 1, \ \sum_{l=0}^{L_j-1} \tilde{g}_{j,l}^2 = 2^{-j}, \text{ and } \sum_{l=0}^{L_j-1} \tilde{g}_{j,l}\tilde{h}_{j,l} = 0.$$

The Jth-order scaling filters enable us to define the level-J MODWT scaling coefficients in terms of the filter:

$$\tilde{V}_{J,t,T} = \sum_{l=0}^{L_J-1} \tilde{g}_{J,l} X_{t-l \bmod T,T} \quad t = 1, \ldots, T. \tag{10}$$

Because the wavelet filter $\{\tilde{h}_{j,l}\}$ sums to zero, has the same length, L_j, as the scaling filter $\{\tilde{g}_{j,l}\}$, and the two filters are orthogonal, the wavelet filters represent the difference between two windowed weighted averages each with bandwidths of effective width 2^j; i.e., a MODWT coefficient tells how much a weighted moving average over a particular time period of length 2^j changes to the next. The level-J MODWT scaling coefficients are associated with the output from a weighted moving average with a window of length 2^J that captures the variation in $X_{t,T}$ over time periods associated with scales J and higher.

By multiresolution analysis, the level-j wavelet "details" associated with the MODWT coefficients are defined as:

$$\tilde{\mathscr{D}}_{j,t,T} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} \tilde{W}_{j,t+l \bmod T,T} \quad t = 1, \ldots, T, \tag{11}$$

and the level-J wavelet "smooths" equal:

$$\tilde{\mathscr{S}}_{J,t,T} = \sum_{l=0}^{L_J-1} \tilde{g}_{J,l} \tilde{V}_{J,t+l \bmod T,T} \quad t = 1, \ldots, T. \tag{12}$$

Since information about the original series is lost in calculating the weighted moving averages, the "details" are the portion of the MODWT synthesis associated with the changes at scale j, whereas the "smooth" is the portion attributable to variation

of scale J and higher. Together the "details" and "smooths" form an additive decomposition:

$$X_{t,T} = \sum_{j=1}^{J} \tilde{\mathscr{D}}_{j,t,T} + \tilde{\mathscr{S}}_{J,t,T}.$$

3.1 Locally-Stationary MODWT Coefficients

Because the MODWT coefficients, $\tilde{W}_{j,t,T}$, capture the information that is lost as the MODWT *intertemporally pans out* on the locally stationary process, $X_{t,T}$, it follows that the MODWT wavelet coefficients at each scale are themselves locally-stationary processes. This is the result of the following theorem:

Theorem 2. *Let $X_{t,T}$ be a locally-stationary process as defined in Definition 1 where the 2π-periodic function $A(u,\omega)$ has third derivative with respect to u that is uniformly bounded on $u \in (0,1)$ and $|\omega| \leq \pi$, and $\{\tilde{h}_{j,l} \,|\, l = 0,\ldots,L_j\}$, $j = 1,\ldots,J \leq \log_2 T$, be maximal overlap discrete wavelet transform filters that satisfy Assumption A.4, then the MODWT wavelet coefficient for a given scale, j, that do not involve the circularity assumption are locally-stationary processes with spectral representation:*

$$\tilde{W}_{j,t,T} = \int e^{i\omega t} A^{o}_{j,t,T}(\omega)\, dZ(\omega) \tag{13}$$

with $\sup_{t,\omega} \left| A^{o}_{j,t,T}(\omega) - A_j(t/T,\omega) \right| \leq KT^{-1}$, *where* $A_j(t/T,\omega) = \tilde{\mathscr{H}}_j(\omega) A(t/T,\omega)$, *and time-varying spectral density function* $f_j(u,\omega) = |\tilde{\mathscr{H}}_j(\omega)|^2 f(u,\omega)$.

4 Estimating Time-Varying Long-Memory in Volatility

Applying Eq. (8) to $h_{t,T}$, we define the level-j MODWT wavelet coefficients of the locally stationary, long-memory process as:

$$\tilde{W}^{(h)}_{j,t,T} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} h_{t-l \bmod T,T} \quad t = 1,\ldots,T.$$

Using the definition of $\tilde{h}_{j,l}$ transfer function found in Appendix 2, Jensen and Whitcher (1999) show that $\tilde{W}^{(h)}_{j,t,T}$ is a locally stationary process with mean zero and time-varying variance

$$\mathrm{Var}(\tilde{W}^{(h)}_{j,t,T}) = \nu^2_h(u,\tau_j) \sim \sigma^2(u)\, 2^{j(2d(u)-1)} \qquad \text{as } j \to \infty, \tag{14}$$

where

$$\sigma(u)^2 = \frac{2^{1-d(u)}\sigma_\epsilon^2\pi^{-2d(u)}(2^{1-d(u)}-1)}{1-2d(u)}.$$

Taking the logarithmic transformation of Eq. (14), we obtain the log-linear relationship

$$\log \nu_h^2(u,\tau_j) = \alpha(u) + D(u)\log 2^j \tag{15}$$

where $D(u) = 2d(u) - 1$.

Jensen and Whitcher (1999) use the time-scale nature of $\tilde{W}_{j,t,T}^{(h)}$ to estimate the time-varying variance, $\nu_h^2(u,\tau_j)$, with the sample variance calculated from the wavelet coefficients whose support contain the point t. Because the support of the level-j wavelet filter includes several filter coefficients with a value close to zero, a 'central portion' of the wavelet filter $\Lambda_j = \{-[g(L_j)/2],\ldots,[g(L_j)/2]\}$, where $0 < g(L_{j-1}) < g(L_j)$, is utilized.[7] In other words, $\nu_h^2(u,\tau_j)$ was estimated with the time-varying sample variance of the wavelet coefficients

$$\tilde{\nu}_h^2(u,\tau_j) = \frac{1}{\#\Lambda_j}\sum_{l\in\Lambda_j}(\tilde{W}_{j,t+l,T}^{(h)})^2. \tag{16}$$

where $\#\Lambda_j$ is the number of elements in the set Λ_j.

Given the interpretation of wavelet coefficients as a measure of the information lost to lower and lower sampling rates of a series, the time-varying sample variance in Eq. (16) makes intuitive sense. At small (large) values of the scaling parameter j, in other words measuring $X_{t,T}$ more (less) frequently, the time support of Λ_j is small (large). So as information is lost to lower levels of sampling frequency, the series becomes smoother as it loses the high frequency dynamics seen in data sampled at shorter time intervals. The series short-lived behavior associated with the high-frequency dynamics of singularities, jumps and cusps no longer occur at the larger values of j and as a result are not found in the behavior of the corresponding wavelet coefficients. Hence, it make sense to take a large time bandwidth when calculating the time-varying sample variance of the wavelet coefficients with a large j, and a tight time bandwidth when computing the time-varying sample variance of the wavelet coefficients at small j.

Jensen and Whitcher (1999) prove $\tilde{\nu}_h^2$ to be a consistent estimator of ν_h^2 and also show that by replacing $\log \nu_h^2$ in Eq. (15) with $\log \tilde{\nu}_h^2$, the OLS estimator of $D(u)$ is a consistent estimator. It follows that the OLS estimator $d(u) = (D(u)+1)/2$ will be a consistent estimator of the time-varying differencing parameter, $d(u)$.

[7] Since the 'central portion' is dependent on the particular family and order of the wavelet filter, there is no closed-form expression for $g(L_j)$. However, Whitcher and Jensen (2000) do tabulate the time width of Λ_j for the Daubechies family of wavelets.

Although $y^*_{t,T}$ is comprised of the locally stationary, long-memory process, $h_{t,T}$, and the white noise processes, z_t, the local second order properties of $y^*_{t,T}$ and $h_{t,T}$ are asymptotically equivalent. The ratio of the time-varying spectrum of $y^*_{t,T}$ and $h_{t,T}$

$$\frac{f_{y^*}(u,\omega)}{f_h(u,\omega)} \to K, \qquad \text{as } \omega \to 0,$$

for some $K < \infty$. Furthermore, it is well known that wavelets are well equipped to filter out unwanted white noise (Donoho and Johnstone 1994, 1995, 1998; Jensen 2000). The variance of $y^*_{t,T}$ MODWT coefficients equals

$$\nu^2_{y^*}(u,\tau_j) \sim \sigma(u)^2 2^{j(2d(u)-1)} + \sigma_z^2 \qquad \text{as } j \to \infty,$$

Since there is no structure to be found in the white noise process, z_t, neither the wavelet coefficients from the white noise process, nor their variances will exhibit any systematic decay or relationship with the scaling parameter. Thus, the OLS wavelet estimator will provide a good estimator of the TVLMSV model.[8]

5 Intraday, Long-Memory Behavior of the DM-US Dollar

The Deutsche mark-US dollar (DM-$) exchange rate has been extensively used to investigate the intra- and inter-daily behavior of foreign exchange rates (Andersen et al. 2003, 2001; Andersen and Bollerslev 1997a,b, 1998; Müller et al. 1990; Baillie and Bollerslev 1990). In its time the interbank spot DM-$ market had the largest turnover of any financial market, was highly liquid, and had low transaction costs. Furthermore, the spot market for DM-$ was a 24-hour market comprised of sequential but not mutually exclusive regional trading centers. Except for the endogenous slow downs in the level of trading that occurred on weekends and regional holidays, the spot DM-$ market was essentially always open. This makes the market for the DM-$ ideal for analyzing the time-varying, long-memory behavior of volatility.

5.1 Data

We use the tick-by-tick DM-$ spot exchange rate bid and ask quotes recorded by Olsen and Associates from the interbank Reuters network over the time period

[8] A possible alternative to our semi-parametric OLS estimator of $d(u)$ is the MCMC methodology of Jensen (2004), but this would first require developing a Bayesian sampling method for locally stationary processes. This is a topic for future research.

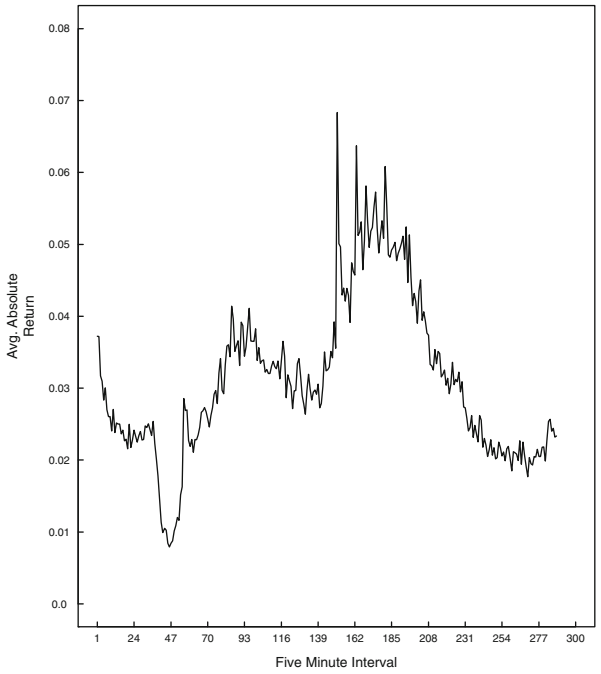


Fig. 1 Intraday average of the absolute value for the 288 daily five-minute returns for the DM-$ exchange rate

October 1, 1992 to September 30, 1993 to construct 74,880 five-minute returns. These returns are computed with the linear interpolation method of Andersen and Bollerslev (1997b) where the quote immediately before and after the end of a five-minute interval are weighted inversely by their distance to the five-minute point. The price is then defined as the midpoint of the logarithmic bid and ask at the five-minute tick. Because of the inactivity found in the market all returns from Friday 21.00 Greenwich mean time (GMT) through Sunday 21.00 GMT are excluded. This leaves us with 260 daily trading cycles in a year, with each day consisting of 288 observations.

In order to determine if the time-varying, long-memory parameter follows the typical intradaily pattern found in the volatility of the DM-$ return, we plot in Fig. 1 the average of each of the 288 five-minute absolute returns over all 260 trading days. As originally found by Andersen and Bollerslev (1997b, 1998), Fig. 1 shows the time-of-day effects caused by the openings and closings of the Asian, European and U.S. markets, the drop in volatility during Hong Kong and Tokyo's lunch hours (between intervals 40 and 60), and the sharp increase in volatility experienced in the afternoon trading sessions of the Europe market and the opening of the U.S. market (interval 156).

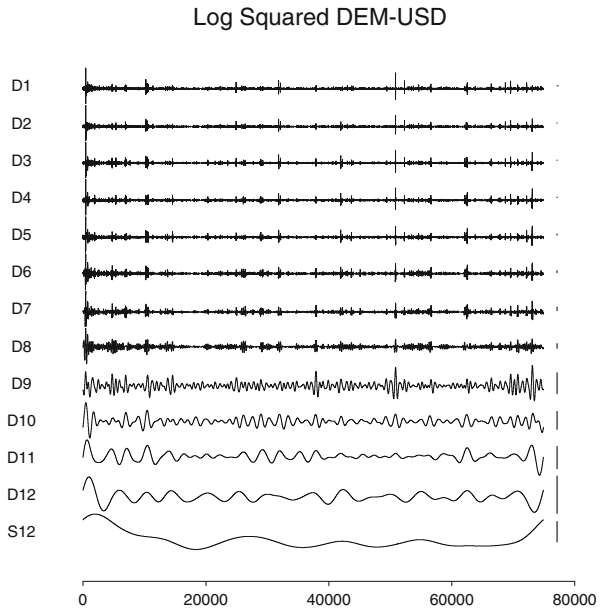


Fig. 2 Multiresolution analysis for the DM-$ log-squared return series wavelet and scaling coefficients over the time period October 1, 1992 through September 29, 1993

5.2 MODWT of Log-Squared Returns

The MODWT coefficients of the log-squared DM-$ returns are computed using the Daubechies (1992) "least asymmetric" class of wavelet filters with 8 nonzero coefficients (LA(8)); i.e., $L = L_1 = 8$ in Eq. (8). Since the LA(8) wavelet has a near zero phase shift, the resulting wavelet coefficients will line up nicely in time with the events that noticeably affect the volatility of the DM-$ return. Because its filter length is long enough to ensure the near bandpass behavior needed for measuring long-memory, the LA(8) wavelet is also a logical choice for the estimation of $d(t)$,

If at the beginning and end of the time period y^* behavior is similar, the circularity assumption of Eq. (8) will have less of an effect on those $\tilde{W}^{(y^*)}$ near the borders. When a noticeable difference is apparent, such as when an upward or downward trend is found, the impact on the wavelet coefficients near the beginning and end can be mitigated by reflecting the series. Since it is difficult to tell if the log-squared returns are the same at the boundaries, and because the news events at the two time periods will not be identical, we apply the MODWT to the reflected series, $y^*_{1,T}, y^*_{2,T}, \ldots, y^*_{T,T}, y^*_{T,T}, y^*_{T-1,T}, \ldots, y^*_{1,T}$.

In Fig. 2 we plot the multiresolution analysis of the DM-$ log-squared returns, $y^*_{t,T}$. This figure consists of individual plots of the wavelet details, $\tilde{\mathscr{D}}_{j,t,T}$, over the entire year, at the scaling levels $j = 1, \ldots, 12$. Each $\tilde{\mathscr{D}}_{j,t,T}$ represents the

Table 1 Translation of levels in a MODWT to appropriate time scales for the DM-$ return series ($\Delta t = 5$ min)

Level j	Scale τ_j
1	5 min.
2	10 min.
3	20 min.
4	40 min.
5	80 min. = 1 h. 20 min.
6	160 min. = 2 h. 40 min.
7	320 min. = 5 h. 20 min.
8	640 min. = 10 h. 40 min. $\approx$ 1/2 day
9	1280 min. = 21 h. 20 min. $\approx$ 1 day
10	2560 min. = 42 h. 40 min. $\approx$ 2 days
11	5120 min. = 85 h. 20 min. $\approx$ 3.5 days
12	10,240 min. = 170 h. 40 min. $\approx$ 7 days

variation of $y^*_{t,T}$ at a localized interval of time where the scales $j = 1, 2, 3, 4$ correspond to variations at the 5, 10, 20, 40 minute level, the scales $j = 5, 6, 7, 8$ correspond to approximately the 1.5, 2.5, 5 hour level, and scales $j = 9, 10, 11, 12$ to approximately the $1/2, 1, 2, 3.5, 7$ day level.[9] For example, the wavelet details, $\tilde{\mathscr{D}}_{1,t,T}$, are associated with changes (differences in weighted averages) in the original series at the 5 minute interval, whereas the $\tilde{\mathscr{D}}_{10,t,T}$ capture the information that is lost when volatility is computed from a two-day return rather than a daily return. Because $j = 9$ is associated with approximately daily variations, the details, $\tilde{\mathscr{D}}_{j,t,T}$, $j = 1, \ldots, 8$, measure the intradaily variation of $y^*_{t,T}$.

The wavelet smooths, $\tilde{\mathscr{S}}_{12,t,T}$, are weighted averages of $y^*_{t,T}$, approximately a week in length, that measure the log-squared returns long-term variation at time scales associated with a week and greater in length. Because the MODWT is a dyadic transform $y^*_{t,T}$ could be decomposed up to scales equal to the integer value of $\log_2 T$. However, we choose $j = 12$ as the largest scale because intraday behavior at frequencies lower than a day-of-the-week effects have not be present in high-frequency exchange rate data (see Harvey and Huang 1991).

Focusing on the $\tilde{\mathscr{S}}_{12,t,T}$ between the beginning of January (interval 19,009) and the end of July (interval 56,448), there is evidence of a slow, moderate, quarterly cycle in the log-squared returns. The plot of the smooths $\tilde{\mathscr{S}}_{12,t,T}$ over the January to July time period reveals periodicity of approximately one and half cycles in length. It is possible that this cycle continues in the data both before January and after July, but because of October's US Employment Report (Oct. 2, interval 439), the US stock market crash (Oct. 5, interval 816), and the Russian crises (Sept. 21, interval 73,098), the value of $\tilde{\mathscr{S}}_{12,t,T}$ before January and after July may be artificially inflated because the above events may be compounding the boundary effects of the wavelet

[9] Table 1 provides the actual conversion between the scaling parameter, j, of the MODWT and the time scale of the DM-$ time series.

decomposition. These boundary effects may also explain the relatively large values of $\widetilde{\mathscr{D}}_{j,t,T}$, for $j = 10, 11, 12$, found at the beginning and end of the sample.

Cyclical behavior in the log-squared returns is also visible at the weekly scale. In the plot of $\widetilde{\mathscr{D}}_{12,t,T}$, found in Fig. 2, a pattern of four cycles occurs each quarter. This cycle is fairly robust over the entire year, flattening out slightly during February, March, and May.

The details $\tilde{\mathscr{D}}_{j,t,T}$ at the scales $j = 1, \ldots, 9$ reveal the transient effect anticipated and unanticipated news events have on volatility. Some note worthy events that our data set encompasses are the election of Bill Clinton to the US presidency (Nov. 11, interval 6,951), the floating of the Swedish krona followed by the realignment of Europe's exchange mechanism (Nov. 19 & 22), and the military confrontation between then Russian president Yeltsin and the old guard of the Russian Parliament (Sept. 21, interval 73,098). There are also macroeconomic news events, most notably the monthly US Employment report (first week of each month), especially the US Employment report for the month of May (June 4, interval 50,876), and also the Bundesbank meeting report (March 4, interval 32,157).

The impact of new information on log-squared returns is captured by the details size relative to those around it at the point in time of the event. In Fig. 2 every time the absolute value of $\tilde{\mathscr{D}}_{1,t,T}$ deviates noticeably from zero it corresponds with the arrival of new information to the market. A similar pattern is also found in the details at scales $j = 2, \ldots, 9$. This pattern in the details suggestions anticipated and unanticipated announcements affect the behavior of volatility over the course of the day the news is released. Two days later, however, the market has assimilated the news and volatility has reverted back to its typical behavior as quantified by the variation in volatility over a two day period. For instance, those details at the five minute to one day level of variation ($j = 1, \ldots, 9$) that correspond with the May's US Employment report (June 4, interval 50,876) are clearly larger than those before and after this date. But the early June details at the scales $j = 10, 11, 13$ are not visibly different from those details at the same scales during other time periods when the market did not experience any news.

5.3 *Time-Varying Differencing Parameters*

We now take the MODWT coefficients of $y^*_{t,T}$ and calculate the time-varying wavelet variances, $\tilde{\nu}^2(t/T, \tau_j)$, for $t = 1, \ldots, 74{,}880$ and $j = 1, \ldots, 16$. For each value of t, we take the log of $\tilde{\nu}^2(t/T, \tau_j)$, where $j = 5, \ldots, 12$, and regress these time-varying wavelet variances on a constant and $\log 2^j$. This produces our OLS estimates of $D(t)$ from which $d(t) = (D(t) + 1)/2$ are calculated. Since $d(t)$ measures the low frequency behavior of $h_{t,T}$, it is common practice to exclude the first few scales of the wavelet variances from the OLS regression (see Jensen 1999b). These excluded time-varying wavelet variances at the scales $j = 1, \ldots, 4$ measure the energy in $h_{t,T}$ over those frequencies associated with behavior on the order of

five to forty minutes in length. We also exclude the wavelet variances at the four largest scales, $j = 13, 14, 15, 16$, from the regression. This helps to reduce the adverse impact the MODWT boundary effects might have on the estimates of $d(t)$.

Our results for $d(t)$ are plotted in Fig. 3 (middle line). The upper and lower lines found in the figure represent the 90% confidence interval of $d(t)$ that we construct by bootstrapping 1,000 replications where the estimated residuals are randomly sampled with replacement. In all but a few instances, the time-varying, long-memory parameter is positive. Since a global estimate of the differencing parameter is the average of the time-varying differencing parameter estimates over the entire time period, finding a positive differencing parameter supports the conclusions of others that the volatility of intradaily foreign exchange rates exhibit long-memory behavior and is not a spurious artifact of structural breaks in the series.

In Fig. 3, the two largest values of $d(t)$ can be ignored. The first occurrence corresponds with Christmas Day and the other with New Years Day. Because these days are effectively "weekends" where low quote activity occurs, the level of volatility during these days is meaningless with regards to the time-varying long-memory parameter. The three most negative values, $d(t) = -0.3275, -0.3080, -0.2537$ correspond to the second highest (June 4), eighth highest (Sept. 21) and highest (Oct. 2) volatility levels, respectively. The first and third most negative values of $d(t)$ occur at the US Employment report announcements. The second smallest value occurs on the day of the Russian crisis (Sept. 21). On the 12 days when the monthly US employment report is released, $d(t)$ only stays positive during the December announcement. In all the other months, the employment report causes $d(t)$ to fall below zero.

In every instance where the value of $d(t)$ rapidly declines to zero or becomes negative and then quickly rebounds back to positive values, the date corresponds with either a macroeconomic announcement, a political event (US presidential election), a meeting of the Bundesbank, or a plunge in the US stock market. This behavior in $d(t)$ suggests that volatility becomes anti-persistent in response to the release of scheduled economic news and to expected and unexpected political events, in the sense that while volatility is still highly persistent and strongly correlated with past observations it is now negatively correlated. Although the new information from these events clearly affects volatility, the rapid increase in $d(t)$ suggests the market quickly trades the asset to its new price and then relies on its long-term dynamics of risk and volatility's inherent property of long-memory in carrying out trades when information is not being disseminated.

5.4 Intraday periodicity

In Fig. 4 we plot the average value of $d(t)$ at each of the daily 288 five-minute intervals. Unlike the regional U-shape behavior of volatility found in the average absolute returns of Fig. 1, where volatility increases and the market thickens due to the opening and closing of a regional trading center (Müller et al. 1990; Baillie and

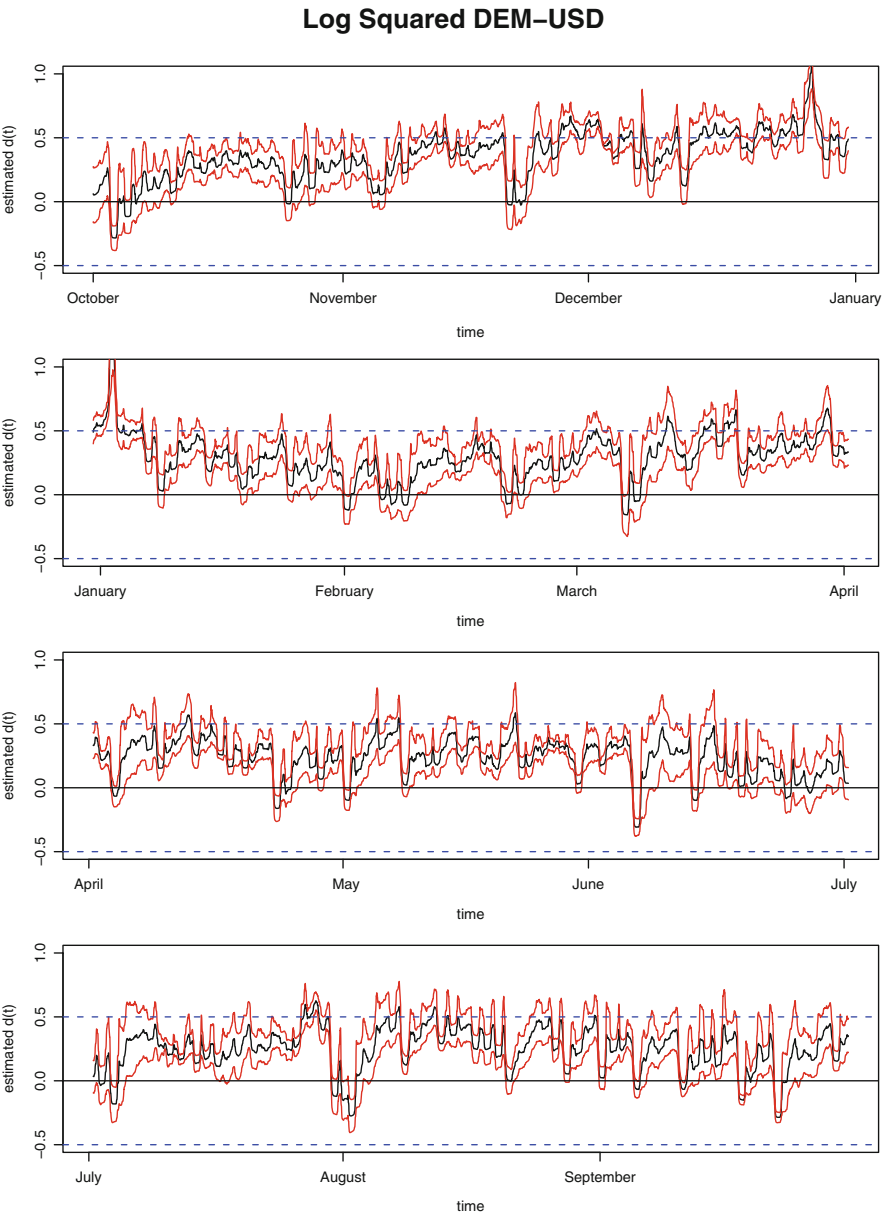


Fig. 3 Semiparametric wavelet estimates (*middle line*) of the time-varying differencing parameter, $d(t)$, and their bootstrapped 90% confidence interval (*upper and lower line*) for the DM-$ log-squared returns using wavelet coefficients ($\tilde{W}_{j,t,T}$, $j = 5, \ldots, 12$, $t = 1, \ldots, 74{,}880$) calculated with the LA(8) wavelet filter

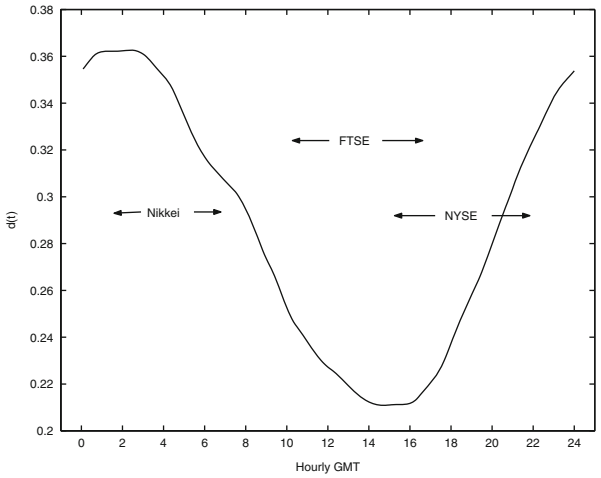


Fig. 4 Intraday average over all 260 trading days for each of the 288 daily five-minute intervals of $d(t)$. The *arrows* show when the particular market is open, the Nikkei, 1:00–7:00 GMT, the FTSE, 10:00–16:30 GMT, and the NYSE, 14:30–22:00 GMT

Bollerslev 1990; Andersen and Bollerslev 1997b, 1998), Fig. 4 reveals an intradaily pattern of time-varying, long-memory being highest when the first and third most active trading centers (London and New York) are closed, and the lowest average value of the time-varying, long-memory parameter corresponds with when the London market is closing.

Because a larger long-memory parameter leads to a smoother or less volatile process, the period of the day when both the London and New York markets are closed is a tranquil time with small and infrequent changes in the DM-$ exchange rate. Whereas, the market is on average the most turbulent during those hours when the London and New York markets are both open (14:30–16:30 GMT). Since the volume during the operation of the London market is of the largest level of the three regional markets and is nearly twice that of the New York market, the small long-memory average associated with the closing of the London market (16:30 GMT) suggests that the heavier market activity leads to lower degrees of long-memory with its accompanying large and frequent changes in volatility.

It is difficult to determine if the decline in the degree of long-memory that occurs when London and New York are simultaneously trading is the result of public or private information. In the equity markets, French and Roll (1986) argue that private price information held by informed traders must be exploited prior to market closing, and thus, higher levels of volatility are to be expected immediately before the market closes. On the other hand, Harvey and Huang (1991) and Ederington and Lee (1993) argue that, unlike the equity markets, the foreign exchange market is a continuous market, giving informed traders a liquid market in which to capitalize on their private information almost 24 hour out of every day. As a result, the increase

in volatility during trading versus non-trading hours is due to the concentration of public information that is released during the trading hours.

Both hypotheses are plausible explanations for the small long-memory parameter associated with concurrent trading in London and New York. The private information hypothesis holds since the level of long-memory in Fig. 4 continually decreases up to the closing of the London market, before beginning to increase. However, the decline in the long-memory parameter may also be the result of the large number of U.S. macroeconomic announcements that take place during the opening hours of the New York market. To distinguish between the two hypothesis requires further research.

The intradaily pattern of $d(t)$ in Fig. 4 also suggest that the three major financial markets, Tokyo, London, and New York, may be fully integrated. Baillie and Bollerslev (1990) surmise that the regional U-shape of volatility found in Fig. 1 is caused by local market makers holding open position during the day but few overnight positions. Such behavior could explain the increase in the level of trading that occurs during the opening and closing hours of each regional market and be a reason for why the regional markets are not fully integrated.

From Fig. 4, the average level of persistence in volatility as measured by $d(t)$ is at its highest point as the East Asian markets open and then declines monotonically through the Tokyo and Hong Kong market's lunch hour, and the opening of the European and New York markets. The average time-varying, long-memory parameter reaches its lowest point of the financial day exactly as the London market closes. It is at this point that Andersen and Bollerslev (1998) find the volatility of the DM-$ exchange rate to be at its high for the day. The average level of long-memory in volatility then steadily increases through the closing of the New York market and the opening of the Pacific markets. Through the entire 24 hour trading day the only market openings and closings to affect the intradaily average of $d(t)$ are the opening of Tokyo and the closing of London.

Finding the smallest average value of the long-memory parameter to occur as the London market closes adds strength to Andersen and Bollerslev (1997a) argument that long-memory is a fundamental component of volatility and is not an artifact of external shocks or regime shifts posited by Lamoureux and Lastrapes (1990). Most new information comes when both Europe's and New York's market are operating (Harvey and Huang 1991). However, it is during these hours that the degree of long-memory and long-term persistence in volatility is at its lowest point, but still positive.

6 Conclusion

In this chapter we have combined the interdaily long-memory behavior of volatility with its intradaily time-of-the-day effect by modeling volatility with a time-varying, long-memory, stochastic volatility model. This model of volatility has a long-memory parameters that varies over time, enabling it to capture the effect

unannounced market crashes, anticipated news announcements, and the opening and closing of regional market has on the level of long-memory in volatility. This model is also general enough to nest within it the stationary, long-memory stochastic volatility models.

To estimate volatility's time-varying long-memory parameter, we introduce a ordinary least squares estimator based on the log-linear relationship between the local wavelet variance and the wavelet scale. We applied our estimator to a years worth of tick-by-tick Deutsche mark-US dollar return data measured at five-minute intervals and found the time-varying, long-memory parameter to be positive over a large percentage of the sample. The circumstances behind those periods where the long-memory parameter was negative were associated with rescheduled new announcements and unexpected market crashes or political upheavals. In addition, we found evidence of cohesion between the Tokyo, Europe and New York markets. On average the long-memory parameter declines over the course of a day, starting at its highest point as the Tokyo market opens and declines until the London market closes. After the close of the London market, the long-memory parameter on average increases through the rest of the day. This daily pattern of the long-memory parameter suggests that the only time-of-the-day effect associated with the long-memory dynamics of volatility is the opening of the Tokyo market and the closing of the London market. Whether this behavior is caused by the predominance of scheduled news announcements occurring near the opening of the New York market, or is the result of informed traders capitalizing on private information before the close of the London market is still open to debate.

Acknowledgements Mark Jensen would like to personally thank James Ramsey for his guidance and advice and for his openness to wavelet analysis and the inference it makes possible in economics, finance and econometrics. Both authors thank the seminar and conference participants at the University of Kansas, Brigham Young University, the Federal Reserve Bank of Atlanta, the Midwest Economic Meetings, the Symposium on Statistical Applications held at the University of Missouri–Columbia, the Conference on Financial Econometrics held at the Federal Reserve Bank of Atlanta, and the James Ramsey Invited Session of the 2014 Symposium on Nonlinear Dynamics and Econometrics held in New York, The views expressed here are ours and are not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

Appendix 1

For clarity and guidance in understanding the class of locally stationary models we first prove the following lemma.

Lemma 1. *If* $|d(u)| < 1/2$ *and* $d(u)$ *and* $\sigma_\eta(u)$ *are continuous on* $\mathbb{R}$ *with* $d(u) = d(0)$, $\sigma_\eta(u) = \sigma_\eta(0)$ *for* $u < 0$, *and* $d(u) = d(1)$, $\sigma_\eta(u) = \sigma_\eta(0)$ *for* $u > 1$, *and differentiable for* $u \in (0, 1)$ *with bounded derivatives, then the triangular process* $h_{t,T}$ *defined in Eq. (6) is a locally stationary process with transfer function:*

$$A(u,\omega) = \frac{\sigma_\eta(u)}{\sqrt{2\pi}}(1-e^{-i\omega})^{-d(u)}$$

and time-varying spectral density function:

$$f(u,\omega) = \frac{\sigma_\eta(u)}{2\pi}|1-e^{-i\omega}|^{-2d(u)}$$

Proof of Lemma 1. From Stirling's formula, $\Gamma(x) \sim \sqrt{2\pi}e^{-x+1}(x-1)^{x-1/2}$ as $x \to \infty$, it follows that as $l \to \infty$, $\Gamma(l+d(t/T))/(\Gamma(l+1)\Gamma(d(t/T))) \sim l^{d(t/T)-1}/\Gamma(d(t/T))$,

$$\sum_{l=0}^{\infty}\left(\frac{\Gamma(l+d(t/T))}{\Gamma(l+1)\Gamma(d(t/T))}\right)^2 < \infty,$$

and,

$$\sum_{l=0}^{T} e^{-il\omega}\frac{\Gamma(l+d(t/T))}{\Gamma(l+1)\Gamma(d(t/T))} = (1-e^{-i\omega})^{-d(t/T)}, \qquad \text{as } T \to \infty.$$

It then follows by Theorem 4.10.1 in Brockwell and Davis (1991) that the triangular process:

$$h_{t,T} = \sigma_\eta(t/T)\sum_{l=0}^{\infty}\frac{\Gamma(l+d(t/T))}{\Gamma(l+1)\Gamma(d(t/T))}\eta_{t-l}. \tag{17}$$

is a well-defined process. Since η_t is Gaussian white noise process, it also follows that $h_{t,T}$ has the spectral representation:

$$h_{t,T} = \int e^{i\omega t}A^0_{t,T}(\omega)\,dZ(\omega)$$

where,

$$A^0_{t,T}(\omega) = \frac{\sigma_\eta(t/T)}{\sqrt{2\pi}}(1-e^{-i\omega})^{-d(t/T)}.$$

Now define:

$$A(u,\omega) = \frac{\sigma_\eta(u)}{\sqrt{2\pi}}(1-e^{-i\omega})^{-d(u)}.$$

Since the spectral representation of $h_{t,T}$ only involves evaluating $d(\cdot)$ and $\sigma_\eta(\cdot)$ at t/T, $A^0_{t,T}(\omega) = A(u,\omega)$ must hold. □

Proof of Theorem 1. By Lemmas 1:

$$\log y_{t,T}^2 = \log \sigma^2(t/T) + \int e^{i\omega t} A_{t,T}^0(\omega)\, dZ(\omega) + \log \epsilon_t^2, \tag{18}$$

where $A_{t,T}^0(\omega) = A(u,\omega) = (\sigma_\eta(u)/\sqrt{2\pi})(1 - e^{-i\omega})^{-d(u)}$. By adding and subtracting the a mean of $\log \epsilon_t^2$, -1.2704, to Eq. (18) we arrive at:

$$\log y_{t,T}^2 = \mu(t/T) + \int e^{i\omega t} A_{t,T}^0(\omega)\, dZ(\omega) + \log \epsilon_t^2 + 1.2704,$$

where $\mu(t/T) = \log \sigma^2(t/T) - 1.2704$. Since $\log \epsilon_t^2 + 1.2704$ is independent and identically distributed with mean zero and spectral density, $4.93/2\pi$, by the Spectral Representation Theorem [Brockwell and Davis (1991), p. 145]:

$$\log \epsilon_t^2 + 1.2704 = \sqrt{4.93/2\pi} \int e^{i\omega t}\, dZ_\epsilon(\omega).$$

so that:

$$\log y_{t,T}^2 = \mu(t/T) + \int e^{i\omega t} A_{t,T}^o(\omega) dZ(\omega) + \sqrt{4.93/2\pi} \int e^{i\omega t} dZ_\epsilon(\omega). \tag{19}$$

It then follows that:

$$\begin{aligned} f(u,\omega) &= |A(u,\omega)|^2 + 4.93/2\pi \\ &= \frac{\sigma_\eta^2(u)}{2\pi} |1 - e^{-i\omega}|^{-2d(u)} + 4.93/2\pi. \end{aligned}$$

□

Proof of Theorem 2. Let $X_{t,T}$ be a locally-stationary process with spectral representation

$$X_{t,T} = \int e^{i\omega t} A_{t,T}^o(\omega) dZ(\omega) \tag{20}$$

where $Z(\omega)$ is an orthonormal increment process. By the definition of a locally-stationary process there exists a continuous, smooth, even function $A(t/T,\omega)$ such that

$$\sup_{u,\omega} \left| A_{t,T}^o(\omega) - A(u,\omega) \right| \leq KT^{-1}.$$

where $u = t/T$. Assume that $\partial^3 A(u,\omega)/\partial u^3$ is uniformly bounded on $u \in (0,1)$ and $|\omega| \leq \pi$. Now define the MODWT of $X_{t,T}$ as

$$\tilde{W}_{j,t,T} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l,T}. \tag{21}$$

Substituting the definition of $X_{t,T}$ from Eq. (20) into Eq. (21) the MODWT coefficient equals

$$\tilde{W}_{j,t,T} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} \int e^{i\omega(t-l)} A^o_{t-l,T}(\omega) dZ(\omega).$$

Since $A^o_{t,T}(\omega) = A(u,\omega) + \mathcal{O}(T^{-1})$ holds uniformly on $u \in (0,1)$ and $|\omega| \leq \pi$

$$\tilde{W}_{j,t,T} = \int e^{i\omega t} \left\{ \sum_{l=0}^{L_j-1} e^{i\omega l} \tilde{h}_{j,l} \left[A(u - l/T, \omega) + \mathcal{O}(T^{-1}) \right] \right\} dZ(\omega).$$

Next, from the Taylor series expansion

$$A(u+v,\omega) = A(u,\omega) + v\frac{\partial}{\partial u} A(u,\omega) + \frac{v^2}{2}\frac{\partial^2}{\partial u^2} A(u,\omega) + \mathcal{O}(v^3)$$

we find

$$\begin{aligned} \tilde{W}_{j,t,T} = \int e^{i\omega t} \Bigg\{ \sum_{l=0}^{L_j-1} e^{-i\omega l} \tilde{h}_{j,l} \Bigg[A(u,\omega) + \left(-\frac{l}{T}\right) \frac{\partial}{\partial u} A(u,\omega) \\ + \left(-\frac{l}{T}\right)^2 \frac{\partial^2}{\partial u^2} A(u,\omega) + \mathcal{O}\left(\left(-\frac{l}{T}\right)^3\right) + \mathcal{O}(T^{-1}) \Bigg] \Bigg\} dZ(\omega). \end{aligned} \tag{22}$$

By the power scaling rule where if $g(t) = th(t)$, then $G(\omega) = (-2\pi i)^{-1} H'(\omega)$, where $G(\cdot)$ and $H(\cdot)$ are the respective Fourier transforms Eq. (22) becomes

$$\begin{aligned} \tilde{W}_{j,t,T} = \int e^{i\omega t} \Bigg\{ \tilde{\mathscr{H}}_j(\omega) \left[A(u,\omega) + \mathcal{O}(T^{-1}) \right] + (2\pi i)^{-1} \Bigg[\frac{1}{T} \frac{\partial}{\partial \omega} \tilde{\mathscr{H}}_j(\omega) \frac{\partial}{\partial u} A(u,\omega) \\ - \frac{1}{2} \frac{1}{T^2} \frac{\partial^2}{\partial \omega^2} \tilde{\mathscr{H}}_j(\omega) \frac{\partial^2}{\partial u^2} A(u,\omega) \Bigg] + \mathcal{O}(T^{-3}) \Bigg\} dZ(\omega) \end{aligned} \tag{23}$$

where $\tilde{\mathscr{H}}_j(\omega) = \sum_{l=0}^{L_j-1} e^{-i\omega l}\tilde{h}_{j,l}$. Since

$$\frac{1}{T}\frac{\partial}{\partial\omega}\tilde{\mathscr{H}}_j(\omega)\frac{\partial}{\partial u}A(u,\omega) = \mathscr{O}(T^{-1})$$
$$\frac{1}{2}\frac{1}{T^2}\frac{\partial^2}{\partial\omega^2}\tilde{\mathscr{H}}_j(\omega)\frac{\partial^2}{\partial u^2}A(u,\omega) = \mathscr{O}(T^{-2})$$

it follows that $\tilde{W}_{j,t,T}$ is a locally-stationary process with spectral representation

$$\tilde{W}_{j,t,T} = \int e^{i\omega t}A^o_{j,t,T}(\omega)dZ(\omega) \tag{24}$$

with

$$\sup_{u,\omega}\left|A^o_{j,t,T}(\omega) - A_j(u,\omega)\right| \le KT^{-1}$$

and $A_j(u,\omega) = \tilde{\mathscr{H}}_j(\omega)A(u,\omega)$. □

Appendix 2

To determine the MODWT frequency domain properties define the transfer function of the filter $\{\tilde{h}_{1,l}\}$ as:

$$\tilde{\mathscr{H}}(\omega) \equiv \sum_{l=0}^{L-1}\tilde{h}_{1,l}e^{-i\omega l}.$$

Since the filtered output of the MODWT wavelet filter $\{\tilde{h}_{1,l}\}$ produces the information lost by filtering the series with a weighted moving average, $\{\tilde{h}_{1,l}\}$ is a high-pass filter whose transfer function $\tilde{\mathscr{H}}(\cdot)$ is supported on the nominal band-pass set of frequencies $[-\pi,-\pi/2) \cup \ (\pi/2,\pi]$. From Eq. (9) it follows that $\{\tilde{g}_{1,l}\}$ is a low-pass filter whose transfer function:

$$\tilde{\mathscr{G}}(\omega) \equiv \sum_{l=0}^{L-1}\tilde{g}_{1,l}e^{-i\omega l} = e^{-i\omega(L-1)}\tilde{\mathscr{H}}(\pi-\omega),$$

has support on $[-\pi/2,\pi/2]$.

We are now in a position to define the higher ordered MODWT wavelet filters in terms of the transfer functions $\tilde{\mathscr{H}}$ and $\tilde{\mathscr{G}}$. The jth-ordered MODWT wavelet filters are found by synthesizing the transfer function:

$$\tilde{\mathscr{H}}_j(\omega) \equiv \tilde{\mathscr{H}}(2^{j-1}\omega) \prod_{k=0}^{j-2} \tilde{\mathscr{G}}(2^k\omega)$$

which by the definition of $\tilde{\mathscr{H}}$ has support on the octave $(\pm 2\pi/2^{j+1}, \pm 2\pi/2^j]$. The Jth-ordered scaling filter has the transfer function:

$$\tilde{\mathscr{G}}_J(\omega) \equiv \prod_{k=0}^{J-1} \tilde{\mathscr{G}}(2^k\omega),$$

whose support is $(-\pi/2^J, \pi/2^J)$.

References

Andersen TG, Bollerslev T (1997a) Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. J Finance 52:975–1005

Andersen TG, Bollerslev T (1997b) Intraday periodicity and volatility persistence in financial markets. J Empir Finance 4:115–158

Andersen TG, Bollerslev T (1998) Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. J Finance 53:219–265

Andersen TG, Bollerslev T, Diebold FX, Labys P (2001) The distribution of realized exchange rate volatility. J Am Stat Assoc 96:42–55

Andersen TG, Bollerslev T, Diebold FX, Labys P (2003) Modeling and forecasting realized volatility. Econometrica 71:579–625

Baillie RR, Bollerslev T (1990) Intra-day and inter-market volatility in foreign exchange rates. Rev Econ Stud 58:565–585

Breidt FJ, Crato N, de Lima P (1998) The detection and estimation of long memory in stochastic volatility. J Econometrics 83:325–348

Brockwell PJ, Davis RA (1991) Time series: theory and methods, 2nd edn. Springer, New York

Crowley PM (2007) A guide to wavelets for economicst. J Econ Surv 21:207–267

Dahlhaus R (1996) On the Kullback-Leibler information divergence of locally stationary processes. Stoch Process Appl 62:139–168

Dahlhaus R (1997) Fitting time series models to nonstationary processes. Ann Stat 25:1–37

Daubechies I (1992) Ten lectures on wavelets. SIAM, Philadelphia

Diebold FX, Atsushi I (2001) Long memory and structural change. J Econometrics 105:131–159

Donoho DL, Johnstone IM (1994) Ideal spatial adaptation via wavelet shrinkage. Biometrika 81:425–455

Donoho DL, Johnstone IM (1995) Adapting to unknown smoothness by wavelet shrinkage. J Am Stat Assoc 90:1200–1224.

Donoho DL, Johnstone IM (1998) Minimax estimation via wavelet shrinkage. Ann Stat 26:879–921

Ederington LH, Lee JH (1993) How markets process information: News releases and volatility. J Finance 48:1161–1191

French KR, Roll R (1986) Stock return variances: The arrival of information and the reaction to traders. J Financ Econ 17:5–26

Gallegati M, Ramsey JB, Semmler W (2014) Interest rate spreads and output: A time scale decomposition analysis using wavelets. Comput Stat Data Anal 76:283–290

Gençay R, Selçuk F, Whitcher B (2001) An introduction to wavelets and other filtering methods for finance and economics. Academic Press, San Diego
Gençay R, Selçuk F, Whitcher B (2005) Mutiscale systemic risk. J Int Money Finance 24:55–70
Harvey AC (2002) Forecasting volatility in the financial markets, 2nd edn, chap Long Memory in Stochastic Volatility. Butterworth-Heinemann, Oxford, pp 307–320
Harvey CB, Huang RD (1991) Volatility in the foreign currency future market. Rev Financ Stud 4:543–569
Hong Y, Kao C (2004) Wavelet-based testing for serial correlation of unknown form in panel models. Econometrica 72:1519–1563
Hong Y, Lee J (2001) One-sided testing for ARCH effects using wavelets. Econom Theory 6:1051–1081
Jensen MJ (1999a) An approximate wavelet MLE of short and long memory parameters. Stud Nonlinear Dynam Econometrics 3:239–253
Jensen MJ (1999b) Using wavelets to obtain a consistent ordinary least squares estimator of the fractional differencing parameter. J Forecast 18:17–32
Jensen MJ (2000) An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets. J Econ Dynam Control 24:361–387
Jensen MJ (2004) Semiparametric Bayesian inference of long-memory stochastic volatility. J Time Ser Anal 25:895–922
Jensen MJ, Liu M (2006) Do long swings in the business cycle lead to strong persistence in output? J Monetary Econ 53:597–611
Jensen MJ, Whitcher B (1999) A semiparametric wavelet-based estimator of a locally stationary long-memory model. Tech. rep., Department of Economics, University of Missouri
Lamoureux CG, Lastrapes WD (1990) Persistence in variance, structural change and the GARCH model. J Bus Econ Stat 8:225–234
Lastrapes WD (1989) Exchange rate volatility and U.S. monetary policy: An ARCH application. J Money Credit Bank 21:66–77
Mallat S (1989) A theory of multiresolution signal decomposition: The wavelet representation. IEEE Trans Pattern Anal Mach Intell 11:674–693
Müller U, Dacorogna M, Olsen R, Pictet O, Schwarz M, Morgenegg C (1990) Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. J Bank Finance 14:1189–1208
Percival DB, Walden AT (2000) Wavelet methods for time series analysis. Cambridge University Press, Cambridge
Phillips PCP (1987) Time series regression with a unit root. Econometrica 55:277–301
Ramsey JB (1999) The contribution of wavelets to the analysis of economic and financial data. Phil Trans R Soc Lond A 357:2593–2606
Ramsey JB, Lampart C (1998a) The decomposition of economic relationships by time scale using wavelets: Expenditure and income. Stud Nonlinear Dynam Econometrics 3:23–42
Ramsey JB, Lampart C (1998b) Decomposition of economic relationships by time scale using wavelets: Money and income. Macroeconomic Dynamics 2:49–71
Rua A, Nunes LC (2009) International comovement of stock market returns: A wavelet analysis. J Empir Finance 16:632–639
Russell JR, Ohanissian A, Tsay RS (2008) True or spurious long memo? A new test. J Bus Econ Stat 26:161–175
Stăriciă C, Granger C (2005) Nonstationarities in stock returns. Rev Econ Stat 87:503–522
Whitcher B, Jensen MJ (2000) Wavelet estimation of a local long-memory parameter. Explor Geophys 31:94–103.

Wavelet Analysis and the Forward Premium Anomaly

Michaela M. Kiermeier

Abstract Forward and corresponding spot rates on foreign exchange markets differ so that forward rates cannot be used as unbiased predictors for future spot rates. This phenomenon has entered the literature under the heading of the Forward Premium Anomaly. We argue that standard econometric analyses implicitly assume that the relationship is time scale independent. We use wavelet analysis to decompose the exchange rate changes, and the forward premia, using the maximal overlap discrete wavelet transform (MODWT). Then we estimate the relationship on a scale-by-scale basis, thereby allowing for market inefficiencies such as noise, technical, and feedback trading as well as fundamental and rational trading. The results show that the forward premia serve as unbiased predictors for exchange rate changes (unbiasedness hypothesis) for certain time scales only. Monthly and weekly data concerning Euro, US-dollar and British Pound for forward periods from 1 month to 5 years is analysed. We find that the unbiasedness hypothesis cannot be rejected if the data is reconstructed using medium-term and long term components. This is most prevalent in the forward transaction periods up to 1 year.

1 Introduction

Spot and forward exchange rates are determined by current expectations about future events. The theory of rational expectations links expectations about future inflation rates, interest rates, with changes in prices in currency markets. Currency price adjustments result from various attempts of market participants to manage

M.M. Kiermeier (✉)
University of Applied Sciences Darmstadt, Darmstadt, Germany

Fachbereich Wirtschaft, Hochschule Darmstadt, Dieburg, Germany
e-mail: michaela.kiermeier@h-da.de

M. Gallegati and W. Semmler (eds.), *Wavelet Applications in Economics and Finance*, Dynamic Modeling and Econometrics in Economics and Finance 20,
DOI 10.1007/978-3-319-07061-2_6,

risks and returns. The Interest Rate Parity states that investors demand a premium or a discount in forward exchange markets according to differentials in interest rates, see for example Shapiro (2009). Since forward rates and interest rates are theoretically linked through the Uncovered Interest Rate Parity (UIP), we investigate the rational expectation theory by focusing on the forward rate's ability to forecast future exchange rate changes. We thereby test if current forward rates provide unbiased predictors for next periods' spot exchange rates, which is what we call the unbiasedness hypothesis throughout this paper. Tests for the floating exchange rate era indicate that future exchange rates are negatively correlated with current forward rates. Its interpretation being, that the forward rate serves a better purpose as a contrary indicator than as an unbiased predictor for future spot rates. According to Fama (1984) these empirical results are referred to as the "Forward Premium Puzzle" or the "Forward Premium Anomaly". Engel (1996) stresses that the unbiasedness hypothesis is routinely rejected in empirical tests which is also the finding of related research, see for example Hodrick (1987), MacDonald and Taylor (1992), Taylor (1995), and Wang and Jones (2002).

Cutler et al. (1990) summarize three important characteristics of data concerning forward premia and exchange rate changes. They find that monthly returns exhibit positive autocorrelation with regards to previous months. Additionally, they point to the fact that there is negative auto-correlation in the medium or long term, and that returns are best explained by fundamentals in the long run. Analyses of survey data support these findings. Allen and Taylor (1990) find that for the London foreign exchange market 90 % of intraday and short term traders use technical analysis for their decisions. In the long run, however, they argue that fundamentals are used by 85 % of market participants in forming expectations. Similar results are found by Cheung and Wong (2000) on the Asian interbank markets for foreign exchange.

Positive feedback traders continue to buy when prices increase, and sell, when prices decrease, whereas negative feedback traders do exactly the opposite. They sell when prices increase and buy when prices decrease. Motivation for positive feedback trading can arise from strategies that include portfolio insurance, positive wealth elasticity, or simply from technical trading. Negative feedback trading, on the other hand, can be induced by profit taking or investment strategies that ask for a constant share of wealth in pre-defined asset classes. Cutler et al. (1990) explain the positive 1 month autocorrelation of exchange rates by assuming that fundamental or positive feedback traders only learn about the fundamentals with a time lag. The use of technical analysis (i.e. noise trading) also results in positive auto-correlation of exchange rates. The negative medium, or long term, auto-correlation is then a direct result from misperceptions in the short run which are corrected in the medium or long term. For these time periods fundamentals become the main driving force. Other possible explanations for overshooting prices and deviations from long term fundamental equilibrium are outlined by Black (1988), and Frankel and Froot (1986), who assume that investors change their willingness to take on risks according to non-fundamental factors, or as a result of the success of forecasting models in the previous period. This way Frankel and Froot (1986) are able to explain the continued deviation of the US$ from its equilibrium rate during the time period 1980 to 1985.

If exchange rate markets are efficient past observations of exchange rate changes, or forward premia, cannot be significant in explaining current exchange rates. However, De Long et al. (1990, 1991) demonstrate that even rational investors can have different opinions on the data generating process with regards to the near future and the long run. They argue that rational investors can correctly perceive positive feedback trading as the driving force behind price changes in the near future, and at the same time acknowledge a reversion to a fundamental equilibrium in the long run.

In this paper we apply wavelet analysis to be able to allow for various types of trading as outlined above. The wavelet decomposition allows us to specifically distinguish short, medium, and long run periods. At the same time we can allow information from past observations to continue to be of importance for the respective time periods. Within these time periods investors can either learn about the relevant information with a time delay, or use feedback, noise, technical, fundamental, rational trading as their respective strategy. We investigate if averaging over various time periods veils the fact that the unbiasedness hypothesis holds true for certain time scales only, i.e. that the fundamental relationship between forward premia and exchange rate changes holds true only at certain time horizons. For that purpose we decompose exchange rate changes, and the forward premia, into their time-scale components using the maximal overlap discrete wavelet transform (MODWT). We thereby restrict the variation of the data to be of influence for a certain time period only. Decomposing weekly and monthly data to their respective time scales allows us to distinguish one short term, three medium term, and three long term periods. We then proceed by estimating the impact of the forward premia on exchange rate changes on a scale-by-scale basis. The robustness of the results is tested by analyzing forward transaction time periods that vary from 1 month to 5 years.

Only recently researchers analyze relationships to hold for various time periods and not only for the short and the long run. This is why wavelet analysis has been applied to macro-economic and financial theories, see for example Ramsey and Lampart (1996), Kiermeier (2014), Kim and Haueck In (2003), Raihan et al. (2005), Gallegati et al. (2011), Gencay et al. (2009).

This paper is organized in the following way. In Sect. 2, we review shortly the underlying theories and attempts to explain the forward premium anomaly in Sect. 2. In Sect. 3 we introduce the basic ideas of wavelet analysis and motivate its use to test for the unbiasedness hypothesis. Section 4 describes the data and the results from performing regression analyses on a scale-by-scale basis. Section 5 concludes.

2 The Forward Premium Anomaly

To test the hypothesis if the forward rate (F) is an unbiased predictor for future spot rates (S) Eq. (1) has to be analyzed econometrically.

$$s_{t+1} - s_t = \mathrm{a} + \mathrm{b}\ (f_t - s_t) + u_{t+1} \tag{1}$$

with u_t being a white noise error term. The lower letter case s and f indicate the logarithmic transformations of the variables S and F.

For the unbiasedness hypothesis to hold, "a" needs to be equal to zero and "b" to one. If "b" equals one the above specification becomes equal to Eq. (2).

$$s_{t+1} = \mathrm{a} + \mathrm{b}\,(f_t) + \epsilon_{t+1} \tag{2}$$

The empirical evidence rejects the unbiasedness hypothesis. In general, the slope coefficient in a regression of the ex post exchange rate changes on a constant and the forward rate differential is significantly negative, see Engel (1996). Therefore, an approach of time varying risk premia was introduced (see Hodrick and Srivastava 1986; Kaminski and Peruga 1990; Bensberg 2012, among others). Backus et al. (1996) conclude that these models have serious shortcomings since they are not in line with market data.

Froot and Thaler (1990) summarize models that attempt to explain the forward premium anomaly including the peso problem and give an outlook for a possible explanation. They argue that the assumption of efficient currency exchange markets cannot be made because in practice investors have different response times to new information. They therefore include past interest rate changes in their econometric specification which results in some positive estimates of the coefficient "b". Chinn and Meredith (2004) use a macro-economic model to give a theoretical foundation for the necessity that different time periods have to be considered in the econometric analysis.

In this paper we extend the idea of analyzing different time horizons, and allow for inefficiencies, such as delayed learning about relevant information, or other forms of feedback, or technical trading as outlined above.

Standard econometric estimation techniques are able to distinguish between short and long term dynamics only. Non-stationary features of the data are usually removed prior to performing an analysis, resulting in the known problem that relationships seem to change in times of financial distress. Different data generating processes (regimes) seem to govern price movements in financial markets. We do not adjust the data prior to the regression analyses. We decompose the data with wavelet analysis which allows for various forms of non-stationarities to be present in the data, and nonetheless, do not cause problems in our analysis when we estimate the relationship on a scale-by-scale basis.

3 Estimation Techniques

Time series analysis and standard econometric methods cannot account for changes in frequency behavior. We use wavelets as a time–frequency analysis that provides information about the frequency behavior of time series at a given point in time. Wavelet analysis estimates the frequency structure of a time series (forward premium and exchange rate changes). In addition to that it keeps the information

when an event of the time series takes place. Wavelet analysis can be understood as a rotation in the function space. The basis functions used in that transformation are wavelets which have finite support on the time axis, i.e. are small waves. For the purpose of transforming the time series, the basis function (wavelet) is dilated, or compressed, to capture frequency behavior, and is shifted along the time axis to capture the date when a certain event takes place. This is how it is possible for a disturbance to be of influence for certain frequencies, or finite time periods only. The result is a representation of the time series in the time and frequency domain. The wavelet approach can allow an analysis of processes whose behaviors differ across scales, i.e. depict different behavior with regards to different time horizons. This is most likely the case for the (forward) currency exchange market due to the aforementioned reasons.

For the purpose of allowing different behavior for different time horizons, the variables exchange rate change and forward premia are decomposed into their time-scale components applying the maximal overlap discrete wavelet transform (MODWT). This procedure allows for any length of a time series and is able to get robust estimators. Wavelets ($\psi_{j,k}$ and $\varphi_{J,k}$) when multiplied with their respective coefficients at a certain level "j" or "J" are called atoms $D_{j,k}$ and $S_{J,k}$ (i.e. $d_{j,k} * \psi_{j,k} = D_{j,k}$ and $s_{J,k} * \varphi_{J,k} = S_{J,k}$) with $\psi_{j,k}$ and $\varphi_{J,k}$ being the wavelet and scaling functions at level "j" or "J" and "k" indicating the location of the wavelet on the time axis. The sums of all atoms, $S_{J,k}(t)$ and $D_{j,k}(t)$, over all locations on the time axis $k = 1, \ldots, \frac{n}{2^j}$ at a certain level "j" or "J" are called crystals and are given by Eqs. (3) and (4).

$$S_J = \sum_{k=1}^{\frac{n}{2^J}} S_{J,k}\varphi_{J,k} \tag{3}$$

$$D_j = \sum_{k=1}^{\frac{n}{2^j}} D_{j,k}\psi_{j,k} \quad \forall j = 1, \ldots, J \tag{4}$$

Defining the importance of information to be valid for a specific time period only, the time series are decomposed to their respective resolutions in time (time scales). The time series forward premia and exchange rate changes are then approximated using only parts of the coefficients and their respective wavelets. To analyze the impact of information for a certain time period the multiresolution decomposition is applied to the time series $(s_{t+1} - s_t)$ and $(f_t - s_t)$ which is defined in Eq. (5):

$$\begin{aligned}
&(f_t - s_t)_j = D_{pj}(t) \quad \forall j = 1, \ldots J-1 \\
&(s_{t+1} - s_t)_j = D_{ej}(t) \quad \forall j = 1, \ldots J-1 \\
&\text{and} \\
&(f_t - s_t)_J = S_{pJ}(t) \quad \text{at level J} \\
&(s_{t+1} - s_t)_J = S_{eJ}(t) \quad \text{at level J}
\end{aligned} \tag{5}$$

The wavelets used in the analysis are "symmlets", which are smooth and comparatively symmetric filter functions. The decomposition is performed sequentially from smallest (high frequencies) to largest (low frequencies) scales. The support width on the time axis doubles in size with each following level i.e. scale. The number of scales used in this analysis equals five (i.e. J = 5) which is a direct result of the number of observations available (see Crowley 2005). We then perform the regression analysis at each level. Changes in exchange rates are regressed on the forward premia at different time scales, i.e. Eq. (1) is estimated at every time scale 1, ..., J using the reconstructed time series as outlined in Eqs. (6) and (7):

$$(s_{t+1} - s_t)\left[\mathrm{D_{ej}}\,(\mathrm{t})\right] = \mathrm{a} + \mathrm{b}\ (f_t - s_t)\ \left[\mathrm{D_{pj}}\,(\mathrm{t})\right] + u_{t+1} \quad \forall D1 - D5 \tag{6}$$

$$(s_{t+1} - s_t)\,[\mathrm{S_{e5}}\,(\mathrm{t})] = \mathrm{a} + \mathrm{b}\ (f_t - s_t)\ \left[\mathrm{S_{p5}}\,(\mathrm{t})\right] + u_{t+1} \quad S5 \tag{7}$$

The unbiasedness hypothesis is then tested by imposing and testing the linear restrictions on the estimated parameter as in the aggregate analysis, i.e. the linear restriction is imposed on "b" to equal one.

4 Empirical Analysis

4.1 The Data

The data used in this analysis is taken from Bank of America/Merril Lynch.

Weekly and monthly Eurocurrency rates for Euro, UK, and the US are taken. Forward Rates are calculated for time periods of 1, 3, 6, 12 months, 2, and 5 years. The weekly rates are available from January 2000 to March 2012, the monthly closing data is available from January 1998 until June 2013. The exchange rates for the currencies US$/Euro and US$/British Pound are available as weekly and monthly observations for all estimation periods from the same source. We begin our analysis with monthly observations and a 1-month forward transaction period. The forward rates are calculated according to the interest rate parity from the monthly data for spot exchange rates, and interest rates, using Eq. (8):

$$F_t = S_t \frac{1 + i_f}{1 + i_d} \tag{8}$$

with

F_t = forward exchange rate (foreign currency per one unit domestic currency) at t
S_t = spot exchange rate at time t
i_f = foreign 1-month Eurocurrency rate
i_d = domestic 1-month Eurocurrency rate

We then proceed by applying wavelet analysis to the data to allow for the possibility that averaging over time scales veils the fact that the forward rate is an unbiased predictor for future spot rates with regards to certain time periods only.

4.2 Wavelet Analysis

We calculate the maximal overlap discrete wavelet transform (MODWT) of the time series forward differential and exchange rate changes for the US$/Euro and US$/BPD exchange rates. In case of the 1 month forward period the number of monthly (weekly) observations is 184 (620). To achieve appropriate resolution we choose the number of scales to be five. The transform results in the estimation of "d1"–"d5" wavelet coefficients and "s5" scaling coefficients.

The variation of the time series that can be explained by various scales i.e. crystals are summarized in Table 1.

The forward premia are well explained by coarse scales (low frequencies) only, whereas for the exchange rate changes all scales contribute significantly to the explanation of the variation in the time series. We find that the forward premia are best explained by time scales ranging from "D4" to "S5", whereas the exchange rate changes by time scales D1–S5. At each scale "j" the coefficients are associated with time periods $[2^j, 2^{j+1}]$. The decomposition of the monthly data allows us to extract components of the data that prevail in the medium or long term whereas the weekly data yields insights for the short term behavior. At the highest frequency of the monthly data—at scale "D1"—coefficients approximate reactions to information for the time period of 2–4 months. At scale two, three, four, and five, the respective time periods are 4–8 months, 8–16 months, 16–32 months, and 32–64 months. Therefore, we associate the first three scales with the medium term (short medium term equals 2–4 months, medium term 4–8 months, and longer medium term 8–16 months). The remaining three scales at the lower frequencies represent long term behavior ("D4" 1.3–2.6 years, "D5" and "S5" represent behavior from 2.6 to 5 years and longer). Extracting the components of the data that are influential in the medium or long term allows us to detect patterns that can be a result of different investment behavior or information used in forming expectations. We perform a similar analysis for the weekly data. In the case of weekly data scale "D1" represents short term behavior (2–4 weeks).

We then regress changes in exchange rates on forward differentials on a scale by scale basis, i.e. we restrict features of the data to be of importance in the medium ("D1"–"D3") or long term ("D4"–"S5"). After decomposing the regression variables we reconstruct the time series using features of the time series at the respective resolutions 1, ..., J only. By testing for the significance of the coefficients estimated in regressions of the decomposed data at various time scales we can infer which of the above outlined possible expectation formations is significant in short, medium and long run. The following Table 2 summarizes the regression results for regressing the US$–Euro and US$–British Pound exchange rate changes

Table 1 Variation of the time series explained by crystals (in %) for forward transaction period of 1 month

	Forward premium (US/Euro)	Exchange rate change (US/Euro)	Forward premium (US/BP)	Exchange rate change (US/BP)
D1	0.455	50.065	0.718	52.012
D2	0.461	23.744	0.508	16.781
D3	1.013	13.291	0.934	14.841
D4	4.164	7.469	5.295	9.321
D5	23.457	2.692	19.533	4.160
S5	70.450	2.738	73.013	2.886

Table 2 Regression results for the US$–Euro and US$–British Pound exchange rate changes regressed on forward premia using reconstructed time series (1 month forward rates)

US$/Euro				US$/British Pound			
Crystals	Intercept	Forward prem.	R2	Crystals	Intercept	Forward prem.	R2
D1	0.00*	3.1*	0.02	D1	0.00*	−1.08	0.01
D2	0.00*	1.17	0.01	D2	0.00*	1.7*	0.03
D3	0.00*	0.33	0.00	D3	0.00*	1.1*	0.02
D4	0.00*	1.7*	0.21	D4	0.00*	1.3*	0.26
D5	0.00*	0.067	0.01	D5	0.00*	0.43*	0.25
S5	0.00*	−0.23*	0.3	S5	0.00*	−0.47*	0.6

*Significance at a 5 % confidence level

on the respective forward premia using the reconstructed time series at scales "D1"–"S5" for a forward transaction period of 1 month.

For the monthly data of US$–Euro exchange rate we find a significant influence of forward premia in explaining exchange rate changes at scales "D1", "D4", and "S5". The significant influence that we find at scales "D1" and "D4" is positive. They represent short medium term (2–4 months) and short long term (1.3–2.6 years), respectively. At level "S5" (more than 5 years) the relationship is, however, significantly negative. The amount of variation explained is highest at levels "D4" and "S5", in all three cases the F-statistics support the estimation design. We therefore conclude that if information from the forward rate is allowed to be of influence for 2–4 months and 1.3–2.6 years respectively then the variables from Eqs. (6)–(7) are significantly, positively linked. This means if we allow information from the previous 2–4 months at scale "D1" and from previous 1.3–2.6 years at scale "D4" to be relevant in explaining respective adjustment time periods of the exchange rate changes that the estimated relationship is positive as is predicted by rational expectation theory. In the long run (more than 5 years) however (at level "S5") the relation is significantly negative. This indicates that at this level a reversion to a mean is the main driving force for market prices. The mean reversion in the long run can be a result from error corrections.

For the US$–British Pound changes in exchange rates can be significantly explained by forward premia at levels "D2", "D3", "D4", "D5", and "S5". Again,

Table 3 Wald-test of the unbiasedness hypothesis for forward transaction time period of 1 month

US$/Euro		US$/British Pound	
Crystals	Test of unbiasedness hypothesis	Crystals	Test of unbiasedness hypothesis
D1	Not rejected	D1	Not rejected
D2	Not rejected	D2	Not rejected
D3	Not rejected	D3	Not rejected
D4	Rejected	D4	Not rejected
D5	Rejected	D5	Rejected
S5	Rejected	S5	Rejected

with the exception of "S5" the estimated relationship is significantly positive in the medium term. For the medium ("D4") and long term ("D5" and "S5") the amount of variation in exchange rate changes explained by respective components of the forward premia is highest. The F-test supports the estimation set up. For the US$–British Pound the rational expectation theory is supported at two medium term scales and two long term scales. In general, the statistical evidence for the rational expectation theory to be the main driving force behind exchange rates changes is higher than in the case of the US$–Euro market because the estimated relationship is significantly positive at all scales except "D1" and "S5". At the highest frequency "D1" the data is not conclusive. This can be a result of the continued importance of technical analysis for that time period as was pointed out by Cheung and Wong (2000). As is the case in the US$–Euro exchange rate, at the longest time period (more than 5 years) the relationship is significantly negative. This supports the idea that in the very long run reversions to a fundamental equilibrium take place, which is one of the stylized facts about capital markets put forth by Cutler et al. (1990).

Determining significant components gives us insights into how long time periods for processing information are. To test the unbiasedness hypothesis, however, we need to impose the linear restriction that the estimated coefficients from Eqs. (6)–(7) comply with rational expectation theory. We apply the Wald test for linear restrictions in a regression model. The null hypothesis (unbiasedness hypothesis) requires the estimated coefficients of the forward premia "b" to be equal to one. Under the null hypothesis the Wald-statistic is distributed as an F-distribution, the degrees of freedom are given by the number of restrictions (i.e. one) and the number of observations in the respective regression analyses, see Wald (1943). The results of these tests are summarized in Table 3.

For the US$–Euro exchange rate we find that the unbiasedness hypothesis is not rejected at levels "D1", "D2", "D3". A significant positive influence of forward premia for these scales is however only given at scale "D1". In other words, the null hypothesis that the estimated coefficient is equal to one is statistically meaningful only at the significant level "D1" (medium term 2–4 months). Although the hypothesis is not rejected at scales two and three as well, regression results indicate that at these time scales the forward premium is not significant in explaining changes in exchange rates. Therefore we conclude, for the US$–Euro exchange rate the forward premium is significant and the unbiasedness-hypothesis is not rejected

only at a time scale where characteristics of the data are influential for 2–4 months (short medium term). At time scales where information is of importance for a longer time period either the forward premium is not significant or the unbiasedness hypothesis is rejected.

For the US$–British Pound exchange rate the forward premia are significant in explaining future exchange rate changes at levels "D2"–"S5". The hypothesis that the estimated coefficient of the forward premia equals one is not rejected for significant levels "D2", "D3", and "D4". At time scale "D1" the hypothesis is not rejected as well, but the regression results indicate that the forward premium is not significant as explanatory variable. The probability for the unbiasedness hypothesis to hold is highest at level "D4". The US$–British Pound exchange rate depicts different characteristics than the US$–Euro exchange rate. In case of US$–British Pound rate we find a significant influence at three levels ("D2", "D3", and "D4"), in addition to that the unbiasedness hypothesis cannot be rejected at these levels. The time scale "D2" represents data characteristics that prevail for 4–8 months, "D3" represents 8–16 months, and "D4" 1.3–2.6 years respectively.

We conclude that aggregating over time scales "D1"–"S5" results in misleading interpretations of the influence of the forward premia in explaining future exchange rate changes because the data demonstrates different behavior in the medium and long term. Only at time scales that represent medium terms the premium is of significant, positive influence for future exchange rate changes. We find different time scales to be significant for different exchange rates.

Finally, to analyze the short term period (2–4 weeks) as well, the above analysis is repeated using weekly data which allows the definition of a short term period. The hypothesis is not supported by statistical inference in the short run, which is what the survey data suggests (see Allen and Taylor 1990; Cheung and Wong 2000). At the short horizon technical trading is perceived to be the most important influence in forming expectations, therefore the insignificance of the forward rate to explain exchange rate changes for that time period is in line with previous results and market data.

In order to check for the robustness of our findings we also analyze the forward premium anomaly for time to maturities of the forward contract of 3 months, 6 months, and 1, 2, and 5 years in the same manner that has been described above. The MODWT shows similar influences of the crystals "D1"–"S5" in explaining the variance of the time series exchange rate changes and forward premia in the case of the three, 6–12 months forward transaction period. Again, the exchange rate changes are best explained by information at every time scale, whereas for the forward premia lower scales carry more influence. Once again, we reconstruct the time series exchange rate change and forward premium for the various forward transaction time periods with information captured at the various time scales "D1"–"S5" for the two exchange rates US$–Euro and US$–British Pound. With the reconstructed time series we estimate Eqs. (6) and (7). In case of the 3 and 6 months forward transaction periods we find similar results as in the 1 month forward transaction period. However, there are differences with regards to the significantly, positive relationships for the medium terms. The relationships are less significant. In case

of the US$–Euro exchange rate the unbiasedness hypothesis is supported at scale "D1" for forward transaction time periods up to 1 year. For the US$–British Pound the unbiasedness hypothesis continues to be supported at scale "D4" and "D5". For forward transaction time periods of 1 year and above, nearly all the significant relationships become negative, i.e. a correction to overshooting and reversions to mean become the main driving forces for forecasting periods of more than 1 year.

5 Conclusion

In this paper we argue that the assumptions made in standard econometric procedures to test for the unbiasedness hypothesis might be responsible for the failure of the theory to be validated in practice. We use the maximal overlap discrete wavelet transform to decompose the data into their time-scale components to allow for inefficiencies in the exchange rate markets. We assume feedback, noise, technical, fundamental and rational trading to be present. The decomposition of the time series exchange rate changes and forward premia allows information to continue to be relevant in the price formation for specific, pre-defined time periods. This way we analyze the forward premium anomaly at different time scales. We then test the unbiasedness hypothesis at the respective scales and find that the hypothesis cannot be rejected at certain time scales. We get different results from analyzing the US$–Euro and the US$–British Pound exchange rates. In case of the forward transaction time period of 1 month a significant positive relationship between the variables forward premia and exchange rate changes can be found in the medium terms for both currencies. It is more pronounced in the case of US$–British Pound exchange rate. The unbiasedness hypothesis is supported for the US$–Euro exchange rate for a time scale of 2–4 months. In case of the US$–British Pound exchange rate the unbiasedness hypothesis gets supported at medium term time scales and for the time period of 1.3–2.6 years (i.e. long term). The analysis of weekly data allows for the definition of a short term period. We find that the unbiasedness hypothesis is not supported in the short run which is in line with survey data for exchange rate markets. The findings are similar when the forward transaction period is extended from 1 month to 5 years. For forecasting periods above a year the influence of the forward market is mostly significantly negative. We conclude that the adjustment time period to new information is crucial for the validity of the unbiasedness hypothesis. Aggregating over the time scales veils the fact, that the theory seems appropriate for certain time periods only. The unbiasedness hypothesis is supported for the medium term.

References

Allen H, Taylor MP (1990) Charts, noise and fundamentals in the London foreign exchange market. Econ J 100:49–59

Backus D, Foresi S, Telmer C (1996) Affine models of currency pricing. NBER Working Paper 5623

Bensberg D (2012) Das forward premium puzzle als ergebnis adverser selektion: eine untersuchung auf theoretischer basis. SVH-Verlag, Saarbruecken

Black F (1988) An equilibrium model of the crash. NBER Macroeconomics Annual, pp 269–395

Cheung Y, Wong C (2000) A survey of market practitioners' views on exchange rate dynamics. J Int Econ 51:401–419

Chinn MD, Meredith G (2004) Monetary policy and long horizon uncovered interest parity, IMF Staff Papers 51(3)

Crowley PM (2005) An intuitive guide to wavelets for economists. Bank of Finland Research Discussion Papers

Cutler DM, Poterba JM, Summers LH (1990) Speculative dynamics and the role of feedback traders. Am Econ Rev 80(2):63–68

De Long JB, Shleifer A, Summers LH, Waldmann RJ (1990) Positive-feedback investment strategies and destabilizing rational speculation. J Finance 45(2):379–395

De Long JB, Shleifer A, Summers LH, Waldmann RJ (1991) The survival of noise traders in financial markets. J Bus 64:1–19

Engel C (1996) The forward discount anomaly and the risk premium: a survey of recent evidence. J Empirical Finance 3(2):123–192

Fama E (1984) Forward and spot exchange rates. J Monetary Econ 14:319–338

Frankel JA, Froot KA (1986) Understanding the US dollar in the eighties: the expectations of chartists and fundamentalists. Economic Record Special Issue, 24–38

Froot KA, Thaler RH (1990) Anomalies: foreign exchange. J Econ Perspect 4(3):179–192

Gallegati M, Gallegati M, Ramsey JB, Semmler W (2011) The US wage Phillips curve across frequencies and over time. Oxf Bull Econ Stat 73(4):489–508

Gencay R, Selçuk R, Whitcher B (2009) An introduction to wavelets and other filtering methods in finance and economics. Academic, Philadelphia

Hodrick RJ (1987) The empirical evidence on the efficiency of forward and futures foreign exchange market. Harwood Academic Publisher, Switzerland

Hodrick RJ, Srivastava S (1986) The covariation of risk premia and expected future spot rates. J Int Money Finance 3:5–30

Kaminski G, Peruga R (1990) Can a time varying risk premium explain excess returns in the forward market for foreign exchange? J Int Econ 28:47–70

Kim S, Haueck In F (2003) The relationship between financial variables and real economic activity: evidence from spectral and wavelet analysis. Stud Nonlinear Dynamics Econ 7(4):1–18

Kiermeier MM (2014) Essay on wavelet analysis and the European term structure of interest rates. Business and Economic Horizons. 9(4):18–26. Doi: 10.15208/beh.2013.19.

MacDonald R, Taylor MP (1992) Exchange rate economics: a survey. IMF Staff Papers 39:1–57

Raihan S, Wen Y, Zeng B (2005) Wavelet: a new tool for business cycle analysis. The Federal Reserve Bank of St. Louis, Working Paper 2005-050A

Ramsey JB, Lampart C (1996) The decomposition of economic relationships by time scale using wavelets. New York University, New York

Shapiro AC (2009) Multinational financial management. Wiley, Hoboken

Taylor MP (1995) The economics of exchange rates. J Econ Lit 33(1):13–47

Wald A (1943) Tests of statistical hypothesis concerning several parameters when the number of observations is large. Trans Am Math Soc 54:426–482

Wang P, Jones T (2002) Testing for efficiency and rationality in foreign exchange markets: a review of the literature and research on foreign exchange market efficiency and rationality with comments. J Int Money Finance 21:223–239

Oil Shocks and the Euro as an Optimum Currency Area

Luís Aguiar-Conraria, Teresa Maria Rodrigues, and Maria Joana Soares

Abstract We use wavelet analysis to study the impact of the Euro adoption on the oil price macroeconomy relation in the Euroland. We uncover evidence that the oil-macroeconomy relation changed in the past decades. We show that after the Euro adoption some countries became more similar with respect to how their macroeconomies react to oil shocks. However, we also conclude that the adoption of the common currency did not contribute to a higher degree of synchronization between Portugal, Ireland and Belgium and the rest of the countries in the Euroland. On the contrary, in these countries the macroeconomic reaction to an oil shock became more asymmetric after adopting the Euro.

1 Introduction

The literature on business cycle synchronization is related to the literature on optimal currency areas. If several countries delegate on some supranational institution the power to perform a common monetary policy, then they lose this policy stabilization instrument. Obviously, business cycle synchronization is not sufficient to guarantee that a monetary union is desirable; however, it is, arguably, a

L. Aguiar-Conraria (✉)
NIPE and Economics Department, University of Minho, Braga, Portugal
e-mail: lfaguiar@eeg.uminho.pt

T.M. Rodrigues
Economics Department, University of Minho, Braga, Portugal
e-mail: id3662@alunos.uminho.pt

M.J. Soares
NIPE and Department of Mathematics and Applications, University of Minho, Braga, Portugal
e-mail: jsoares@math.uminho.pt

M. Gallegati and W. Semmler (eds.), *Wavelet Applications in Economics and Finance*, Dynamic Modeling and Econometrics in Economics and Finance 20,
DOI 10.1007/978-3-319-07061-2_7,

necessary condition: a country with an asynchronous business cycle will face several difficulties in a monetary union, because of the 'wrong' stabilization policies.

In the economics literature, to test if a group of countries form an Optimum Currency Area (OCA), it is common to check if the different countries face essentially symmetric or asymmetric exogenous shocks (e.g. see Peersman 2011). In the latter case, it is more difficult to argue for a monetary union. However, even if the shock is symmetric, one still has to check if its impact is similar across countries. If this is not the case, the symmetric shock will have asymmetric effects, which deteriorates the case for a monetary union.

There is a caveat to the previous argument. Some authors argue that even if a region is not *ex ante* an OCA it may, *ex post*, become one. The argument for this endogenous OCA is simple and intuitive: by itself the creation of a common currency area will create the conditions for the area to become an OCA. For example, Frankel and Rose (1998) and Rose and Engel (2002) argue that, because currency union members have more trade, business cycles are more synchronized across currency union countries. Imbs (2004) makes a similar argument for financial links. After the creation of a currency area, the finance sector will become more integrated and hence business cycles will become more synchronized. In effect, Inklaar et al. (2008) conclude that convergence in monetary and fiscal policies has a significant impact on business cycle synchronization. However, Baxter and Kouparitsas (2005) conclude otherwise and Camacho et al. (2008) present evidence that differences between business cycles in Europe have not been disappearing.

We tackle this issue by focusing on one shock that every country faces: oil price changes. We study the relation between oil and the macroeconomy in the 11 countries that first joined the Euro in 1999. We investigate how this relation changed after the adoption of the Euro and test if it became more or less asymmetric after the Euro adoption. The analysis is performed in the time-frequency domain, using wavelet analysis.

We are not the first authors to use wavelets to analyse the oil price-macroeconomy relationship. Naccache (2011) and Aguiar-Conraria and Soares (2011a) have already relied on this technique to assess this relation. Actually, wavelet analysis is particularly well suited for this purpose for several reasons. First, because oil price dynamics is highly nonstationary, it is important to use a technique, such as wavelet analysis, that does not require stationarity. Second, wavelet analysis is particularly useful to study how relations evolve not only across time, but also across frequencies, as it is unlikely that these relations remain invariant. Third, Kyrtsou et al. (2009) presented evidence showing that several energy markets display consistent nonlinear dependencies. Based on their analysis, the authors call for nonlinear methods to analyze the impact of oil shocks. Wavelet analysis is one such method. We should also add that wavelets have already proven to be insightful when studying business cycles synchronizations, e.g. see Aguiar-Conraria and Soares (2011b) and Crowley and Mayes (2008).

We use data on the Industrial Production for the first countries joining the Euro and estimate the coherence between this variable and oil prices. The statistical procedure is similar to the one used by Vacha and Barunick (2012) to study

co-movements in the time-frequency space between energy commodities. By itself, this analysis will allow us to characterize how the relationship evolved and how the 2000s are different from the 1980/1990s. We will see that in late 1980s and early 1990s, the strongest coherence is for cycles with periods that range between 4 and 8 years, while in more recent times it became a shorter run relation, with coherence being higher for cycles with periods between 2 and 4 years.

After estimating, for each country, the coherencies between industrial production and oil prices, we propose a metric to compare these coherencies and measure and test the degree of synchronization among countries. Interestingly, we show that the relation between oil and the macroeconomy in some countries was more similar before than after the euro adoption. This is particularly true for Portugal, Ireland and Belgium. It seems that, at least for these three countries, the endogenous OCA theory is refuted.

This chapter follows a very simple structure. In Sect. 2, we provide a brief introduction to the mathematics of the continuous wavelet transform and explain how to derive the metric that we use to compare the oil-macroeconomy relation in the different countries. We also discuss the advantages of choosing a complex wavelet function. We describe the data and present our results in section three and section four concludes.

2 The Continuous Wavelet Transform

Wavelet analysis performs the estimation of the spectral characteristics of a time-series as a function of time, revealing how the different periodic components of a particular time-series evolve over time. This technical presentation is, necessarily brief. For a detailed technical overview, the reader can check Aguiar-Conraria and Soares (2014). Alternatively, for a thorough intuitive discussion on these concepts, the reader is referred to Cazelles et al. (2007) and Aguiar-Conraria et al. (2012).

A wavelet is simply a rapid decaying oscillatory function. Mathematically, for $\psi(t)$ to be called a wavelet, it must satisfy $\int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty$ and a certain technical condition which, for functions with sufficient decay, is equivalent to requiring that it has zero mean, i.e. $\int_{-\infty}^{\infty} \psi(t)\, dt = 0$. The continuous wavelet transform (CWT) of a given time-series x is given by

$$W_x(\tau, s) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{|s|}} \overline{\psi}\left(\frac{t-\tau}{s}\right) dt, \tag{1}$$

where s is a scaling or dilation factor that controls the width of the wavelet and τ is a translation parameter controlling the location of the wavelet. Here, and throughout, the bar denotes complex conjugate.

When the wavelet $\psi(t)$ is chosen as a complex-valued function, as we do, the wavelet transform $W_x(\tau, s)$ is also complex-valued. In this case, the transform can be

separated into its real part, $\Re(W_x)$, and imaginary part, $\Im(W_x)$, or in its amplitude, $|W_x(\tau,s)|$, and phase, $\phi_x(\tau,s)$: $W_x(\tau,s) = |W_x(\tau,s)|\, e^{i\phi_x(\tau,s)}$.[1] For real-valued wavelet functions, the imaginary part is constantly zero and the phase is, therefore, undefined.

When one is interested in studying the oscillatory behavior of a variable, or a set of variables, it is almost mandatory to use a complex wavelet, because the phase yields important information about the position of the variable in the cycle. In particular, if one is comparing two time-series, one can compute the phases and the phase-difference of the wavelet transform of each series and thus obtain information about the possible delays in the oscillations of the two series as a function of time and frequency.

In order to describe the time-frequency localization properties of the CWT, we have to assume that both the wavelet ψ and its Fourier transform $\hat{\psi}$ are well localized functions. More precisely, these functions must have sufficient decay to guarantee that the quantities defined below are all finite.[2] In what follows, for simplicity, assume that the wavelet has been normalized so that $\int_{-\infty}^{\infty} |\psi(t)|^2\, dt = 1$. With this normalization, $|\psi(t)|^2$ defines a probability density function. The mean and standard deviation of this distribution are called, respectively, the center, μ_ψ, and radius, σ_ψ, of the wavelet. They are, naturally, measures of localization and spread of the wavelet. The center $\mu_{\hat{\psi}}$ and radius $\sigma_{\hat{\psi}}$ of $\hat{\psi}$, the Fourier transform of the wavelet ψ, are defined in a similar manner. The interval $\left[\mu_\psi - \sigma_\psi, \mu_\psi + \sigma_\psi\right]$ is the set where $\psi(t)$ attains its "most significant" values whilst the interval $\left[\mu_{\hat{\psi}} - \sigma_{\hat{\psi}}, \mu_{\hat{\psi}} + \sigma_{\hat{\psi}}\right]$ plays the same role for $\hat{\psi}(f)$. The rectangle $H_\psi := \left[\mu_\psi - \sigma_\psi, \mu_\psi + \sigma_\psi\right] \times \left[\mu_{\hat{\psi}} - \sigma_{\hat{\psi}}, \mu_{\hat{\psi}} + \sigma_{\hat{\psi}}\right]$ in the $(t,f)-$plane is called the Heisenberg box or window for the function ψ. We then say that ψ is localized around the point $\left(\mu_\psi, \mu_{\hat{\psi}}\right)$ of the time-frequency plane, with uncertainty given by $\sigma_\psi \sigma_{\hat{\psi}}$. The Heisenberg uncertainty principle establishes that the uncertainty is bounded from below by the quantity $1/2$.

The Morlet wavelet became the most popular of the complex valued wavelets for several reasons.[3] Among then we highlight two: (1) the Heisenberg box area reaches its lower bound with this wavelet, i.e. the uncertainty attains the minimum possible value; (2) the time radius and the frequency radius are equal, $\sigma_\psi = \sigma_{\hat{\psi}} = \frac{1}{\sqrt{2}}$,and, therefore, this wavelet represents the best compromise between time and frequency concentration. The Morlet wavelet is given by

[1] The phase-angle $\phi_x(\tau,s)$ of the complex number $W_x(\tau,s)$ can be obtained from the formula: $\tan(\phi_x(\tau,s)) = \frac{\Im(W_x(\tau,s))}{\Re(W_x(\tau,s))}$, using the information on the signs of $\Re(W_x)$ and $\Im(W_x)$ to determine to which quadrant the angle belongs to.

[2] The precise requirements are that $|\psi(t)| < C(1+|t|)^{-(1+\epsilon)}$ and $|\hat{\psi}(f)| < C(1+|f|)^{-(1+\epsilon)}$, for $C < \infty, \epsilon > 0$.

[3] Actually, it is also common to call it the Gabor wavelet. Authors who do this, usually reserve the name Morlet to the real part of Eq. (2).

$$\psi_{\omega_0}(t) = \pi^{-\frac{1}{4}} e^{i\omega_0 t} e^{-\frac{t^2}{2}}, \tag{2}$$

where ω_0 is a localization parameter. Strictly speaking $\psi_{\omega_0}(t)$ is not a true wavelet, however, for sufficiently large ω_0 (e.g. $\omega_0 > 5$), for all practical purposes can be considered as such. For the most common choice of ω_0, $\omega_0 = 6$, we have that $f \simeq \frac{1}{s}$ facilitating the conversion from scales to frequencies. To our knowledge, in economics, every paper that uses the continuous wavelet transform uses $\omega_0 = 6$.

Another important family of analytic wavelets is the Generalized Morse Wavelets. This family of wavelets is increasingly popular in physical sciences. Like in Aguiar-Conraria and Soares (2011b), we checked if our results are robust to the use of this other wavelet family. For a range of reasonable parameter values, namely values that imply that the Heisenberg box area was close to its lower bound, our results were quite similar.

2.1 Wavelet and Cross Wavelet Power

In analogy with the terminology used in the Fourier case, the (local) wavelet power spectrum (sometimes called scalogram or wavelet periodogram) is defined as $WPS_x(\tau, s) = |W_x(\tau, s)|^2$. This gives us a measure of the variance distribution of the time-series in the time-scale (frequency) plane.

In our applications, we are interested in detecting and quantifying relationships between two time series. The concepts of cross-wavelet power, cross-wavelet coherency and wavelet phase-difference are natural generalizations of the basic wavelet analysis tools that enable us to deal with the time-frequency dependencies between two time-series.

The cross-wavelet transform of two time-series, $x(t)$ and $y(t)$, is defined as

$$W_{xy}(\tau, s) = W_x(\tau, s)\,\overline{W_y}(\tau, s), \tag{3}$$

where W_x and W_y are the wavelet transforms of x and y, respectively. The cross-wavelet power is simply given by $\left|W_{xy}(\tau, s)\right|$. While we can interpret the wavelet power spectrum as depicting the local variance of a time-series, the cross-wavelet power of two time-series depicts the local covariance between these time-series at each time and frequency.

In analogy with the concept of coherency used in Fourier analysis, given two time series $x(t)$ and $y(t)$ one can define their *complex wavelet coherency* ϱ_{xy} by:

$$\varrho_{xy} = \frac{S\left(W_{xy}\right)}{\left[S\left(|W_x|^2\right) S\left(|W_y|^2\right)\right]^{1/2}}, \tag{4}$$

where S denotes a smoothing operator in both time and scale; smoothing is necessary, because, otherwise, coherency would have modulus one at all scales and

times.[4] Time and scale smoothing can be achieved by convolution with appropriate windows; see Aguiar-Conraria and Soares (2014), for details.

The absolute value of the complex wavelet coherency is called the *wavelet coherency* and is denoted by R_{xy}, i.e.

$$R_{xy} = \frac{|S\left(W_{xy}\right)|}{\left[S\left(|W_x|^2\right) S\left(|W_y|^2\right)\right]^{1/2}}, \tag{5}$$

with $0 \leq R_{xy}(\tau, s) \leq 1$.

The complex wavelet coherency can be written in polar form, as $\varrho_{xy} = \left|\varrho_{xy}\right| e^{i\phi_{xy}}$. The angle ϕ_{xy} is called the *phase-difference* (phase lead of x over y), i.e.

$$\phi_{xy} = \text{Arctan}\left(\frac{\Im\left(S\left(W_{xy}\right)\right)}{\Re\left(S\left(W_{xy}\right)\right)}\right). \tag{6}$$

A phase-difference[5] of zero indicates that the time series move together at the specified time-frequency; if $\phi_{xy} \in (0, \frac{\pi}{2})$, then the series move in phase, but the time series x leads y; if $\phi_{xy} \in (-\frac{\pi}{2}, 0)$, then it is y that is leading; a phase-difference of π (or $-\pi$) indicates an anti-phase relation; if $\phi_{xy} \in (\frac{\pi}{2}, \pi)$, then y is leading; time series x is leading if $\phi_{xy} \in (-\pi, -\frac{\pi}{2})$.

To test for statistical significance of the wavelet coherency we rely on Monte Carlo simulations. However, there are no such tests for the phase-differences, because there is no consensus on how to define the null hypothesis. The advice is that we should only interpret the phase-difference on the regions where coherency is statistically significant.

2.2 Complex Wavelet Coherency Distance Matrix

In this section, we adapt a formula derived by Aguiar-Conraria and Soares (2011b) to find a metric for measuring the distance between a pair of matrices of complex coherencies. Given two $F \times T$ matrices C_x and C_y of complex coherencies, let $C_{xy} = C_x C_y^H$, where C_y^H is the conjugate transpose of C_y, be their covariance matrix. Performing the Singular Value Decomposition (SVD) of this matrix yields

$$C_{xy} = U \Sigma V^H, \tag{7}$$

[4] In the above formula and in what follows, we will omit the arguments (τ, s).

[5] Some authors prefer a slightly different definition, $\text{Arctan}\left(\frac{\Im(W_{xy})}{\Re(W_{xy})}\right)$. In this case, one would have $\phi_{xy} = \phi_x - \phi_y$, hence the name phase-difference.

where the matrices U and V are unitary matrices (i.e. $U^H U = V^H V = I$), and $\Sigma = \text{diag}(\sigma_i)$ is a diagonal matrix with non-negative diagonal elements ordered from highest to lowest, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_F \geq 0$. The columns, $\mathbf{u}_k$ of the matrix U and the columns $\mathbf{v}_k$ of V are known, respectively, as the singular vectors for C_x and C_y, and the σ_i are known as the singular values. Let $\mathbf{l}_x^k$ and $\mathbf{l}_y^k$ be the so-called leading patterns, i.e. the $1 \times T$ vectors obtained by projecting each of the matrices C_x and C_y onto the respective kth singular vector (axis):

$$\mathbf{l}_x^k := \mathbf{u}_k^{\mathrm{H}} C_x \quad \text{and} \quad \mathbf{l}_y^k := \mathbf{v}_k^H C_y. \tag{8}$$

It can be shown that each of the matrices C_x and C_y can be written as

$$C_x = \sum_{k=1}^{F} \mathbf{u}_k \mathbf{l}_x^k, \quad C_y = \sum_{k=1}^{F} \mathbf{v}_k \mathbf{l}_y^k, \tag{9}$$

and also that very good approximations can be obtained by using only a small number $K < F$ of terms in the above expressions.

After reducing the information contained in the complex coherency matrices C_x and C_y to a few components, say the K most relevant leading patterns and singular vectors, the idea is to define a distance between the two matrices by appropriately measuring the distances from these components. We compute the distance between two vectors (leading patterns or leading vectors) by measuring the angles between each pair of corresponding segments, defined by the consecutive points of the two vectors, and take the mean of these values. This would be easy to perform if all the values were real. In our case, because we use a complex wavelet, we need to define an angle in a complex vector space. Aguiar-Conraria and Soares (2011b) discuss several alternatives. In this paper, we make use of the Hermitian inner product $\langle \mathbf{a}, \mathbf{b} \rangle_{\mathbb{C}} = \mathbf{a}^H \mathbf{b}$ and corresponding norm $\|\mathbf{a}\| = \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle_{\mathbb{C}}}$ and compute the so-called Hermitian angle between the complex vectors $\mathbf{a}$ and $\mathbf{b}$, $\Theta_H(\mathbf{a}, \mathbf{b})$, by the formula

$$\cos(\Theta_H) = \frac{|\langle \mathbf{a}, \mathbf{b} \rangle_{\mathbb{C}}|}{\|\mathbf{a}\| \|\mathbf{b}\|}, \quad \Theta_H \in [0, \frac{\pi}{2}]. \tag{10}$$

The distance between two complex vectors $\mathbf{p} = (p_1, \ldots, p_M)$ and $\mathbf{q} = (q_1, \ldots, q_M)$ (applicable to the leading patterns and leading vectors) is simply defined by

$$d(\mathbf{p}, \mathbf{q}) = \frac{1}{M-1} \sum_{i=1}^{M-1} \Theta_H \left(\mathbf{s}_i^{\mathbf{p}}, \mathbf{s}_i^{\mathbf{q}} \right) \tag{11}$$

where the ith segment $\mathbf{s}_i^{\mathbf{p}}$ is the two-vector $\mathbf{s}_i^{\mathbf{p}} := (i+1, p_{i+1}) - (i, p_i) = (1, p_{i+1} - p_i)$. To compare the matrix C_x with the complex wavelet coherencies of country x

with the corresponding matrix for country y, C_y, we then compute the following distance:

$$\text{dist}\left(C_x, C_y\right) = \frac{\sum_{k=1}^{K} \sigma_k^2 \left[d\left(\mathbf{l}_x^k, \mathbf{l}_y^k\right) + d\left(\mathbf{u}_k, \mathbf{v}_k\right)\right]}{\sum_{k=1}^{K} \sigma_k^2}, \tag{12}$$

where σ_k are the kth largest singular values that correspond to the first K leading patterns and K leading vectors.

The above distance is computed for each pair of countries and, with this information, we can then fill a matrix of distances.

3 The Oil-Macroeconomy Relationship and the Euro[6]

We analyze the oil price-macroeconomy relation in the Euro area by looking both at the coherency content and phasing of cycles. We look at the first 11 countries joining the Euro: Austria, Belgium, Finland, France, Germany, Ireland, Italy, Luxembourg, Netherlands, Portugal and Spain. For this type of purpose, to measure real economic activity, most studies use either real GDP or an Industrial Production Index. We use the Industrial Production Index because wavelet analysis is quite data demanding, and having monthly data is a plus. We use seasonally adjusted data from the OECD Main Economic Indicators database, from January of 1986 to December of 2011. We have, therefore, 26 years of data. Exactly 13 years before and 13 years after the Euro adoption. The oil price data is the West Texas Intermediate Spot Oil Price taken from the Federal Reserve Economic Data—FRED—St. Louis Fed.

In Fig. 1, we can see the behavior of the Industrial Production Index for three distinct countries: Finland, Germany, and Portugal. These three countries, as we will see next, have distinct behaviors. While Germany is part of the Euro core, Finland and Portugal are not. In particular, while in the frst half of the sample Finland is not synchronized with the rest of Europe, in the second half the convergence is obvious. This convergence is not observed in the case of Portugal. If one computes the correlation between the series, these results can be reasonably predicted. For example, before 1999, the correlation between Finland's and Germany's IP is 0.05, which increases to 0.83 after 1999. Between Portugal and Germany, the correlation between the two series remains relatively constant (0.5 in both samples).

Note, however, that we do not plan to compare the industrial production indexes by themselves. We want to compare their reactions to the same oil price shocks. For each country, we estimate the wavelet coherency between the yearly rate of growth of Industrial Production and the oil price. It is known that oil price increases are

[6]To replicate our results, the reader can use a Matlab wavelet toolbox that we wrote. It is freely available at http://sites.google.com/site/aguiarconraria/joanasoares-wavelets. Our data is also available in that website.

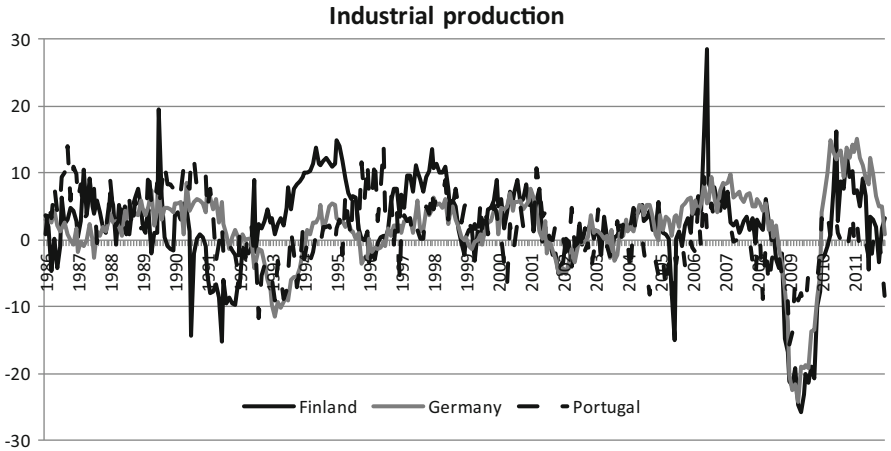


Fig. 1 Industrial Production Index for selected countries

more important than oil price decreases. Because of that, Hamilton (1996 and 2003) proposed a nonlinear transformation of the oil price series. In our computations, we use the Hamilton's Net Oil Price. Because we focus our analysis on business cycle frequencies, we estimate the coherence for frequencies corresponding to periods between 1.5 and 8 years.

In Fig. 2, we have our first set of results. For each country, on the left (a) we have the wavelet coherency between Industrial Production and Oil Prices.[7] On the right, we have the phase-differences: on the top (b), we have the phase-difference in the 2–4 years frequency band (chosen to capture the region of high coherency that appears in most of the countries after 2000); in the bottom (c), we have the phase-difference in the 4–8 years frequency band, which captures the regions of high coherency in the late 1980s and in the first half of the 1990s—recall that it only makes sense to interpret the phase-differences in the regions of high coherency.

For most countries the region with the strongest coherency is located between the mid-1980s and mid-1990s at the 4–8 years frequency band. And for most countries, the phase-difference is consistently between $\pi/2$ and π, suggesting that oil price increases anticipate downturns in the Industrial Production. After the Euro adoption, in 1999, for most of the countries, the strongest region of high coherency is in the 2–4 years frequency band after 2005. Again, the phase differences are located between $\pi/2$ and π, consistent with the idea that negative oil shocks anticipate

[7]The grey contour designates the 5 % significance level, obtained by 1,000 Monte Carlo simulations based on two independent ARMA(1,1) processes as the null. Coherency ranges from white/light grey (low coherency) to black/dark grey (high coherency). The cone of influence, which is the region subject to border distortions is shown with a thick line.

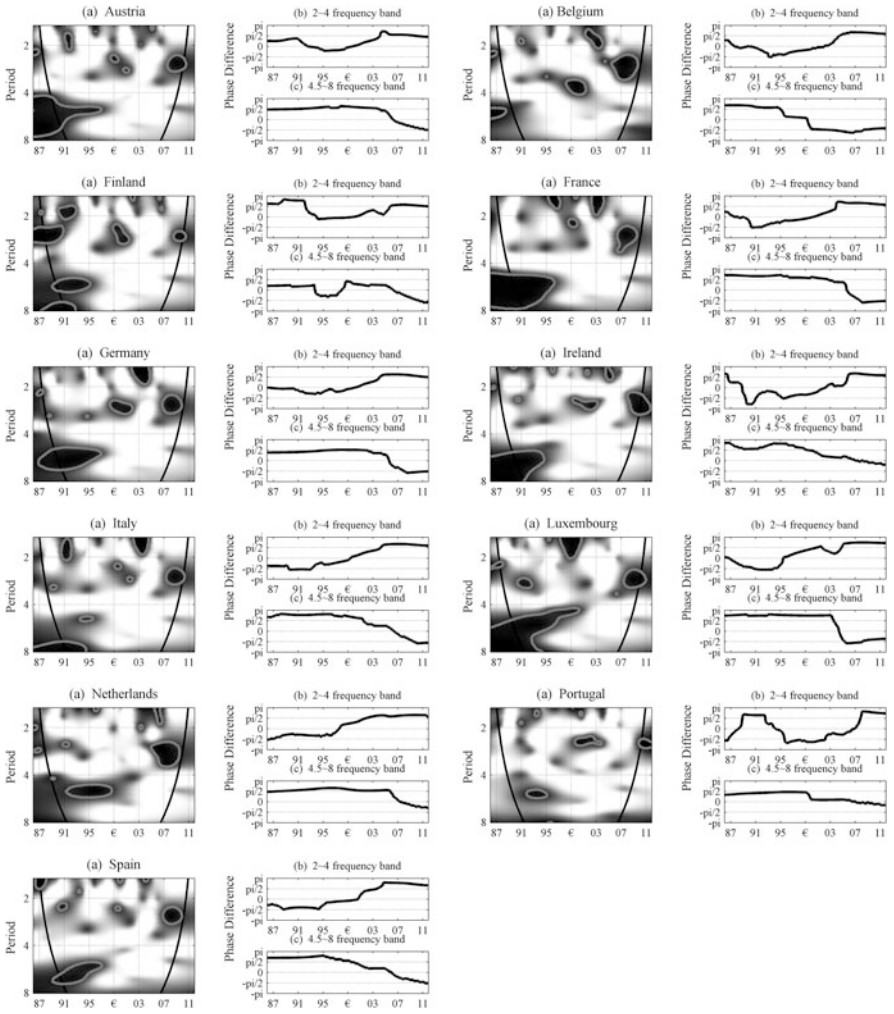


Fig. 2 On the left—wavelet coherency between each country's Industrial Production and Oil Prices. The *grey* scale ranges from *white/light grey* (low coherency) to *black/dark grey* (high coherency). The grey contour designates the 5 % significance level, based on MonteCarlo simulations. On the right—phase-difference between Industrial Production and Oil Prices at 2–4 years (*top*) and 4–8 years (*bottom*) frequency bands

downturns in the Industrial Production. The most interesting aspect is this change in the predominant frequencies.

These results are consistent with the results of other authors, who conclude that, in the more recent times, the negative impact of oil shocks is shorter-lived than before. This may happen because the oil exporting countries follow different pricing strategies—see, for example, Aguiar-Conraria and Wen (2012)—, because

	Au	Be	Fi	Fr	Ge	Ir	It	Lx	Ne	Pt	Sp
Austria		0.091	**0.055**	**0.042**	**0.050**	0.097	**0.048**	**0.066**	**0.056**	0.095	**0.049**
Belgium	**0.056**		0.084	0.085	0.082	0.126	0.077	0.087	**0.074**	**0.070**	0.093
Finland	**0.077**	**0.067**		**0.061**	**0.078**	0.097	**0.053**	**0.073**	**0.059**	0.093	**0.065**
France	**0.049**	0.076	0.075		**0.054**	0.100	**0.049**	**0.069**	**0.049**	0.086	**0.047**
Germany	**0.041**	**0.067**	**0.074**	**0.053**		0.089	**0.043**	**0.061**	**0.062**	0.083	**0.054**
Ireland	**0.056**	**0.063**	0.072	**0.054**	**0.064**		0.075	0.106	0.098	0.104	**0.078**
Italy	**0.060**	**0.066**	0.078	**0.048**	**0.058**	**0.060**		**0.059**	**0.051**	0.084	**0.040**
Luxembourg	**0.056**	**0.066**	0.079	**0.065**	**0.057**	**0.059**	**0.050**		0.072	0.083	**0.067**
Netherlands	**0.059**	0.075	**0.067**	**0.073**	**0.059**	**0.060**	**0.067**	**0.060**		0.080	**0.053**
Portugal	**0.075**	0.077	**0.074**	**0.068**	0.071	0.070	**0.056**	**0.069**	**0.062**		0.092
Spain	**0.063**	0.076	0.078	**0.052**	**0.065**	**0.045**	**0.060**	**0.066**	**0.055**	0.083	
Gray scale code:			**p < 0.01**			**p < 0.05**			**p < 0.10**		

Fig. 3 Lower triangle-complex wavelet dissimilarities before the Euro. Upper triangle-complex wavelet dissimilarities after the Euro

the nature of oil shocks was different—see, for example, Hamilton (2009) or Kilian (2008 and 2009)—or because the western macroeconomies became more flexible—see, for example, Blanchard and Galí (2010) who argue that less rigid wages as well as a smaller share of oil in the production are candidate explanations for the shorter-lived impact of oil shocks.

To assess if the oil price-macroeconomy relation is similar between two countries, we compute the distance between the complex wavelet coherencies matrices associated with both countries, using formula (12). This measure takes into account both the real and the imaginary parts of the complex coherency. A value very close to zero means that (1) the contribution of cycles at each frequency to the total correlation between oil prices and the industrial production is similar in both countries, (2) this contribution happens at the same time in both countries and, finally, (3) the leads and lags between the oil price cycles and the industrial production cycles are similar in both countries. Note that the Anna Karenina principle applies. If the distance is zero, or close to zero, the two series are similar in every regard. If the distance is not zero, the origin of the distance may be any of the three referred aspects. To distinguish between them, one may look at the pictures with the coherency and phase-differences between the two series. To test if the similarity is statistically significant, we again rely on Monte Carlo simulations.

For each pair of countries we estimate two distances: one before the Euro adoption and the other after the adoption. It is as if we divide each of the pictures in Fig. 2 in two halves: left and right. To measure the distances between two countries before the Euro, we compare the left halves. And we compare the right halves to measure the distance after 1999. Given that, by definition, a distance matrix is symmetric, to save space, we use the lower triangle for the distances before the Euro adoption and in the upper-triangle we have the distances after the euro adoption. These results are described in Fig. 3.

It is interesting to note that the endogeneity of the OCAs does not survive our analysis, at least when one considers the case of Portugal, Belgium and, even more

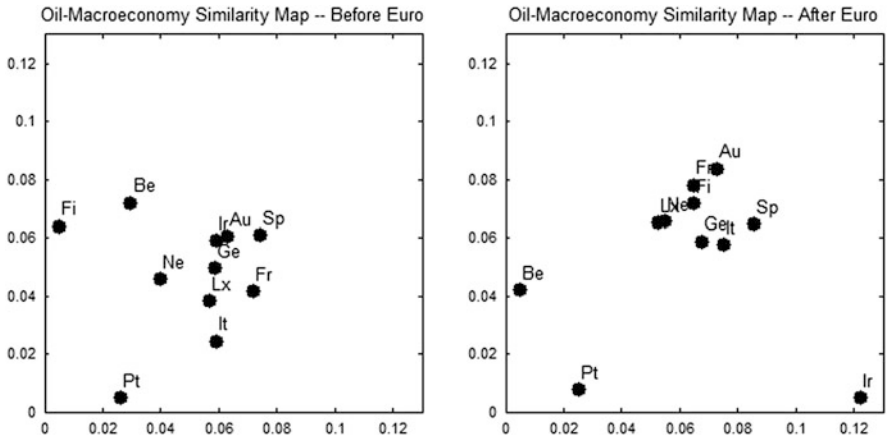


Fig. 4 Multidimensional scaling maps

strongly, Ireland. Before the Euro adoption, Portugal was synchronized with Italy (1 % significance), France, Netherlands, Luxembourg (5 % significance), Austria, Finland, Germany and Ireland (10 % significance). In the second half of the sample, Portugal is only synchronized with Belgium. Similar results hold for Belgium. The case of Ireland is even stronger. Before the birth of the Euro, Ireland was synchronized with every country except Finland. With 1 % significance in the majority of the cases. After the Euro adoption, Ireland is simply synchronized with Italy and Spain at 10 % significance level. The only country that clearly became more synchronized after the Euro adoption was Finland.

The same information is displayed in Fig. 4, where we use the distances of Fig. 3 to plot a map with the countries into a two-axis system—see Camacho et al. (2006).[8] This cannot be performed with perfect accuracy because distances are not Euclidean. In these maps it is clear that while most of the countries became slightly tighter, particularly in the case of Finland who moved to the core after 1999, this was not the case for Belgium, Portugal and Ireland, who now look like three isolated islands with no strong connections to mainland.

4 Conclusions

Unlike most previous studies on OCAs and on business cycle synchronization, which rely on time domain methods—such as VAR, gravity and panel data models—, we relied on time-frequency domain methods. To be more precise, we

[8] Basically, we reduce each of the distance matrices to a two-column matrix, called the configuration matrix, which contains the position of each country in two orthogonal axes.

used wavelet analysis to study the impact of the Euro adoption on the member countries' macroeconomic reaction to one of the most common shocks: oil shocks. Given that energy is such an important production input, and that due to several reasons (including ecological, political and economic reasons) it is such a volatile sector, the transmission mechanism of oil shocks to the macroeconomy is bound to have important effects. If a group of countries have asymmetric responses to the same oil shock, it is highly unlikely that those countries form an OCA.

We estimated the wavelet coherency between the Industrial production of the 11 countries that first joined the Euro and the oil price. We uncovered evidence that shows that the oil-macroeconomy relation changed in the past decades. In the second half of 1980s and in the first half of 1990s, oil price increases preceded macroeconomic downturns. This effect occurred at frequencies with periods around 6 years. However, in the last decade, the regions of high coherencies were located at frequencies that corresponded to shorter-run cycles (cycles with periods around 3 years).

We also showed that after the Euro adoption some countries became more similar with respect to how their macroeconomies react to oil shocks. This is true for Austria, France, Germany, Italy, Luxembourg, Netherlands, and Spain and even more true for Finland, who had a rather asymmetric reaction to oil shocks before the Euro adoption. However, we also showed that at least three countries do not share a common response to oil shocks: Portugal, Ireland and Belgium. Particularly interesting is the conclusion that the adoption of the common currency did not contribute to a higher degree of synchronization between these countries and the rest of the countries in the Euroland. This effect is particular surprising in the case of Ireland, who was highly synchronized before 1999.

Acknowledgements We offer this paper as a token of our intellectual respect for James Ramsey, who, in a series of papers, some of them co-authored with Camille Lampart, got us interested on wavelet applications to Economics. We thank an anonymous referee for his comments. The usual disclaimer applies. Financial support from Fundação para a Ciência e a Tecnologia, research grants PTDC/EGE-ECO/100825/2008 and PEst-C/EGE/UI3182/2013, through Programa Operacional Temático Factores de Competitividade (COMPETE) is gratefully acknowledged.

References

Aguiar-Conraria L, Soares MJ (2011a) Oil and the macroeconomy: using wavelets to analyze old issues. Empir Econ 40(3):645–655

Aguiar-Conraria L, Soares MJ (2011b) Business cycle synchronization and the Euro: a wavelet analysis. J Macroecon 33(3):477–489

Aguiar-Conraria L, Soares MJ (2014) The continuous wavelet transform: moving beyond uni- and bivariate analysis. J Econ Surv. 28(2):344–375

Aguiar-Conraria L, Wen Y (2012) OPEC's oil exporting strategy and macroeconomic (in)stability. Energy Econ 34(1):132–136

Aguiar-Conraria L, Magalhães PC, Soares MJ (2012) Cycles in politics: wavelet analysis of political time-series. Am J Polit Sci 56(2):500–518

Baxter M, Kouparitsas M (2005) Determinants of business cycle comovement: a robust analysis. J Monet Econ 52(1):113–157

Blanchard O, Galí J (2010) The macroeconomic effects of oil price shocks: why are the 2000s so different from the 1970s? In: Galí, J, Gertler M (eds) International dimensions of monetary policy. University of Chicago Press, Chicago, pp 373–421

Camacho M, Perez-Quirós G, Saiz L (2006) Are European business cycles close enough to be just one? J Econ Dyn Control 30(9–10):1687–1706

Camacho M, Perez-Quirós G, Saiz, L (2008) Do European business cycles look like one? J Econ Dyn Control 32(7):2165–2190

Cazelles B, Chavez M, de Magny GC, Guégan J-F, Hales S (2007) Time-dependent spectral analysis of epidemiological time-series with wavelets. J R Soc Interface 4(15):625–636

Crowley P, Mayes D (2008) How fused is the Euro area core?: an evaluation of growth cycle co-movement and synchronization using wavelet analysis. J Bus Cycle Meas Anal 4:63–95

Frankel J, Rose A (1998) The endogeneity of the optimum currency area criteria. Econ J 108(449):1009–1025

Hamilton J (1996) This is what happened to the oil price-macroeconomy relationship. J Monet Econ 38(2):215–220

Hamilton J (2003) What is an oil shock? J Econom 113(2):363–398

Hamilton J (2009) Causes and consequences of the oil shock of 2007–08. Brookings Pap Econ Act 40(1):215–283

Imbs J (2004) Trade, finance, specialization, and synchronization. Rev Econ Stat 86(3):723–734

Inklaar R, Jong-A-Pin R, de Haan J (2008) Trade and business cycle synchronization in OECD countries—a re-examination. Eur Econ Rev 52(4):646–666

Kilian L (2008) Exogenous oil supply shocks: how big are they and how much do they matter for the U.S. economy? Rev Econ Stat 90(2):216–240

Kilian L (2009) Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil market. Am Econo Rev 99(3):1053–1069

Kyrtsou C, Malliaris A, Serletis A (2009) Energy sector pricing: on the role of neglected nonlinearity. Energy Econ 31(3):492–502

Naccache T (2011) Oil price cycles and wavelets. Energy Econ 33(2):338–352

Peersman G (2011) The relative importance of symmetric and asymmetric shocks: the case of United Kingdom and Euro area. Oxf Bull Econ Stat 73(1):104–118

Rose A, Engel C (2002) Currency unions and international integration. J Money Credit Bank 34(4):1067–1089

Vacha L, Barunick J (2012) Co-movement of energy commodities revisited: evidence from wavelet coherence analysis. Energy Econ 34(1):241–247

Wavelet-Based Correlation Analysis of the Key Traded Assets

Jozef Baruník, Evžen Kočenda and Lukas Vacha

Abstract This chapter reveals the time-frequency dynamics of the dependence among key traded assets—gold, oil, and stocks, in the long run, over a period of 26 years. Using both intra-day and daily data and employing a variety of methodologies, including a novel time-frequency approach combining wavelet-based correlation analysis with high-frequency data, we provide interesting insights into the dynamic behavior of the studied assets. We account for structural breaks and reveal a radical change in correlations after 2007–2008 in terms of time-frequency behavior. Our results confirm different levels of dependence at various investment horizons indicating heterogeneity in stock market participants' behavior, which has not been documented previously. While these key assets formerly had the potential to serve as items in a well-diversified portfolio, the events of 2007–2008 changed this situation dramatically.

J. Baruník • L. Vacha (✉)
Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodarenskou Vezi 4, 18200 Prague, Czech Republic

Institute of Economic Studies, Charles University, Opletalova 21, 11000 Prague, Czech Republic
e-mail: barunik@utia.cas.cz; vachal@utia.cas.cz

E. Kočenda
CERGE-EI, Charles University and the Czech Academy of Sciences, Politickych veznu 7, 11121 Prague, Czech Republic

CESifo, Munich, IOS Regensburg, Germany

The William Davidson Institute at the University of Michigan Business School, Ann Arbor, MI 48109, USA

CEPR, London, UK

Euro Area Business Cycle Network, London, UK
e-mail: evzen.kocenda@cerge-ei.cz

M. Gallegati and W. Semmler (eds.), *Wavelet Applications in Economics and Finance*, Dynamic Modeling and Econometrics in Economics and Finance 20,
DOI 10.1007/978-3-319-07061-2_8,

1 Introduction, Motivation, and Related Literature

In this chapter, we contribute to the literature by studying the dynamic relationship among gold, oil, and stocks in a time-frequency domain by employing a wavelet-based methodology. Considering the time-frequency domain offers new perspectives on the relationships among the assets and differentiates our contribution from much of the related literature. The time-frequency approach also enables us to uncover patterns underpinned by the investment potential derived from the ongoing financialization of commodities.

Traders in financial markets make their decisions over various horizons, for example, minutes, hours, days, or even longer periods such as months and years, as discussed by Ramsey (2002). Nevertheless, majority of the empirical literature studies the relationships in the time domain only and aggregating the behavior across all investment horizons. Our analysis includes both time and time-frequency methods. The time domain tools we apply to measure correlations are the parametric DCC GARCH and nonparametric realized volatility. Although these two methods are fundamentally different, they both average the relationships over the full range of available frequencies and suffer from restricted application when analyzing non-stationary time series. In contrast, wavelets allow us to analyze time series within a time-frequency domain framework that allows for various forms of localization. Thus, when analyzing non-homogeneous and non-stationary time series, wavelet analysis is preferred because it is more flexible. For example, when considering stock markets, we can work with prices and thus study the dynamics of the dependencies at various investment horizons or frequencies at the same moment, where the lowest frequency will contain the trend component of the data. Therefore, we can determine short and long-term dependence structures. Wavelets are able to deliver valuable and unorthodox inferences in the fields of economics and finance, as evinced in recent applications, for example, by Faÿ et al. (2009), Gallegati et al. (2011), Vacha and Barunik (2012), Aguiar-Conraria et al. (2012), or Graham et al. (2013).

Our analysis is performed using data from a long period, 26 years, from 1987 to 2012. We conduct a thorough, wavelet-based analysis and uncover rich time-frequency dynamics in the relationships among gold, oil, and stocks. The selection of the three assets is motivated by the fact that gold and oil are the most actively traded commodities in the world. Similarly, to represent stocks, we use the S&P 500, which is one of the most actively traded and comprehensive stock indices in the world. Gold is traditionally perceived as a store of wealth, especially with respect to periods of political and economic insecurity (Aggarwal and Lucey 2007). However, gold is both a commodity and a monetary asset. Approximately 40 % of newly mined gold is used for investment (Thompson Reuters GFSM, 2012). Unlike gold, oil is essential component of contemporary industrial economies, as reflected by the 88 million barrels consumed daily worldwide. As oil is a vital production input, its price is driven by distinct demand and supply shocks (Hamilton 2009). Oil has also become financialized over time, as documented in Büyükşahin and

Robe (2013). Fratzscher et al. (2013) show that oil was not correlated with stocks until 2001, but as oil began to be used as a financial asset, the link between oil and other assets strengthened. Finally, stocks reflect the economic and financial development of firms and market perceptions of a companys standing; they also represent investment opportunities and a link to perceptions of aggregate economic development. Further, stock prices provide helpful information on financial stability and can serve as an indicator of crises (Gadanecz and Jayaram 2009). Thus, a broad market index can be used to convey information on the status and stability of the economy. In our analysis, we consider the S&P 500, which is frequently used as a benchmark for the overall U.S. stock market. In our analysis, stocks complement the commodities of gold and oil to represent the financial assets traded by the modern financial industry.

What motivates our analysis of the links among the three assets above? The literature analyzing the dynamic correlations among assets proposes a number of important reasons that the issue should be investigated. An obvious motivation for analyzing co-movements is that substantial correlations among assets greatly reduce their potential to be included in a portfolio from the perspective of risk diversification. Even if assets in a portfolio initially exhibit low correlation, a potential change in correlation patterns represents an imperative to redesign such a portfolio. Both issues are also linked to the Modern Portfolio Theory (MPT) of Markowitz (1952). MPT assumes, among other things, that correlations between assets are constant over time. However, correlations between assets may well depend on the systemic relationships between them and change when these relationships change. Thus, evidence of time-varying correlations between assets substantially undermines MPT results and, more important, its use to protect investors from risk.

Empirical evidence on co-movements among assets may well depend on the choice of assets, technique employed, and the period under study. In a seminal study on co-movements in the monthly prices of unrelated commodities, Pindyck and Rotemberg (1990) find excess co-movement among seven major commodities, including gold and oil. However, the co-movements are measured in a rather simple manner as individual cross-correlations over the entire period (1960–1985). The excess co-movements were attributed to irrational or herding behavior in markets. Using a concordance measure, Cashin et al. (1999) analyze the same set of commodities over the same period as Pindyck and Rotemberg (1990) and find no evidence for co-movements in the prices of the analyzed commodities. When they extend the period to 1957–1999, the co-movements are again absent and they contend that the entire notion of co-movements in the prices of unrelated commodities is a myth. A single exception is the co-movement in gold and oil prices that Cashin et al. (1999) credit to inflation expectations and further provide evidence that booms in oil and gold prices often occur at the same time (Cashin et al. 2002). Still, it has to be noted that gold may well be traded independently from other assets on the pretext of being a store of value during downward market swings. Hence, it does not necessarily co-move with related or unrelated commodities.

An extension of the co-movement analysis to the time-frequency domain offers the potential for an interesting comparison of how investment horizons influence the diversification of market risk. The importance of various investment horizons

for portfolio selection has been recognized by Samuelson (1989). In this respect, Marshall (1994) demonstrates that investor preferences for risk are inversely related to time and different investment horizons have direct implications for portfolio selection. Graham et al. (2013) provide empirical evidence related to the issues studied in this chapter by studying the co-movements of various assets using wavelet coherence and demonstrating that at the long-term investment horizon co-movement among stocks and commodities increased at the onset of the 2007–2008 financial crisis. Thus, the diversification benefits of using these assets are rather limited.

With the above motivations and findings in mind, in this chapter we adopt a comprehensive approach and contribute to the literature by analyzing the prices of three assets that have unique economic and financial characteristics: the key commodities gold and oil and important stocks represented by the S&P 500 index. To this end, we consider a long period (1987–2012) at both intra-day and daily frequencies and an array of investment horizons to deliver a comprehensive study in the time-frequency domain based on wavelet analysis. Our key empirical results can be summarized as follows: (1) correlations among the three assets are low or even negative at the beginning of our sample but subsequently increase, and the change in the patterns becomes most pronounced after decisive structural breaks take place (breaks occur during the 2006–2009 period at different dates for specific asset pairs); (2) correlations before the 2007–2008 crisis exhibit different patterns at different investment horizons; (3) during and after the crisis, the correlations exhibit large swings and their differences at shorter and longer investment horizons become negligible. This finding indicates vanishing potential for risk diversification based on these assets: after the structural change, gold, oil, and stocks could not be combined to yield effective risk diversification during the post-break period studied.

The chapter is organized as follows. In Sect. 2, we introduce the theoretical framework for the wavelet methodology we use to perform our analysis. Our large data set is described in detail in Sect. 3 with a number of relevant commentaries. We present our empirical results in Sect. 4. Section 5 briefly concludes.

2 Theoretical Framework for the Methodologies Employed

In the following section, we introduce the methodologies employed. While standard approaches (e.g., DCC GARCH and realized volatility) allow us to study the covariance matrix solely in the time domain, we are interested in studying its time-frequency dynamics. In other words, we are interested in determining how the correlations vary over time and various investment horizons. We are able to do so by using the innovative time-frequency approach of wavelet analysis. Wavelets are a relatively new method in economics, despite their potential benefits to economists (Ramsey 2002; Gençay et al. 2002).

We continue with a brief introduction of the methodologies used to estimate the dynamic correlations, namely: (1) the parametric DCC GARCH approach; (2)

non-parametric realized measures; and (3) a time-frequency approach in the form of a wavelet analysis.

2.1 Time-Varying Correlations: DCC GARCH Methodology

In this section, we introduce the Dynamic Conditional Correlation Generalized Autoregressive Conditional Heteroscedasticity (DCC GARCH) model for estimating dynamic correlations in a multivariate setting. The DCC GARCH was proposed by Engle (2002) and is a logical extension of Bollerslev's constant conditional correlation (CCC) model (Bollerslev 1990), in which the volatilities of each asset were allowed to vary over time but the correlations were time invariant. The DCC version, however, also allows for dynamics in the correlations.

Engle (2002) defines the covariance matrix as:

$$H_t = D_t R_t D_t. \tag{1}$$

where R_t is the conditional correlation matrix and $D_t = diag\{\sqrt{h_{i,t}}\}$ is a diagonal matrix of time varying standard variation from the i-th univariate (G)ARCH(p, q) processes $h_{i,t}$. Parameter n represents the number of assets at time $t = 1, \ldots, T$. The correlation matrix is then given by the transformation

$$R_t = diag(\sqrt{q_{11,t}}, \ldots, \sqrt{q_{nn,t}}) Q_t diag(\sqrt{q_{11,t}}, \ldots, \sqrt{q_{nn,t}}), \tag{2}$$

where $Q_t = (q_{ij,t})$ is

$$Q_t = (1 - \alpha - \beta)\overline{Q} + \alpha \eta_{t-1} \eta'_{t-1} + \beta Q_{t-1}, \tag{3}$$

where $\eta_t = \varepsilon_{i,t}/\sqrt{h_{i,t}}$ denotes the standardized residuals from the (G)ARCH model, $\overline{Q} = T^{-1} \sum \eta_t \eta'_t$ is a $n \times n$ unconditional covariance matrix of η_t, and α and β are non-negative scalars such that $\alpha + \beta < 1$.

We estimate the DCC GARCH using the standard quasi-maximum likelihood method proposed by Engle (2002). Further, we assume Gaussian innovations. The DCC model can be estimated consistently by estimating the univariate GARCH models in the first stage and the conditional correlation matrix in the second stage. The parameters are also estimated in stages. This two-step approach avoids the dimensionality problem encountered in most multivariate GARCH models (Engle 2002; Engle and Sheppard 2001).[1] Furthermore, the DCC model is parsimonious and ensures that time-varying correlation matrices between the stock exchange returns are positive definite.

[1] Bauwens and Laurent (2005) demonstrate that the one-step and two-step methods provide very similar estimates.

2.2 Time-Varying Correlations: Realized Volatility Approach

Due to increased availability of high-frequency data, a simple technique for estimating the covariance matrix was recently developed. In contrast to the DCC GARCH, this method is non-parametric. It is based on the estimating the covariance matrix analogously to the realized variation by taking the sum outer product of the observed high-frequency returns over the period considered. Following Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004) we define the realized covariance over the time interval $[t-h,t]$ for $0 \leq h \leq t \leq T$ as

$$\widehat{RC}_{t,h} = \sum_{i=1}^{M} \mathbf{r}_{t-h+\left(\frac{i}{M}\right)h} \mathbf{r}'_{t-h+\left(\frac{i}{M}\right)h}, \tag{4}$$

where M denotes the number of observations in the interval $[t-h,t]$. Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004) demonstrate that the *ex-post* realized covariance $\widehat{RC}_{t,h}$ is an unbiased estimator of the *ex-ante* expected covariation. Furthermore, given increasing sampling frequency, i.e. $h > 0$ and $M \to \infty$, the realized covariance is a consistent estimator of the covariation. In practice, we only observe discrete prices, hence discretization bias is unavoidable. More serious damage to the estimator is also caused by market microstructure effects such as the bid-ask bounce, price discreteness, and the bid-ask spread. The literature advises employing rather sparse sampling when applying the estimator in practice; however this entails discarding a large amount of the available data. Following the suggestion by Andersen and Benzoni (2007) to obtain the best trade-off between reduced bias and information loss, we use 5-min data to calculate the realized covariances.[2] An important assumption regarding the price processes is that the data are synchronized, which implies collecting the prices simultaneously. However, this is not an issue in our analysis, as all three examined assets are paired using equal time-stamps matching.

2.3 Time-Frequency Dynamics in Correlations: Wavelet Approach

As we are interested in how the correlations vary over time and at different investment horizons, we need to conduct a wavelet analysis that allows us to work simultaneously in the time and frequency domains. The DCC GARCH and realized volatility methods outlined above do not allow the researcher to extend the analysis

[2]This is the optimal sampling frequency determined based on the substantial research on the noise-to-signal ratio. The literature is well surveyed by Hansen and Lunde (2006), Bandi and Russell (2006), McAleer and Medeiros (2008), and Andersen and Benzoni (2007).

to the frequency domain; hence we are only able to study the covariance matrix in the time domain.

Wavelet time-frequency domain analysis is very powerful tool when we expect changes in economic relationships such as structural breaks. Wavelet analysis can react to these changes because the wavelet transform uses a localized function with finite support for the decomposition—a wavelet. In contrast, when using a pure frequency approach, represented by the Fourier transform, one obtains information on all of the frequency components, but because the amplitude is fixed throughout the period considered, the time information is completely lost. Thus, in the event of sudden changes in economic relationships or the presence of breaks during the period studied, one is unable to locate precisely where this change occurs. Additionally, due to the non-stationarity induced by such breaks, Fourier transform-based estimates may not be precise. Therefore, the wavelet transform has substantial advantages over the Fourier transform when the time series is non-stationary or is only locally stationary (Roueff and Sachs 2011).

An important feature of wavelet analysis is the decomposition of the economic relationship into time-frequency components. Wavelet analysis often uses scale instead of frequency, as scale typically characterizes frequency bands. The set of wavelet scales can be further interpreted as investment horizons at which we can study the economic relationships separately. Thus, every scale describes the development of the economic relationship at a particular frequency while retaining the time dynamics. Subsequently, the wavelet decomposition generally provides a more complex picture compared to the time domain approach, which aggregates all investment horizons. Therefore, if we expect that economic relationships follow different patterns at various investment horizons, then a wavelet analysis can uncover interesting characteristics of the data that would otherwise remain hidden. An introduction to the wavelet methodology with a remarkable application to economics and finance is provided in Gençay et al. (2002) and Ramsey (2002).

2.4 Wavelet Transform

While we use a discrete version of the wavelet transform, we begin our introduction with the continuous wavelet transform (CWT), as it is the cornerstone of the wavelet methodology. Next, we continue by describing a special form of discrete wavelet transform named the "maximal overlap discrete wavelet transform" (MODWT). Following standard notation, we define the continuous wavelet transform $W_x(j, s)$ as a projection of a wavelet function[3] $\psi_{j,s}(t) = \frac{1}{\sqrt{j}}\psi\left(\frac{t-s}{j}\right) \in L^2(\mathbb{R})$ onto the time series $x(t) \in L^2(\mathbb{R})$,

[3]We use the least asymmetric wavelet with length L=8, denoted as LA(8).

$$W_x(j,s) = \int_{\infty}^{\infty} x(t)\frac{1}{\sqrt{j}}\overline{\psi\left(\frac{t-s}{j}\right)}dt, \tag{5}$$

where s determines the position of the wavelet in time. The scaling, or dilatation parameter j controls how the wavelet is stretched or dilated. If the scaling parameter j is low (high), then the wavelet is more (less) compressed and able to detect high (low) frequencies. One of the most important conditions a wavelet must fulfill is the admissibility condition: $C_\psi = \int_0^\infty \frac{|\Psi(f)|^2}{f} df < \infty$, where $\Psi(f)$ is the Fourier transform of a wavelet $\psi(.)$. The decomposed time series $x(t)$ can be subsequently recovered using the wavelet coefficients as follows

$$x(t) = \frac{1}{C_\psi}\int_0^\infty \left[\int_{-\infty}^{\infty} W_x(j,s)\psi_{j,s}(t)ds\right]\frac{dj}{j^2}, \quad s>0. \tag{6}$$

Further, the continuous wavelet transform preserves the energy or variance of the analyzed time series; hence

$$x^2 = \frac{1}{C_\psi}\int_0^\infty \left[\int_{-\infty}^{\infty} |W_x(j,s)|^2 ds\right]\frac{dj}{j^2}. \tag{7}$$

Equation (7) is an important property that allows us to work with the wavelet variance, covariance and the wavelet correlation. For a more detailed introduction to continuous wavelet transform and wavelets, see Daubechies (1992), Chui (1992), and Percival and Walden (2000).

As we study discrete time series, we only require a limited number of scales, and some form of discretization is needed. The counterpart of the continuous wavelet transform in discrete time is the discrete wavelet transform (DWT),[4] which is a parsimonious form of the continuous transform, but it has some limiting properties that make its application to real time series relatively difficult. These limitations primarily concern the restriction of the sample size to the power of two and the sensitivity to the starting point of the transform. Therefore, in our analysis, we use a modified version of the discrete wavelet transform—MODWT—which has some advantageous properties that are summarized below.

In contrast to the DWT, the MODWT does not use downsampling, as a consequence the vectors of the wavelet coefficients at all scales have equal length, corresponding to the length of transformed time series. Thus, the MODWT is not restricted to sample sizes that are powers of two. However, the MODWT wavelet coefficients are no longer orthogonal to each other at any scale. Additionally, the MODWT is a translation-invariant type of transform; therefore, it is not sensitive

[4]For a definition and detailed discussion of the discrete wavelet transform, see Mallat (1998), Percival and Walden (2000), and Gençay et al. (2002).

to the choice of the starting point of the examined process. Both the DWT and MODWT wavelet and scaling coefficients can be used for energy decomposition and analysis of variance of a time series in the time-frequency domain, however Percival (1995) demonstrates the dominance of the MODWT estimator of variance over the DWT estimator. Furthermore, Serroukh et al. (2000) analyze the statistical properties of the MODWT variance estimator for non-stationary and non-Gaussian processes and show its statistical properties. For additional details on the MODWT, see Mallat (1998) and Percival and Walden (2000).

2.5 *Maximal Overlap Discrete Wavelet Transform*

This section demonstrates an application of the pyramid algorithm to obtain the MODWT wavelet and scaling coefficients. The method is based on filtering time series with MODWT wavelet filters; the output after filtering is then filtered again in a subsequent stage to obtain other wavelet scales.

Let us begin with the first stage. The wavelet coefficients at the first scale ($j = 1$) are obtained via circular filtering of time series x_t using the MODWT wavelet and scaling filters $h_{1,l}$ and $g_{1,l}$ (Percival and Walden 2000):

$$W_x(1,s) \equiv \sum_{l=0}^{L-1} h_{1,l}\, x(s-l \bmod N), \quad V_x(1,s) \equiv \sum_{l=0}^{L-1} g_{1,l}\, x(s-l \bmod N). \tag{8}$$

The second step of the algorithm uses the scaling coefficients $V_x(1,s)$ instead of x_t. The wavelet and scaling filters have a width $L_j = 2^{j-1}(L-1)+1$; therefore, for the second scale, the length of the filter is $L_2 = 15$. After filtering, we obtain the wavelet coefficients at scale $j = 2$:

$$W_x(2,s) \equiv \sum_{l=0}^{L-1} h_{2,l}\, V_x(1,s-l \bmod N), \quad V_x(2,s) \equiv \sum_{l=0}^{L-1} g_{2,l}\, V_x(1,s-l \bmod N). \tag{9}$$

After the two steps of the algorithm we have two vectors of the MODWT wavelet coefficients at scale $j = 1$ and $j = 2$; $W_x(1,s), W_x(2,s)$ and one vector of the MODWT wavelet scaling coefficients at scale two $V_x(2,s)$, where $s = 0, 1, \ldots, N-1$ is the same for all vectors. The vector $W_x(1,s)$ represents wavelet coefficients that reflect variations at the frequency band $f[1/4, 1/2]$, $W_x(2,s)$: $f[1/8, 1/4]$ and $V_x(2,s)$: $f[0, 1/8]$.

The transfer function of the filter $h_l : l = 0, 1, \ldots, L-1$, where L is the width of the filter, is denoted as $H(.)$. The pyramid algorithm exploits the fact that if we

increase the width of the filter to $2^{j-1}(L-1)+1$, then the filter with the impulse response sequence in the form[5]:

$$\{h_0, \underbrace{0,\ldots,0}_{2^{j-1}-1 \text{ zeros}}, h_1, \underbrace{0,\ldots,0}_{2^{j-1}-1 \text{ zeros}}, h_{L-2}, \underbrace{0,\ldots,0}_{2^{j-1}-1 \text{ zeros}}, h_L\}, \tag{10}$$

has a transfer function defined as $H\left(2^{j-1}f\right)$. Using this feature of the filters, we can write the pyramid algorithm simply in the following form

$$W_x(j,s) \equiv \sum_{l=0}^{L-1} h_l V_x\left(j-1, s-2^{j-1}l \ mod N\right), \ s=0,1,\ldots,N-1, \tag{11}$$

$$V_x(j,s) \equiv \sum_{l=0}^{L-1} g_l V_x\left(j-1, s-2^{j-1}l \ mod N\right), \ s=0,1,\ldots,N-1, \tag{12}$$

where for the first stage we set $x = V_x(0,s)$. Thus, after performing the MODWT, we obtain $J^m \leq log_2(N)$ vectors of wavelet coefficients and one vector of scaling coefficients. The j-th level wavelet coefficients in vector $W_x(j,s)$ represent frequency bands $f[1/2^{j+1}, 1/2^j]$, while the j-th level scaling coefficients in vector $V_x(j,s)$ represent $f[0, 1/2^{j+1}]$. In the subsequent analysis of the wavelet correlations we apply the MODWT with the wavelet filter LA(8), with reflecting boundary conditions.

2.6 Wavelet Correlation

Applying the wavelet transform allows for a scale-by-scale decomposition of a time series, where every scale represents an investment horizon. In the bivariate case, where we decompose two time series, we can study correlations at investment horizons represented by scales. The method is called wavelet correlation and offers an alternative means of studying the dependence between two time series, as it can uncover different dependencies across available scales.

When the MODWT is used, all vectors of the wavelet coefficients have the same length. Thus, for a time series x_t, $t = 1,2,\ldots,N$, we obtain $j=1,\ldots,J^m$ vectors of wavelet coefficients and one vector of scaling coefficients of length N. The maximal level of wavelet decomposition is denoted J^m, $J^m \leq log_2(N)$. The

[5]The number of zeros between filter coefficients is $2^{j-1}-1$, i.e., for the filter at the first stage, we have no zeros, and for the second stage there is just one zero between each coefficient; hence the width of the filter is 15.

wavelet correlation $\rho_{xy}(j)$ between time series x_t and y_t at scale j is then defined as (Whitcher et al. 2000):

$$\rho_{xy}(j) = \frac{cov\left(W_x(j,s)W_y(j,s)\right)}{\left[Var\left(W_x(j,s)\right) var\left(W_y(j,s)\right)\right]^{\frac{1}{2}}} \equiv \frac{\gamma_{xy}(j)}{\nu_x(j)\,\nu_y(j)}, \tag{13}$$

where $\nu_x^2(j)$ and $\gamma_{xy}(j)$ denote wavelet variance and covariance, respectively. Additional details on wavelet variance and covariance are provided in Appendices "Wavelet Variance" and "Wavelet Covariance". The wavelet correlation estimator directly uses the definition of the wavelet correlation Eq. (13), thus we can write:

$$\hat{\rho}_{xy}(j) \equiv \frac{\hat{\gamma}_{xy}(j)}{\hat{\nu}_x(j)\hat{\nu}_y(j)}, \tag{14}$$

where $\hat{\gamma}_{xy}(j)$ denotes the wavelet covariance estimator and $\hat{\nu}_x(j)^2$ and $\hat{\nu}_y(j)^2$ are estimators of wavelet variance at scale j for time series x_t and y_t. Whitcher et al. (1999) established the central limit theorem for estimator Eq. (14), as well as the approximate confidence intervals; empirical values are reported in Sect. 4.1. Additional details on this topic can be found in Serroukh et al. (2000).

3 Data

In the empirical section, we analyze the prices of gold, oil, and a representative U.S. stock market index, the S&P 500. The data set contains the tick prices of the S&P 500 and the futures prices of gold and oil, where we use the most active rolling contracts from the pit (floor traded) session. All of the assets are traded on the platforms of the Chicago Mercantile Exchange (CME).[6]

We restrict our study to the intraday 5-min and daily data sampled during the business hours of the New York Stock Exchange (NYSE), as most of the liquidity of the S&P 500 comes from the period when the U.S. markets are open. The sample period runs from January 2, 1987 until December 31, 2012.[7] To synchronize the data, we employ Greenwich Mean Time (GMT) stamp matching. Further, we exclude transactions executed on Saturdays and Sundays, U.S. federal holidays, December 24 to 26, and December 31 to January 2, as the low activity on these days could lead to estimation bias. Therefore, we use data from 6,472 trading

[6]Oil (Light Crude) is traded on the New York Mercantile Exchange (NYMEX) platform, gold is traded on the Commodity Exchange, Inc. (COMEX), a division of NYMEX, and the S&P 500 is traded at the CME in Chicago. All data were acquired from Tick Data, Inc.

[7]The CME introduced the Globex(R) electronic trading platform in December 2006 and began to offer nearly continuous trading.

Table 1 Descriptive statistics for high-frequency and daily gold, oil and, stock (S&P 500) returns over the sample period extending from January 2, 1987 until December 31, 2012

	High-frequency data			Daily data		
	Gold	Oil	Stocks	Gold	Oil	Stocks
Mean	1.00e−06	3.19e−06	−2.46e−06	2.22e−04	2.42e−04	2.70e−04
St. dev.	0.001	0.002	0.001	0.010	0.023	0.012
Skewness	−0.714	1.065	0.326	−0.147	−1.063	−0.392
Kurtosis	47.627	104.561	32.515	10.689	19.050	11.474
Minimum	−0.042	−0.045	−0.024	−0.077	−0.384	−0.098
Maximum	0.023	0.163	0.037	0.103	0.136	0.107

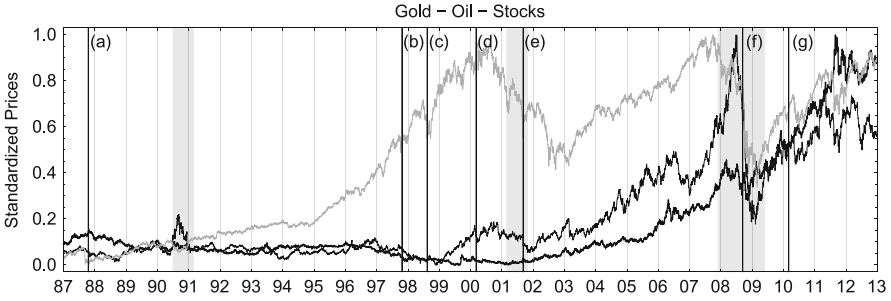


Fig. 1 Normalized prices of gold (*thin black*), oil (*black*), and stocks (*gray*). The figure highlights several important recession periods in *gray* (described in greater detail in the text), and crashes using *black lines*: (*a*) Black Monday; (*b*) the Asian crisis; (*c*) the Russian ruble devaluation; (*d*) the dot-com bubble burst; (*e*) the WTC 9/11 attacks; (*f*) the Lehman Brothers Holdings bankruptcy; and (*g*) the Flash Crash

days. Descriptive statistics of the intra-day and daily returns of the data that form our sample are presented in Table 1. Overall, the statistics are standard with the remarkable exception of a very high excess kurtosis of 104.561 for oil. This is mainly a consequence of a single positive price change of 16.3 % (January 19, 1991), when the worst deliberate environmental damage in history was caused by Iraqi leader Saddam Hussein, who ordered a large amount of oil to be spilled into the Persian Gulf (Khordagui and Al-Ajmi 1993).

Figure 1 depicts the development of the prices of the three assets, in which several recessions and crisis periods can be detected. Following the National Bureau of Economic Research (NBER),[8] there were three recessions in the U.S. during the period studied: July 1990 to March 1991, March 2001 to November 2001, and December 2007 to June 2009. These recessions are highlighted by gray bands. Furthermore, black lines depict 1-day crashes associated with large price drops. Specifically, Black Monday (October 19, 1987), the Asian crisis (October 27, 1997),

[8]US Business Cycle Expansions and Contractions, NBER, accessed April 5, 2013 (http://www.nber.org/cycles.html).

the Russian ruble devaluation (August 17, 1998), the dot-com bubble burst (March 10, 2000), the World Trade Center attacks (September 11, 2001), the Lehman Brothers Holdings bankruptcy (September 15, 2008), and the Flash Crash (March 6, 2010). The largest 1-day drops in the studied sample occurred on the following dates, with percentage declines given in parentheses: October 19, 1987 (20.47 %), October 26, 1987 (8.28 %), September 29, 2008 (8.79 %), October 9, 2008 (7.62 %), October 15, 2008 (9.03 %), and December 1, 2008 (8.93 %).

The above crashes differ in nature, and we discuss them briefly below. On Monday, October 19, 1987, known as Black Monday, stock markets around the world dropped in a very short time and recorded the largest 1-day crash in history. After this extreme event, many expected the most troubled years since the 1930s. Nevertheless, stock markets quickly recovered from the losses and closed 1987 in positive territory. There is still no consensus on the cause of the crash; potential reasons include illiquidity, program trading, overvaluation and market psychology.[9]

For many consecutive years stock markets did not record large shocks until 1996 when the Asian financial crisis emerged. Investors were leaving emerging overheated Asian shares that on October 27, 1997 resulted in a mini-crash of the U.S. markets. On August 17, 1998 the Russian government devalued the ruble, defaulted on domestic debt and declared a moratorium on payment to foreign creditors, which also caused an international crash. The 1996 and 1997 crashes are believed to be exogenous shocks to U.S. stock markets. The inflation of the so-called dot-com bubble emerged in the period 1997–2000, when several internet-based companies entered the markets and fascinated many investors confident in their future profits, while overlooking the companies' fundamental value. Ultimately, this resulted in a gradual collapse, or bubble burst, during the years 2000–2001. The World Trade Centre was attacked on September 11, 2001. Although markets recorded a sudden drop, the shock was exogenous and should not be attributed to internal market forces. The recent financial crisis of 2007–2008, also called as the global financial crisis (for a detailed treatment, see Bartram and Bodnar (2009)), was initiated by the bursting of the U.S. housing-market bubble. Consequently, in September and October 2008, stock markets experienced large declines. On May 6, 2010, financial markets witnessed the largest intraday drop in history known as the Flash Crash or The Crash of 2:45. The Dow Jones Industrial Average declined by approximately 1,000 points (9 %), but the loss was recovered within a few minutes. The crash was likely caused by high-frequency trading or large directional bets.

[9]For additional information on the crash, see Waldrop (1987) and Carlson (2007).

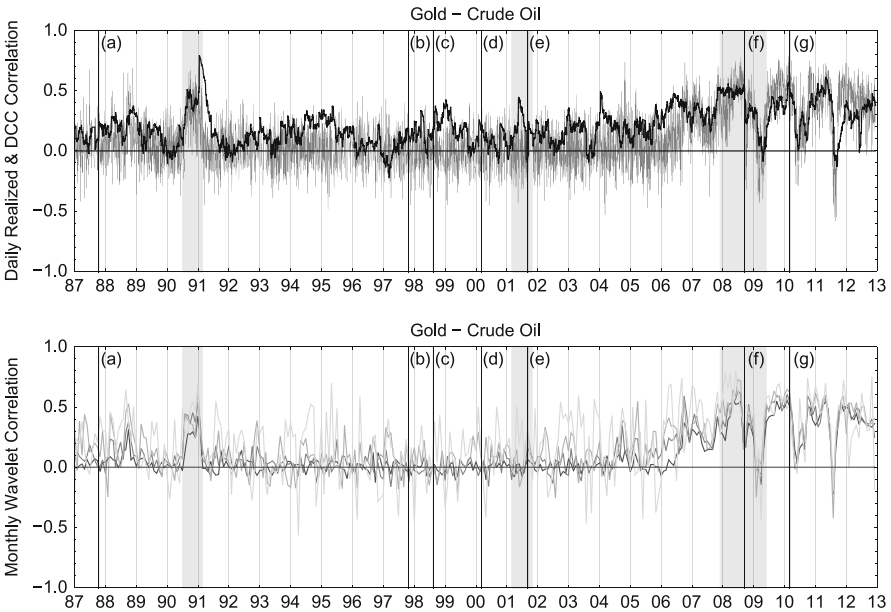


Fig. 2 Dynamics in gold-oil correlations. The upper plot of the panel contains the realized correlation for each day of the sample and daily correlations estimated from the DCC GARCH model. The lower plot contains time-frequency correlations based on the wavelet correlation estimates from high-frequency data for each month separately. We report correlation dynamics at 10-min, 40-min, 2.66 h (approximate), and 1.6-day (approximate) horizons depicted by the *thick black to thin black lines*. The plots highlight several important recession periods in *gray* (described in greater detail in the text), and crashes using *black lines*: (*a*) Black Monday; (*b*) the Asian crisis; (*c*) the Russian ruble devaluation; (*d*) the dot-com bubble burst; (*e*) the WTC 9/11 attacks; (*f*) the Lehman Brothers Holdings bankruptcy; and (*g*) the Flash Crash

4 Empirical Analysis of the Relationships Among Gold, Oil, and Stocks

4.1 Dynamic Correlations

Dynamic correlations for each pair of assets are depicted in Figs. 2, 3, 4. Each figure consists of two panels that plot correlations obtained by the three methods described in Sect. 2. The upper panels of the figures display realized volatility-based correlations computed on 5-min returns for each day and daily correlations from the parametric DCC GARCH(1,1) estimates. The lower panels depict the evolution of time-frequency correlations obtained through a wavelet decomposition

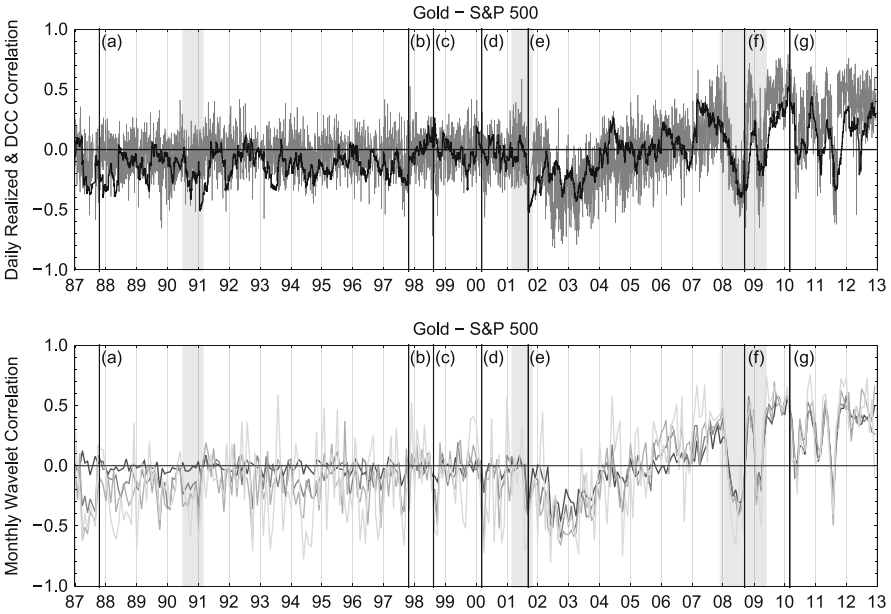


Fig. 3 Dynamics in gold-stocks correlations. The upper plot of the panel contains the realized correlation for each day of the sample and daily correlations estimated from the DCC GARCH model. The lower plot contains time-frequency correlations based on the wavelet correlation estimates from high-frequency data for each month separately. We report correlation dynamics at 10-min, 40-min, 2.66 h (approximate), and 1.6-day (approximate) horizons depicted by the *thick black to thin black lines*. The plots highlight several important recession periods in *gray* (described in greater detail in the text), and crashes using *black lines*: (*a*) Black Monday; (*b*) the Asian crisis; (*c*) the Russian ruble devaluation; (*d*) the dot-com bubble burst; (*e*) the WTC 9/11 attacks; (*f*) the Lehman Brothers Holdings bankruptcy; and (*g*) the Flash Crash

of 5-min data.[10] The panel displays only four investment horizons as examples: 10, 40, 160 min and 1.6 days are depicted in the figures.

The correlations between asset pairs exhibit stable and similar patterns, where majority of time the correlations are low or even negative, until 2001 between gold and stocks, until 2004 between oil and stocks, and until 2005 between gold and oil. After these stable years, the pattern of the correlations fundamentally changes. The general pattern of dynamic correlations between the pairs of variables is the same regardless of what method is used. Nevertheless, there are noticeable differences. Correlations based on realized volatility provide very rough evidence. More contoured correlation patterns are inferred from the DCC GARCH method. The wavelet correlations illustrate the methods advantages over the two previous

[10]For the sake of clarity in the plot, we report monthly correlations, computed on monthly price time series.

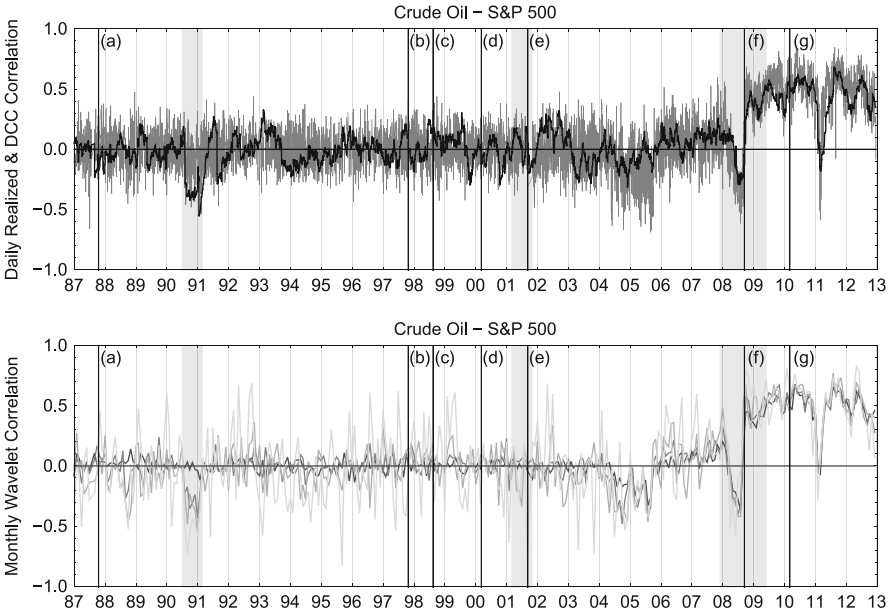


Fig. 4 Dynamics in oil-stocks correlations. The upper plot of the panel contains the realized correlation for each day of the sample and daily correlations estimated from the DCC GARCH model. The lower plot contains time-frequency correlations based on the wavelet correlation estimates from high-frequency data for each month separately. We report correlation dynamics at 10-min, 40-min, 2.66 h (approximate), and 1.6-day (approximate) horizons depicted by the *thick black to thin black lines*. The plots highlight several important recession periods in *gray* (described in greater detail in the text), and crashes using *black lines*: (*a*) Black Monday; (*b*) the Asian crisis; (*c*) the Russian ruble devaluation; (*d*) the dot-com bubble burst; (*e*) the WTC 9/11 attacks; (*f*) the Lehman Brothers Holdings bankruptcy; and (*g*) the Flash Crash

methods, as it allows us to observe individual correlation patterns for a number of investment horizons, providing time-frequency research output.[11]

In addition to the graphical illustration, the dynamic correlation results are summarized in Tables 2, 3, 4. The correlations for each asset pair are presented in individual tables containing the summarized correlations over a period of 1 year. The tables have two main parts: the results in the left panels are based on high-frequency intraday data for different investment horizons ranging from 10 ($j = 1$) to 80 ($j = 4$) min, whereas the right panels contain daily correlations with investment horizons ranging from 2 to 32 days. Both panels also include a low-frequency component (approximately 1 year). With the aim of supporting the results, we compute confidence intervals around the reported point correlation estimates. The 95 % confidence intervals of the estimates are nearly symmetric, with maximum

[11]While the wavelet method is superior to the other two methods in terms of dynamic correlation analysis, we employ the other two methods as a benchmark.

Table 2 Time-frequency correlation estimates for the gold–oil pair

	High-frequency data (minutes)					Daily data (days)					
	10	20	40	80	160-y.	2	4	8	16	32	64-y.
1987	0.02	0.03	0.08	0.15	0.80	0.12	0.00	0.14	0.07	−0.13	0.77
1988	0.11	0.19	0.19	0.23	0.42	0.23	0.25	0.32	0.38	0.55	0.93
1989	0.02	0.03	0.06	0.06	0.53	0.02	0.11	0.05	0.09	0.74	−0.14
1990	0.16	0.27	0.30	0.29	0.43	0.51	0.47	0.40	0.33	0.76	0.69
1991	0.21	0.32	0.31	0.32	0.63	0.01	0.00	0.47	0.48	−0.31	−0.47
1992	0.02	0.08	0.03	0.01	−0.56	0.06	0.01	−0.18	0.25	−0.05	−0.41
1993	0.00	0.01	0.04	0.02	0.60	0.12	0.04	0.20	0.30	−0.26	−0.77
1994	0.02	0.02	0.03	0.03	−0.13	0.16	0.37	0.22	−0.09	−0.28	0.33
1995	0.01	0.00	0.03	0.07	0.05	0.23	0.17	0.07	−0.02	0.16	0.39
1996	0.01	0.02	0.00	0.04	−0.62	−0.09	−0.03	0.13	−0.34	−0.31	−0.68
1997	0.00	−0.01	0.00	0.06	0.33	0.00	−0.22	0.04	−0.13	0.09	0.57
1998	0.00	−0.02	−0.01	−0.01	0.65	0.14	0.28	0.40	0.21	0.65	0.18
1999	0.01	0.01	−0.01	0.02	−0.58	−0.02	0.12	0.31	−0.17	0.17	−0.80
2000	0.00	0.00	0.01	0.07	−0.68	0.16	0.03	0.32	−0.12	0.01	0.44
2001	0.00	0.01	0.01	0.02	−0.83	0.23	0.04	0.11	−0.25	−0.10	0.10
2002	−0.01	−0.01	0.04	0.07	0.62	0.10	0.03	−0.17	0.08	−0.64	0.86
2003	0.01	0.02	0.04	0.06	0.68	0.24	0.05	−0.08	0.12	0.47	0.54
2004	0.04	0.08	0.10	0.11	0.40	0.17	0.38	0.23	0.13	−0.58	−0.74
2005	0.01	0.07	0.09	0.11	−0.42	0.08	0.07	0.22	0.42	0.27	0.40
2006	0.11	0.17	0.30	0.35	0.74	0.37	0.53	0.47	0.58	0.57	0.92
2007	0.26	0.30	0.33	0.35	0.29	0.49	0.38	0.07	0.41	0.42	0.48
2008	0.32	0.35	0.39	0.39	0.74	0.44	0.45	0.55	0.67	0.41	0.27
2009	0.19	0.21	0.22	0.22	−0.21	0.19	0.20	0.53	−0.03	−0.12	0.45
2010	0.33	0.34	0.36	0.37	−0.30	0.29	0.35	0.48	0.57	0.07	−0.37
2011	0.26	0.27	0.31	0.29	0.22	0.20	0.18	0.20	0.37	0.62	0.72
2012	0.40	0.42	0.42	0.41	−0.36	0.37	0.40	0.63	0.43	−0.19	0.71

The high-frequency set contains Wavelet correlation estimates based on high-frequency data. The daily set contains Wavelet correlation estimates based on daily data

values ranging from ± 0.014 for the first scale $j = 1$ up to ± 0.04 for the last scale $j = 4$.[12] Thus, based on the 95 % confidence intervals, all reported correlation point estimates are statistically significant.

4.1.1 Gold–Oil

The analysis of the intraday data for the gold-oil pair reveals a short period (1990–1991) of higher correlations, corresponding to the spike visible in Figs. 2, 3, 4, which should be associated with the economic downturn in the U.S. from July

[12]For the sake of brevity, we do not report confidence intervals for all estimates. These results are available from the authors upon request.

Table 3 Time-frequency correlation estimates for the gold-stocks pair

	High-frequency data (minutes)					Daily data (days)					
	10	20	40	80	160-y.	2	4	8	16	32	64-y.
1987	0.05	−0.04	−0.11	−0.22	−0.54	−0.22	−0.22	−0.24	−0.39	−0.58	0.64
1988	−0.02	−0.05	−0.14	−0.22	−0.21	−0.25	0.08	−0.06	−0.17	−0.12	−0.54
1989	−0.03	−0.11	−0.19	−0.15	−0.59	0.03	−0.26	−0.23	0.06	−0.67	−0.92
1990	−0.04	−0.15	−0.20	−0.25	−0.77	−0.32	−0.33	−0.28	−0.10	−0.36	−0.84
1991	−0.01	−0.07	−0.10	−0.09	−0.55	−0.16	−0.14	0.11	0.18	0.14	−0.59
1992	−0.03	−0.04	−0.03	−0.10	0.52	0.01	−0.01	−0.21	−0.28	0.37	0.21
1993	−0.02	−0.06	−0.10	−0.13	0.49	−0.24	−0.15	−0.23	0.03	−0.12	0.43
1994	−0.02	−0.10	−0.17	−0.21	0.41	−0.26	−0.18	0.05	0.00	0.38	0.31
1995	−0.01	−0.04	0.01	−0.04	−0.37	−0.17	−0.06	−0.02	0.01	−0.39	0.71
1996	−0.04	−0.12	−0.08	−0.13	−0.26	−0.20	−0.27	−0.24	0.47	0.57	−0.77
1997	−0.03	−0.06	−0.07	−0.11	−0.49	−0.15	−0.17	−0.26	−0.05	0.03	−0.93
1998	−0.05	−0.07	−0.13	−0.11	0.80	−0.03	0.17	0.28	−0.05	0.49	0.43
1999	−0.01	−0.01	−0.04	0.01	0.54	−0.03	0.10	0.05	0.15	0.45	−0.75
2000	−0.03	−0.07	−0.10	−0.20	0.54	−0.03	0.00	0.24	0.32	−0.25	−0.80
2001	−0.01	−0.01	0.00	0.01	−0.49	−0.24	−0.11	0.07	0.04	0.16	0.43
2002	−0.27	−0.34	−0.38	−0.37	−0.58	−0.21	−0.24	−0.37	−0.28	−0.34	−0.66
2003	−0.26	−0.35	−0.38	−0.42	0.46	−0.42	−0.12	−0.07	−0.51	−0.49	0.18
2004	−0.07	−0.09	−0.08	−0.09	0.65	0.03	0.14	0.38	0.14	0.17	0.29
2005	−0.02	−0.02	0.02	0.00	0.08	−0.08	0.11	0.09	−0.02	0.40	0.13
2006	0.05	0.11	0.17	0.20	0.30	0.10	−0.01	0.20	0.34	0.65	0.16
2007	0.20	0.26	0.29	0.27	−0.18	0.39	0.28	0.42	0.42	0.85	0.39
2008	0.11	0.14	0.10	0.09	0.87	−0.03	−0.16	−0.09	−0.16	−0.68	−0.89
2009	0.14	0.13	0.15	0.17	−0.17	0.01	−0.05	0.28	0.31	−0.05	−0.38
2010	0.25	0.25	0.28	0.29	0.06	0.14	0.29	0.46	0.38	−0.08	−0.20
2011	0.13	0.14	0.18	0.13	−0.40	−0.17	−0.18	−0.08	−0.04	0.20	0.49
2012	0.40	0.39	0.37	0.38	0.67	0.42	0.26	0.62	0.58	0.06	0.05

The high-frequency set contains Wavelet correlation estimates based on high-frequency data. The daily set contains Wavelet correlation estimates based on daily data

1990 to March 1991. During the period 1992–2005, the intraday correlations are remarkably low at short and longer horizons; see Table 2. In 2006, a significant increase in correlation begins, reaching its maximum in 2012 at all investment horizons. In contrast to the period 1990–1991, the recent financial crisis changed the correlation structure of the gold and oil pair, indicating the existence of an important structural break in the correlation structure. This result is in line with the detected structural break on September 8, 2006 (Sect. 4.2). Therefore, in terms of risk diversification, the situation changed dramatically for traders active at short-term investment horizons, as there is a significant increase in correlation after 2008 at all available investment horizons.

Dynamic correlations based on daily data reveal a more complex pattern. From 1987 until just before the global financial crisis erupted, correlations at diverse investment horizons seem quite heterogeneous (Table 2). We observe very low correlations at short investment horizons measured in days, whereas at longer invest-

Table 4 Time-frequency correlation estimates for the oil-stocks pair

	High-frequency data (minutes)					Daily data (days)					
	10	20	40	80	160-y.	2	4	8	16	32	64-y.
1987	0.03	0.01	0.05	0.04	−0.64	−0.11	0.21	−0.07	−0.09	−0.03	0.66
1988	0.03	−0.03	−0.05	−0.11	0.30	−0.06	0.13	−0.09	−0.15	−0.304	−0.74
1989	0.01	0.00	0.03	−0.02	0.13	−0.08	0.12	−0.06	−0.10	−0.74	0.06
1990	−0.02	−0.12	−0.18	−0.20	−0.54	−0.38	−0.46	−0.62	−0.25	0.04	−0.84
1991	−0.04	−0.10	−0.17	−0.19	−0.49	−0.06	−0.06	0.38	0.34	−0.51	0.58
1992	0.03	0.01	0.04	−0.02	−0.51	0.08	0.04	0.20	−0.20	−0.52	0.50
1993	0.00	0.00	−0.01	−0.02	0.73	−0.05	−0.11	0.24	0.42	0.10	−0.63
1994	0.00	0.01	−0.03	−0.04	−0.77	−0.23	−0.05	0.01	0.13	−0.47	−0.58
1995	−0.02	0.01	0.02	0.00	−0.47	−0.05	0.04	0.06	0.37	−0.14	−0.20
1996	0.00	0.00	0.00	−0.03	−0.02	−0.02	0.05	−0.18	−0.15	−0.38	0.56
1997	0.00	0.00	0.07	0.08	−0.61	−0.15	0.00	0.02	0.08	0.19	−0.50
1998	0.00	−0.02	0.02	0.02	0.64	0.04	0.13	0.15	0.21	0.52	−0.79
1999	−0.01	0.02	0.03	0.00	−0.94	−0.03	0.01	0.09	0.17	0.73	0.99
2000	0.02	−0.01	−0.02	−0.05	−0.18	−0.11	−0.09	0.07	−0.11	−0.53	−0.24
2001	0.01	0.04	0.03	−0.05	0.51	−0.12	−0.04	0.08	0.03	0.84	0.81
2002	−0.01	−0.01	−0.02	−0.03	−0.70	0.17	0.19	0.41	0.24	0.36	−0.53
2003	−0.01	−0.03	−0.04	−0.05	0.57	−0.24	0.08	−0.49	−0.36	−0.58	−0.54
2004	−0.07	−0.15	−0.18	−0.25	0.13	−0.13	0.01	0.05	−0.08	−0.68	−0.71
2005	−0.16	−0.19	−0.17	−0.18	0.00	−0.09	0.14	0.04	−0.46	−0.20	0.32
2006	0.04	0.07	0.09	0.10	−0.08	0.07	0.07	0.08	0.24	0.46	−0.12
2007	0.09	0.13	0.13	0.12	−0.63	0.17	0.07	−0.04	0.04	0.18	0.74
2008	0.26	0.27	0.31	0.33	0.70	0.42	0.32	0.09	0.05	0.12	−0.47
2009	0.42	0.46	0.48	0.50	0.92	0.53	0.28	0.61	0.10	−0.05	0.51
2010	0.57	0.59	0.58	0.62	0.69	0.70	0.71	0.51	0.58	0.91	0.86
2011	0.50	0.53	0.53	0.56	0.32	0.53	0.57	0.62	0.53	−0.03	0.74
2012	0.49	0.48	0.47	0.46	0.26	0.52	0.54	0.74	0.53	0.12	0.30

The high-frequency set contains Wavelet correlation estimates based on high-frequency data. The daily set contains Wavelet correlation estimates based on daily data

ment horizons of approximately 1 month, the correlations are higher. Beginning in 2008, the pattern changes significantly. Markets become quite homogeneous in perceptions of time because correlations at shorter and longer investment horizons become less diversified. Thus, the differences between short and long investment horizons diminish. One of the possible explanations is increased uncertainty in financial markets and poor economic performance in many developed countries during that period.

4.1.2 Gold-Stocks and Oil-Stocks

In comparison to the gold-oil pair, the gold-stocks and oil-stocks pairs provide a rather different picture (Tables 3 and 4). During the period 1991–1992, negative correlations dominate, especially at longer horizons. The negative correlations are

quite frequent for the two pairs, but they occur more often for the gold-stocks pair. Since 2001, the gold-stocks pair exhibits very rich correlation dynamics. The period of negative correlation begins in 2001, reaching its minimum in 2003, followed by a steady increase. After 2005, this pair exhibits significantly higher correlation, except for two short periods in 2008 and 2009.[13] In the 2012, we observe a significant increase in the correlation between gold and stocks at all available scales. There is an increase in magnitude that is three times larger relative to the previous year. This finding indicates a very limited possibility to diversify risk between stocks and gold in 2012.

The correlations of the oil-stocks pair also increased after the recent financial crisis began. Nevertheless, unlike the other two pairs, the correlation between oil and stocks before the crisis was considerably lower than after the crisis. This implies that the developments in 2008 had the strongest impact on the correlation structure of this pair. Further, from 2008 on, this pair has the highest correlation of the three examined pairs and highly homogeneous correlations at all scales. Therefore, after 2008 until the end of our sample, we only observe a negligible possibility for risk diversification in the sense of various investment horizons.

4.2 Risk Diversification

A wavelet methodology allows us to decompose mutual dependencies into different investment horizons; subsequently, we are able to generalize inferences related to risk diversification. When correlations are heterogeneous in their magnitudes at various investment horizons, market participants are able to diversify risk across these investment horizons, represented by scales. However, negligible or even no differences in correlation magnitudes at different investment horizons prevent effective risk diversification. For our set of assets, there was room for risk diversification until 2001; whereas problems with risk diversification arise after 2001.

Structural breaks cause important change with respect to the heterogeneity of correlations. In our analysis, we test for structural breaks in the correlations by employing the supF test (Hansen 1992; Andrews 1993; Andrews and Ploberger 1994) with p-values computed based on Hansen (1997); the results are summarized in Table 5. An illustrative example of a pre-structural break period is the gold-oil pair, with the break detected on September 8, 2006. We can observe a significant increase in the overall correlation estimated by the DCC GARCH during the periods 1994–1996 and 1998–2000. The DCC GARCH estimates aggregate the correlation over all investment horizons. However, using wavelet correlations, we obtain additional information that this increase might be caused particularly by the long-term correlations, as the correlations at short investment horizons are close

[13] On an annual basis, there was only a small decrease in 2011, as shown in Table 3.

Table 5 Values of the supF test with corresponding p-values

	Gold–oil	Gold-stocks	Oil-stocks
Date of the break	September 8, 2006	May 5, 2009	September 26, 2008
supF	3390.3	2544.3	7284.9
p-value	< 2.2e−16	< 1.1e−16	< 2.2e−16

The break dates dividing the period into pre-break and post-break. The full period covers January 2, 1987 to December 31, 2012

to zero. Similar patterns are observed for the gold-stocks and oil-stocks pairs, for which structural breaks were detected on May 5, 2009 and September 26, 2008, respectively.

Thus, we observe that the correlations between asset pairs were very heterogeneous across investment horizons before the structural break. Conversely, after the structural break, the correlation pattern became mostly homogeneous, which implies that gold, oil and stocks could no longer be simultaneously included in a single portfolio for risk diversification purposes. This finding contradicts the results of Baur and Lucey (2010), who find gold to be a good hedge against stocks and therefore a safe haven during financial market turmoil. However, our result is in line with the argument of Bartram and Bodnar (2009) that diversification provided little help for investors during the financial crisis.

The change in the correlation structure described above can also be attributed to changes in investors' beliefs,[14] which become mostly homogeneous across investment horizons after the structural break. The homogeneity can be partially induced by broader uncertainty regarding financial markets pricing fundamentals.[15] Investors tendency to favor more aggressive strategies may be one of the reasons that we observe increased homogeneity in the correlations across investment horizons. Furthermore, the homogeneity in correlations may have been increased by the introduction of completely electronic trading on exchange platforms in 2005, which was accompanied by an increased volume of automatic trading.

5 Conclusions

In this chapter, we study dynamic co-movements among key traded assets by employing the realized volatility and DCC GARCH approaches as a benchmark and the wavelet methodology, a novel time-frequency approach. In terms of the dynamic method, the wavelet-based correlation analysis enables us to analyze co-movements among assets, not only from a time series perspective, but also from

[14]Additional information on the role of investors' beliefs can be found in Ben-David and Hirshleifer (2012).

[15]Connolly et al. (2007) study the importance of time-varying uncertainty on asset correlation that subsequently influences the availability of diversification benefits.

the investment horizon perspective. Thus, we are able to provide unique evidence on how correlations among major assets vary over time and at different investment horizons. We analyze dynamic correlations in the prices of gold, oil, and a broad U.S. stock market index, the S&P 500, over 26 years from January 2, 1987 until December 31, 2012. The analysis is performed on both intra-day and daily data.

Our findings suggest that the wavelet analysis outperforms the standard benchmark approaches. Further, it offers a crucial message based on the evidence of very different patterns in linkages among assets over time. During the period before the pairs of assets suffered from structural breaks, our results revealed very low, even negative, but heterogeneous correlations for all pairs. After the breaks, the correlations for all pairs increased on average, but their magnitudes exhibited large positive and negative swings. Surprisingly, despite this strongly varying behavior, the correlations between pairs of assets became homogeneous and did not differ across distinct investment horizons. A strong implication emerges. Prior to the structural break, it was possible to use all three assets in a well-diversified portfolio. However, after the structural changes occurred, gold, oil, and stocks could not be used in conjunction for risk diversification purposes during the post-break period studied.

Acknowledgements We benefited from valuable comments we received from Abu Amin, Ladislav Krištoufek, Brian Lucey, Paresh Narayan, Lucjan Orlowski, Perry Sadorsky, Yi-Ming Wei, and Yue-Jun Zhang. The usual disclaimer applies. The support from the Czech Science Foundation (GAČR) under Grants GA13-24313S and GA14-24129S is gratefully acknowledged.

Appendix 1: Wavelet Variance

The variance of a time series can be decomposed into its frequency components, which are called scales in the wavelet methodology. Using wavelets, we can identify the portion of variance attributable to a specific frequency band of the examined time series. In this section, we demonstrate how to estimate wavelet variance and demonstrate that the summation of all of the components of wavelet variance yields the variance of the time series.

Let us suppose a real-valued stochastic process x_i, $i = 1, \ldots, N$, whose $\frac{L}{2}$th backward difference is a covariance stationary stochastic process with mean zero. Then, the sequence of the MODWT wavelet coefficients $W_x(j, s)$, unaffected by the boundary conditions, for all scales $j = 1, 2, \ldots, J^m$ is also a stationary process with mean zero. As we use the least asymmetric wavelet of length $L = 8$, we can expect stationarity of the MODWT wavelet coefficients. Following Percival (1995), we define the wavelet variance at scale j as the variance of wavelet coefficients at scale j as:

$$v_x(j)^2 = var\left(W_x(j, s)\right). \tag{15}$$

For coefficients unaffected by the boundary conditions, which are defined for each scale separately $M_j = N - L_j + 1 > 0$, the unbiased estimator of wavelet variance at scale j reads:

$$\nu_x(j)^2 = \frac{1}{M_j} \sum_{s=L_j-1}^{N-1} W_x(j,s)^2. \tag{16}$$

As the variance of a covariance stationary process x_i is equal to the integral of the spectral density function $S_x(.)$, the wavelet variance at a scale j is the variance of the wavelet coefficients $W_x(j,s)$ with spectral density function $S_x(j)(.)$:

$$\nu_x(j)^2 = \int_{-1/2}^{1/2} S_x(j)(f)df = \int_{-1/2}^{1/2} \mathscr{H}_j(f) S_x(j)(f)df, \tag{17}$$

where $\mathscr{H}_j(f)$ is the squared gain function of the wavelet filter h_j (Percival and Walden 2000). As the variance of a process x_i is the sum of the contributions of the wavelet variances at all scales we can write:

$$var(x) = \sum_{j=1}^{\infty} \nu_x(j)^2. \tag{18}$$

In case we have only a finite number of scales, we have to add also variance of the scaling coefficients vector; thus we can write:

$$var(x) = \int_{-1/2}^{1/2} S_x(f)df = \sum_{j=1}^{J^m} \nu_x(j)^2 + var\left(V_x(J^m, s)\right). \tag{19}$$

Appendix 2: Wavelet Covariance

The wavelet covariance of two processes x_t and y_t is estimated in a similar manner as the wavelet variance. As a first step, we perform the MODWT to obtain vectors of wavelet and scaling coefficients at all scales $j = 1, 2, \ldots, J^m$. While we use the LA(8) wavelet with length $L = 8$, we can use non-stationary processes, which are stationary after the d-th difference, where $d \leq L/2$. The wavelet covariance of x_t and y_t at scale j is defined as:

$$\gamma_{xy}(j) = Cov\left(W_x(j,s), W_y(j,s)\right). \tag{20}$$

Taking into consideration the MODWT wavelet coefficients unaffected by boundary conditions denoted $M_j = N - L_j + 1 > 0$, then for processes x_t and y_t defined above, the estimator of the wavelet covariance at a scale j is defined as

$$\hat{\gamma}_{xy}(j) = \frac{1}{M_j} \sum_{s=L_j-1}^{N-1} W_x(j,s) W_y(j,s), \tag{21}$$

When processes x_t and y_t are Gaussian, the MODWT estimator of wavelet covariance is unbiased and asymptotically normally distributed (Whitcher et al. 1999). When we have an infinite time series, the number of available scales goes to infinity, $J^m \to \infty$, then the sum of all available wavelet covariances $\gamma_{xy}(j)$ yields the covariance of x_t and y_t:

$$Cov(x_t, y_t) = \sum_{j=1}^{\infty} \gamma_{xy}(j). \tag{22}$$

For a finite real time series, the number of scales is limited by $J^m \leq log_2(T)$, the covariance of x_t and y_t is a sum of covariances of the MODWT wavelet coefficients $\gamma_{xy}(j)$ at all scales $j = 1, 2, \ldots, J^m$ and the covariance of the scaling coefficients $V_x(J,s)$ at scale J^m:

$$Cov(x_t, y_t) = Cov\left(V_x(J^m, s), V_y(J^m, s)\right) + \sum_{j=1}^{J^m} \gamma_{xy}(j). \tag{23}$$

Appendix 3: MODWT Wavelet Filters

Let us introduce the MODWT scaling and wavelet filters g_l and h_l, $l = 0, 1, \ldots, L-1$, where L denotes the length of the wavelet filter. For example, the Least Asymmetric (LA8) wavelet filter has length $L = 8$ (Daubechies 1992). Generally, the scaling filter is a low-pass filter whereas the wavelet filter is a high-pass filter. There are three basic properties that both MODWT filters must satisfy. Let us describe these properties for the MODWT wavelet filter

$$\sum_{l=0}^{L-1} h_l = 0, \quad \sum_{l=0}^{L-1} h_l^2 = 1/2, \quad \sum_{l=-\infty}^{\infty} h_l h_{l+2N} = 0, \quad N \in \mathbb{Z}_N \tag{24}$$

and for the MODWT scaling filter

$$\sum_{l=0}^{L-1} g_l = 1, \quad \sum_{l=0}^{L-1} g_l^2 = 1/2, \quad \sum_{l=-\infty}^{\infty} g_l g_{l+2N} = 0, \quad N \in \mathbb{Z}_N. \tag{25}$$

The transfer function of a MODWT filter $\{h_l\}$ at frequency f is defined via the Fourier transform as

$$H(f) = \sum_{l=-\infty}^{\infty} h_l e^{-i2\pi f l} = \sum_{l=0}^{L-1} h_l e^{-i2\pi f l} \tag{26}$$

with the squared gain function defined as: $\mathscr{H}(f) = |H(f)|^2$.

References

Aggarwal R, Lucey BM (2007) Psychological barriers in gold prices? Rev Financ Econ 16(2):217–230

Aguiar-Conraria L, Martins M, Soares MJ (2012) The yield curve and the macro-economy across time and frequencies. J Econ Dyn Control 36:1950–1970

Andersen T, Benzoni L (2007) Realized volatility. In: Andersen T, Davis R, Kreiss J, Mikosch T (eds) Handbook of financial time series. Springer, Berlin

Andersen T, Bollerslev T, Diebold F, Labys P (2003) Modeling and forecasting realized volatility. Econometrica 71(2):579–625

Andrews DW (1993) Tests for parameter instability and structural change with unknown change point. Econometrica 61:821—856

Andrews DW, Ploberger W (1994) Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62:1383–1414

Bandi F, Russell J (2006) Volatility. In: Birge J, Linetsky V (eds) Handbook of financial engineering. Elsevier, Amsterdam

Barndorff-Nielsen O, Shephard N (2004) Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. Econometrica 72(3):885–925

Bartram SM, Bodnar GM (2009) No place to hide: the global crisis in equity markets in 2008/2009. J Int Money Finance 28(8):1246–1292

Baur DG, Lucey BM (2010) Is gold a hedge or a safe haven? an analysis of stocks, bonds and gold. Financ Rev 45(2):217–229

Bauwens L, Laurent S (2005) A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. J Bus Econ Stat 23(3):346–354

Ben-David I, Hirshleifer D (2012) Are investors really reluctant to realize their losses? trading responses to past returns and the disposition effect. Rev Financ Stud 25(8):2485–2532

Bollerslev T (1990) Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch approach. Rev Econ Stat 72(3):498–505

Büyükşahin B, Robe MA (2013) Speculation, commodities and cross-market linkages. J Int Money Finance 42:38–70

Carlson M (2007) A brief history of the 1987 stock market crash with a discussion of the federal reserve response. Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board

Cashin P, McDermott C, Scott A (1999) The myth of co-moving commodity prices. IMF working paper WP/99/169 international monetary fund, Washington

Cashin P, McDermott CJ, Scott A (2002) Booms and slumps in world commodity prices. J Dev Econ 69(1):277–296

Chui C (1992) An inroduction to wavelets. Academic, New York

Connolly RA, Stivers C, Sun L (2007) Commonality in the time-variation of stock–stock and stock–bond return comovements. J Financ Mark 10(2):192–218
Daubechies I (1992) Ten lectures on wavelets. SIAM, Philadelphia
Engle R (2002) Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. J Bus Econ Stat 20(3):339–350
Engle RF, Sheppard K (2001) Theoretical and empirical properties of dynamic conditional correlation multivariate garch. Technical report, National Bureau of Economic Research
Faÿ G, Moulines E, Roueff F, Taqqu M (2009) Estimators of long-memory: Fourier versus wavelets. J Econom 151(2):159–177
Fratzscher M, Schneider D, Van Robays I (2013) Oil prices, exchange rates and asset prices. CESifo working paper no 4264
Gadanecz B, Jayaram K (2009) Measures of financial stability–a review. Bank for international settlements, IFC bulletin 3
Gallegati M, Gallegati M, Ramsey JB, Semmler W (2011) The us wage phillips curve across frequencies and over time. Oxf Bull Econ Stat 74(4):489–508
Gençay R, Selçuk F, Whitcher B (2002) An introduction to wavelets and other filtering methods in finance and economics. Academic, San Diego
Graham M, Kiviaho J, Nikkinen J (2013) Short-term and long-term dependencies of the s&p 500 index and commodity prices. Quant Finance 13(4):583–592
Hamilton J (2009) Causes and consequences of the oil shock of 2007-08. Brookings papers in economic activity 40(1):215–283
Hansen BE (1992) Tests for parameter instability in regressions with i (1) processes. J Bus Econ Stat 10:321–335
Hansen BE (1997) Approximate asymptotic p values for structuras-change tests. J Bus Econ Stat 15(1):60–67
Hansen P, Lunde A (2006) Realized variance and market microstructure noise. J Bus Econ Stat 24(2):127–161
Khordagui H, Al-Ajmi D (1993) Environmental impact of the gulf war: an integrated preliminary assessment. Environ Manage 17(4):557–562
Mallat S (1998) A wavelet tour of signal processing. Academic, San Diego
Markowitz H (1952) Portfolio selection. J Finance 7(1):77–91
Marshall JF (1994) The role of the investment horizon in optimal portfolio sequencing (an intuitive demonstration in discrete time). Financ Rev 29(4):557–576
McAleer M, Medeiros MC (2008) Realized volatility: a review. Econom Rev 27(1–3):10–45
Percival DB (1995) On estimationof the wavelet variance. Biometrika 82:619–631
Percival D, Walden A (2000) Wavelet methods for time series analysis. Cambridge University Press, Cambridge
Pindyck RS, Rotemberg JJ (1990) The excess co-movement of commodity prices. Econ J 100(403):1173–89
Ramsey JB (2002) Wavelets in economics and finance: past and future. Stud Nonlin Dyn Econom 6(3): Article 1, 1–27
Roueff F, Sachs R (2011) Locally stationary long memory estimation. Stoch Process Appl 121(4):813–844
Samuelson PA (1989) The judgment of economic science on rational portfolio management: indexing, timing, and long-horizon effects. J Portf Manage 16(1):4–12
Serroukh A, Walden AT, Percival DB (2000) Statistical properties and uses of the wavelet variance estimator for the scale analysis of time series. J Am Stat Assoc 95:184–196
Vacha L, Barunik J (2012) Co-movement of energy commodities revisited: evidence from wavelet coherence analysis. Energy Econ 34(1):241–247
Waldrop MM (1987) Computers amplify black monday: the sudden stock market decline raised questions about the role of computers; they may not have actually caused the crash, but may well have amplified it. Science 238(4827):602

Whitcher B, Guttorp P, Percival DB (1999) Mathematical background for wavelets estimators for cross covariance and cross correlation. Technical report 38, National Research Center for Statistics and the Environment

Whitcher B, Guttorp P, Percival DB (2000) Wavelet analysis of covariance with application to atmosferic time series. J Geophys Res 105:941–962

Part III
Forecasting and Spectral Analysis

Forecasting via Wavelet Denoising: The Random Signal Case

Joanna Bruzda

Abstract In the paper we evaluate the usability of certain wavelet-based methods of signal estimation for forecasting economic time series. We concentrate on extracting stochastic signals embedded in white noise with the help of wavelet scaling based on the non-decimated version of the discrete wavelet transform. The methods used here can be thought of as a type of smoothing, with weights depending on the frequency content of the examined processes. Both our simulation study and empirical examination based on time series from the M3-JIF Competition database show that the suggested forecasting procedures may be useful in economic applications.

1 Introduction

Time-scale (wavelet) analysis is well known for its ability to examine the frequency content of the processes under scrutiny with a good joint time–frequency resolution. The endogenously varying time window, which underlies this type of frequency analysis, makes this approach efficient computationally, enables a precise timing of events causing or influencing economic fluctuations and makes it possible to analyze economic relationships decomposed according to time horizons. The latter characteristic of wavelet analysis makes use of the fact that this type of studies is not limited to the short and long run, thus making it possible to conduct an examination for octave frequency bands, as is the case for the discrete wavelet transform (DWT) and the so-called continuous discrete (non-decimated) wavelet methodology, or even any arbitrary frequency band when continuous wavelet methods are used. As Ramsey and Lampart (1998) noticed, economists have long recognized that

J. Bruzda (✉)
Nicolaus Copernicus University, Toruń, Poland
e-mail: Joanna.Bruzda@uni.torun.pl

M. Gallegati and W. Semmler (eds.), *Wavelet Applications in Economics and Finance*, Dynamic Modeling and Econometrics in Economics and Finance 20,
DOI 10.1007/978-3-319-07061-2_9,

economic decision-making takes place at different time horizons and, furthermore, not only does the strength of a relationship change along with the scale of the analysis, but also the whole set of causes which influence the dependent variable. On the other hand, in certain applications some scales seem to be more important than other; for example, as was discussed in Ramsey (1996), forecasting, especially at longer horizons, involves global properties of time series, whereas in-sample fitting explores local properties.

In this paper we want to examine in more detail the predictive abilities of wavelets. Although wavelet analysis is not a forecasting technique *per se*, its distinguishing features, such as computational efficiency, excellent temporal resolution and the possibility to decompose time series according to time scales—which we mentioned above—as well as overall methodological simplicity and conventional DWT's good decorrelation property allow one to believe that it can be helpful in forecasting economic processes. From the theoretical point of view, wavelets, on the one hand, should enable one to conduct a more precise study via building models specified within or across different frequency bands and computing forecasts as aggregates of the forecasted values for the component series, while, on the other hand, they may significantly simplify the analysis via transforming the predicted variables in such a manner that it may be possible to find better forecast models. In the latter case it is assumed that the wavelet transform leads to an analysis of less complicated uni- and multivariate processes, to which tailor-made forecasting techniques can be applied (*cf.* Li and Hinich 2002; Ramsey 2002; Kaboudan 2005). One should keep in mind, however, that there are also certain obvious disadvantages of forecasting with wavelets resulting from the possibility of overparametrization and the arbitrariness of the choice of decomposition level and wavelet function.[1]

Further in the text we will focus exclusively on modeling univariate time series, although it is worth mentioning that some steps in the direction of modeling and forecasting multivariate time series utilizing the properties of wavelets as mentioned above have also been taken in the literature (see, e.g., Bruzda 2013b, Chap. VI, and references therein).

By far one of the most important characteristics of wavelets is their denoising property, i.e., their potential in signal estimation. Our contribution in this paper consists in suggesting two forecasting procedures that make explicit use of nonparametric wavelet estimation of *random signals* and evaluate them on some simulated and real economic data. To the best of our knowledge, this has not been done yet in the literature on time series. In previous applications of wavelet denoising for forecasting purposes, researchers focused exclusively on wavelet thresholding (see Alrumaih and Al-Fawzan 2002; Ferbar et al. 2009; Schlüter and Deuschle 2010). We argue that applying thresholding rules may not be the best way to

[1] For the forecasting procedures suggested further in the text the best outcomes are usually produced by the shortest Daubechies wavelets and a small number (1–2) of decomposition stages. In any case, however, a reasonable strategy may be to use an optimized forecasting setup—see also our remarks in the concluding section.

proceed if the underlying signal is stochastic, as this approach transforms Gaussian processes into non-Gaussian ones, preserves outliers and effectively reduces the high frequency spectra to zero. Instead, we suggest a sort of rescaling of the examined process's spectrum. To this end, we use the idea of wavelet scaling (see Percival and Walden 2000, §10.3), which we apply on the basis of the non-decimated discrete wavelet transform (see Bruzda 2013a). This approach can be thought of as a type of smoothing, with weights depending on the spectral properties of the process under study. Nonparametric signal estimation is then followed by certain simple forecasting techniques, such as the naïve (no-change) method or the ARIMA model building strategy, although it can also be used in combination with, e.g., some nonparametric methods, such as neural networks. Both our simulation studies and empirical examinations confirm that using denoising techniques prior to Box–Jenkins ARIMA modeling can lower forecast error variance in short- and medium-sized samples. In addition, the high computational efficiency of this approach makes it attractive for forecasting in automatic systems that are applied to large datasets.

The rest of the paper is organized as follows: in the next section we briefly describe the conventional DWT and its non-decimated variant known as the maximal overlap discrete wavelet transform (MODWT). In Sect. 3 we give a short overview of the procedures used in wavelet forecasting of univariate time series; in Sect. 4 we present the methodology of the signal extraction used in this study, which we also refer to as wavelet smoothing, and we introduce the forecasting procedures that are based on it. In Sect. 5 we evaluate the performance of the suggested forecasting methods with some simulated data. Finally, Sect. 6 presents a real data example based on 16 time series from the M3-JIF Competition database (see Makridakis and Hibon 2000). Section 7 offers brief conclusions.

2 DWT and MODWT

The wavelet transform consists in decomposing a signal into rescaled and shifted versions of a function, called the mother wavelet, which integrates to 0 and whose energy is equal to 1. It is assumed that dilated and translated copies of the mother wavelet (called the wavelet atoms or the daughter wavelets) are well localized on both the time and frequency axes. This way the wavelet transform is well suited to examine phenomena exhibiting various periodicities evolving with time. In contrast to the short-time Fourier transform, which uses the same data windows to analyze different frequencies, the wavelet transform is a windowing technique with a varying window size, i.e., it examines high frequency oscillations in small data windows and low frequency components in large windows. This distinguishing property of wavelets, which makes it possible to 'see both the forest and the trees', is referred to as the wavelet zoom and stands behind the success of wavelets in applied science. Further in the text we will be working with discretely sampled wavelets (i.e., we will conduct cascade filtering known as the discrete wavelet transform—DWT)

because—following Percival and Walden (2000)—we feel that they are a natural tool for analyzing discrete time series.

Let us consider a discrete signal of length $N = 2^{J_0}$ of the form: $\mathbf{x} = (x_0, x_1, \ldots, x_{N-1})^T$. The DWT of $\mathbf{x}$ is defined via a couple of quadrature mirror filters: the (half-band) high-pass wavelet filter $\{h_l\}_{j=0,\ldots,L-1}$ and the corresponding (half-band) low-pass scaling filter $\{g_l\}_{j=0,\ldots,L-1}$, where L is the filters' width and L is even. The two filters fulfill the so-called quadrature mirror relationship, i.e., $g_l = (-1)^{l+1} h_{L-1-l}$, they are even-shift orthogonal, the coefficients of the wavelet filter integrate to 0 and those of the scaling filter to $\sqrt{2}$ while, in both cases, their squares sum up to 1 (see Percival and Walden 2000, Chap. IV). In our simulation and empirical studies in Sects. 5 and 6 we concentrate exclusively on the two shortest filters belonging to the family of Daubechies filters. These are the Haar filters defined as:

$$h_0 = \tfrac{1}{\sqrt{2}}, \quad h_1 = -\tfrac{1}{\sqrt{2}}, \quad g_0 = \tfrac{1}{\sqrt{2}}, \quad g_1 = \tfrac{1}{\sqrt{2}}, \tag{1}$$

and the Daubechies' extremal phase filters of width 4 (denoted $d4$), for which:

$$h_0 = \tfrac{1-\sqrt{3}}{4\sqrt{2}}, \quad h_1 = \tfrac{-3+\sqrt{3}}{4\sqrt{2}}, \quad h_2 = \tfrac{3+\sqrt{3}}{4\sqrt{2}}, \quad h_3 = \tfrac{-1-\sqrt{3}}{4\sqrt{2}}, \tag{2}$$

and $\{g_l\}$ is defined according to the quadrature mirror relationship.

For $N = 2^{J_0}$ it is possible to perform up to J_0 stages of a wavelet decomposition, and the numbers of the (conventional) wavelet and scaling coefficients (W_{jt} and V_{jt}, respectively) for the consecutive decomposition levels $j = 1, 2, \ldots, J_0$ are then: $N/2, N/4, \ldots, 1$. By contrast, due to the lack of downsampling by 2, the MODWT produces exactly the same number (N) of both types of coefficients (denoted $\tilde{W}_{jt}$ and $\tilde{V}_{jt}$) at each decomposition stage. The wavelet coefficients W_{jt} and $\tilde{W}_{jt}$ are associated with variations at scales $\lambda_j = 2^{j-1}$, i.e., with oscillations whose periods are approximately in the intervals $[2, 4), [4, 8), \ldots, \left[2^J, 2^{J+1}\right)$, whereas the scaling coefficients V_{jt} and $\tilde{V}_{jt}$ correspond to the low frequency components of the signals, i.e., oscillations with periods of length at least 2^j ($j = 1, 2, \ldots, J_0$).

The coefficients of the DWT and MODWT are computed via the so-called pyramid algorithm as follows:

$$W_{jt} = \sum_{l=0}^{L-1} h_l V_{j-1,[2t+1-l] \bmod N_{j-1}}, \quad V_{jt} = \sum_{l=0}^{L-1} g_l V_{j-1,[2t+1-l] \bmod N_{j-1}} \tag{3}$$

for $t = 0, \ldots, N_j - 1$, $N_j = N/2^j$, $j = 1, \ldots, J_0$, where $V_{0t} = x_t$ ($t = 0, \ldots, N-1$), and:

$$\tilde{W}_{jt} = \sum_{l=0}^{L-1} \tilde{h}_l \tilde{V}_{j-1,[t-2^{j-1}l] \bmod N}, \quad \tilde{V}_{jt} = \sum_{l=0}^{L-1} \tilde{g}_l \tilde{V}_{j-1,[t-2^{j-1}l] \bmod N} \tag{4}$$

for $t = 0, \ldots, N-1, j = 1, \ldots, J_0$, where $\tilde{V}_{0t} = x_t$ $(t = 0, \ldots, N-1)$ and $\tilde{h}_l = h_l/\sqrt{2}, \tilde{g}_l = g_l/\sqrt{2}$.

In the definitions above the circular convolution is used in order to define the boundary coefficients, although according to our forecast procedures described in further sections of this paper the coefficients affected by the extrapolation method are removed or replaced with coefficients computed with backcasted values of the signal.

In our further considerations we will adopt the following matrix notation. First, the coefficients of the DWT at level $J \leq J_0$ can be written in the form of a vector:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_J \\ \mathbf{V}_J \end{bmatrix}_{N \times 1},$$

where $\mathbf{W}_j$ $(j = 1, \ldots, J)$ and $\mathbf{V}_J$ are column vectors of length N_j $(j = 1, \ldots, J)$ and N_J of the appropriate wavelet and scaling coefficients, W_{jt} and V_{Jt}, respectively. These are obtained by an orthonormal transform of the form:

$$\mathbf{W} = \mathcal{W}\,\mathbf{x},$$

where the square matrix $\mathcal{W}$ can be partitioned as follows:

$$\mathcal{W} = \begin{bmatrix} \mathcal{W}_1 \\ \mathcal{W}_2 \\ \vdots \\ \mathcal{W}_J \\ \mathcal{V}_J \end{bmatrix}_{N \times N}$$

and $\mathbf{W}_j = \mathcal{W}_j\mathbf{x}$ $(j = 1, \ldots, J)$, $\mathbf{V}_J = \mathcal{V}_J\mathbf{x}$. Then, by using the inverse wavelet transform, the original signal is recovered as follows:

$$\mathbf{x} = \mathcal{W}^T\mathbf{W} = \sum_{j=1}^{J} \mathcal{W}_j^T\mathbf{W}_j + \mathcal{V}_J^T\mathbf{V}_J = \sum_{j=1}^{J} \mathbf{D}_j + \mathbf{S}_J. \tag{5}$$

The $N \times 1$ dimensional vectors $\mathbf{D}_j$ $(j = 1, \ldots, J)$ and $\mathbf{S}_J$ are known as the details and the Jth level smooth (approximation), while the whole additive decomposition in Eq. (5) defines the so-called multiresolution analysis (MRA).

For the MODWT, the coefficient vectors $\tilde{\mathbf{W}}_j$ and $\tilde{\mathbf{V}}_J$ of length N are obtained via multiplying $\mathbf{x}$ by the appropriate square matrices $\tilde{\mathcal{W}}_j$ and $\tilde{\mathcal{V}}_J$, respectively, i.e., the following matrix operations are executed:

$$\tilde{\mathbf{W}}_j = \tilde{\mathcal{W}}_j \mathbf{x}, \quad \tilde{\mathbf{V}}_J = \tilde{\mathcal{V}}_J \mathbf{x}.$$

Then the recovery formula takes the form:

$$\mathbf{x} = \sum_{j=1}^{J} \tilde{\mathcal{W}}_j^T \tilde{\mathbf{W}}_j + \tilde{\mathcal{V}}_J^T \tilde{\mathbf{V}}_J = \sum_{j=1}^{J} \tilde{\mathbf{D}}_j + \tilde{\mathbf{S}}_J. \tag{6}$$

We finish this section with two further remarks. First, for our considerations in Sect. 4 we notice that, exclusively in the case of the Haar filters, the following additive decomposition also takes place:

$$\mathbf{x} = \tilde{\mathbf{W}}_1 + \tilde{\mathbf{W}}_2 + \cdots + \tilde{\mathbf{W}}_J + \tilde{\mathbf{V}}_J. \tag{7}$$

This becomes obvious by noting that the following relationships hold for $J \leq J_0$:

$$\tilde{W}_{J,t} = \tfrac{1}{2}\left(\tilde{V}_{J-1,t} - \tilde{V}_{J-1,t-\lambda_J}\right), \quad \tilde{V}_{J,t} = \tfrac{1}{2}\left(\tilde{V}_{J-1,t} + \tilde{V}_{J-1,t-\lambda_J}\right),$$

where we define $V_{0,t} = x_t$ $(t = 0, 1, \ldots, N-1)$.

Second, in our theoretical discussion, simulation study and empirical evaluation we concentrate on the MODWT-based wavelet and scaling coefficients and the MODWT-based details and smooths. The reasons for this are the following: first, computing signal estimates with the inverse non-decimated wavelet transform results in a lower mean squared error (MSE) than for the inverse DWT (see Bruzda 2013a). Furthermore, the non-decimated transform can be applied to signals of a length which is not a multiple of a power of 2 and—what seems to be important from the forecasting point of view—due to its time invariance property it treats all observations in a similar manner. As a result, forecasts are always the same functions of previous observations. Finally, the MODWT-based wavelet coefficients produce more efficient estimators of wavelet variance (see Percival 1995; Percival and Walden 2000, Chap. VIII).

3 Univariate Forecasting with Wavelets

Among the methods of wavelet forecasting of univariate time series, Schlüter and Deuschle (2010) distinguished the following:

1. Forecasting signals estimated via wavelet thresholding
2. Structural time series modeling with component processes obtained through the wavelet MRA
3. Modeling wavelet and scaling coefficients
4. Forecasting locally stationary wavelet processes.

The first approach makes use of the fact that a white noise stochastic component in a representation of the form 'signal + noise' affects wavelet coefficients from all decomposition levels in the same way. As a result, for deterministic signals it is suggested to remove all the coefficients that are less in magnitude than a certain threshold. The modified wavelet coefficients constitute the building blocks of the signal estimate, which is eventually obtained via the inverse wavelet transform.

In order to compute the modified coefficients, the so-called hard and soft thresholding rules are usually used. The former is defined as:

$$W'_{jt} = W_{jt}\mathbf{1}_{\{|W_{jt}|>\lambda\}}, \tag{8}$$

and the latter is given as:

$$W'_{jt} = \text{sig}\left(W_{jt}\right)\ \left(\left|W_{jt}\right| - \lambda\right)\quad \mathbf{1}_{\{|W_{jt}|>\lambda\}}, \tag{8a}$$

where λ is the assumed threshold value. For DWT-based denoising, the threshold λ can be set globally, while in the case of the non-decimated wavelet transform the threshold changes with the decomposition stage.[2] So far, DWT-based thresholding has mainly been used for forecasting purposes (see Alrumaih and Al-Fawzan 2002; Ferbar et al. 2009; Schlüter and Deuschle 2010) and the thresholds were those introduced by Donoho and Johnstone (1994, 1995).

In Schlüter and Deuschle (2010), the authors document that wavelet denoising improves short-term forecasts from ARMA and ARIMA models. This conclusion was obtained for an oil price series and a Euro/Dollar exchange rate. Alrumaih and Al-Fawzan (2002) came to similar conclusions for the Saudi Stock Index. Finally, Ferbar et al. (2009) showed in a simulation study that wavelet thresholding is an attractive alternative to exponential smoothing for forecasting with an asymmetric loss function.

The second approach to wavelet forecasting makes use of the wavelet multiresolution analysis defined via Eqs. (5) and (6) and consists in a separate modeling and forecasting of details and smooths. The final forecast is obtained as a sum of predictions for the component series. The approach was suggested in Arino (1995) and further applied in different forms by, among others, Yu et al. (2001), Zhang et al. (2001), Wong et al. (2003) and Fernandez (2008). In particular, Arino (1995) analyzes monthly car sales in Spain by defining two component processes: seasonal and irregular fluctuations $D_1 + D_2 + D_3 + D_4$ and the trend component S_4. (S)ARIMA models are then estimated for each of these series. By contrast, Zhang et al. (2001) use MODWT-based multiresolution analysis with five decomposition levels and combine it with neural network processing, while in Yu et al. (2001)

[2]For a presentation and discussion of thresholding rules and methods of threshold selection, see Ogden (1997), Chaps. VII and VIII; Vidakovic (1999), Chap. VI; Percival and Walden (2000), Chap. X; and Nason (2008), Chap. III.

a similar approach is advocated, but the authors additionally suggest an iterative forecasting procedure to compute boundary coefficient values.

The third method of forecasting with wavelets is used by, e.g., Renaud et al. (2002), Kaboudan (2005), Chen et al. (2010), and Minu et al. (2010). In each of these papers the authors build linear or nonlinear forecast models based on wavelet and scaling coefficients. The final forecast is usually obtained via an appropriate inverse wavelet transform although, in the case of the Haar wavelet, Eq. (7) is often used. Such an approach implicitly assumes that wavelet and scaling coefficients can be employed to build simpler models than those based on the original data. In Renaud et al. (2002) the so-called multiscale autoregressive (MAR) models are introduced. These are linear models for the original series having MODWT-based Haar wavelet and scaling coefficients and their specific lagged values as regressors. Minu et al. (2010) transform the MAR model into a nonlinear form, i.e., to the form of a neural network. A combination of neural network modeling and wavelet decompositions is also considered in Kaboudan (2005). This time it is suggested to build separate nonlinear models for conventional (DWT-based) wavelet and scaling coefficients. In Chen et al. (2010) the WAW procedure (abbr. from wavelet, ARMAX, Winters) was introduced, according to which non-decimated wavelet and scaling coefficients are used to define trend, seasonal and high frequency components. These components are further forecasted with some conventional techniques such as exponential smoothing, harmonic regression and ARMAX models.

The fourth approach to univariate forecasting with wavelets, introduced by Fryźlewicz et al. (2003), is based on the notion of locally stationary wavelet processes. The method uses time-varying autoregressive representations of data, the parameters of which are estimated by solving wavelet variants of the Yule–Walker equations. It makes use of the non-decimated wavelet transform and, in practice, requires an estimation of the so-called evolutionary wavelet spectrum with the help of the corrected wavelet periodogram (see, e.g., Nason 2008, and references therein).

According to the classification recalled here, the two forecasting procedures which we introduce below expand on the first approach.

4 Signal Estimation and Forecasting via Wavelet Smoothing

Signal estimation based on wavelet thresholding assumes that the observed process is composed of a deterministic signal and a random (white or colored) noise. Such an assumption results, in practice, in elimination of the high frequency spectrum of the process, i.e., the variance of wavelet coefficients from lower decomposition levels (especially from the first level) is entirely attributed to the noise. This, however, will be inappropriate when the estimated signal is stochastic (but see wavelet denoising for the so-called sparse signals discussed in Percival and Walden 2000, Chap. X).

Assuming the signals' stochastic character is in accordance with the concept of structural time series modeling or the popular exponential smoothing methods, and seems to be more adequate in describing economic processes, which are usually modelled under the stochastic dependence paradigm. Moreover, the spectral shape of an economic process often does not correspond closely with the representation of the form 'deterministic function + white noise'.

Let us consider a process y_t generated according to:

$$y_t = x_t + \eta_t, \tag{9}$$

where x_t is a random signal and η_t is a white noise process, uncorrelated with x_t at all leads and lags. We assume for the moment that the signal x_t has a mean value of zero and that the DWT- and MODWT-based wavelet coefficients are covariance stationary. This means in particular that the wavelet filters applied here have enough vanishing moments to stabilize the variance of x_t.

Percival and Walden (2000), §10.3, consider a scaling (shrinkage) estimator of the following form:

$$\widehat{W}_{jt}^{x} = a_j W_{jt}^{y}, \tag{10}$$

where $\widehat{W}_{jt}^{x}$ and W_{jt}^{y} denote the jth elements of $\widehat{\mathbf{W}}_j^{x}$ and $\mathbf{W}_j^{y}$, respectively, with $\mathbf{W}_j^{y}$ being the vector of the jth level conventional DWT coefficients of the observed process y_t, and $\widehat{\mathbf{W}}_j^{x}$ being the appropriate vector for the signal estimate. Then, by minimizing the risk defined as:

$$E\|\mathbf{x} - \widehat{\mathbf{x}}\|^2, \tag{11}$$

where $\mathbf{x}$ and $\widehat{\mathbf{x}}$ are column vectors of length N of the signal and its estimate, the following solution is obtained:

$$a_j = \frac{E\left(W_{jt}^{x} W_{jt}^{y}\right)}{E\left(W_{jt}^{y}\right)^2} = \frac{E\left(W_{jt}^{x}\right)^2}{E\left(W_{jt}^{y}\right)^2}. \tag{12}$$

The final signal estimate is computed via the inverse DWT, assuming that the scaling coefficients of y_t are left unchanged.

In Bruzda (2013a), in order to estimate the signal x_t a similar idea to the one above was applied to the MODWT coefficients, leading to two kinds of estimators. The first among them has the following form for a given $J \leq J_0 = \log_2 N$:

$$\tilde{\mathbf{x}} = \sum_{j=1}^{J} \tilde{\mathcal{W}}_j^T \widehat{\tilde{\mathbf{W}}}_j^{x} + \tilde{\mathcal{V}}_J^T \widehat{\tilde{\mathbf{V}}}_J^{x} = \sum_{j=1}^{J} \widehat{\tilde{\mathbf{D}}}_j^{x} + \widehat{\tilde{\mathbf{S}}}_J^{x} = \sum_{j=1}^{J} a_j \tilde{\mathbf{D}}_j^{y} + b_J \tilde{\mathbf{S}}_J^{y}, \tag{13}$$

where $\widehat{\tilde{W}}_{jt}^{x}$ is defined similarly to $\widehat{W}_{jt}^{x}$ in Eq. (10), i.e.:

$$\widehat{\tilde{W}}_{jt}^{x} = a_j \tilde{W}_{jt}^{y}, \tag{10a}$$

while $\widehat{\tilde{V}}_{J}^{x}$ is again equal to $\tilde{V}_{J}^{y}$ (i.e., b_J is set to 1) or is obtained via rescaling $\tilde{V}_{J}^{y}$ as in Eq. (13) (i.e., we can also set $b_J < 1$).

The second signal estimator is defined as:

$$\tilde{\tilde{\mathbf{x}}} = \widehat{\tilde{\mathbf{W}}}_1^{x} + \cdots + \widehat{\tilde{\mathbf{W}}}_J^{x} + \widehat{\tilde{\mathbf{V}}}_J^{x} = a_1 \tilde{\mathbf{W}}_1^{y} + \cdots + a_J \tilde{\mathbf{W}}_J^{y} + b_J \tilde{\mathbf{V}}_J^{y}, \tag{14}$$

where this time we exclusively consider the Haar filters defined in Eq. (1). The reasons for choosing the Haar wavelet are the following. First, other wavelet filters are associated with larger phase shifts and do not necessarily generate decreasing smoothing weights. Moreover, for the Haar wavelet the additive decomposition defined in Eq. (7) holds. This results in leaving signals not corrupted by noise unchanged.

The smoothing constants a_j are obtained as in Eq. (12), although now it is more convenient to define them on the basis of the non-decimated coefficients, i.e.:

$$a_j = \frac{E\left(\tilde{W}_{jt}^{x}\tilde{W}_{jt}^{y}\right)}{E\left(\tilde{W}_{jt}^{y}\right)^2} = \frac{E\left(\tilde{W}_{jt}^{x}\right)^2}{E\left(\tilde{W}_{jt}^{y}\right)^2} = 1 - \frac{E\left(\tilde{W}_{jt}^{\eta}\right)^2}{E\left(\tilde{W}_{jt}^{y}\right)^2}. \tag{15}$$

Assuming that the wavelet variance of the noise at the first stage of the wavelet decomposition (i.e., for scale $\lambda_1 = 1$) is given as:

$$\sigma_{\eta}^2(\lambda_1) = E\big(\tilde{W}_{1t}^{\eta}\big)^2 = h\,\sigma_y^2(\lambda_1) = h\,E\big(\tilde{W}_{1t}^{y}\big)^2$$

for certain $h \in [0,\ 1]$, from our assumption about the noise term η_t we finally obtain:

$$\widehat{\tilde{W}}_{j}^{x} = \left[1 - \frac{h}{2^{j-1}}\frac{\sigma_y^2(\lambda_1)}{\sigma_y^2(\lambda_j)}\right]\tilde{W}_{j}^{y}, \tag{16}$$

where $\sigma_y^2(\lambda_j) = E\left(\tilde{W}_{jt}^{y}\right)^2$ is the wavelet variance of y_t at scale $\lambda_j = 2^{j-1}$ ($j = 1, 2, \ldots, J$) and it is assumed that the expression in the square brackets is positive.

A similar reasoning can be applied to the scaling coefficients. If the scaling coefficients are (trend-)stationary with a mean value m_t, the smoothing formula becomes:

$$\widehat{\tilde{V}}_{Jt}^{x} = m_t + \left[1 - \frac{h}{2^{J-1}}\frac{\sigma_y^2(\lambda_1)}{E\left(\tilde{V}_{Jt}^{y} - m_t\right)^2}\right]\cdot\left[\tilde{V}_{Jt}^{y} - m_t\right]. \tag{17}$$

Alternatively, the coefficients $\tilde{V}_{Jt}^{y}$ can be left unchanged. Such an approach will be justified especially for integrated processes.

The estimators in Eqs. (13) and (14) are defined via linear time-invariant filters, which makes them more appropriate in forecasting applications. Furthermore, in Bruzda (2013a) it was shown that the estimator $\tilde{\mathbf{x}}$ defined in Eq. (13) has a smaller risk than its DWT-based counterpart $\widehat{\mathbf{x}}$. On the other hand, the risk of the estimator $\tilde{\tilde{\mathbf{x}}}$ defined in Eq. (14) will usually be higher than that of $\widehat{\mathbf{x}}$. This comes as no surprise since $\tilde{\tilde{\mathbf{x}}}$ is based on a causal filter, while the other two estimators ($\widehat{\mathbf{x}}$ and $\tilde{\mathbf{x}}$) are not. We hypothesize, however, that such an estimator can be of interest for forecasting purposes, especially for low values of correlations between scales, since it provides a very simple nonparametric method of signal extraction based on current data. For example, for level four decomposition the weighting scheme is the following:

$$\begin{aligned}\tilde{\tilde{x}}_t = {} & \theta_1 y_t + \theta_2 y_{t-1} + \theta_3 \left(y_{t-2} + y_{t-3}\right) + \theta_4 \left(y_{t-4} + y_{t-5} + y_{t-6} + y_{t-7}\right) + \\ & + \theta_5 \left(y_{t-8} + y_{t-9} + y_{t-10} + y_{t-11} + y_{t-12} + y_{t-13} + y_{t-14} + y_{t-15}\right).\end{aligned} \tag{18}$$

Because the wavelet coefficients $\tilde{W}_{jt}$ are obtained via filtering which embeds difference operators, while the scaling coefficients $\tilde{V}_{Jt}$ are computed via cascade filtering with filters whose coefficients sum up to 1, for level J decomposition the sum of the smoothing weights θ_j is always equal to b_J. Because the details $\tilde{D}_{jt}$ and smooths $\tilde{S}_{Jt}$ share the above-mentioned properties of $\tilde{W}_{jt}$ and $\tilde{V}_{Jt}$, then also the weights of the filter which defines the estimator $\tilde{\mathbf{x}}$ sum up to b_J, but this time the filter is symmetric and has wider support. For example, for wavelet filters which will be considered further in the simulation and in the empirical part of the paper—provided in Eqs. (1) and (2)—level $J = 1, 2, 3, 4$ decomposition results in filters of width 3, 7, 15, 31 in the case of the Haar wavelet and 7, 19, 43, 91 in the case of the $d4$ wavelet, respectively. On the other hand, the width of the causal filter defining the estimator in Eq. (14) is equal to 2^J.

The scaling estimator of Percival and Walden requires that wavelet filters be (relatively) good at decorrelating processes, both within and between scales.[3] For the estimators in Eqs. (13) and (14) we assume that the MODWT coefficients corresponding to different dyadic frequency intervals can be treated as approximate band-pass white noises and that between-scale wavelet decorrelation is relatively effective.

In order to use the methods suggested here in real data applications, the wavelet variance $\sigma_Y^2\left(\lambda_j\right)$ must be replaced with its estimate. To this end we propose to use an estimator utilizing, again, the MODWT coefficients and excluding all boundary values. This is due to the fact that such an estimator is unbiased and efficient.[4] It was shown in Bruzda (2013a) via extensive computer simulations that combining

[3]For a discussion of the Daubechies filters' decorrelation properties, see, e.g., Bruzda (2013b), Chap. II.

[4]This estimator is more efficient than its DWT-based counterpart (for details see Percival 1995; Percival and Walden 2000, pp. 308–310). There are, however, instances when a biased MODWT-

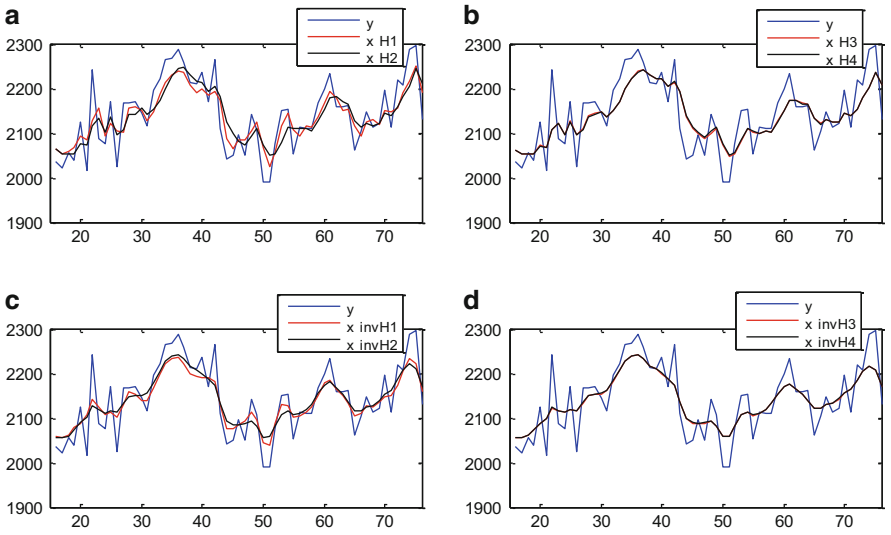


Fig. 1 Signal estimates obtained with the Haar wavelet and the estimators $\tilde{\tilde{\mathbf{x}}}$ [graphs (**a**) and (**b**); estimates H1–H4 correspond to $J = 1, \ldots, 4$] and $\tilde{\mathbf{x}}$ [graphs (**c**) and (**d**); estimates invH1–invH4 correspond to $J = 1, \ldots, 4$]; series N2883 from the M3-IJF-Competition database; $h = 1$

this estimator of the wavelet variance with the formulae in Eqs. (13) and (14) leads, in small samples, to signal approximations which are often associated with smaller MSEs than a signal extraction based on parametric models and maximum likelihood estimation.

As an illustration we applied our two signal estimators, using the Haar filters in each case, to one of the series analyzed further in Sect. 5—series N2883 of length 76 from the M3-IJF Competition database (see Makridakis and Hibon 2000)—see Figs. 1, 2, and 3. For the estimator based on the inverse wavelet transform the mean value of the series was used to extrapolate the data at the end of the sample, while the values at the beginning of the signal were not estimated. Both the wavelet and scaling coefficients were rescaled assuming that the smoothing parameter h is equal to 1, 0.75 and 0.5, respectively. Up to four decomposition levels were considered. Clearly, the 'non-inverse' method produces a phase shift which is, however, relatively small. Besides, this example seems to suggest that assuming $h = 1$ may result in too smooth signal estimates for short-term forecasting.

We end this section with an outline of the forecasting procedures that are used further in Sects. 5 and 6. The 'non-inverse' estimator $\tilde{\tilde{\mathbf{x}}}$ is directly applicable to forecasting, i.e., after signal estimation we can, e.g., build some parametric models and use them to compute forecasts, whereas the estimator $\tilde{\mathbf{x}}$ requires considering a

based estimator can outperform the unbiased estimator—see, e.g., Bruzda (2011) for time delay estimation at higher scales as well as references in Percival and Walden (2000), p. 378.

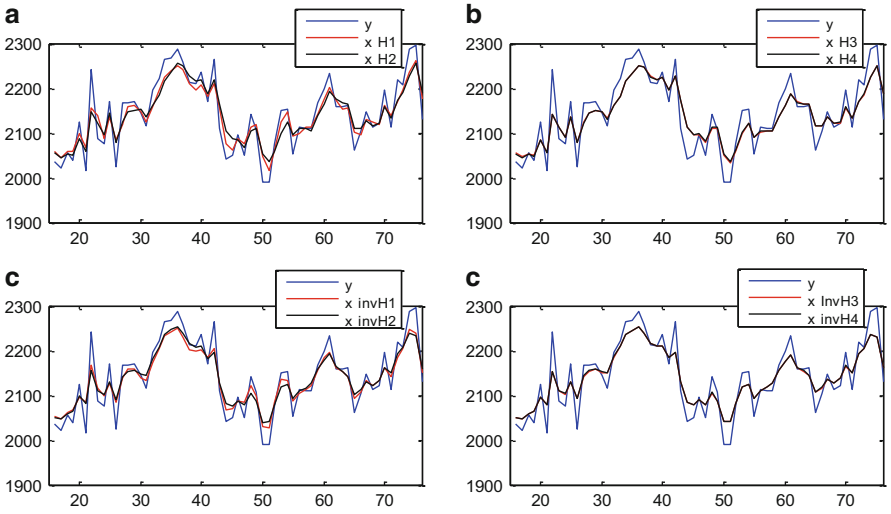


Fig. 2 Signal estimates obtained with the Haar wavelet; $h = 0.75$; see Fig. 1 for more details

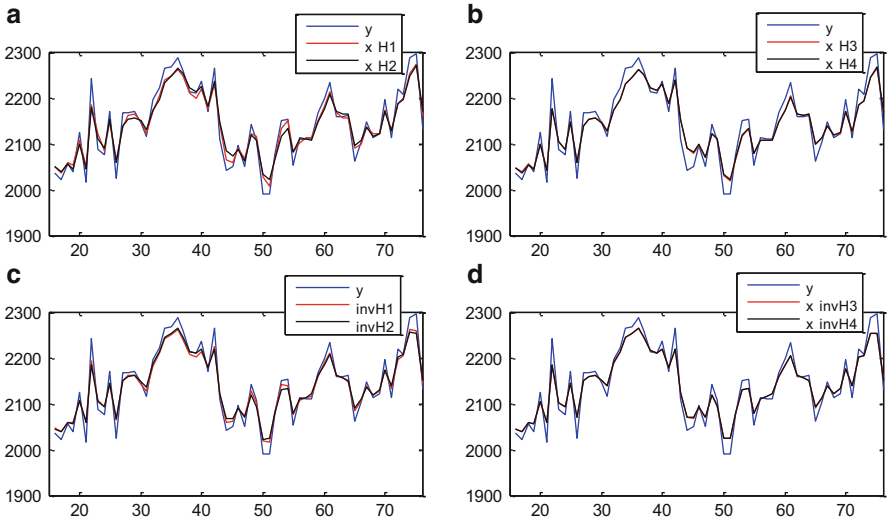


Fig. 3 Signal estimates obtained with the Haar wavelet; $h = 0.5$; see Fig. 1 for more details

method of boundary wavelet and scaling coefficients extrapolation that is already at the stage of signal estimation. In what follows we use a two-step procedure where in the first step we construct AR(I)MA models for the original data and use them to compute forecasts in a horizon necessary for a signal estimation with the inverse non-decimated wavelet transform, while in the second we apply our 'inverse' estimator $\tilde{\mathbf{x}}$ and build models to forecast the signal. Because we expect that the

suggested approach can reduce forecast error variance in relatively small samples, short Daubechies wavelets seem to be the most promising devices in order to reach this aim. In what follows, we refer to the suggested methods of signal estimation and forecasting as wavelet smoothing because they resemble exponential smoothing but use a specific weighting scheme depending, additionally, on the spectral content of the processes under scrutiny.

In the following two sections we concentrate on examining the predictive abilities of the procedures proposed here with some simulated and real data.

5 A Simulation Study

The simulation study aimed to examine if wavelet smoothing can help to reduce forecast error variance in the case of short-term predictions based on relatively small samples. Below we report our results obtained for samples of length $N = 50$, although a portion of all of the data-generating processes (DGPs) considered here was also examined for $N = 35$ and 100, and we shortly comment on these simulations as well. Forecasts up to five steps ahead were computed, i.e., series of length $N + 5$ were generated and the last five observations were used in the forecast evaluation. All computations were performed in Matlab R2008b with Optimization Toolbox Version 4.1.[5] The DGPs were the following:

$$y_t = x_t + \eta_t,$$

$$\eta_t \sim n.i.i.d.\left(0, \sigma^2\right),$$

where the signal x_t was defined according to:

(a) $$x_t = 0.9x_{t-1} + \xi_t,$$

(b) $$x_t = x_{t-1} - 0.6x_{t-2} + 0.4x_{t-3} + \xi_t,$$

and $\xi_t \sim n.i.i.d.(0, 1)$. The parameter σ^2 took on two values: 1 and 4. Also, we experimented with other parameter values (e.g., with 0.9 replaced by 0.6 in the DGP (a)) as well as certain nonstationary specifications, especially the following signal

[5]The Matlab codes used in the simulation and empirical studies are available upon request.

models: $\Delta x_t = 1.3\ \Delta x_{t-1} - 0.5\ \Delta x_{t-2} + \xi_t$ and $\Delta x_t = \Delta x_{t-1} - 0.5\ \Delta x_{t-2} + \xi_t$, where $\xi_t \sim n.i.i.d.\ (0, 1)$). Each time all initial observations were set to 0.

The series y_t was forecasted up to five steps ahead with the following methods:

(I) Simple exponential smoothing with a given/estimated value of the smoothing parameter α and the starting value equal to the first observation in the sample;
(II) ARMA(p, q) models with (p, q) chosen according to the AIC (BIC) criterion and maximum orders (1, 1) in the case of the signal (a) and (3, 3) for the second DGP, estimated by the maximum likelihood method with the Matlab function `armax` under the default settings[6];
(III) ARIMA(p, 1, q) models, identified and estimated as in (II);
(IV) The naïve (no-change) method
and the procedures suggested here:
(V) The 'non-inverse' method with $h = 1$ (0.75) and the maximum level of decomposition from 1 to 4, combined with (II)–(IV);
(VI) The 'inverse' method based on the Haar wavelet with $h = 1$ (0.75) and the maximum level of decomposition from 1 to 4, where the same forecasting methods were used in both steps, (II) and (III), respectively, or an ARIMA model in the first step was combined with the no-change method (IV) in the second;
(VII) The 'inverse' method based on the Daubechies $d4$ wavelet with $h = 1$ (0.75) and the maximum level of decomposition 1 and 2, combined with (II)–(IV) as in (VI).

Two methods of scaling coefficient treatment were considered: leaving them unchanged and scaling according to the formula in Eq. (17), used under the assumption of an estimated constant mean value. Since the latter approach produced better outcomes on average, this one is exclusively reported here. Also, we experimented with wavelet smoothing under $h = 0.5$, but the results were usually less satisfying than those presented here. The MODWT-based estimators were used in the estimation of the wavelet variance and the variance of scaling coefficients, and they were computed after removing all of the boundary values. This means, in particular, that estimation of the wavelet variance was unbiased. The ARMA models in (II) as well as the ARMA models for the original data and the estimated signals in (V)–(VII) were all estimated assuming a constant mean value, while the appropriate ARIMA(p, 1, q) models—assuming no deterministic terms.

Tables 1, 2, 3, and 4 present the ratios of the forecast MSEs to the benchmark forecast MSE, defined as the best result among those obtained with the benchmark (I)–(IV) methods, computed on the basis of 5,000 and 2,000 replications for signal (a) and (b), respectively. Exponential smoothing considered among the benchmark models was the one with an estimated α. The case with a constant α (set equal to 0.5 or 0.25) is presented in the notes to Tables 1, 2, 3, and 4. To save space, each time

[6]The default settings mean, in particular, automatic choices of a search method and treatment of initial conditions.

Table 1 Simulation results for the signal (a); ratios of forecast MSEs for forecasts from 1 to 5 steps ahead; $\sigma^2 = 1$

Forecasting methods			(A1)	(B1)	(C1)	(A2)	(B2)	(C2)	(A3)	(B3)
Horizon H	$H=1$	$h=0.75$	**0.954**	1.029	**0.972**	**0.961**	1.000	**0.969**	**0.971**	1.011
		$h=1$	**0.991**	1.057	1.000	1.020	1.016	**0.995**	1.042	1.031
	$H=2$	$h=0.75$	**0.943**	1.024	**0.955**	**0.951**	**0.984**	**0.956**	**0.959**	**0.997**
		$h=1$	**0.961**	1.071	**0.966**	**0.987**	**0.996**	**0.969**	1.004	1.007
	$H=3$	$h=0.75$	**0.940**	**0.997**	**0.950**	**0.950**	**0.965**	**0.954**	**0.955**	**0.975**
		$h=1$	**0.944**	1.050	**0.946**	**0.969**	**0.974**	**0.958**	**0.981**	**0.983**
	$H=4$	$h=0.75$	**0.937**	**0.982**	**0.944**	**0.947**	**0.951**	**0.951**	**0.952**	**0.960**
		$h=1$	**0.931**	1.026	**0.939**	**0.955**	**0.958**	**0.952**	**0.965**	**0.969**
	$H=5$	$h=0.75$	**0.982**	**0.982**	**0.987**	**0.990**	**0.966**	**0.993**	**0.992**	**0.976**
		$h=1$	**0.968**	1.022	**0.978**	**0.989**	**0.971**	**0.992**	**0.995**	**0.980**

Forecasting methods			(C3)	(A4)	(B4)	(C4)	(D1)	(E1)	(F1)	(D2)
Horizon H	$H=1$	$h=0.75$	**0.982**	**0.974**	1.023	**0.987**	**0.962**	1.059	**0.993**	**0.959**
		$h=1$	1.012	1.047	1.038	1.022	**0.974**	1.065	1.011	**0.972**
	$H=2$	$h=0.75$	**0.965**	**0.960**	1.010	**0.965**	**0.947**	1.024	**0.970**	**0.952**
		$h=1$	**0.982**	1.006	1.018	**0.988**	**0.945**	1.016	**0.986**	**0.954**
	$H=3$	$h=0.75$	**0.961**	**0.954**	**0.986**	**0.960**	**0.943**	**0.991**	**0.962**	**0.953**
		$h=1$	**0.966**	**0.981**	**0.989**	**0.969**	**0.932**	**0.981**	**0.971**	**0.947**
	$H=4$	$h=0.75$	**0.957**	**0.950**	**0.977**	**0.956**	**0.943**	**0.974**	**0.961**	**0.955**
		$h=1$	**0.963**	**0.964**	**0.980**	**0.964**	**0.924**	**0.961**	**0.964**	**0.941**
	$H=5$	$h=0.75$	**0.996**	**0.990**	**0.989**	**0.994**	**0.986**	**0.974**	1.003	1.001
		$h=1$	**0.999**	**0.993**	**0.990**	1.000	**0.960**	**0.964**	**0.997**	**0.981**

Forecasting methods			(E2)	(F2)	(D3)	(E3)	(F3)	(D4)	(E4)	(F4)
Horizon H	$H=1$	$h=0.75$	1.015	**0.981**	**0.964**	1.014	**0.986**	**0.967**	1.028	**0.988**
		$h=1$	1.023	**0.993**	**0.981**	1.028	1.000	**0.994**	1.037	1.010
	$H=2$	$h=0.75$	**0.984**	**0.972**	**0.959**	**0.988**	**0.978**	**0.954**	1.000	**0.970**
		$h=1$	**0.980**	**0.990**	**0.965**	**0.984**	**0.998**	**0.967**	**0.992**	**0.996**
	$H=3$	$h=0.75$	**0.961**	**0.969**	**0.961**	**0.964**	**0.977**	**0.952**	**0.974**	**0.969**
		$h=1$	**0.952**	**0.991**	**0.958**	**0.956**	**0.999**	**0.955**	**0.960**	**0.992**
	$H=4$	$h=0.75$	**0.950**	**0.972**	**0.962**	**0.952**	**0.979**	**0.950**	**0.967**	**0.969**
		$h=1$	**0.937**	**0.994**	**0.952**	**0.941**	1.003	**0.944**	**0.947**	**0.990**
	$H=5$	$h=0.75$	**0.964**	1.019	1.009	**0.969**	1.024	**0.995**	**0.983**	1.011
		$h=1$	**0.955**	1.037	**0.992**	**0.960**	1.047	**0.982**	**0.965**	1.032

Forecasting methods			(G1)	(H1)	(I1)	(G2)	(H2)	(I2)
Horizon H	$H=1$	$h=0.75$	**0.963**	1.058	1.014	**0.962**	1.014	1.001
		$h=1$	**0.970**	1.050	1.025	**0.979**	1.013	1.002
	$H=2$	$h=0.75$	**0.952**	1.022	**0.990**	**0.958**	**0.984**	**0.999**
		$h=1$	**0.948**	1.001	1.024	**0.963**	**0.974**	1.020
	$H=3$	$h=0.75$	**0.950**	**0.991**	**0.981**	**0.961**	**0.958**	**0.996**
		$h=1$	**0.939**	**0.968**	1.014	**0.958**	**0.949**	1.034

(continued)

Table 1 (continued)

Forecasting methods			(G1)	(H1)	(I1)	(G2)	(H2)	(I2)
	$H=4$	$h=0.75$	**0.950**	**0.974**	**0.982**	**0.961**	**0.949**	1.000
		$h=1$	**0.932**	**0.950**	1.012	**0.951**	**0.937**	1.051
	$H=5$	$h=0.75$	**0.996**	**0.975**	1.025	1.010	**0.962**	1.049
		$h=1$	**0.972**	**0.956**	1.048	**0.994**	**0.958**	1.109

Note: The forecast models were selected according to the AIC criterion; for the benchmark methods (I)–(IV) the ratios of the forecast MSEs to the forecast MSE for the benchmark method are: for $H=1$—1, 1.041, 1.029, 1.153; for $H=2$—1, 1.045, 1.035, 1.130; for $H=3$—1, 1.020, 1.026, 1.114; for $H=4$—1.003, 1, 1.034, 1.119; and for $H=5$—1.051, 1, 1.077, 1.174, respectively, i.e., the best results among the benchmark (I)–(IV) methods were obtained with exponential smoothing (I) for shorter-term predictions, while stationary ARMA models produced the best forecasts 4 and 5 steps ahead; the appropriate relative MSEs for forecasts from 1 to 5 steps ahead with exponential smoothing with a constant α equal to 0.5 (0.25) are, respectively: 0.972 (1.124), 0.981 (1.061), 0.982 (1.017), 0.984 (0.993), 1.036 (1.015); (A1), (A2), (A3), (A4)—(V) used in combination with the naïve method (IV); (B1), (B2), (B3), (B4)—(V) used in combination with (II); (C1), (C2), (C3), (C4)—(V) used in combination with (III); (D1), (D2), (D3), (D4)—(VI) used in combination with the naïve method (IV); (E1), (E2), (E3), (E4)—(VI) used in combination with (II); (F1), (F2), (F3), (F4)—(VI) used in combination with (III); (G1), (G2)—(VII) used in combination with the naïve method (IV); (H1), (H2)—(VII) used in combination with (II); (I1), (I2)—(VII) used in combination with (III); numbers in *brackets* in the heading row denote maximum decomposition levels; all results better than the benchmark are in *bold*; the best results for each horizon are *underlined*

Table 2 Simulation results for the signal (a); ratios of forecast MSEs for forecasts from 1 to 5 steps ahead; $\sigma^2=4$

Forecasting methods			(A1)	(B1)	(C1)	(A2)	(B2)	(C2)	(A3)	(B3)
Horizon H	$H=1$	$h=0.75$	**0.993**	1.031	1.007	**0.967**	**0.992**	**0.981**	**0.969**	**0.995**
		$h=1$	1.011	1.069	1.024	**0.989**	1.002	**0.990**	**0.999**	1.008
	$H=2$	$h=0.75$	**0.982**	1.047	**0.991**	**0.964**	**0.996**	**0.968**	**0.966**	1.000
		$h=1$	**0.987**	1.096	**0.994**	**0.974**	1.001	**0.976**	**0.984**	1.004
	$H=3$	$h=0.75$	**0.972**	1.016	**0.980**	**0.964**	**0.979**	**0.970**	**0.968**	**0.988**
		$h=1$	**0.963**	1.061	**0.975**	**0.964**	**0.987**	**0.966**	**0.975**	**0.994**
	$H=4$	$h=0.75$	**0.965**	**0.997**	**0.969**	**0.959**	**0.966**	**0.961**	**0.961**	**0.977**
		$h=1$	**0.949**	1.031	**0.961**	**0.951**	**0.967**	**0.953**	**0.959**	**0.978**
	$H=5$	$h=0.75$	**0.977**	**0.996**	**0.982**	**0.976**	**0.973**	**0.975**	**0.979**	**0.982**
		$h=1$	**0.953**	1.024	**0.965**	**0.962**	**0.973**	**0.964**	**0.972**	**0.983**

Forecasting methods			(C3)	(A4)	(B4)	(C4)	(D1)	(E1)	(F1)	(D2)
Horizon H	$H=1$	$h=0.75$	**0.986**	**0.971**	1.001	**0.986**	**0.999**	1.045	1.011	**0.982**
		$h=1$	**0.999**	1.006	1.014	1.007	1.005	1.061	1.020	**0.983**
	$H=2$	$h=0.75$	**0.973**	**0.969**	1.008	**0.974**	**0.988**	1.042	**0.993**	**0.977**
		$h=1$	**0.985**	**0.992**	1.009	**0.996**	**0.981**	1.043	**0.997**	**0.970**
	$H=3$	$h=0.75$	**0.971**	**0.970**	**0.995**	**0.974**	**0.976**	1.010	**0.985**	**0.975**
		$h=1$	**0.978**	**0.981**	**0.997**	**0.987**	**0.958**	1.008	**0.972**	**0.958**

(continued)

Table 2 (continued)

Forecasting methods			(C3)	(A4)	(B4)	(C4)	(D1)	(E1)	(F1)	(D2)
	$H=4$	$h=0.75$	**0.963**	**0.960**	**0.983**	**0.963**	**0.969**	**0.991**	**0.974**	**0.968**
		$h=1$	**0.963**	**0.962**	**0.981**	**0.968**	**0.944**	**0.986**	**0.956**	**0.945**
	$H=5$	$h=0.75$	**0.981**	**0.977**	**0.988**	**0.979**	**0.982**	**0.990**	**0.986**	**0.987**
		$h=1$	**0.976**	**0.971**	**0.987**	**0.981**	**0.949**	**0.985**	**0.963**	**0.958**

Forecasting methods			(E2)	(F2)	(D3)	(E3)	(F3)	(D4)	(E4)	(F4)
Horizon H	$H=1$	$h=0.75$	1.005	**0.990**	**0.982**	1.004	**0.989**	**0.990**	1.011	**0.998**
		$h=1$	1.023	**0.993**	**0.986**	1.022	**0.996**	1.010	1.032	1.020
	$H=2$	$h=0.75$	**0.998**	**0.980**	**0.980**	**0.998**	**0.984**	**0.982**	1.004	**0.984**
		$h=1$	1.001	**0.985**	**0.976**	1.000	**0.990**	**0.991**	1.007	1.005
	$H=3$	$h=0.75$	**0.980**	**0.980**	**0.981**	**0.982**	**0.984**	**0.979**	**0.987**	**0.981**
		$h=1$	**0.977**	**0.976**	**0.968**	**0.977**	**0.986**	**0.977**	**0.982**	**0.994**
	$H=4$	$h=0.75$	**0.968**	**0.972**	**0.973**	**0.971**	**0.976**	**0.970**	**0.975**	**0.972**
		$h=1$	**0.959**	**0.960**	**0.953**	**0.960**	**0.971**	**0.960**	**0.962**	**0.977**
	$H=5$	$h=0.75$	**0.971**	**0.989**	**0.995**	**0.976**	**0.999**	**0.988**	**0.976**	**0.990**
		$h=1$	**0.962**	**0.980**	**0.971**	**0.964**	**0.994**	**0.973**	**0.963**	**0.995**

Forecasting methods			(G1)	(H1)	(I1)	(G2)	(H2)	(I2)
Horizon H	$H=1$	$h=0.75$	1.005	1.051	1.020	**0.986**	1.009	1.000
		$h=1$	1.009	1.060	1.031	**0.993**	1.026	1.007
	$H=2$	$h=0.75$	**0.994**	1.040	1.008	**0.983**	**0.995**	**0.994**
		$h=1$	**0.988**	1.036	1.016	**0.981**	1.002	1.006
	$H=3$	$h=0.75$	**0.982**	1.009	**0.995**	**0.980**	**0.975**	**0.991**
		$h=1$	**0.963**	**0.998**	**0.991**	**0.968**	**0.976**	1.004
	$H=4$	$h=0.75$	**0.975**	**0.989**	**0.991**	**0.973**	**0.964**	**0.982**
		$h=1$	**0.950**	**0.980**	**0.980**	**0.954**	**0.959**	**0.995**
	$H=5$	$h=0.75$	**0.988**	**0.990**	1.000	**0.993**	**0.965**	1.001
		$h=1$	**0.955**	**0.978**	**0.984**	**0.969**	**0.959**	1.021

Note: The forecast models were selected according to the AIC criterion; for the benchmark methods (I)–(IV) the ratios of the forecast MSEs to the forecast MSE for the benchmark method are: for $H=1$—1, 1.064, 1.014, 1.390; for $H=2$—1, 1.051, 1.011, 1.371; for $H=3$—1, 1.028, 1.017, 1.361; for $H=4$—1, 1.006, 1.004, 1.343; and for $H=5$—1.024, 1, 1.039, 1.377, respectively, i.e., the best results among the benchmark (I)–(IV) methods were obtained with exponential smoothing (I), except for forecasts five steps ahead, in which case ARMA models produced better outcomes; the appropriate relative MSEs for forecasts from 1 to 5 steps ahead with exponential smoothing with a constant α equal to 0.5 (0.25) are, respectively: 1.012 (0.972), 1.022 (0.972), 1.037 (0.976), 1.036 (0.967), 1.068 (0.994); see note to Table 1 for the description of forecast methods and other details

we report only the outcomes obtained with one of the two information criteria—those producing the smaller forecast MSE for the best benchmark method (or the subsequent best when the best was exponential smoothing or the naïve approach). The other results, also for our experimentations with other data lengths (as well as parameter values), are available upon request.

The most important finding from the simulations is that wavelet smoothing is able to produce lower forecast MSEs of both short- and longer-term predictions than

Table 3 Simulation results for the signal (b); ratios of forecast MSEs for forecasts from 1 to 5 steps ahead; $\sigma^2 = 1$

Forecasting methods			(A1)	(B1)	(C1)	(A2)	(B2)	(C2)	(A3)	(B3)
Horizon H	$H=1$	$h=0.75$	**0.962**	**0.978**	**0.945**	**0.939**	**0.953**	**0.938**	**0.941**	**0.965**
		$h=1$	**0.997**	**0.965**	**0.942**	**0.978**	**0.951**	**0.944**	**0.984**	**0.950**
	$H=2$	$h=0.75$	1.023	**0.972**	**0.973**	**0.996**	**0.960**	**0.980**	1.000	**0.974**
		$h=1$	**0.967**	**0.968**	**0.942**	**0.940**	**0.946**	**0.958**	**0.949**	**0.953**
	$H=3$	$h=0.75$	1.011	**0.995**	**0.992**	**0.997**	**0.976**	**0.994**	1.000	**0.990**
		$h=1$	**0.962**	**0.998**	**0.967**	**0.951**	**0.978**	**0.983**	**0.958**	**0.983**
	$H=4$	$h=0.75$	**0.967**	**0.989**	**0.963**	**0.963**	**0.969**	**0.971**	**0.963**	**0.979**
		$h=1$	**0.937**	1.003	**0.957**	**0.938**	**0.984**	**0.971**	**0.942**	**0.981**
	$H=5$	$h=0.75$	1.073	1.002	1.021	1.062	**0.988**	1.047	1.059	**0.995**
		$h=1$	1.039	1.012	1.002	1.032	1.009	1.030	1.032	1.006

Forecasting methods			(C3)	(A4)	(B4)	(C4)	(D1)	(E1)	(F1)	(D2)
Horizon H	$H=1$	$h=0.75$	**0.936**	**0.942**	**0.974**	**0.941**	**0.939**	**0.996**	**0.959**	**0.936**
		$h=1$	**0.947**	**0.989**	**0.957**	**0.956**	**0.948**	**0.997**	**0.962**	**0.946**
	$H=2$	$h=0.75$	**0.976**	**0.999**	**0.994**	**0.994**	1.032	**0.988**	1.001	1.025
		$h=1$	**0.966**	**0.950**	**0.957**	**0.976**	**0.964**	**0.982**	**0.964**	**0.958**
	$H=3$	$h=0.75$	1.008	**0.997**	**0.987**	**0.994**	1.023	**0.995**	1.007	1.023
		$h=1$	**0.984**	**0.956**	**0.986**	1.000	**0.964**	**0.999**	**0.978**	**0.966**
	$H=4$	$h=0.75$	**0.978**	**0.960**	**0.996**	**0.959**	**0.974**	**0.987**	**0.979**	**0.978**
		$h=1$	**0.963**	**0.940**	**0.995**	**0.981**	**0.932**	1.001	**0.960**	**0.940**
	$H=5$	$h=0.75$	1.049	1.055	1.017	1.031	1.060	**0.999**	1.044	1.066
		$h=1$	1.022	1.028	1.022	1.036	1.007	1.008	**0.995**	1.018

Forecasting methods			(E2)	(F2)	(D3)	(E3)	(F3)	(D4)	(E4)	(F4)
Horizon H	$H=1$	$h=0.75$	**0.977**	**0.954**	**0.938**	**0.977**	**0.961**	**0.935**	**0.982**	**0.959**
		$h=1$	**0.978**	**0.962**	**0.952**	**0.979**	**0.969**	**0.961**	**0.986**	**0.975**
	$H=2$	$h=0.75$	**0.985**	1.021	1.031	**0.990**	1.027	1.018	1.000	1.017
		$h=1$	**0.971**	**0.991**	**0.967**	**0.976**	**0.995**	**0.962**	**0.973**	**0.989**
	$H=3$	$h=0.75$	**0.982**	1.026	1.028	**0.990**	1.040	1.015	**0.999**	1.023
		$h=1$	**0.994**	**0.996**	**0.975**	1.001	1.008	**0.968**	**0.993**	1.001
	$H=4$	$h=0.75$	**0.979**	**0.989**	**0.982**	**0.981**	1.007	**0.969**	**0.988**	**0.977**
		$h=1$	**0.997**	**0.980**	**0.947**	**0.992**	**0.987**	**0.940**	**0.996**	**0.980**
	$H=5$	$h=0.75$	**0.997**	1.062	1.071	**0.997**	1.083	1.054	1.009	1.053
		$h=1$	1.011	1.040	1.026	1.002	1.044	1.016	1.009	1.034

Forecasting methods			(G1)	(H1)	(I1)	(G2)	(H2)	(I2)
Horizon H	$H=1$	$h=0.75$	**0.952**	1.008	**0.974**	**0.946**	**0.984**	**0.972**
		$h=1$	**0.957**	**0.995**	**0.960**	**0.957**	**0.985**	**0.972**
	$H=2$	$h=0.75$	1.049	1.012	1.011	1.039	**0.999**	1.038
		$h=1$	**0.981**	1.005	**0.964**	**0.973**	**0.995**	**0.985**
	$H=3$	$h=0.75$	1.038	1.005	1.009	1.036	**0.993**	1.035
		$h=1$	**0.978**	1.049	**0.999**	**0.980**	1.037	1.003

(continued)

Table 3 (continued)

Forecasting methods			(G1)	(H1)	(I1)	(G2)	(H2)	(I2)
	$H=4$	$h=0.75$	**0.987**	1.000	**0.993**	**0.989**	**0.990**	1.007
		$h=1$	**0.944**	1.029	**0.962**	**0.952**	1.058	**0.981**
	$H=5$	$h=0.75$	1.081	1.004	1.047	1.082	**0.997**	1.076
		$h=1$	1.029	1.031	1.023	1.037	1.080	1.039

Note: The forecast models were selected according to the BIC criterion; for the benchmark methods (I)–(IV) the ratios of the forecast MSEs to the forecast MSE for the benchmark method are: for $H=1$—1, 1.012, 1.028, 1.217; for $H=2$—1.025, 1, 1.071, 1.475; for $H=3$—1.034, 1, 1.062, 1.420; for $H=4$—1.016, 1, 1.049, 1.317; and for $H=5$—1.112, 1, 1.126, 1.441, respectively, i.e., the best results among the benchmark (I)–(IV) methods were obtained with the ARMA models (II), except for horizon 1, in which case exponential smoothing produced better outcomes; the appropriate relative MSEs for forecasts from 1 to 5 steps ahead with exponential smoothing with a constant α equal to 0.5 (0.25) are, respectively: 0.991 (0.982), 1.064 (0.959), 1.063 (0.978), 1.032 (0.974), 1.163 (1.055); see note to Table 1 for the description of forecast methods and other details

Table 4 Simulation results for the signal (b); ratios of forecast MSEs for forecasts from 1 to 5 steps ahead; $\sigma^2=4$

Forecasting methods			(A1)	(B1)	(C1)	(A2)	(B2)	(C2)	(A3)	(B3)
Horizon H	$H=1$	$h=0.75$	**0.980**	**0.967**	**0.987**	**0.977**	**0.965**	**0.987**	**0.980**	**0.973**
		$h=1$	**0.977**	**0.967**	**0.962**	**0.983**	**0.970**	**0.979**	**0.993**	**0.978**
	$H=2$	$h=0.75$	1.031	**0.981**	1.004	1.026	**0.975**	**0.990**	1.027	**0.983**
		$h=1$	**0.985**	**0.980**	**0.966**	**0.987**	**0.974**	**0.982**	**0.994**	**0.969**
	$H=3$	$h=0.75$	1.022	**0.974**	**0.972**	1.022	**0.963**	**0.973**	1.031	**0.969**
		$h=1$	**0.969**	**0.983**	**0.955**	**0.978**	**0.977**	**0.967**	**0.996**	**0.988**
	$H=4$	$h=0.75$	1.013	**0.988**	**0.992**	1.007	**0.981**	**0.985**	1.009	**0.988**
		$h=1$	**0.976**	**0.995**	**0.970**	**0.976**	**0.989**	**0.973**	**0.985**	**0.989**
	$H=5$	$h=0.75$	1.056	**0.995**	1.018	1.052	**0.986**	1.015	1.059	**0.993**
		$h=1$	1.004	**0.996**	**0.983**	1.009	**0.998**	1.004	1.023	**0.999**

Forecasting methods			(C3)	(A4)	(B4)	(C4)	(D1)	(E1)	(F1)	(D2)
Horizon H	$H=1$	$h=0.75$	**0.996**	**0.989**	**0.988**	1.008	**0.975**	**0.973**	**0.984**	**0.976**
		$h=1$	**0.987**	1.013	**0.994**	1.011	**0.962**	**0.978**	**0.965**	**0.969**
	$H=2$	$h=0.75$	1.007	1.032	**0.988**	1.028	1.030	**0.978**	**0.993**	1.032
		$h=1$	**0.985**	1.008	**0.983**	**0.999**	**0.977**	**0.981**	**0.972**	**0.984**
	$H=3$	$h=0.75$	**0.992**	1.039	**0.986**	1.009	1.025	**0.976**	**0.973**	1.026
		$h=1$	**0.987**	1.014	**0.998**	1.011	**0.967**	**0.984**	**0.953**	**0.972**
	$H=4$	$h=0.75$	**0.999**	1.012	**0.985**	**0.999**	1.014	**0.984**	**0.993**	1.012
		$h=1$	**0.986**	**0.996**	**0.990**	1.000	**0.970**	**0.992**	**0.967**	**0.972**
	$H=5$	$h=0.75$	1.031	1.068	**0.998**	1.057	1.055	**0.991**	1.020	1.058
		$h=1$	1.013	1.042	1.005	1.045	**0.996**	**0.995**	**0.983**	1.004

Forecasting methods			(E2)	(F2)	(D3)	(E3)	(F3)	(D4)	(E4)	(F4)
Horizon H	$H=1$	$h=0.75$	**0.968**	**0.995**	**0.980**	**0.979**	1.001	**0.985**	**0.992**	1.010
		$h=1$	**0.976**	**0.975**	**0.978**	**0.977**	**0.984**	1.000	**0.995**	1.007

(continued)

Table 4 (continued)

Forecasting methods			(E2)	(F2)	(D3)	(E3)	(F3)	(D4)	(E4)	(F4)
	$H=2$	$h=0.75$	**0.978**	1.004	1.032	**0.991**	1.013	1.033	**0.992**	1.025
		$h=1$	**0.978**	**0.987**	**0.988**	**0.976**	**0.990**	1.003	**0.984**	1.009
	$H=3$	$h=0.75$	**0.968**	**0.985**	1.029	**0.974**	**0.990**	1.032	**0.987**	1.005
		$h=1$	**0.974**	**0.968**	**0.981**	**0.979**	**0.968**	**0.999**	**0.991**	**0.996**
	$H=4$	$h=0.75$	**0.982**	**0.998**	1.012	**0.986**	1.005	1.014	**0.989**	**0.999**
		$h=1$	**0.989**	**0.979**	**0.977**	**0.993**	**0.982**	**0.992**	1.000	**0.994**
	$H=5$	$h=0.75$	**0.985**	1.025	1.062	**0.988**	1.036	1.064	**0.996**	1.044
		$h=1$	**0.992**	**0.997**	1.015	1.001	1.011	1.031	1.004	1.024

Forecasting methods			(G1)	(H1)	(I1)	(G2)	(H2)	(I2)
Horizon H	$H=1$	$h=0.75$	**0.984**	**0.984**	**0.996**	**0.988**	**0.978**	1.007
		$h=1$	**0.970**	**0.980**	**0.965**	**0.989**	**0.985**	**0.985**
	$H=2$	$h=0.75$	1.040	**0.988**	1.008	1.041	**0.986**	1.016
		$h=1$	**0.987**	**0.986**	**0.977**	1.000	**0.981**	**0.988**
	$H=3$	$h=0.75$	1.038	**0.975**	**0.984**	1.036	**0.970**	**0.983**
		$h=1$	**0.980**	**0.991**	**0.963**	**0.989**	1.005	**0.973**
	$H=4$	$h=0.75$	1.021	**0.988**	1.003	1.017	**0.978**	1.000
		$h=1$	**0.977**	**0.999**	**0.970**	**0.982**	1.017	**0.979**
	$H=5$	$h=0.75$	1.064	**0.989**	1.021	1.062	**0.985**	1.032
		$h=1$	1.005	1.002	**0.994**	1.014	1.024	1.008

Note: The forecast models were selected according to the BIC criterion; for the benchmark methods (I)–(IV) the ratios of the forecast MSEs to the forecast MSE for the benchmark method are: for $H=1$—1, 1.025, 1.038, 1.499; for $H=2$—1.023, 1, 1.046, 1.617; for $H=3$—1.009, 1, 1.011, 1.636; for $H=4$—1.008, 1, 1.006, 1.569; and for $H=5$—1.071, 1, 1.053, 1.655, respectively, i.e., the best results among the benchmark (I)–(IV) methods were obtained with the ARMA models (II), except for horizon 1, in which case exponential smoothing produced better outcomes; the appropriate relative MSEs for forecasts from 1 to 5 steps ahead with exponential smoothing with a constant α equal to 0.5 (0.25) are, respectively: 1.067 (0.971), 1.140 (1.009), 1.128 (1.005), 1.112 (0.995), 1.195 (1.065); see note to Table 1 for the description of forecast methods and other details

the AR(I)MA models. Applying the wavelet signal estimation before constructing AR(I)MA models or forecasting via the no-change method can lead to a ca. 5 % reduction of the forecast MSEs relative to the best conventional approach. On the other hand, when comparing the forecast MSEs of the ordinary ARMA (ARIMA) models with those for the ARMA (ARIMA) models estimated on smoothed data, we can find reductions even up to 8–10 %.

To assess if the predictive gains are statistically significant, chosen forecast outcomes were examined with the Diebold and Mariano (DM) test for equal predictive ability (see Diebold and Mariano 1995), assuming quadratic loss functions. For example, for our first simulation, whose outcomes are presented in Table 1, comparing the method denoted in the tables as (D1) under $h=0.75$ (which produced the following relative forecast MSEs for forecasts from 1 to 5 steps ahead: 0.962, 0.947, 0.943, 0.943, 0.986—see Table 1) with exponential smoothing with an estimated parameter α (and the relative forecast MSEs equal to 1 for forecasts

from 1 to 4 steps ahead and 1.051 for predictions five steps ahead), the following DM statistics (and the corresponding p-values) are obtained: 5.486 (0.000), 8.964 (0.000), 10.45 (0.000), 12.12 (0.000), and 13.37 (0.000). Comparing the same implementation scheme of wavelet smoothing with exponential smoothing with a constant α equal to 0.5 (with the relative forecasts MSEs, reported in the note to Table 1, equal to 0.972, 0.981, 0.982, 0.984, 1.034) gives the following test results: 1.412 (0.158), 5.310 (0.000), 6.830 (0.000), 7.743 (0.000), and 9.305 (0.000). This means that wavelet signal estimation (alone or combined with some parametric models) is able to produce significantly better results than the other smoothing methods that are used in forecasting as well as AR(I)MA models, which in the first half of our simulation exercises were generally outperformed at shorter horizons by exponential smoothing.

Depending on the DGP and the particular way of implementation of wavelet smoothing, the decrease in the forecast MSE can result from very good signal estimation with wavelets in terms of the MSE in small samples, and, in particular, from good signal estimation under the assumption of the smallest possible value of the variance of the signal's error term (see the simulation studies in Bruzda 2013a), better performance of information criteria in discovering the true structure of the signal, and in some cases also better small-sample properties of maximum likelihood estimators of parameters (for example, quite often a lower bias of the estimators of autoregressive parameters) or a combination of these phenomena. However, using wavelet smoothing in practice will usually require many arbitrary decisions or a sort of optimizing the settings as for the value of the smoothing constant h and the number of decomposition stages J as well as the other implementation details of the method (i.e., the choice of the wavelet signal estimator, the kind of parametric model, the method of scaling coefficient treatment and the wavelet filter).

Other conclusions from the experiments can be summarized as follows:

- In the case of shorter-term forecasts (usually 1 and 2 steps ahead) the simple 'non-inverse' approach is able to outperform the 'inverse' method. On the other hand, the 'inverse' estimator produces better forecasts at longer horizons. Moreover, this approach slightly beats the 'non-inverse' method in samples of length 100.
- The optimal value of the smoothing constant h depends *inter alia* on the prediction horizon. When forecasting 1–2 steps ahead, better outcomes are obtained under $h = 0.75$. Longer prediction horizons are often associated with higher optimal values of h.
- In our experiments we do not observe systematic predictive gains from using the longer wavelet filter than the Haar, i.e., the Daubechies $d4$ filter. However, it is worth adding that the Haar filter embeds just one difference operator and, as such, it may not always be well suited to the case of nonstationary signals, also if it is used in the 'inverse' method. Its application is limited not only due to the high dynamic ranges (i.e., the ratios of the maximum and minimum values of the spectral densities) of the conventional within-scale Haar wavelet coefficients, but also because of the high between-scale correlations. Also, the other Daubechies

filters, such as $d4$ and $la8$, give then worse results as compared to the stationary case. However, as was noted by Craigmile and Percival (2005), in the class of processes with stationary backward differences the between-scale correlation of wavelet coefficients diminishes as the width of the wavelet filter increases. Nevertheless, the within-scale correlations remain substantial (see the discussion in Bruzda 2013b, Chap. II).

- Most often the lowest (1–2) maximum decomposition levels J produce the best outcomes. This results from a decreasing number of non-boundary wavelet coefficients available at higher decomposition stages and, at the same time, an increasing number of parameters to be estimated. The first part of this remark suggests that backcasting applied prior to wavelet signal estimation may further improve the relative performance of our methods. Our experimentation with nonstationary DGPs points out that, in this case, often higher decomposition stages should be considered.
- The relative forecast MSEs on average clearly increase (and the relative performance of wavelet smoothing on average deteriorates) with the length of the series. However, the gains from wavelet smoothing are present even in samples of length 100. By contrast, for samples of length 35 the relative MSEs are often smaller than those for $N = 50$.
- For stationary processes, smoothing of scaling coefficients usually lowers the forecast MSEs. The problem with high dynamic ranges of the coefficients $\tilde{V}_{Jt}$ can, to some extent, be mitigated by applying higher maximum levels of decomposition.
- The predictive gains from our methods applied to nonstationary DGPs are often rather modest. Because of this we consider wavelet smoothing as an alternative to (and a generalization of) exponential smoothing in the case of forecasting mainly stationary processes, but not exclusively then.[7]

6 An Empirical Illustration

To practically verify the approach suggested here, wavelet smoothing in its two variants was applied to compute forecasts of 16 time series from the M3-IJF Competition database (see Makridakis and Hibon 2000). To simplify matters, we chose series N2868–N2883 of a length in the range 76–79 without a clear increasing or decreasing tendency (see Fig. 4). As previously, two approaches to handling the scaling coefficients were considered, and we report the (slightly) better results obtained via smoothing the coefficients. This remains in accordance with our previous findings, especially as a more careful examination with unit root tests points to the stationarity of about 12 of the series. Each series was forecasted up to

[7]For example, under structural instability short wavelet filters may provide a better tool for smoothing than other procedures.

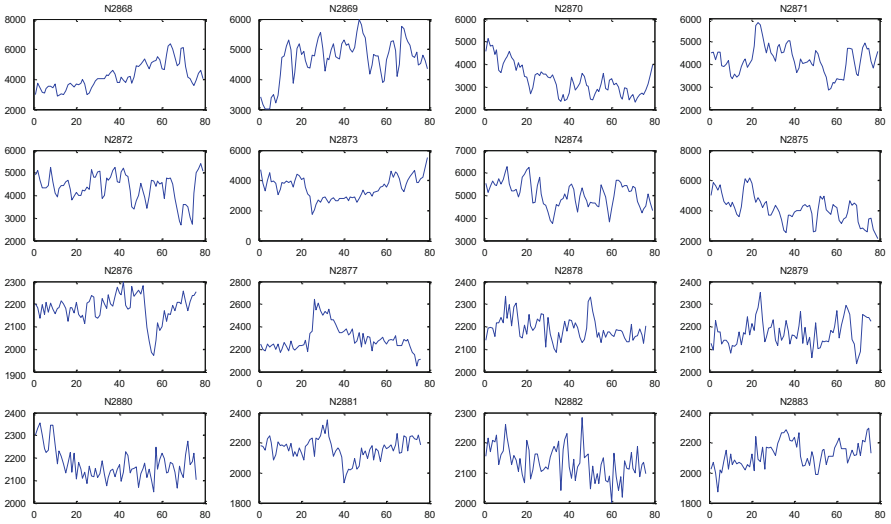


Fig. 4 Series N2868–N2883 from the M3-IJF competition database

five steps ahead, starting with forecasts computed on the basis of samples of length 52–55 and then increasing the samples by one observation, recomputing all of the wavelet quantities and re-estimating the parametric models. In this way, 20 forecasts for each horizon, method and time series were obtained. The forecasting methods were as in the simulation study, except for the fact that the smoothing constant h took on the values: 0.5, 0.75 and 1, whereas the maximum orders p and q for the ARMA(p, q) and ARIMA(p, 1, q) models were set to 4.[8] All the other settings (e.g., the numbers of decomposition levels and the information criteria used) were as in the simulation study, except that the function `armax` was run with the option `'InitialState = Backcast'`, which on average gave better forecasts from the ordinary AR(I)MA models.

Table 5 summarizes the results obtained with the BIC criterion and $h = 0.75$. It contains the ratios of the average MSPEs (mean squared percentage errors) to the average benchmark MSPE chosen as the best result among the four standard approaches, denoted in the table as 'RW' (the random walk—no-change model),

[8] In our earlier study based on a similar dataset (see Bruzda 2013b, Chap. VI), we exclusively considered purely autoregressive specifications for the original and smoothed data, thus reaching conclusions that differed from those here in some respects (in particular, the average predictive gains from wavelet smoothing clearly rose with the forecast horizon). Here we evaluate the suggested forecasting procedures in a different class of processes, utilizing a different estimation procedure, and we also deepen our analysis of the empirical results.

Table 5 Ratios of MSPEs; real data example; BIC criterion and $h = 0.75$

	(A1) (D1) (G1)	(B1) (E1) (H1)	(C1) (F1) (I1)	(A2) (D2) (G2)	(B2) (E2) (H2)	(C2) (F2) (I2)	(A3) (D3) ES1	(B3) (E3) ES2	(C3) (F3) RW	(A4) (D4) ARMA	(B4) (E4) ARIMA	(C4) (F4) ES
All (16)												
1-step	1.208	0.957	0.968	1.230	0.999	0.990	1.217	0.950	0.908	1.216	0.974	0.953
	1.035	0.991	0.977	1.044	0.968	0.977	1.043	0.962	0.931	1.037	0.960	0.934
	1.049	0.970	0.983	1.056	0.999	0.989	***1.276***	***1.571***	**1.017**	**1**	**1.044**	**1.015**
2-step	1.112	0.981	1.015	1.099	0.996	1.045	1.092	0.966	0.988	1.090	1.004	1.069
	1.058	1.005	1.025	1.056	0.977	1.033	1.055	0.966	0.999	1.043	1.008	0.992
	1.087	0.995	1.125	1.082	1.006	1.123	***1.080***	***1.054***	**1.147**	**1**	**1.096**	**1.145**
3-step	1.117	0.964	0.987	1.090	0.969	1.047	1.088	0.951	1.027	1.086	0.970	1.119
	1.102	0.995	0.978	1.096	0.964	1.012	1.096	0.942	1.036	1.083	1.032	1.028
	1.083	1.032	1.028	1.130	1.207	1.114	***1.057***	***0.961***	**1.225**	**1**	**1.068**	**1.225**
4-step	1.153	0.975	1.013	1.119	0.971	1.036	1.120	0.965	1.064	1.117	0.979	1.147
	1.158	1.002	0.966	1.150	0.979	1.035	1.150	0.952	1.083	1.137	1.063	1.078
	1.196	0.990	1.074	1.183	0.967	1.089	***1.086***	***0.975***	**1.311**	**1**	**1.057**	**1.305**
5-step	1.071	0.988	1.020	1.053	0.969	0.993	1.054	0.958	1.052	1.051	0.980	1.079
	1.099	1.017	0.985	1.094	0.982	1.027	1.095	0.957	1.084	1.083	1.048	1.063
	1.125	0.989	1.057	1.118	0.962	1.044	***1.040***	***0.968***	**1.247**	**1**	**1.027**	**1.238**
All stationary models												
1-step (6)	1.444	0.992	1.070	1.466	1.058	1.099	1.447	1.044	1.069	1.446	1.034	1.080
	1.219	1.004	1.013	1.228	1.006	1.032	1.224	0.976	1.014	1.216	1.001	1.014
	1.243	0.925	0.934	1.252	1.042	1.080	***1.509***	***1.868***	**1.183**	**1**	**1.137**	**1.191**

(continued)

Table 5 (continued)

	(A1) (D1) (G1)	(B1) (E1) (H1)	(C1) (F1) (I1)	(A2) (D2) (G2)	(B2) (E2) (H2)	(C2) (F2) (I2)	(A3) (D3) ES1	(B3) (E3) ES2	(C3) (F3) RW	(A4) (D4) ARMA	(B4) (E4) ARIMA	(C4) (F4) ES
2-step (8)	1.178	0.999	1.081	1.177	1.014	1.099	1.167	0.974	1.056	1.166	1.027	1.145
	1.111	1.031	1.062	1.118	1.012	1.092	1.117	0.975	1.061	1.100	1.046	1.036
	1.145	0.977	1.203	1.149	0.977	1.173	***1.164***	***1.127***	**1.190**	**1**	**1.150**	**1.191**
3-step (8)	1.203	0.944	1.060	1.168	0.977	1.153	1.168	0.988	1.120	1.166	1.013	1.211
	1.197	0.993	1.012	1.191	0.974	1.042	1.191	0.952	1.094	1.174	1.098	1.103
	1.242	0.970	1.096	1.229	1.036	1.124	***1.131***	***1.021***	**1.341**	**1**	**1.143**	**1.340**
4-step (10)	1.269	0.943	1.122	1.230	0.975	1.194	1.235	1.013	1.216	1.232	1.040	1.317
	1.296	1.014	1.072	1.289	1.007	1.110	1.291	0.996	1.236	1.270	1.169	1.262
	1.341	0.995	1.148	1.327	1.125	1.200	***1.203***	***1.083***	**1.486**	**1**	**1.210**	**1.481**
5-step (9)	1.339	0.898	1.203	1.327	0.943	1.229	1.336	0.993	1.325	1.333	0.961	1.347
	1.376	1.008	1.156	1.382	0.987	1.183	1.387	0.980	1.358	1.360	1.047	1.295
	1.420	0.940	1.172	1.419	0.976	1.103	***1.339***	***1.247***	**1.600**	**1**	**1.280**	**1.586**
First part (8)												
1-step	1.228	0.960	0.972	1.251	1.004	0.992	1.237	0.952	0.907	1.236	0.976	0.955
	1.046	0.995	0.983	1.055	0.971	0.983	1.054	0.965	0.933	1.047	0.963	0.937
	1.060	0.974	0.988	1.067	1.004	0.997	***1.298***	***1.606***	**1.015**	**1**	**1.053**	**1.018**
2-step	1.119	0.983	1.016	1.105	0.998	1.048	1.098	0.968	0.989	1.096	1.006	1.074
	1.063	1.006	1.027	1.061	0.979	1.037	1.059	0.966	1.001	1.047	1.009	0.995
	1.093	0.997	1.127	1.088	1.002	1.131	***1.083***	***1.056***	**1.147**	**1**	**1.099**	**1.148**
3-step	1.120	0.964	0.984	1.091	0.968	1.045	1.089	0.950	1.026	1.087	0.967	1.121
	1.104	0.995	0.976	1.099	0.963	1.011	1.098	0.940	1.035	1.085	1.030	1.028
	1.143	0.993	1.090	1.132	0.957	1.116	***1.055***	***0.957***	**1.226**	**1**	**1.067**	**1.227**

4-step	1.152	0.974	1.009	1.118	0.969	1.032	1.119	0.963	1.061	1.116	0.975	1.148
	1.158	1.002	0.960	1.149	0.977	1.031	1.149	0.948	1.081	1.136	1.059	1.077
	1.197	0.985	1.070	1.183	0.954	1.089	***1.082***	***0.970***	**1.307**	**1**	**1.053**	**1.307**
5-step	1.068	0.987	1.016	1.049	0.968	0.987	1.051	0.956	1.048	1.048	0.976	1.077
	1.096	1.017	0.980	1.091	0.980	1.022	1.092	0.953	1.081	1.080	1.042	1.060
	1.123	0.983	1.052	1.116	0.944	1.040	***1.035***	***0.964***	**1.239**	**1**	**1.023**	**1.237**
Second part (8)												
1-step	0.986	0.993	0.998	0.997	1.011	1.044	0.998	1.013	1.017	1.006	1.034	1.013
	0.963	1.022	0.976	0.964	1.002	0.985	0.969	1.005	0.988	0.971	1.001	0.973
	0.960	0.991	0.996	0.967	1.025	0.948	***1.030***	***1.124***	**1.152**	**1.098**	**1**	**1.067**
2-step	0.918	0.922	0.990	0.951	0.939	0.974	0.948	0.935	0.970	0.950	0.982	0.958
	0.923	0.974	0.976	0.937	0.935	0.951	0.941	0.954	0.946	0.938	1.004	0.935
	0.924	0.968	1.063	0.940	1.106	0.934	***0.998***	***1.009***	**1.140**	**1.005**	**1**	**1.051**
3-step	1.037	0.965	1.103	1.052	0.990	1.092	1.047	0.991	1.080	1.046	1.046	1.051
	1.012	1.001	1.052	1.029	0.999	1.041	1.033	1.004	1.053	1.027	1.080	1.017
	1.024	1.048	1.115	1.040	1.328	1.035	***1.120***	***1.094***	**1.195**	**1**	**1.097**	**1.169**
4-step	1.176	0.987	1.172	1.156	1.031	1.175	1.147	1.030	1.163	1.145	1.099	1.121
	1.158	1.023	1.164	1.161	1.050	1.165	1.163	1.086	1.168	1.152	1.173	1.124
	1.174	1.161	1.214	1.171	1.434	1.081	***1.229***	***1.121***	**1.445**	**1**	**1.192**	**1.245**
5-step	1.194	1.008	1.179	1.181	1.039	1.218	1.169	1.026	1.193	1.165	1.148	1.161
	1.197	1.013	1.176	1.197	1.068	1.211	1.199	1.097	1.201	1.188	1.271	1.175
	1.209	1.211	1.276	1.209	1.650	1.191	***1.250***	***1.124***	**1.537**	**1**	**1.207**	**1.273**

(continued)

Table 5 (continued)

	(A1) (D1) (G1)	(B1) (E1) (H1)	(C1) (F1) (I1)	(A2) (D2) (G2)	(B2) (E2) (H2)	(C2) (F2) (I2)	(A3) (D3) ES1	(B3) (E3) ES2	(C3) (F3) RW	(A4) (D4) ARMA	(B4) (E4) ARIMA	(C4) (F4) ES
Nonstationary models—first part												
1-step (3)	1.194	1.053	1.011	1.222	1.086	1.028	1.212	0.997	0.888	1.211	1.052	0.971
	1.027	1.111	1.078	1.040	1.064	1.064	1.040	1.078	0.983	1.034	1.052	0.991
	1.036	1.137	1.157	1.043	1.097	1.050	***1.280***	***1.582***	**1**	**1.124**	**1.113**	**1.004**
2-step (2)	0.996	0.987	0.875	0.943	1.007	0.953	0.944	1.004	0.841	0.940	0.992	0.912
	0.974	0.982	0.970	0.945	0.932	0.927	0.941	0.994	0.872	0.945	0.944	0.924
	0.995	1.118	0.953	0.959	1.142	1.052	***0.898***	***0.898***	**1.092**	**1.053**	**1**	**1.087**
3-step (2)	1.042	1.150	0.915	1.032	1.078	0.895	1.022	0.975	0.907	1.018	0.970	1.022
	0.997	1.136	1.014	0.993	1.065	1.074	0.990	1.036	1.018	0.986	0.978	0.965
	1.022	1.200	1.231	1.017	0.874	1.242	***1.000***	***0.917***	**1.082**	**1.133**	**1.000**	**1.080**
4-step (2)	1.228	1.373	1.058	1.197	1.272	0.958	1.188	1.149	1.008	1.183	1.122	1.078
	1.179	1.294	0.999	1.163	1.221	1.176	1.159	1.136	1.031	1.159	1.109	0.946
	1.210	1.299	1.234	1.196	0.843	1.149	***1.136***	***1.011***	**1.280**	**1.321**	**1.000**	**1.280**
5-step (3)	1.041	1.324	1.066	1.011	1.231	0.976	1.004	1.148	1.008	1.001	1.229	1.051
	1.066	1.271	1.037	1.048	1.213	1.108	1.045	1.157	1.048	1.044	1.287	1.068
	1.080	1.281	1.188	1.065	1.163	1.229	***0.969***	***0.899***	**1.164**	**1.240**	**1.000**	**1.163**
Nonstationary models—second part												
1-step (7)	0.998	1.016	1.004	0.991	1.010	1.048	0.988	1.009	1.010	0.999	1.034	1.006
	0.979	1.052	1.004	0.970	1.018	0.995	0.971	1.019	0.994	0.977	1.014	0.980
	0.976	1.015	1.010	0.969	1.041	0.958	***1.009***	***1.089***	**1.188**	**1.142**	**1**	**1.030**
2-step (6)	0.880	0.988	0.998	0.903	0.962	0.938	0.910	0.971	0.963	0.918	1.060	0.954
	0.904	1.069	1.043	0.912	0.984	0.925	0.916	1.015	0.952	0.923	1.104	0.944
	0.903	1.011	1.109	0.910	1.217	0.996	***0.932***	***1.001***	**1.084**	**1.120**	**1.035**	**1**

3-step (6)	0.923	0.997	1.022	0.933	0.981	0.975	0.939	0.989	1.007	0.942	1.089	0.961
	0.877	1.013	0.961	0.886	0.960	0.894	0.890	0.979	0.940	0.896	1.097	0.894
	0.885	0.984	1.003	0.891	1.371	1.034	***0.977***	***1.041***	**1**	**1.058**	**1.066**	**1.036**
4-step (4)	1.069	1.198	1.078	0.998	1.174	1.065	0.997	1.188	1.124	1.008	1.426	1.040
	1.082	1.220	1.058	1.042	1.228	1.063	1.041	1.353	1.142	1.059	1.640	1.051
	1.082	1.133	1.084	1.041	1.616	1.168	***0.998***	***0.959***	**1.360**	**1.275**	**1.084**	**1**
5-step (4)	0.955	1.191	1.057	0.926	1.156	1.037	0.932	1.173	1.107	0.938	1.516	1.024
	0.997	1.208	1.054	0.970	1.228	1.002	0.974	1.325	1.086	0.990	1.825	1.046
	0.990	1.140	1.135	0.967	1.843	1.253	***0.908***	***0.924***	**1.193**	**1.232**	**1.046**	**1**

Note: Forecast methods are as defined below Table 1, except for RW—the naïve method; ARMA—ARMA(p, q) models; ARIMA—ARIMA (p.1, q) models; ES—exponential smoothing with an estimated smoothing constant; ES1—exponential smoothing with $\alpha = 0.5$, ES2—exponential smoothing with $\alpha = 0.25$ (the case of $\alpha = 0.75$ was excluded from the presentation—generally, it produced worse results than the other values of α); all results that are better than the benchmark are *underlined*; the four benchmark outcomes are in *bold*, and ES1 and ES2 are presented in *bold italics*; the numbers of the series are given in *brackets*

'ARMA', 'ARIMA' and 'ES' (exponential smoothing with an estimated α). The MSPE for a single series was computed as:

$$MSPE = \frac{\sum_{t=1}^{20} (y_t - y_{tp})^2}{\sum_{t=1}^{20} y_t^2},$$

i.e., as Theil's U coefficient.[9] The smoothing constant h set equal to 0.5 or 1 resulted in a smaller number of cases when wavelet smoothing beat the other methods, although, especially if h was set to 1, the predictive gains from wavelet smoothing were sometimes higher (see the results collated in Table 6). Neither of the two information criteria uniformly outperformed the other in choosing better forecast models, although a proportion of the targeted indications (to models producing lower MSPEs among the best benchmark ARMA/ARIMA models) somewhat favored the BIC criterion (with the proportion being about 9:7). Because of this, the outcomes obtained with the BIC criterion are exclusively presented here. We note in passing that the AIC criterion gave slightly more arguments in favor of wavelet smoothing, thus leading more often to predictive gains from the wavelet approach.[10]

Because a more careful examination pointed out that the individual MSPEs in the first half of our data are much higher than those in the second half, the empirical results are also presented separately for these two groups of series. In fact, the aggregated outcomes for all 16 series closely resemble those for the first eight series only (see Tables 5 and 6). It is also worth mentioning that the series in these two groups seem to have different dynamic properties. In particular, the majority (5 out of 8) of the series in the second dataset can be identified as ARMA(1.1) with a negative MA parameter, while those in the first group are either purely autoregressive, sometimes exhibiting a sort of periodicity (i.e., certain significant higher-order autoregressive terms), or have a more complicated structure than the simple 'AR + noise'. Also, we note that in the second dataset there are series with evident level shifts. Besides, to have some further insight into possible gains from wavelet smoothing, the outcomes are also presented separately for the series for which the best benchmark forecasts were obtained with ARMA models and for all the others divided into those from the first and second half of the series. The subgroups are denoted as: 'All stationary models', 'Nonstationary models—first part' and 'Nonstationary models—second part', respectively.

[9]The change of evaluation criterion from MSE to MSPE was dictated by the aggregation of the forecasting results.

[10]Detailed results are available upon request.

Table 6 Ratios of MSPEs; real data example; BIC criterion and $h = 1$

	(A1)	(B1)	(C1)	(A2)	(B2)	(C2)	(A3)	(B3)	(C3)	(A4)	(B4)	(C4)
	(D1)	(E1)	(F1)	(D2)	(E2)	(F2)	(D3)	(E3)	(F3)	(D4)	(E4)	(F4)
	(G1)	(H1)	(I1)	(G2)	(H2)	(I2)	ES1	ES2	RW	ARMA	ARIMA	ES
All (16)												
1-step	1.347	0.955	0.967	1.399	0.981	1.003	1.381	0.943	0.960	1.381	0.936	0.955
	1.076	1.012	0.985	1.092	1.013	1.008	1.092	1.017	1.024	1.088	1.025	0.989
	1.085	0.993	1.001	1.098	0.969	1.009	***1.276***	***1.571***	**1.017**	**1**	**1.044**	**1.015**
2-step	1.137	0.993	1.031	1.131	0.968	1.053	1.122	0.923	1.019	1.120	0.931	1.030
	1.045	1.022	1.026	1.045	0.970	1.057	1.044	0.957	1.069	1.033	1.005	0.994
	1.080	1.002	1.107	1.075	1.014	1.074	***1.080***	***1.054***	**1.147**	**1**	**1.096**	**1.145**
3-step	1.110	0.991	1.022	1.082	0.964	1.083	1.079	0.922	1.075	1.076	0.942	1.086
	1.074	1.022	0.980	1.068	0.944	1.099	1.068	0.943	1.117	1.056	1.042	1.038
	1.120	1.070	1.131	1.109	1.027	1.126	***1.057***	***0.961***	**1.225**	**1**	**1.068**	**1.225**
4-step	1.126	1.005	1.055	1.088	0.980	1.103	1.089	0.939	1.126	1.085	0.951	1.130
	1.119	1.014	1.035	1.109	0.946	1.132	1.109	0.975	1.149	1.096	1.075	1.068
	1.166	1.201	1.136	1.150	1.024	1.142	***1.086***	***0.975***	**1.311**	**1**	**1.057**	**1.305**
5-step	1.036	0.997	1.061	1.018	0.961	1.053	1.020	0.927	1.084	1.016	0.934	1.095
	1.060	1.039	1.093	1.055	0.982	1.127	1.057	1.072	1.152	1.044	1.083	1.054
	1.093	1.365	1.135	1.084	1.077	1.102	***1.040***	***0.968***	**1.247**	**1**	**1.027**	**1.238**
All stationary models												
1-step (6)	1.616	1.084	1.129	1.674	1.105	1.176	1.649	1.060	1.123	1.648	1.083	1.118
	1.263	1.043	1.064	1.279	1.047	1.086	1.275	1.056	1.091	1.270	1.084	1.078
	1.288	1.064	1.128	1.305	1.028	1.114	***1.509***	***1.868***	**1.183**	**1**	**1.137**	**1.191**

(continued)

Table 6 (continued)

	(A1) (D1) (G1)	(B1) (E1) (H1)	(C1) (F1) (I1)	(A2) (D2) (G2)	(B2) (E2) (H2)	(C2) (F2) (I2)	(A3) (D3) ES1	(B3) (E3) ES2	(C3) (F3) RW	(A4) (D4) ARMA	(B4) (E4) ARIMA	(C4) (F4) ES
2-step (8)	1.212	0.999	1.073	1.221	0.984	1.114	1.207	0.954	1.073	1.205	0.980	1.098
	1.102	1.030	1.080	1.113	0.968	1.130	1.112	0.964	1.144	1.096	1.034	1.036
	1.141	1.014	1.176	1.149	1.020	1.102	***1.164***	***1.127***	**1.190**	**1**	**1.150**	**1.191**
3-step (8)	1.184	1.004	1.105	1.146	0.992	1.138	1.146	0.955	1.142	1.144	0.986	1.154
	1.161	0.931	1.021	1.155	0.913	1.145	1.155	0.940	1.139	1.139	1.081	1.141
	1.218	1.021	1.255	1.203	1.006	1.146	***1.131***	***1.021***	**1.341**	**1**	**1.143**	**1.340**
4-step (10)	1.223	1.005	1.160	1.180	1.008	1.191	1.186	0.982	1.227	1.183	1.003	1.252
	1.244	0.949	1.128	1.237	0.933	1.250	1.240	1.020	1.257	1.220	1.167	1.241
	1.301	1.200	1.301	1.285	1.142	1.207	***1.203***	***1.083***	**1.486**	**1**	**1.210**	**1.481**
5-step (9)	1.281	0.954	1.229	1.274	0.965	1.265	1.286	0.960	1.293	1.282	0.968	1.351
	1.315	0.954	1.286	1.325	0.973	1.348	1.333	1.208	1.409	1.306	1.271	1.339
	1.369	1.236	1.463	1.370	0.918	1.282	***1.339***	***1.247***	**1.600**	**1**	**1.280**	**1.586**
First part (8)												
1-step	1.374	0.957	0.972	1.427	0.986	1.010	1.408	0.946	0.965	1.407	0.938	0.957
	1.088	1.017	0.992	1.105	1.019	1.017	1.104	1.024	1.032	1.099	1.031	0.996
	1.098	0.997	1.010	1.111	0.970	1.008	***1.298***	***1.606***	**1.015**	**1**	**1.053**	**1.018**
2-step	1.146	0.994	1.034	1.137	0.968	1.055	1.128	0.922	1.022	1.125	0.932	1.032
	1.050	1.024	1.029	1.050	0.971	1.061	1.048	0.957	1.073	1.037	1.006	0.996
	1.086	1.002	1.112	1.081	1.004	1.062	***1.083***	***1.056***	**1.147**	**1**	**1.099**	**1.148**
3-step	1.112	0.991	1.023	1.082	0.962	1.082	1.079	0.920	1.075	1.076	0.940	1.086
	1.076	1.023	0.979	1.069	0.943	1.101	1.068	0.942	1.119	1.056	1.044	1.038
	1.124	1.067	1.134	1.111	1.012	1.108	***1.055***	***0.957***	**1.226**	**1**	**1.067**	**1.227**

4-step	1.125	1.004	1.054	1.087	0.978	1.101	1.088	0.936	1.125	1.084	0.948	1.129
	1.119	1.013	1.034	1.109	0.943	1.132	1.109	0.973	1.148	1.096	1.076	1.067
	1.167	1.198	1.137	1.151	1.006	1.118	***1.082***	***0.970***	**1.307**	**1**	**1.053**	**1.307**
5-step	1.033	0.996	1.060	1.015	0.958	1.049	1.017	0.924	1.081	1.013	0.931	1.092
	1.058	1.039	1.093	1.053	0.980	1.125	1.054	1.072	1.151	1.042	1.082	1.051
	1.091	1.364	1.136	1.083	1.055	1.075	***1.035***	***0.964***	**1.239**	**1**	**1.023**	**1.237**
Second part (8)												
1-step	1.025	1.002	0.964	1.056	0.995	0.993	1.065	0.994	0.985	1.081	0.996	1.022
	0.975	1.025	0.969	0.983	1.003	0.965	0.993	1.008	0.983	1.004	1.024	0.973
	0.967	1.038	0.952	0.990	1.034	1.129	***1.030***	***1.124***	**1.152**	**1.098**	**1**	**1.067**
2-step	0.917	0.974	0.950	0.972	0.954	1.008	0.975	0.941	0.968	0.982	0.933	0.978
	0.908	0.964	0.955	0.932	0.953	0.936	0.940	0.952	0.967	0.944	0.980	0.941
	0.906	1.002	0.978	0.938	1.269	1.428	***0.998***	***1.009***	**1.140**	**1.005**	**1**	**1.051**
3-step	1.051	1.008	1.006	1.082	1.026	1.119	1.081	1.006	1.073	1.084	1.000	1.091
	1.003	0.996	1.011	1.031	0.977	1.037	1.040	0.983	1.064	1.040	0.999	1.022
	1.017	1.151	1.051	1.049	1.522	1.748	***1.120***	***1.094***	**1.195**	**1**	**1.097**	**1.169**
4-step	1.149	1.035	1.074	1.132	1.063	1.179	1.126	1.046	1.155	1.126	1.044	1.156
	1.111	1.032	1.078	1.120	1.028	1.143	1.125	1.039	1.169	1.120	1.057	1.121
	1.131	1.289	1.106	1.136	1.644	1.952	***1.229***	***1.121***	**1.445**	**1**	**1.192**	**1.245**
5-step	1.143	1.026	1.085	1.137	1.081	1.214	1.126	1.061	1.183	1.124	1.074	1.189
	1.133	1.042	1.086	1.139	1.076	1.187	1.143	1.076	1.214	1.138	1.119	1.173
	1.148	1.408	1.120	1.157	1.902	2.147	***1.250***	***1.124***	**1.537**	**1**	**1.207**	**1.273**

(continued)

Table 6 (continued)

	(A1) (D1) (G1)	(B1) (E1) (H1)	(C1) (F1) (I1)	(A2) (D2) (G2)	(B2) (E2) (H2)	(C2) (F2) (I2)	(A3) (D3) ES1	(B3) (E3) ES2	(C3) (F3) RW	(A4) (D4) ARMA	(B4) (E4) ARIMA	(C4) (F4) ES
Nonstationary models—first part												
1-step (3)	1.330	0.965	0.957	1.386	1.006	0.994	1.371	0.966	0.949	1.371	0.931	0.939
	0.995	1.044	1.045	1.001	1.022	1.049	1.001	1.051	1.079	1.004	1.052	1.058
	0.995	1.085	1.074	0.998	1.097	1.058	***1.021***	***1.153***	**1**	**1.262**	**1.164**	**1.029**
2-step (2)	1.000	1.030	0.970	0.940	0.977	0.934	0.942	0.880	0.920	0.938	0.832	0.886
	0.946	1.058	0.924	0.912	1.034	0.902	0.907	0.988	0.909	0.911	0.970	0.928
	0.973	1.019	0.969	0.929	1.029	0.996	***0.898***	***0.898***	**1.092**	**1.053**	**1**	**1.087**
3-step (2)	1.064	1.085	0.928	1.057	1.009	1.075	1.043	0.948	1.037	1.037	0.940	1.046
	0.983	1.418	0.994	0.979	1.155	1.127	0.974	1.074	1.215	0.971	1.075	0.893
	1.012	1.343	0.944	1.007	1.187	1.153	***1.000***	***0.917***	**1.082**	**1.133**	**1**	**1.080**
4-step (2)	1.242	1.314	1.119	1.207	1.213	1.248	1.192	1.119	1.240	1.187	1.109	1.190
	1.158	1.509	1.129	1.138	1.279	1.194	1.132	1.165	1.243	1.132	1.172	0.961
	1.196	1.593	1.071	1.179	1.020	1.269	***1.134***	***0.993***	**1.290**	**1.309**	**1**	**1.290**
5-step (3)	1.022	1.281	1.135	0.987	1.185	1.075	0.978	1.109	1.118	0.974	1.115	1.080
	1.041	1.380	1.147	1.019	1.225	1.161	1.014	1.178	1.153	1.013	1.130	0.998
	1.058	1.839	1.053	1.038	1.487	1.121	***0.969***	***0.899***	**1.164**	**1.240**	**1**	**1.163**
Nonstationary models—second part												
1-step (7)	1.032	1.025	0.977	1.038	0.992	0.980	1.041	0.988	0.974	1.061	0.994	1.015
	0.983	1.036	0.972	0.979	1.014	0.964	0.984	1.014	0.978	1.001	1.035	0.970
	0.976	1.057	0.955	0.984	1.028	1.160	***1.009***	***1.089***	**1.188**	**1.142**	**1**	**1.030**
2-step (6)	0.891	1.073	0.989	0.936	0.989	0.999	0.951	0.981	0.977	0.967	0.983	0.984
	0.904	1.051	1.001	0.920	1.029	0.962	0.927	1.036	0.989	0.945	1.083	0.971
	0.900	1.104	1.012	0.922	1.332	1.709	***0.932***	***1.001***	**1.084**	**1.120**	**1.035**	**1**

3-step (6)	0.934	1.029	0.969	0.958	1.000	1.021	0.970	0.988	0.985	0.978	0.998	0.991
	0.899	1.018	0.975	0.915	0.982	0.956	0.923	0.993	0.977	0.938	1.017	0.937
	0.907	1.178	1.001	0.926	1.477	1.951	***0.942***	***1.004***	**0.965**	**1.021**	**1.028**	**1**
4-step (4)	1.063	1.289	1.110	0.969	1.296	1.161	0.970	1.301	1.181	0.990	1.387	1.175
	1.060	1.349	1.136	1.008	1.310	1.086	1.006	1.336	1.135	1.038	1.461	1.118
	1.059	1.380	1.150	1.009	1.819	3.172	***0.998***	***0.959***	**1.360**	**1.275**	**1.084**	**1**
5-step (4)	0.955	1.259	1.076	0.917	1.283	1.139	0.926	1.317	1.141	0.940	1.417	1.148
	0.994	1.327	1.107	0.958	1.346	1.090	0.963	1.394	1.133	0.993	1.543	1.121
	0.984	1.432	1.103	0.957	2.305	3.433	***0.908***	***0.924***	**1.193**	**1.232**	**1.046**	**1**

See note to Table 5

The main findings from the empirical study are as follows:

- The empirical results divided as described above reveal that if a series is best forecasted with an ARMA model, even better predictions may be obtained if an ARMA model is applied to the series transformed via wavelet smoothing. Also, if the best forecasts are generated with certain methods in the group of nonstationary models, better outcomes are often produced by the nonstationary ('ARIMA', 'RW') procedures applied to the smoothed series. In the group 'Nonstationary models—second part', in fact, most often 'ES' and 'RW' give the best forecasts among those produced by the standard approaches. This may explain why a forecast improvement is then achieved with methods denoted as (A1)–(A4), (D1)–(D4) and (G1)–(G2), i.e., with the no-change method applied to series smoothed via wavelet denoising. On the other hand, in the group 'Nonstationary models—first part', ARIMA models are usually the best predictors among the conventional models. This may stand behind the relative success of the forecasts obtained with ARIMA models estimated on the smoothed series, i.e., the methods (C1)–(C4), (F1)–(F4), and (I1)–(I2).
- If h is set to 0.75, the forecast gains from using wavelet smoothing generally do not rise with the forecast horizon. Moreover, when forecasting five steps ahead, simple exponential smoothing with a constant (and relatively small) value of α outperforms wavelet smoothing in both groups of series denoted as 'nonstationary models'. This may be explained by the fact that, for longer-term forecasting, the series grouped under the name 'nonstationary models' usually require better smoothing than that produced by wavelet denoising with a reasonable number of estimated parameters (and decomposition stages). Setting $h = 1$ often slightly improves longer-term predictions from the wavelet approach. For example, the relative MSPEs for five-step ahead forecasts in the middle of the left part of Table 5, obtained with the methods (B1)–(B4), equal respectively to 0.987, 0.968, 0.956, 0.976, are replaced in Table 6 with the following values: 0.996, 0.958, 0.924 and 0.931. On the other hand, however, it turns out that the higher the smoothing constant, the more variable are the outcomes produced by different models, thus more often leading to clearly suboptimal results. This gives an additional argument for considering smaller than 1 values of h in practical applications.
- Although the aggregated results for the first and second group of series seem to suggest that a level 1 decomposition may be the best (or at least the safest) choice, a more careful examination reveals that the more structurally stable among the series labeled as 'Nonstationary models', i.e., those in the group 'Nonstationary models—first part', are better forecasted if a higher level wavelet decomposition is applied. This confirms our previous findings from the simulation study.
- The aggregated results in the upper part of Table 5 suggest that, in empirical studies, we may benefit from wavelet smoothing mainly via better short-term (1- or 2-step ahead) forecasts. However, as was already mentioned, setting $h = 1$ slightly improves the aggregated results at the longest horizon considered here. The possibility of improving longer-term predictions through wavelet smoothing

holds true even for purely autoregressive processes. This has been confirmed in another simulation exercise in which we generated nearly nonstationary AR(1) processes (with the autoregressive parameter set to 0.9 and 0.95) and, under $h = 1$, obtained gains from wavelet smoothing for forecasts from 2–3 to 5 steps ahead (with the maximum horizon set to 5). This may result from the fact that setting h equal to 1 leads to very good estimates of signals with the lowest possible variance of the error term,[11] which makes it possible to reduce the forecast error variance. Finally, the good outcomes obtained for the first portion of our data, consisting mainly of purely autoregressive (and sometimes periodic) processes, or processes with a more complicated structure than 'AR + noise', can also be explained by the fact that wavelet smoothing captures the influence of higher-order terms in a representation of these processes, thus leading to less complicated parametric models.

It is also worth adding that a portion of all the forecasting results obtained with the wavelet methods was tested for equal predictive ability with those produced by the best benchmark approaches, using the DM test under the assumption of the quadratic loss function, but no significant predictive gains were found. This, at least partially, results from a relatively small number (20) of forecasts considered for each series and forecast horizon and, at the same time, from large forecast error variances. On the other hand, the repeatability of changes in forecast MSEs across different methods and forecast horizons, which we discussed above, observed especially at the aggregate level, seems to support our findings from the simulation studies.

7 Conclusions

Random signal estimation based on wavelet shrinkage combined with the MODWT can be interesting for extracting components of economic processes as well as for forecasting purposes. It relies, however, on the assumptions that the time series under study are composed of a *stochastic* signal and an observational (white) noise and that the conventional wavelet transform is relatively effective at within- and between-scale decorrelation of these series. Although these assumptions are certainly restrictive, both our empirical and simulation studies document that wavelet random signal estimation applied prior to constructing parametric AR(I)MA models can moderately reduce the forecast MSEs in the case of short- and medium-sized samples. Because of the conceptual simplicity of the approach and due to the fact that the computational complexity of the pyramid algorithm used to perform the MODWT is quite low (strictly speaking, it is of the same order as the famous fast Fourier transform), the method may be useful in automatic forecasting systems

[11] Under $h = 1$ our signal estimates roughly correspond to those obtained via the so-called canonical decomposition (see, e.g., Kaiser and Maravall 2005).

applied to large datasets comprising time series with relatively similar dynamic properties. In fact, it takes only about 10 h to compute one million forecasts on a medium-class personal computer. However, any application of wavelet smoothing will usually require making many arbitrary decisions concerning the procedure's configuration (e.g., choosing the value of the smoothing constant and the maximum level of decomposition). Splitting a single time series into estimation and validation subsamples or, alternatively, applying a wavelet classification based on the (normalized) wavelet variance at different scales to find the most similar cluster of historical time series should help to optimize the settings in practical applications.

The forecasting procedures based on random signal estimation with wavelets can be compared with analogous methods relying on wavelet thresholding (see Alrumaih and Al-Fawzan 2002; Ferbar et al. 2009; Schlüter and Deuschle 2010). In our opinion, the methods suggested here are better suited for short-term forecasting of economic time series because wavelet thresholding builds on the assumption of deterministic signals, transforms Gaussian processes into non-Gaussian ones, preserves outliers and reduces the high frequency spectra to 0. The latter characteristic is certainly problematic in the presence of certain high frequency oscillations in the data. By contrast, wavelet smoothing is based on linear time-invariant filters and offers more flexibility as to the level of noise reduction. In conclusion, we recommend this approach for short-term forecasting based on wavelet denoising.

Acknowledgments The author acknowledges financial support from the Polish National Science Center (Decision no. DEC-2013/09/B/HS4/02716). The author would also like to thank the anonymous Reviewer for valuable comments and suggestions which helped improve the paper.

References

Alrumaih RM, Al-Fawzan MA (2002) Time series forecasting using wavelet denoising. J King Saud Univ Eng Sci 14:221–234

Arino M (1995) Time series forecasts via wavelets: an application to car sales in the Spanish market. Discussion Paper No. 95-30, Institute of Statistics and Decision Sciences, Duke University. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.34.9279&rep=rep1&type=pdf. Accessed 18 Feb 2014

Bruzda J (2011) Some aspects of the discrete wavelet analysis of bivariate spectra for business cycle synchronisation. Economics 16:1–46

Bruzda J. (2013a) On simple wavelet estimators of random signals and their small-sample properties. Journal of Statistical Computation and Simulation, in press, http://dx.doi.org/10.1080/00949655.2014.941843.

Bruzda J (2013b) Wavelet analysis in economic applications. Toruń University Press, Toruń

Chen H, Nicolis O, Vidakovic B (2010) Multiscale forecasting method using ARMAX models. Curr Dev Theory Appl Wavelets 4:267–287

Craigmile PF, Percival DB (2005) Asymptotic decorrelation of between-scale wavelet coefficients. IEEE T Inf Theory 51:1039–1048

Diebold FX, Mariano RS (1995) Comparing predictive accuracy. J Bus Econ Stat 13:253–263

Donoho DL, Johnstone IM (1994) Ideal spatial adaptation by wavelet shrinkage. Biometrika 81:425–455

Donoho DL, Johnstone IM (1995) Adapting to unknown smoothness via wavelet shrinkage. J Am Stat Assoc 90:1200–1224
Ferbar L, Čreslovnik D, Mojškerc B, Rajgelj M (2009) Demand forecasting methods in a supply chain: smoothing and denoising. Int J Prod Econ 118:49–54
Fernandez V (2008) Traditional versus novel forecasting techniques: how much do we gain? J Forecasting 27:637–648
Fryźlewicz P, Van Bellegem S, von Sachs R (2003) Forecasting nonstationary time series by wavelet process modelling. Ann I Stat Math 55:737–764
Kaboudan M (2005) Extended daily exchange rates forecasts using wavelet temporal resolution. New Math Nat Comput 1:79–107
Kaiser R, Maravall A (2005) Combining filter design with model-based filtering (with an application to business-cycle estimation). Int J Forecasting 21:691–710
Li TH, Hinich MJ (2002) A filter bank approach for modeling and forecasting seasonal patterns. Technometrics 44:1–14
Makridakis S, Hibon M (2000) The M3-competition: results, conclusions and implications. Int J Forecasting 16:451–476
Minu KK, Lineesh MC, Jessy John C (2010) Wavelet neural networks for nonlinear time series analysis. Appl Math Sci 4:2485–2495
Nason GP (2008) Wavelet methods in statistics with R. Springer-Business Media, New York
Ogden RT (1997) Essential wavelets for statistical applications and data analysis. Birkhäuser, Boston
Percival DB (1995) On estimation of the wavelet variance. Biometrika 82:619–631
Percival DB, Walden AT (2000) Wavelet methods for time series analysis. Cambridge University Press, Cambridge
Renaud O, Starck J-L, Murtagh F (2002) Wavelet-based forecasting of short and long memory time series. Working Paper No. 2002.04, University of Geneva. http://www.unige.ch/ses/metri/cahiers/2002_04.pdf. Accessed 18 Feb 2014
Ramsey JB (1996) If nonlinear models cannot forecast, what use are they? Stud Nonlinear Dyn E 1:65–86
Ramsey JB (2002) Wavelets in economics and finance: past and future. Stud Nonlinear Dyn E 6(3):1–27, article 1
Ramsey JB, Lampart C (1998) The decomposition of economic relationships by time scale using wavelets: expenditure and income. Stud Nonlinear Dyn E 3:23–42
Schlüter S, Deuschle C (2010) Using wavelets for time series forecasting—does it pay off? Diskussionspapier No. 4/2010, Institut für Wirtschaftspolitik und Quantitative Wirtschaftsforschung, Friedrich-Alexander-Universität. http://www.econstor.eu/obitstream/10419/36698/1/626829879.pdf. Accessed 18 Feb 2014
Vidakovic B (1999) Statistical modeling by wavelets. Wiley, New York
Wong H, Ip W-C, Xie Z, Lui X (2003) Modelling and forecasting by wavelets, and the application to exchange rates. J Appl Stat 30:537–553
Yu P, Goldenberg A, Bi Z (2001) Time series forecasting using wavelets with predictor-corrector boundary treatment. In: Proceedings of the 7th ACM SIGKDD international conference on knowledge discovery and data mining, San Francisco, 26–29 August 2001
Zhang B-L, Coggins R, Jabri MA, Dersch D, Flower B (2001) Multiresolution forecasting for futures trading using wavelet decompositions. IEEE T Neural Networ 12:765–775

Short and Long Term Growth Effects of Financial Crises

Fredrik N.G. Andersson and Peter Karpestam

Abstract Growth theory predicts that poor countries will grow faster than rich countries. Yet, growth in developing countries has been consistently lower than growth in developed countries. The poor economic performance of developing countries coincides with both long-lasting and short-lived financial crises. In this paper, we analyze to what extent financial crises can explain low growth rates in developing countries. We distinguish between inflation, currency, banking, debt, and stock-market crises and separate the short- and long-run effects of them. Our results show that financial crises have reduced growth and that policy decisions have caused them to be worsened and/or extended.

1 Introduction

From 1973 to 2007, the labor productivity growth of developed countries averaged 2 % per year. Over the same period, the average labor productivity growth in Africa and Latin America averaged 0.5 and 0.8 % per year, respectively. Only developing countries in Asia were able to match (and exceed) growth in the developed world (3.2 % per year).[1] During this period, Africa and Latin America, in particular, faced several financial crises (Wilson et al. 2000; Reinhart and Rogoff 2011). For example, Latin America suffered economically due to persistent financial crises throughout most of the 1970s and the 1980s (De Gregorior and Guidotti 1995), while large

[1] http://www.conference-board.org/.

F.N.G. Andersson (✉) • P. Karpestam
Department of Economics, Lund University, P.O. Box 7082, 220 07 Lund, Sweden
e-mail: ngf.andersson@nek.lu.se; peter.karpestam@nek.lu.se

M. Gallegati and W. Semmler (eds.), *Wavelet Applications in Economics and Finance*, Dynamic Modeling and Econometrics in Economics and Finance 20,
DOI 10.1007/978-3-319-07061-2_10,

parts of Africa faced "near-permanent banking-stress" for 20 years (Kane and Rice 2001).

In this paper, we analyze to what extent the poor economic performance in 30 developing countries between 1973 and 2007 can be explained by the occurrence of financial crises. Previous studies of the effects of financial crisis have given inconclusive results: the short-run growth effects are mostly negative (Norman and Romain 2006; Ramey and Ramey 1995; Hausmann and Gavin 1996; Easterly et al. 2001), but estimates of the long-term growth effects of financial crises are less conclusive. Some studies show that long-run growth is reduced by financial crises (Englebrecht and Langley 2001; Rousseau and Wachtel 2002; Bordo et al. 2010; Eichengreen and Hausman 1999) while other suggests that long-run growth effects is even enhanced by financial crises (Bruno and Easterly 1998; Ranciere et al. 2008).

Arguably, how severe the growth effects are of a financial crisis depends on (1) what kind of financial crisis it is, (2) if more than one crisis is coinciding, (3) through which growth channel it affects the economy and (4) the time horizon. The literature on financial crises commonly distinguishes between five different types of financial crises: inflation, currency, banking, debt, and stock market crises (see e.g. Reinhart and Rogoff 2011). The respective types of financial crises are sometimes linked. Sovereign debt crises, for example, are often preceded by a banking crisis, forcing the national government to take over debts in the banking sector (Velasco 1987; Reinhart and Rogoff 2011). In turn, debt crises often spill over into currency crises (Kaminsky and Reinhardt 1999) and countries facing insolvency sometimes inflate the economy to reduce the debt burden (Labán and Sturzenegger 1994). This action may, in turn, cause an inflation crisis as well.

Each type of financial crisis can affect economic growth, but how much they affect growth is likely to be different. Stock markets crashes for example can affect investments (Tobin 1969; von Furstenberg 1977) and/or private consumption through a wealth effect (Friedman 1957; Paiella 2009). But, in stock markets in developing countries, only a limited number of people own shares (Enisan and Olufisayo 2009). This causes wealth effects to be small at the aggregate level. A currency crisis, however, is likely to have more severe effects on the economy than the stock market crash, especially for developing countries that are dependent on foreign investment capital and technology. Similarly, a banking crisis that affects the channeling of capital from savers to borrowers are likely to affect growth more than an inflation crisis against which agents in the economy can hedge by for example by price indexing contracts (McNelis 1988).

Bonfiglioli (2008) argue that financial crises that only affect productivity growth are worse for a developing country than financial crises that only affect capital accumulation. Most of the income difference among countries is due to differences in productivity and not in capital intensity (Gourinchas and Jeanne 2006). A financial crisis that affects productivity growth causes greater welfare effects than a financial crisis that affects capital accumulation for a developing country because it reduces that country's ability to catch up economically with developed countries (Bonfiglioli 2008). Understanding through which growth channel the crises affects growth is thus important.

The time horizon is also likely to affect the impact of the financial crises on economic growth. An inflation crisis is less likely to affect growth over the long-term given that the agents index their contracts, but if the crisis is unexpected could have negative effects over the short-term. A debt crisis on the other hand that only affects the country's access to capital for a year is unlikely to cause a major reduction in capital accumulation while a debt crisis that continues to restrict the country's access to foreign capital for several years is likely to impact capital accumulation over the long-term.

Most papers only consider the growth effects of one kind of financial crisis at the time and either the short-run or the long-run growth effects of that specific kind of crisis. In this paper, we analyze the growth effects of five different types of financial crises (inflation, currency, banking, debt, and stock market crashes) simultaneously on labor productivity growth and its two growth channels (total factor productivity growth and capital accumulation). The growth effects are separated into short-run effects and long-run effects using a Band Spectrum Regression model (see e.g. Andersson 2011a). Our focus is on developing countries, but we compare and contrast the results with developments among developed countries over the same time period.

Our results show that financial crises have negative growth effects both in the short-run and the long-run. Inflation crises have mostly only short-run effects on growth and persistent inflation crises have no long-run growth effects. Persistent debt and currency crises on the other hand are associated with a decline in the long-run growth rates. The long-run growth effects mainly occur through the total factor productivity channel, although there is an effect on capital accumulation as well. Based on these results we find that growth in Latin America would have kept pace with the growth in developed countries. Growth among African countries would also have been higher had they not suffered from financial crises, but the average growth rate would still have been lower than among developed countries.

The remainder of the paper is organized as follows: Sect. 2 presents the model, Sect. 3 contains the empirical results, and Sect. 4 concludes the paper.

2 Labor Productivity Growth and Financial Crises

The starting point for our model is a Cobb–Douglas production function with Harrod neutral technology and constant returns to scale,

$$Y_{it} = (A_{it}L_{it})^{\alpha} K_{it}^{1-\alpha} \tag{1}$$

where Y is the real GDP, K is real capital, A is technology, L is employment, $\alpha \epsilon (0, 1)$ is the labor output elasticity, i denotes country, and t time. Dividing both sides of (1) by L and then taking the log and first difference, we obtain the following expression of the (log-) labor productivity growth rate,

$$\Delta y_{it} = \alpha \Delta a_{it} + (1 - \alpha)\, \Delta k_{it}, \tag{2}$$

where $\Delta y_{it} = ln(Y_{it}/L_{it}) - ln(Y_{it-1}/L_{it-1})$, $a_{it} = ln(A_{it}) - ln(A_{it-1})$ and $\Delta k_{it} = ln(K_{it}/L_{it}) - ln(K_{it-1}/L_{it-1})$. In Eq. (2) we observe that labor productivity growth comes from two channels: either total factor productivity growth (Δa_{it}) or capital accumulation (Δk_{it}). Financial crises may thus either affect labor productivity growth through total factor productivity, through capital accumulation or through both.

We estimate three regression models using a band spectrum regression (see Engle 1974) to test which financial crises that affect labor productivity growth, through which growth channel and over which time horizon they affect economic growth. A band spectrum regression is an estimation technique to separate between effects over different time horizons. The estimation is carried out in two steps. First, all the variables are transformed to the frequency domain where each time-horizon (in our case the short-run and the long-run) are easily identified. Second, parameter estimates for each time horizon are then obtained by estimating the models on a subset of frequencies representing a given horizon rather than on the entire frequency band (for more information see Engle 1974; Andersson 2011a). In a small sample, the band spectrum regression generally performs better compared to other methods to distinguish between short-run and long-run effects such as cointegration analysis (Corbae et al. 2002; Andersson 2008) or a simple moving average.

Any band pass filter can be used to transform the series to the frequency domain. In this paper we use the Maximal Overlap Discrete Wavelet Transform (MODWT).[2] This transform is chosen because it combines time and frequency resolution, whereby the transform is suitable to transform series that contains nonrecurring events such as structural breaks and outliers (Percival and Walden 2006).[3]

The band spectrum regression is not limited to two time horizons and the model can include several time horizons (e.g. short-run, medium-run, long-run, etc.). But, in line with standard economic theory, we limit the analysis to a short-run business cycle component and the long-run trend component. Baxter and King (1999), Assenmacher-Wesche and Gerlach (2008a, b) and Andersson (2011b) show that the business cycle in general lasts between 4 and 8 years, and we consequently define the short-run as fluctuations lasting up to 8 years.

Specifically, we estimate the following three regression models; one model for labor productivity growth and one model for each of the two growth channels. The model for labor productivity growth is given by,

[2]To employ the maximal overlap discrete wavelet transform one can use several different sets of basis functions. We chose to use Haar wavelet basis functions because they minimize the potential effect of boundary coefficients (see Percival and Walden 2006). Alternative basis have been employed such as the Daubechie (4) and Daubechie (6) wavelets but the results are similar irrespective of filter.

[3]For more information about the MODWT, see e.g., Ramsey and Lampart (1998), Percival and Walden (2006), Crowley (2007), and Andersson (2008).

$$\Delta y = \beta_{y1} + \beta_{y1} F_{it}^{SR} + \beta_{y2} F_{it}^{LR} + \theta_{y1} C_{it}^{SR} + \theta_{y1} C_{it}^{LR} + \varepsilon_{yit}, \tag{3}$$

the model for total factor productivity growth is given by,

$$\Delta a_{it} = \beta_{a1} + \beta_{a1} F_{it}^{SR} + \beta_{a2} F_{it}^{LR} + \theta_{a1} C_{it}^{SR} + \theta_{a1} C_{it}^{LR} + \varepsilon_{ait}, \tag{4}$$

and one model for capital per employee growth is given by,

$$\Delta k_{it} = \beta_{k1} + \beta_{k1} F_{it}^{SR} + \beta_{k2} F_{it}^{LR} + \theta_{k1} C_{it}^{SR} + \theta_{k1} C_{it}^{LR} + \varepsilon_{kit}, \tag{5}$$

where, F_{it} is a vector with dummy variables indicating the respective types of financial crises and C_{it} is a vector with common used control variables (see below), *SR* denotes the short-run, *LR* denotes the long-run, β and θ are the parameters to be estimated, and ε is a stochastic error term.

The variables are decomposed into a short-run and a long-run component using the MODWT. Applying the MODWT we get the following decomposition of F and C,

$$F_{it} = D_{1Fit} + D_{2Fit} + S_{2Fit} \tag{6}$$

and

$$C_{it} = D_{1Cit} + D_{2Cit} + S_{2Cit} \tag{7}$$

where the wavelet details D_1 and D_2 represent 2–4 year-long and 4–8 years-long cycles respectively, and the wavelet smooth S_2 is the trend component representing persistent long-run developments in the economy lasting 8 years and beyond.[4] Given our definition of the length of the business cycle, the short-run components of F and C are represented the two wavelet details,

$$F_{it}^{SR} = D_{1Fit} + D_{2Fit} \tag{8}$$

and

$$C_{it}^{SR} = D_{1Cit} + D_{2Cit}. \tag{9}$$

It then follows that the long-run component of F and C is represented by the wavelet smooth,

$$F_{it}^{LR} = S_{2Fit} \tag{10}$$

[4]The decomposition of the variables is made variable-by-variable and country-by-country. Not just the dependent is decomposed, but all variables are decomposed into time horizons.

and

$$C_{it}^{LR} = S_{2Cit}. \tag{11}$$

As control variables we have included education, Freedom House's political rights index and the KOF index. Education is used as a proxy for human capital affecting productivity. Freedom House political rights index is included as a measure of political institutions—weak and non-democratic political institutions are often an underlying factor in generating financial crises but can also prolong the duration of the crises through erroneous and late policy responses (Kane and Rice 2001; Acemoglu et al. 2003; Tommasi 2004). In contrast, strong and often democratic institutions are better equipped to prevent and solve crises once they occur (Cavallo and Cavallo 2010; Rodrik 2000) whereby the effect of financial crises should be less democratic countries than non-democratic countries. The KOF index is included as a measure of globalization. Increased financial integration and liberalization both increase the probability of financial crises and can increase their severity (see e.g., Kaminsky and Reinhardt 1999; Ranciere et al. 2008).

An econometric concern in this model is reversed causality between the financial crises and economic growth. Financial crises may reduce growth, but a financial crisis may also be outcome of a period of low growth rates. Following Beck (2008), we address this and employ internal instruments to correct for this possible error.[5]

3 Empirical Analysis

3.1 Data

Our data set contains 51 countries (see Table 9 in Appendix) covering the period of 1973–2007. The final year is dictated by availability of real investment data (Penn World Table 6.3) that is needed to generate national capital stock estimates. Of the 51 countries, the World Bank classifies 21 as developed countries, and 30 countries are classified as developing countries.[6] We rely on external data sources for labor productivity, financial crises, institutions, education, and globalization. A detailed description of the data and the data sources are available in Table 10 (Appendix).

Our indicators of financial crises are collected from Reinhart and Rogoff's (2011) database.[7] This database distinguishes between five different types of crises that are indicated with dummy variables: inflation, currency, banking, debt, and stock

[5]Data availability makes it impossible to find external instruments for each of the five financial crises, and we rely instead on internal instruments.

[6]See http://data.worldbank.org/about/country-classifications.

[7]The database can be obtained from Reinhart's webpage: http://terpconnect.umd.edu/~creinhar/Courses.html.

market crises. A detailed definition of these financial crises is available in Reinhart and Rogoff (2011). An inflation crisis occurs when the annual rate of inflation exceeds 20 % per year, whereas a currency crisis occurs when the national currency loses 15 % or more of its value against the USD or some other relevant currency. Additionally, a banking crisis is defined as a bank run that leads to a government takeover of a bank. Lastly, a debt crisis is defined as a country defaulting on its external debt.

National capital stocks are estimated using the perpetual inventory method assuming a fixed 5 % depreciation rate.[8] Larson et al. (2000) have estimated capital stock data for the period of 1967–1997, and we use their estimates for 1967 as our initial capital stock estimate. Short and long-run total factor productivity growth estimates are obtained from first estimating the model,[9]

$$\Delta y_{it}^{h} = \rho_{h1} + \rho_{h2} \Delta k_{it}^{h} + \varepsilon_{it}^{h} \tag{12}$$

where h denotes the time horizon (i.e. *SR* or *LR*)., Using Eq. (12) and the parameter estimates, $\widehat{\rho}$, total factor productivity is then estimated as,

$$\Delta \widehat{a}_{it}^{h} = \Delta y_{it}^{h} - \widehat{\rho}_{yh2} \Delta k_{it}^{h} \tag{13}$$

For education, we use the total years of schooling among the labor force.[10] Education data are only available at a 5-year interval, and without higher frequency data, we cannot include the variable in the short-run models. Therefore, education is only included in the long-run models. To capture the effect of globalization on the financial system and the overall economy, we use the KOF index, which is a combined measure of economic, social and political globalization (Dreher 2006). Recently, the KOF index has been used in empirical research to capture the macroeconomic effects of the current globalization process (see e.g., Bergh and Nilsson 2010). Based on Cavallo and Cavallo's (2010) discussion of the link between democratic institutions and financial crises, we use the Freedom House political rights index to control for institutional quality. Each country is scored by Freedom House between 1 and 7, where countries with a score between 1 and 2.5 are defined as free. Countries with a score between 3.0 and 5.0 are partly free, and countries with a score between 5.5 and 7 are not free. Because we are modeling growth rates, we use the percentage change in education, political rights and the KOF index in the regression models.

[8] We also tested alternative depreciation rates (3 and 7 %), but changing the depreciation rate has only a minor effect on estimated capital output elasticity, and no significant effect on the estimates of the effects of financial crises.

[9] This regression model is derived from Eq (2).

[10] Alternative measures, such as secondary schooling, were also considered, but models including total schooling have better statistical properties than models using secondary schooling.

Table 1 Descriptive statistics: average yearly growth rates

	Developed countries	African countries	Asian developing countries	Latin American countries
Labor productivity growth	2.00	0.51	3.19	0.78
Capital growth	2.67	1.36	3.92	1.56
Financial crisis	0.54	1.21	0.84	1.73
Inflation crisis	0.06	0.20	0.06	0.45
Currency crisis	0.09	0.23	0.17	0.42
Banking crisis	0.12	0.16	0.23	0.19
Debt crisis	0.00	0.28	0.12	0.43
Stock market crisis	0.27	0.34	0.27	0.24
Political rights	1.20	5.03	3.56	2.79
Education	1.23	3.93	2.30	2.06
Globalization	1.17	1.75	2.02	1.46

Labor productivity growth rates: growth is, on average, the highest among developing Asian countries with an average yearly growth rate of 3.19 %, while it is the lowest among African countries, at 0.51 % per year. Among Latin American countries, average labor productivity growth is 0.78 % per year and among developed countries 2.00 %. Labor productivity growth is more volatile among developing countries than among developed countries. While growth remains within a span of −2 to 5 % per year among developed countries, among African countries yearly growth fluctuations of ±15 % points are common.

As can be seen in Table 1, developed countries have experienced fewer financial crises than developing countries (0.54 per year). A stock market crisis is the most common (0.27 per year), and a debt crisis is the least common (0 per year). Among African countries, the average is 1.21 per year, and a stock market crash (0.34 per year) is the most common followed by debt (0.28 per year), currency (0.23 per year) and inflation crises (0.20 per year). Developing Asian countries experience 0.84 crises per year of which a stock market crash (0.27 per year) and bank crisis (0.23 per year) are the two most common types. Latin America has the highest frequency of financial crises (1.73 per year). In Latin America, inflation crises are the most common (0.45 per year), followed by debt (0.43 per year), and currency crises (0.42 per year).

Currency and debt crises often coincide in the long-run, see Table 2. The correlation between inflation and currency crises is 0.77, and the correlation between inflation and debt crises is 0.43. There is, however, no significant correlation between any of the other financial crises. Over the short term (Table 3), the highest correlation is between inflation and currency crises, at 0.14, but this is not significantly different from zero. Although financial crises occur simultaneously over the long term, they are independent over the short term.

The high long-term correlation between inflation and currency crises implies that we can interpret these two crises as a joint monetary crisis instead of two

Table 2 Explanatory variables—long-run correlation matrix

	Inflation crisis	Currency crisis	Banking crisis	Debt crisis	Stock crash	Political rights	Education	Globalization (KOF)
Inflation crisis	1.00							
Currency crisis	0.77***	1.00						
Banking crisis	0.14	0.21**	1.00					
Debt crisis	0.43***	0.45***	0.20**	1.00				
Stock market crash	−0.07	0.07	−0.12	−0.03	1.00			
Political rights	0.24**	0.23**	−0.02	−0.04	0.02	1.00		
Education	0.04	0.10	−0.08	0.05	0.02	0.06	1.00	
Globalization (KOF)	0.00	0.06	0.16*	−0.05	0.05	−0.16*	−0.14	1.00

*Significant level at 10 %; **significant level at 5 %; ***significant level at 1 %

Table 3 Explanatory variables—short-run correlation matrix

	Inflation crisis	Currency crisis	Banking crisis	Debt crisis	Stock crash	Political rights	Education	Globalization (KOF)
Inflation crisis	1.00							
Currency crisis	0.14	1.00						
Banking crisis	−0.01	0.06	1.00					
Debt crisis	0.03	0.09	0.05	1.00				
Stock market crash	0.01	0.00	0.02	0.01	1.00			
Political rights	0.01	−0.04	0.02	−0.08	−0.05	1.00		
Education	0.01	−0.06	0.01	−0.05	−0.04	0.04	1.00	
Globalization (KOF)	−0.00	0.02	−0.00	−0.05	0.03	0.01	0.00	1.00

separate crises (over the long term). The significant and positive correlation with the Freedom House political rights index suggests that policy decisions are at least in part responsible for causing the monetary crises.

3.2 Regression Results

For each regression model, we present two regression results: the results from a complete model that includes all variables and the results from a reduced model where the insignificant variables have been removed. The error term in the model is specified as a two-way error component model that includes fixed effects for both cross-sectional and time effects. We use robust standard errors to account for

heteroskedasticity (see e.g., Arellano 1987; Baltagi 2008). The regression results are available in Table 4 (long run) and Table 5 (short run).[11]

Labor productivity growth responds negatively to a financial crisis both over the long term and the short term. However, the impacts of the different types of crises are not the same in the short and the long run. Inflation, currency, and banking crises affect growth in the short run, but in the long run, only currency and debt crises have significant effects. Stock market crashes have no growth effect at all, irrespective of the time horizon. Banking crises have the largest short-term effect on growth, −1.33 % points per year, and currency crises have the largest long-term effect, −1.27 %.

Overall, short-run growth models explain little of the variation in the data (R^2 is 0.11). Short-term crises have no long-run effect, and their most negative effect comes from increasing volatility in the economy. But, even if financial crises do cause higher short-term volatility, as is indicated by the low R^2-values, most of the short-term volatility in the data is due to other factors. Because of this, the impact of financial crises over the short term is limited. Over the long term, the explanatory power of the models is higher: R^2 is between 0.36 and 0.39.

The high long-term correlation between inflation and currency crises creates a multicolinearity problem in the model, and it is only possible to include one of the two at a time. However, because of the high correlation between the two, we interpret them as representing a monetary crisis. The effect of a long-run monetary (currency) crisis reduces growth by 1.09 % points per year. When occurring jointly with a debt crisis (which is often the case), growth is reduced by another 1.27 % point. Combined, the two crises thus reduce growth by 2.34 % points per year.

Turning to the growth channels, we find a stronger effect of financial crises on total factor productivity than on capital accumulation. This result is in accordance with Bonfiglioli (2008), who found that financial development has a stronger effect on productivity than on capital accumulation. In the short run, financial crises have a negative impact on total factor productivity, but no effect on the capital accumulation. Because these negative effects on productivity capture both demand and productivity effects over the short term and capital accumulation is unaffected by financial crises, these results indicate that aggregate demand is more important than the aggregate supply side in the response to financial crises in the short-run.

In the long run, financial crises (i.e., a debt crisis) have a negative impact on both capital accumulation and total factor productivity. Debt crises reduce capital accumulation growth by −2.14 % points and total factor productivity by −0.98 % points. Total factor productivity is also negatively affected by monetary crises (currency crisis), at −1.18 % points. Considering that these crises often coincide,

[11]We have assumed that the errors in the respective regression models are normally distributed to perform inference on the parameters. The normality hypothesis is supported by a Jarque–Bera normality test for all but one case—the long-run African labor productivity growth model. However, once we include two dummy variables to control for outliers we do not reject the normality assumption for this growth model either.

Table 4 Regression results of the long-run growth models

	Labor productivity growth		Capital growth		Total factor productivity	
	Full model	Reduced model	Full model	Reduced model	Full model	Reduced model
Capital	0.27***(0.05)	0.27***(0.06)	–	–	–	–
Inflation	−0.08(0.92)	–	0.38(1.28)	–	−0.10(0.73)	–
Currency	−1.39*(0.81)	−1.27**(0.58)	−1.80(1.42)	–	−1.28(0.81)	−1.18**(0.58)
Banking	−0.81(0.51)	–	−0.10(0.90)	–	−0.80(0.51)	–
Debt	−0.95**(0.47)	−1.09**(0.48)	−1.79**(0.81)	−2.14***(0.73)	−0.84*(0.46)	−0.98**(0.47)
Stock market	0.96(0.83)	–	1.02(1.46)	–	0.89(0.84)	–
Political rights	0.15(0.12)	–	−0.01(0.21)	–	0.15(0.12)	–
Education	−0.12(0.08)	–	0.12(0.14)	–	−0.13*(0.07)	–
Globalization (KOF)	0.25**(0.12)	0.23**(0.12)	−0.05(0.21)	–	0.25**(0.12)	0.24(0.12)
Adjusted R^2	0.39	0.36	0.10	0.07	0.22	0.38
BIC	−0.44	−0.38	0.68	0.68	−0.43	−0.38

*Significant level at 10 %; **significant level at 5 %; ***significant level at 1 %

Table 5 Regression results of the short-run growth models

	Labor productivity growth		Capital growth		Total factor productivity	
	Full model	Reduced model	Full model	Reduced model	Full model	Reduced model
Capital	0.35***(0.03)	0.35***(0.03)	–	–	–	–
Inflation	−0.92***(0.32)	−0.92***(0.31)	0.39(0.27)	–	−0.92***(0.31)	−0.92***(0.31)
Currency	−0.69***(0.24)	−0.69***(0.24)	0.01(0.20)	–	−0.69***(0.24)	−0.68***(0.24)
Banking	−1.33***(0.26)	−1.33***(0.26)	−0.25(0.22)	–	−1.33***(0.26)	−1.33***(0.26)
Debt	0.01(0.29)	–	−0.46(0.25)	–	0.00(0.29)	–
Stock market	0.25(0.18)	–	0.03(0.15)	–	0.25(0.18)	–
Political rights	−0.09(0.13)	–	−0.09(0.11)	–	−0.10(0.13)	–
Education	0.21***(0.06)	0.20***(0.06)	0.01(0.05)	–	0.21***(0.06)	0.20***(0.06)
Globalization (KOF)	−0.01(0.03)	–	0.03(0.02)	–	−0.01(0.03)	–
Adjusted R^2	0.11	0.11	0.00	–	0.03	0.04
BIC	2.16	2.14	1.86	–	2.24	2.22

*Significant level at 10 %; **significant level at 5 %; ***significant level at 1 %

the combined effect on total factor productivity is −2.08 % points each year the crisis lasts.

All countries have experienced short-run financial crises, but only Africa and Latin America have experienced persistent long-run financial crises. To explore if the crises effects are the same for both continents, we estimate two sub-panels using long-run data: one for African countries and one for Latin American countries. These long-run estimation results are presented in Table 6.

For Africa as well as Latin America, a debt crisis has a significant and negative impact on capital accumulation. However, a debt crisis affects total factor productivity in Africa but not in Latin America. Instead, total factor productivity in Latin America is affected negatively by inflation crises. Further, capital accumulation in Latin America is negatively affected by banking crises, which is not the case for Africa.

A positive and significant correlation between monetary crises and the Freedom House index suggests that monetary crises are partially caused by monetary policy decisions over the long term. For example, debt crises during the early 1980s created a need for many developing countries to become less dependent on foreign sources of capital and adjust their economies. Latin American economies postponed this process by inflating their currency (Labán and Sturzenegger 1994). Not all developing countries have followed this same path (Djikstra 1997). Consequently, they have not suffered as much from the inflation and currency crises that resulted from the policy response. For example, during the Southeast Asian crises in 1996–1997, policy makers responded quickly and inflation never rose to the same levels as in Latin America. As a result, Southeast Asia recovered quickly from the crisis (Pilbeam 2006).

3.3 Potential Labor Productivity Growth

To illustrate the long-run growth effect of financial crises among developing countries, we calculate the average capital growth rate and average total factor productivity growth rate for developed countries, African countries, Asian countries and Latin American countries.[12] For African countries and Latin American countries, we also calculate the potential long-run growth rates, defined as the growth rates that these countries would have been achieved in the absence of long and persistent financial crises. These potential growth rates are estimated as,

$$\Delta y_{it}^{pot} = \Delta y_{it} - \widehat{\beta}_{y1} F_{it}^{SR} - \widehat{\beta}_{y2} F_{it}^{LR}, \tag{14}$$

[12]The sum of the capital accumulation effect and total factor productivity equals labor productivity growth, see Eq. (2).

Table 6 Regression results of the long-run growth models for Africa and Latin America

	Growth		Capital growth		Total factor productivity	
	Full model Africa	Reduced model Latin America	Full model Africa	Reduced model Latin America	Full model Africa	Reduced model Latin America
Capital	0.12(0.12)	0.35***(0.06)	–	–	–	–
Inflation	–	–	–	–	–	−2.00***(0.80)
Currency	–	−1.02**(0.48)	–	–	–	–
Banking	–	−1.77***(0.54)	–	−2.59**(1.21)	–	–
Debt	−2.15***(0.73)	–	−2.17***(0.92)	−1.94**(0.85)	−1.79**(0.72)	–
Stock market	–	–	–	–	–	–
Political rights	−0.53**(0.25)	0.47***(0.24)	–	–	−0.42*(0.25)	0.43***(0.18)
Education	–	–	–	–	–	0.41**(0.20)
Globalization (KOF)	–	–	–	–	–	–
Adjusted R^2	0.38	0.89	0.12	0.14	0.20	0.28
BIC	−0.04	−0.32	0.55	0.42	0.06	−0.47

*Significant level at 10 %; **significant level at 5 %; ***significant level at 1 %

Table 7 Long run labor productivity-, capital- and total factor productivity growth for the average African and the average Latin American country

Average	Growth	Capital growth	Total factor productivity	Growth	Capital growth	Total factor productivity
	Latin America (%)			Africa (%)		
1973–1980	0.94	1.12	−0.19	1.32	1.78	−0.47
1981–1990	−0.42	0.39	−0.81	−0.32	0.24	−0.56
1991–2000	1.19	0.32	0.86	−0.18	−0.34	0.16
2001–2007	1.48	0.13	1.32	1.32	0.23	1.09
1973–2007	0.78	0.49	0.28	0.51	0.48	0.03
	Latin America potential growth (%)			Africa potential growth (%)		
1973–1980	2.19	1.34	0.85	1.60	1.97	−0.37
1981–1990	1.49	1.00	0.49	0.56	0.50	0.06
1991–2000	2.45	0.77	1.68	0.60	−0.14	0.74
2001–2007	2.00	0.33	1.67	2.00	0.44	1.56
1973–2007	2.03	0.86	1.17	1.19	0.69	0.50

Table 8 Long run labor productivity-, capital- and total factor productivity growth for the average developed and the average Asian country

Average	Growth	Capital growth	Total factor productivity	Growth	Capital growth	Total factor productivity
	Developed countries (%)			Asia (%)		
1973–1980	2.54	1.63	0.91	3.08	2.56	0.53
1981–1990	1.98	1.10	0.88	2.76	1.87	0.89
1991–2000	1.93	0.87	1.06	3.01	1.60	1.41
2001–2007	1.51	0.79	0.72	3.50	1.15	2.35
1973–2007	1.99	1.10	0.89	3.19	1.80	1.40

$$\Delta a_{it}^{pot} = \Delta y_{it} - \widehat{\beta}_{a1} F_{it}^{SR} - \widehat{\beta}_{a2} F_{it}^{LR}, \tag{15}$$

$$\Delta k_{it}^{pot} = \Delta y_{it} - \widehat{\beta}_{k1} F_{it}^{SR} - \widehat{\beta}_{k2} F_{it}^{LR}, \tag{16}$$

where Δy_{it}^{pot}, Δa_{it}^{pot}, Δk_{it}^{pot} are the potential labor productivity growth rate, potential total factor productivity growth rate and potential capital growth rate respectively, and $\widehat{\beta}$ are the estimated parameters. We use the parameter estimate from Table 6 for African and Latin American countries and the parameter estimates from Table 3 for Asian and developed countries. The results for African and Latin American countries are presented in Table 7 and the results for developed countries and developing Asian countries are available in Table 8. The results are summarized decade-by-decade.

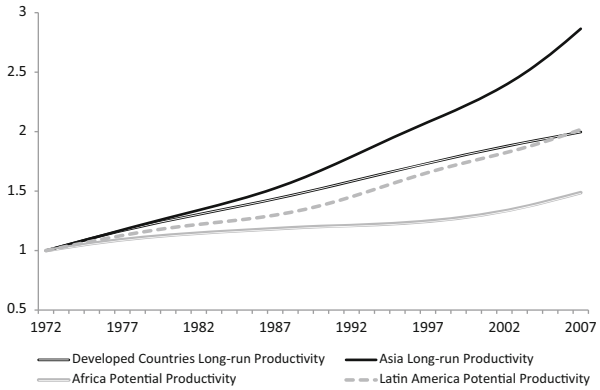


Fig. 1 Estimated long-run labor productivity growth for the average developed, African, Asian and Latin American country

Financial crises are estimated to have reduced average growth in Latin America by 1.25 % points per year and African countries by 0.68 % points per year. The average potential growth for the entire period of 1973–2007 in Latin America is equal to the observed growth for the developed countries: 2.03 % compared to 2.00 %, respectively. Average potential African growth is lower, at 1.19 %. From 2000 to 2007, however, potential African growth exceeded observed growth among developed countries (2.00 % compared to 1.51 %).

On average, growth is the highest in Asia. Table 8 shows that the high Asian growth rates of 50 % can be explained by capital accumulation. Latin American growth is lagging behind observed Asian growth due to lower capital accumulation rates. Additionally, Africa is trailing Asia because of lower potential capital accumulation rates and lower potential total factor productivity growth.

In relation to developed countries, these results show that Latin America would have been falling behind during the 1970s and 1980s had there been no financial crises. They also show that they would have been catching up from the 1990s and onward. Similarly, Africa would have been falling behind from the 1970s and throughout the 1990s but catching up thereafter. Without the financial crises, growth would have been higher, but limited investments (due to other factors than financial crises) would still prevent African and Latin American countries from catching up to developed countries and developing Asian countries.

In Fig. 1, potential African and Latin American labor productivity level is plotted together with the observed long run labor productivity level for Asia and the other developed countries. Because our data set begins in 1973, we set the productivity level to 1 in 1972. As can be seen in Fig. 1, developing Asian countries outpace all other countries. Latin American countries catch up with developed countries in the late 1980s, and both set of countries double their productivity level between 1972 and 2007. African countries, however, still lag behind.

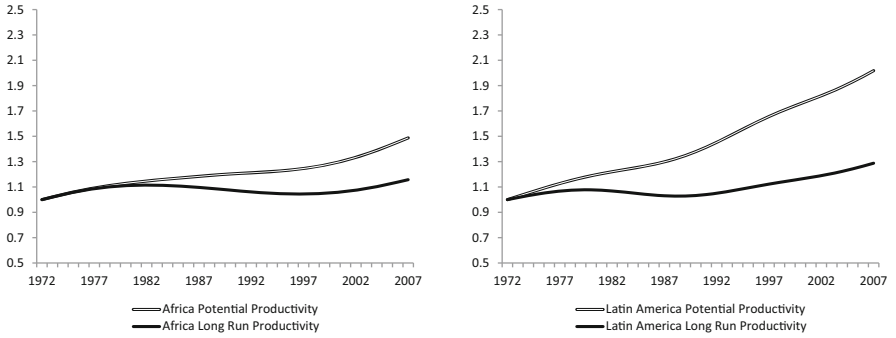


Fig. 2 Observed and potential long-run labor productivity growth for the average African and Latin American country. (Panel **a**) Africa long-run productivity level. (Panel **b**) Latin America long-run productivity level

The difference between the estimated long-run productivity level and the estimated long-run potential labor productivity level are shown in Fig. 2: Africa is in Panel A, and Latin America is in Panel B. As can be seen in the figure, this difference grows persistently over time. In 2007, the actual productivity level was 36.2 % below the potential in Latin America and 22.2 % in Africa. Considering that productivity has been below the potential level since the 1970s, we define, similar to Boyd et al. (2002), the cost of financial crises as the cumulative difference between potential and the actual productivity level,

$$\sum_{i=1973}^{2007} ln\left(potential\ productivity\ level_i\right) - ln\left(long\ run\ productivity\ level_i\right). \tag{17}$$

For African countries, the cumulative cost of financial crises equals 3.92 years of production per employee between 1973 and 2007 and 9.14 years of production per employee for Latin American countries. Despite the fact that financial crises cannot fully explain why Latin American and African countries are lagging behind productivity in developed countries and developing Asian countries, the cost of long term financial crises are substantial over time.

4 Conclusions

Our results show that long-run financial crises can in part explain the poor economic performance of African and Latin American developing countries since the 1970s. Without financial crises over the entire period of 1972–2007, Latin American growth would have equaled that of developed countries. However, Africa would have still lagged behind. Our results suggests that the most influential of all crises are

debt crises, which have affected both African and Latin American countries over the long term. Debt crises are also significantly correlated with inflation and currency crises. Moreover, inflation and currency crises are correlated with Freedom House's political rights index, which suggest that the policy response to the debt crises of the early 1980s made the economic growth consequences of the debt crises worse.

These results also show that even without financial crises, African and Latin American capital accumulation rates would have been lagging behind the rates of developed countries and, in particular, the capital accumulation rates of developing Asian countries. Over the considered period, Asian countries have grown the fastest. Additionally, more than 50 % of their growth is explained by capital growth. Even if financial crises can explain part of the African and Latin American countries poor economic performance, other factors affecting capital growth have contributed significantly.

Bonfiglioli (2008) and Gourinchas and Jeanne (2006) have argued that low productivity growth is worse for a developing country than low capital accumulation rates, as the potential to catch up with rich countries is conditioned on the same level of productivity. Our results show that financial crises, over the long term, affect both capital accumulation and total factor productivity. Our results thus indicate that the crises and their subsequent policy responses have had a severe negative impact on the ability of developing countries to catch up with developed countries.

Financial crises have both short- and long-term economic effects. However, financial crises explain little of the short-term variation in the data. Although financial crises have a negative impact on all countries (not just developing countries), compared to the "normal" short term volatility in the data (caused by non-crises factors), financial crises generate little volatility. The short-term consequences are consequently small compared to the long-term consequences.

Appendix

See Tables 9 and 10.

Table 9 Countries Included in the analysis

Developing countries	Developed countries
Argentina	Australia
Bolivia	Austria
Brazil	Belgium
Chile	Canada
Colombia	Denmark
Costa Rica	Finland
Côte d'Ivoire	France
Dominican Republic	Germany
Ecuador	Greece
Egypt	Ireland
Guatemala	Italy
India	Japan
Indonesia	Netherlands
Kenya	Norway
Malaysia	Portugal
Mexico	South Korea
Morocco	Spain
Nigeria	Sweden
Peru	Switzerland
Philippines	United Kingdom
Singapore	United States
South Africa	
Sri Lanka	
Thailand	
Tunisia	
Turkey	
Uruguay	
Venezuela	
Zambia	
Zimbabwe	

Table 10 Variable description

Variable	Description
Labor productivity	Estimates of labor productivity are collected from the Conference Board's total economy database (http://www.conference-board.org/data/economydatabase)
Capital stock	Capital stock data are estimated using the perpetual inventory method. Real capital investment data come from Penn World Tables 6.3 We assume a fixed depreciation rate of 5 % but also tested a 3 and a 7 % depreciation rate. Changing the depreciation rates has no significant effect on the estimates of the effects of financial crises. We rely on Larson et al. (2000) to obtain an initial capital stock value

(continued)

Table 10 (continued)

Variable	Description
Financial crisis	We rely on Reinhart and Rogoff's (2011) database of financial crises. The database distinguishes between five different crises (inflation, currency, debt, banking and stock market crises). An inflation crisis is defined as annual inflation exceeding 20 %. A currency crisis is defined as the domestic currency losing 15 % of its value against the USD or another relevant currency. A banking crisis is defined as a bank run leading to a bank closure, merger or takeover by the public sector. A banking crisis is also when a bank needs assistance, which spreads to other banking institutions. A debt crisis is when a country defaults on its external debtThis database can be found here: http://terpconnect.umd.edu/~creinhar/Courses-html. The link also contains a detailed description of the data
Education	The education variable measures the increase in the total number of years of schooling among the labor force. The data are collected from the World Development Indicators (http://data.worldbank.org/indicator)
Political rights	We use Freedom House's political rights index. The database can be found here: www.freedomhouse.org
Globalization	To measure globalization, we use the KOF index by Dreher (2006), which combines three dimensions of globalization (economic, social, and political). Economic globalization accounts for 36 % of the index, social globalization for 38 % of the index, and political globalization for 26 % of the index. The database is available from: http://globalization.kof.ethz.ch/

References

Acemoglu D, Johnson S, Robinson J (2003) Institutional causes, macroeconomic symptoms: volatility, crises and growth. J Monetary Econ 50:49–123

Andersson FNG (2008) Wavelet analysis of economic time series. Dissertation, Lund University

Andersson FNG (2011a) Band spectrum regression using wavelet analysis. Lund University Department of Economics Working Paper 2011:22

Andersson FNG (2011b) Monetary policy, asset price inflation and consumer price inflation. Econ Bull 31(1):759–770

Arellano M (1987) Computing robust standard errors for within-groups estimators. Oxf Bull Econ Stat 49:431–34

Assenmacher-Wesche K, Gerlach S (2008a) Money, growth, output gaps and inflation at low and high frequencies: spectral estimates for Switzerland. J Econ Dyn Control 32:411–435

Assenmacher-Wesche K, Gerlach S (2008b) Interpreting Euro area inflation at high and low frequency. Eur Econ Rev 52:964–986

Baltagi BH (2008) Econometric analysis of panel data, 4th edn. Wiley, West Sussex

Baxter M, King RG (1999) Measuring business cycles: approximate band-pass filters for economic time series. Rev Econ Stat 81:575–593

Beck T (2008) The econometrics of finance and growth. The World Bank Policy Research Working Paper 4608

Bergh A, Nilsson T (2010) Good for living? On the relationship between globalization and life expectancy. World Dev 38(9):1191–1203

Bonfiglioli A (2008) Financial integration, productivity and capital accumulation. J Int Econ 76:337–355

Bordo MD, Meissner CM, Stuckler D (2010) Foreign currency debt, financial crises and economic growth: a long-run view. J Int Money Finance 29:642–665
Boyd JH, Kwak S, Smith B (2002) The real output losses associated with modern banking crises. J Money Credit Bank 37:977–999
Bruno M, Easterly W (1998) Inflation crises and long-run growth. J Monetary Econ 41:3–26
Cavallo AF, Cavallo EA (2010) Are crises good for long-term growth? The role of political institutions. J Macroecon 32:838–857
Corbae D, Ouliaris S, Phillips PCB (2002) Band spectral regression with trending data. Econometrica 70:1067–1109
Crowley PM (2007) A guide to wavelets for economists. J Econ Surv 21:207–267
De Gregorior J, Guidotti P (1995) Financial development and economic growth. World Dev 23(1):433–448
Djikstra AG (1997) Fighting inflation in Latin America. Dev Change 28:531–557
Dreher A (2006) Does globalization affect growth? empirical evidence from a new index. Appl Econ 38:1091–1110
Easterly W, Islam R, Stiglitz J (2001) Volatility and macroeconomic paradigm for rich and poor countries: advances in macroeconomic theory. In: Drèze JD (ed) Advances in macroeconomic theory. Palgrave, New York
Eichengreen B, Hausman R (1999) Exchange rates and financial fragility. In: Proceedings Federal Reserve Bank of Kansas City, pp 329–368
Englebrecht H-J, Langley C (2001) Inflation crisis, deflation, and growth: further evidence. Appl Econ 33:1157–1165
Engle RF (1974) Bandspectrum regressions. Int Econ Rev 15(1):1–11
Enisan AA, Olufisayo AO (2009) Stock market development and economic growth: evidence from seven sub-Saharan African countries. J Econ Bus 61:162–171
Friedman M (1957) A theory of the consumption function. Princeton University Press, Princeton
von Furstenberg GM (1977) Corporate investment: does market valuation matter in the aggregate? Brookings Papers Econ Act 8:347–408
Gourinchas P, Jeanne O (2006) The elusive gains from international financial integration. Rev Econ Stud 73(3):715–741
Hausmann R, Gavin M (1996) Securing stability and growth and in a shock prone region. The Policy Challenge for Latin America. Inter-American Development Bank Research Department Working Paper 315
Kaminsky GL, Reinhardt CM (1999) The twin crises: the causes of banking and balance-of-payments problems. Am Econ Rev 89:473–500
Kane EJ, Rice T (2001) Bank runs and banking policies: lessons for African policy makers. J Afr Econ 10:36–71
Labán R, Sturzenegger F (1994) Fiscal conservatism as a response to the debt crisis. J Dev Econ 45:305–324
Larson DF, Butzer R, Mundlak Y, Crego A (2000) A cross-country database for sector investment and capital. World Bank Econ Rev 14:371–91
McNelis PD (1988) Indexation and stabilization: theory and experience. World Bank Res Obs 3:157–169
Norman VL, Romain R (2006) Financial development, financial fragility, and growth. J Money Credit Bank 38:1051–1076
Paiella M (2009) The stock market, housing and consumption spending: a survey of the evidence on wealth effects. J Econ Surv 23:947–973
Percival D, Walden T (2006) Wavelet methods for time series analysis. Cambridge University Press, New York
Pilbeam K (2006) International Finance. Palgrave McMillan, New York
Ranciere R, Tornell A, Westermann F (2008) Systemic crises and growth. Q J Econ 123:359–406
Ramey V, Ramey R (1995) Cross country evidence on the link between volatility and growth. Am Econ Rev 85:1138–1159

Ramsey JB, Lampart C (1998) The decomposition of economic relationships by time scale using wavelets: expenditure and income. Stud Non-Linear Dyn Econom 3:23–42

Reinhart C, Rogoff KS (2011) From financial crisis to debt crisis. Am Econ Rev 101(5):1676–1706

Rousseau PL, Wachtel P (2002) Inflation thresholds and the finance-growth nexus. J Int Money Finance 21:777–793

Rodrik D (2000) Institutions for high-quality growth: what they are and how to acquire them. NBER working paper 7540.

Tobin J (1969) A general equilibrium approach to monetary theory. J Money Credit Bank 1:15–29

Tommasi M (2004) Crisis, political institutions, and policy reform: the good, the bad, and the ugly. In: Tungodden B, Stern N, Kolstad I (eds) Annual World Bank conference on development economic–Europe 2003: toward pro-poor policies: aid, institutions and globalization. World Bank and Oxford University Press, Oxford

Velasco A (1987) Financial crises and balance of payments crises. A simple model of the southern cone experience. J Dev Econ 27:263–283

Wilson B, Saunders A, Gerard CJR (2000) Financial fragility and Mexico's 1994 Peso Crisis: an event-window analysis of market-valuation effects. J Money Credit Bank 32(3):450–468

Measuring Risk Aversion Across Countries from the Consumption-CAPM: A Spectral Approach

Ekaterini Panopoulou and Sarantis Kalyvitis

Abstract Using the Consumption-CAPM, Campbell (2003, Consumption-based asset pricing, Constantinides G, Harris M, Stulz R (eds), Handbook of the economics of finance, Amsterdam, North-Holland) reports cross-country evidence that imply implausibly large coefficients of relative risk aversion, thus confirming the "equity premium puzzle" in an international context. In this paper we adopt a spectral approach to re-estimate the values of risk aversion over the frequency domain. Our findings indicate that at lower frequencies risk aversion falls substantially across countries, thus yielding in many cases reasonable values of the implied coefficient of risk aversion.

1 Introduction and Related Literature

A recurrent puzzle in the macroeconomics and finance literature has been the failure of financial theory to explain the magnitude of excess stock returns by the covariance between the return on stocks and consumption growth over the same period, termed as the "equity premium puzzle" (Mehra and Prescott 1985). Standard asset pricing models, like the Consumption Capital Asset Pricing Model (henceforth C-CAPM), can only match the data if investors are extremely risk averse in order to reconcile the large differential between real equity returns and real returns available on

E. Panopoulou (✉)
Kent Business School, University of Kent, Canterbury CT2 7PE, UK
e-mail: A.Panopoulou@kent.ac.uk

S. Kalyvitis
Department of International and European Economic Studies, Athens University of Economics and Business, Patision Str 76, Athens 10434, Greece
e-mail: skalyvitis@aueb.gr

M. Gallegati and W. Semmler (eds.), *Wavelet Applications in Economics and Finance*, Dynamic Modeling and Econometrics in Economics and Finance 20,
DOI 10.1007/978-3-319-07061-2_11,

short-term debt instruments.[1] Much of the resulting empirical literature has focused on the US markets where longer data series exist, whereas Campbell (1996, 2003) focuses on some smaller stock markets and finds evidence that the "equity premium puzzle" persists. Specifically, Campbell (2003) reports evidence from 11 countries that imply extremely high values of risk aversion, which usually exceed many times the value of 10 considered plausible by Mehra and Prescott (1985), and claims "*. . .that the equity premium puzzle is a robust phenomenon in international data*".

Most empirical studies on the "equity premium puzzle" have focused on relatively short horizons; however, examining the long-run components ("low frequencies") of the puzzle is important because the majority of investors typically have long holding horizons. Indeed, Brainard et al. (1991) have shown that the performance of the C-CAPM improves as the horizon increases, a finding confirmed by Daniel and Marshall (1997) who have found that at lower frequencies aggregate returns and consumption growth are more correlated and the behavior of the equity premium becomes less puzzling. In a series of papers, Parker (2001, 2003) and Parker and Julliard (2005) have allowed for long-term consumption dynamics by focusing on the ultimate risk to consumption, defined as the covariance between an asset's return during a quarter and consumption growth over the quarter of the return and several following quarters, and have found that it explains the cross-sectional variation in returns surprisingly well, but also show that the "equity premium puzzle" is not eliminated.

In this paper we follow step-by-step the approach adopted by Campbell (2003) using the same model and data, in order to re-evaluate over the frequency domain his assessment that the standard, representative agent, consumption-based asset pricing theory based on constant relative risk aversion utility fails to explain the average returns of risky assets in international markets. We choose to proceed using Campbell's (2003) theoretical setup and dataset in order to make our results as comparable as possible and we adopt a spectral approach to re-estimate the values of risk aversion over the frequency domain. According to the spectral representation theorem (Granger and Hatanaka 1964) a time series can be seen as the sum of waves of different periodicity and, hence, there is no reason to believe that economic variables should present the same lead/lag cross-correlation at all frequencies. We incorporate this rationale into Campbell's (2003) approach and dataset in order to separate different layers of dynamic behavior of "equity premium puzzle" by distinguishing between the short run (fluctuations from 2 to 8 quarters), the medium run or business cycle (lasting from 8 to 32 quarters), and the long run (oscillations of duration above 32 quarters). Our findings indicate that in the short run and medium run, the coefficients of risk aversion for the countries at hand are implausibly high, confirming the evidence reported by Campbell (2003). However, at lower frequencies risk aversion falls substantially across countries, thus yielding in many cases reasonable values of the implied coefficient of risk aversion.

[1] Mehra (2003) and Cochrane (2005) provide extensive surveys of the relevant literature.

Our results are in line with evidence from long-run asset pricing. Bansal and Yaron (2004), Bansal et al. (2005) and Hansen et al. (2008) have shown that when consumption risk is measured by the covariance between long-run cashflows from holding a security and long-run consumption growth in the economy, the differences in consumption risk provide useful information about the expected return differentials across assets. Theoretical research on asset pricing using loss aversion theory suggests that time-varying expected asset returns follow a low frequency movement (Barberis et al. 2001; Grüne and Semmler 2008). Semmler et al. (2009) have shown that when there are time-varying investment opportunities, due to low frequency movements in the returns, a buy and hold strategy is not optimal. Readjustments of consumption and rebalancing of the portfolio should therefore follow the low frequency component of the returns from the financial assets in order to increase wealth and welfare.

It is worth noting that the spectral estimation of consumption-based models has also been considered by Berkowitz (2001) and Cogley (2001). Berkowitz (2001) has proposed a one-step Generalized Spectral estimation technique for estimating parameters of a wide class of dynamic rational expectations models in the frequency domain. By applying his method to the C-CAPM he finds that when the focus is oriented towards lower frequencies, risk aversion attains more plausible values at the cost of a risk-free rate puzzle generated by low estimates of the discount factor. Cogley (2001) decomposes approximation errors over the frequency domain from a variety of stochastic discount factor models and finds that their fit improves at low frequencies, but only for high degrees of calibrated risk aversion. Recently, Kalyvitis and Panopoulou (2013) show how low frequencies of consumption risk can be incorporated in the standard (Fama and French 1992) two-step estimation methodology and find that its lower frequencies can explain the cross-sectional variation of expected returns in the U.S. and eliminate the "equity premium puzzle". In this paper we show how low frequencies of consumption risk can be incorporated in Campbell's (2003) empirical setup in an easily implementable way, in order to separate and compare different layers of dynamic behavior of the "equity premium puzzle" across countries by distinguishing between the short run, the medium run (business cycle), and the long run.

We close the introductory section by noting that our approach complements standard time-domain analysis by interpreting (high) low-frequency estimates as the (short) long-run component of the "equity premium puzzle". Yet we stress that the maintained hypothesis is that over any subsegment of the observed time series the precise same frequencies hold at the same amplitudes, resulting in a signal that is homogeneous over time. A straightforward extension therefore to address the empirical limitations of the standard model is to consider state-dependent preferences.[2] As is well known, equity risk premia are higher at business cycles troughs compared to peaks (Campbell and Cochrane 1999). In turn, a number of papers have explored the implications for asset pricing of allowing the coefficient of

[2]We thank an anonymous Referee for pointing out to us this extension.

relative risk aversion to vary with key macroeconomic aggregates. Danthine et al. (2004) allow the pricing kernel to depend on the level of consumption, in addition to its growth rate. In a similar vein, Gordon and St-Amour (2004) provide strong empirical evidence for countercyclical risk aversion, rising during recessions and falling during expansions, by postulating a model with time varying risk aversion depending on per capita consumption. Lettau and Ludvigson (2009) show that the leading asset pricing models fundamentally mischaracterize the observed positive joint behavior of consumption and asset returns in recessions, when aggregate consumption is falling. Another related extension involves the differential impact of structural breaks, crises or 'rare events' in the ex post equity risk premium, which can be correlated in their timing across countries (Barro 2006; Ghosh and Julliard 2012; Nakamura et al. 2013).

Relaxing the assumption of time invariance and allowing for a decomposition of a series into orthogonal components according to scale (time components) gives rise to the wavelet approach, recently applied to economics and finance in the pioneering papers by Ramsey and Lampart (1998a), Ramsey (1999, 2002). Wavelet analysis encompasses both time or frequency domain approaches and can assess simultaneously the strength of the comovement at different frequencies and how such strength has evolved over time.[3] In the context of asset pricing, Gençay et al. (2003, 2005) and Fernandez (2006) have established that the predictions of the Capital Asset Pricing Model model are more relevant at the medium-term, rather than at short-time horizons. Our approach provides a further step towards understanding the frequency components of the "equity premium puzzle" and additional research is warranted to integrate our findings with their time-domain counterpart in the context of wavelet analysis.

The structure of the paper is as follows. Section 2 describes briefly the methodology employed, while Sect. 3 presents the empirical results. Section 4 concludes the paper.

2 Measuring Risk Aversion over the Frequency Domain

We follow Campbell (2003) and we assume a representative investor who faces an intertemporal choice problem in complete and frictionless capital markets. The representative investor can freely trade in some asset i and can obtain a gross return $(1+R_{i,t+1})$ on this asset for the period from time t to $t+1$. Her objective is to

[3]Crowley (2007) and Rua (2012) provide excellent surveys on wavelet analysis, which has been applied to, among others, the examination of foreign exchange data using waveform dictionaries (Ramsey and Zhang 1997), decomposition of the economic relationships of expenditure and income (Ramsey and Lampart 1998a,b), decomposition of the relationship between wage inflation and unemployment (Gallegati et al. 2011) the analysis of the relationship between stock market returns and economic activity (Kim and In 2003).

maximize a time-separable utility function, $U(C_t)$, in consumption, C. The solution to this problem yields the following Euler condition:

$$U'(C_t) = \delta E_t[(1 + R_{i,t+1})U'(C_{t+1})] \tag{1}$$

where δ is the discount factor. The left-hand side of Eq. (1) is the marginal utility cost of consumption while the right-hand side is the expected marginal utility benefit of investing in asset i at time t, selling it at time $t + 1$ and consuming the profits. Given that the investor equates marginal cost and marginal benefit, Eq. (1) describes the optimum. Dividing Eq. (1) by $U'(C_t)$ yields

$$E_t\left[(1 + R_{i,t+1})\delta\frac{U'(C_{t+1})}{U'(C_t)}\right] = 1 \tag{2}$$

where $\delta\frac{U'(C_{t+1})}{U'(C_t)}$ is the intertemporal marginal rate of substitution of the investor, or the stochastic discount factor. Following Rubinstein (1976), Lucas (1978), Breeden (1979), Grossman and Shiller (1981), Mehra and Prescott (1985) and Campbell (2003), we employ a time-separable power utility function $U(C_t) = \frac{C_t^{1-\gamma}}{1-\gamma}$ where γ is the coefficient of relative risk aversion and we get from (2) that:

$$E_t\left[(1 + R_{i,t+1})\delta(\frac{C_{t+1}}{C_t})^{-\gamma}\right] = 1 \tag{3}$$

The power utility specification has many desirable features. Firstly, it is scale-invariant when returns have constant distributions implying that risk premia are not influenced by increases in aggregate wealth or the scale of the economy. Secondly, even when individuals have different initial wealth, we can still aggregate them in a power utility function as long as each individual can be characterized by the same power utility function. The major shortcoming of this traditionally adopted utility function is that it restricts the elasticity of intertemporal substitution to be the reciprocal of the coefficient of relative risk aversion. Weil (1989) and Epstein and Zin (1991) have proposed an alternative utility specification that retains the property of scale invariance without placing any restrictive linkages between the coefficient of relative risk aversion and the elasticity of intertemporal substitution. However, in this study, we concentrate on the power utility specification in order to aid comparison with other studies on developed markets. Furthermore, Kocherlakota (1996) reports that modifications to preferences such as those proposed by Epstein and Zin, habit formation due to Constantinides (1990) or "keeping up with the Joneses" as proposed by Abel (1990) fail to resolve the puzzle.

Following Hansen and Singleton (1983), we assume that the joint conditional distribution of asset returns and consumption is lognormal. With time-varying volatility we get after taking logs that:

$$E_t r_{i,t+1} + \log\delta - \gamma E_t[\Delta c_{t+1}] + 1/2(\sigma_i^2 + \gamma^2\sigma_c^2 - 2\gamma\sigma_{i,c}) = 0 \tag{4}$$

where $c_t \equiv \log(C_t)$, $r_{i,t} \equiv \log(1 + R_{i,t})$, and σ_i^2 and σ_c^2 denote the unconditional variances of log stock return innovations and log consumption innovations respectively, and $\sigma_{i,c}$ represents the unconditional covariance of innovations between log stock returns and consumption growth. Consider now that an asset with a riskless return, $r_{f,t+1}$, exists. For this asset the return innovation variance σ_f^2 and the unconditional covariance of innovations between the log risk free return and consumption growth, $\sigma_{f,c}$, are both zero. Equation (4) becomes:

$$E_t r_{f,t+1} + \log\delta - \gamma E_t[\Delta c_{t+1}] + 1/2\gamma^2\sigma_c^2 = 0 \tag{5}$$

Letting then $e_{i,t+1} \equiv E_t[r_{i,t+1} - r_{f,t+1}]$ denote the excess return over the riskfree rate and subtracting Eq. (5) from Eq. (4) we get:

$$e_{i,t+1} + \frac{\sigma_i^2}{2} = \gamma\sigma_{i,c} \tag{6}$$

Equation (6) suggests that the excess return on any asset over the riskless rate is constant and therefore the risk premium on all assets is linear in expected consumption growth with the slope coefficient, γ, given by:

$$\gamma = \frac{e_{i,t+1} + 0.5\sigma_i^2}{\sigma_{i,c}} \tag{7}$$

Now, departing from the time domain to the frequency domain we can rewrite (7) for each frequency. After dropping the time subscript for simplicity, we get that the coefficient of risk aversion over the whole band of frequencies ω, where ω is a real variable in the range $0 \preceq \omega \preceq \pi$, is given by:

$$\gamma_\omega = \frac{e + 0.5f_{ee}(\omega)}{f_{ec}(\omega)} \tag{8}$$

where e denotes the excess log return of the stock market over the risk-free rate, $f_{ee}(\omega)$ denotes the spectrum of excess returns, and $f_{ec}(\omega)$ denotes the co-spectrum of consumption and excess returns.

The spectrum shows the decomposition of the variance of a series and is defined as the discrete Fourier transform of its autocovariance function:

$$f_{ee}(\omega) = \sum_{k=-\infty}^{\infty} \varphi_k e^{-ik\omega}$$

where ω is a real variable, $0 \preceq \omega \preceq \pi$, and φ_k is the autocovariance function of the series, i.e. $\varphi_k = Cov(e_t, e_{t-k})$.[4] Using the symmetric property of the covariance,

[4]For a detailed analysis, see Hannan (1969), Anderson (1971), Koopmans (1974) and Priestley (1981).

$\varphi_k = \varphi_{-k}$ along with the trigonometric property that $e^{i\omega} + e^{-i\omega} = 2\cos(\omega)$, the spectrum can be rewritten as:

$$f_{ee}(\omega) = \varphi_0 + 2\sum_{k=1}^{\infty}\varphi_k \cos(k\omega)$$

Consider now the bivariate spectrum $F_{ec}(\omega)$ for a bivariate zero mean covariance stationary process $Z_t = [e_t, c_t]^\top$ with covariance matrix $\Phi(\cdot)$, which is the frequency domain analogue of the autocovariance matrix. The diagonal elements of $F_{ec}(\omega)$ are the spectra of the indvidual processes, $f_{ee}(\omega)$ and $f_{cc}(\omega)$, while the off-diagonal ones refer to the cross-spectrum or cross spectral density matrix of e_t and c_t. In detail:

$$F_{ec}(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\Phi(k)e^{-ik\omega} = \begin{bmatrix} f_{cc}(\omega) & f_{ec}(\omega) \\ f_{ce}(\omega) & f_{ee}(\omega) \end{bmatrix} \tag{9}$$

where $F_{ec}(\omega)$ is an Hermitian, non-negative definite matrix, i.e. $F_{ec}(\omega) = F_{ec}^*(\omega)$, where F^* is the complex conjugate transpose of F since $f_{ec}(\omega) = \overline{f_{ce}(\omega)}$. As is well known, the cross-spectrum, $f_{ec}(\omega)$, between e and c is complex-valued and can be decomposed into its real and imaginary components, given here by:

$$f_{ec}(\omega) = C_{ec}(\omega) - iQ_{ec}(\omega), \tag{10}$$

where $C_{ec}(\omega)$ denotes the *co-spectrum* and $Q_{ec}(\omega)$ the *quadrature spectrum.* The measure of comovement between returns and consumption over the frequency domain is then given by:

$$c_{ec}^2(\omega) \equiv \frac{|f_{ec}(\omega)|^2}{f_{ex}(\omega)\, f_{cc}(\omega)} = \frac{C_{ee}^2 + Q_{cc}^2}{f_{ee}(\omega)\, f_{cc}(\omega)} \tag{11}$$

where $0 \leq c_{ec}^2(\omega) \leq 1$ is the squared *coherency*, which provides a measure of the correlation between the two series at each frequency and can be interpreted intuitively as the frequency-domain analog of the correlation coefficient.[5]

The spectra and co-spectra of a vector of time-series for a sample of T observations can be estimated for a set of frequencies $\omega_n = 2\pi n/T$, $n = 1, 2, \ldots, T/2$. The relevant quantities are estimated through the periodogram, which is based on a representation of the observed time-series as a superposition of sinusoidal waves of various frequencies; a frequency of π corresponds to a time period

[5]Engle (1976) gives an early treatment on the frequency-domain analysis and its time-domain counterpart.

of two quarters, while a zero frequency corresponds to infinity.[6] This estimated periodogram is an unbiased but inconsistent estimator of the spectrum because the number of parameters estimated increases at the same rate as the sample size. Consistent estimates of the spectral matrix can be obtained by either smoothing the periodogram, or by employing a lag window approach that both weighs and limits the autocovariances and cross-covariances used. For example, the spectrum of e_t is estimated by:

$$f_{ee}(\omega) = \frac{1}{2\pi} \sum_{k=-(T-1)}^{T-1} w(k)\widehat{\varphi}_k e^{-ik\omega}$$

where the kernel, $w(k)$, is a series of lag windows. We use the Bartlett's window which assigns linearly decreasing weights to the autocovariances and cross-covariances in the neighborhood of the frequencies considered and zero weight thereafter. The number of ordinates, m, is set using the rule $m = 2\sqrt{T}$, as suggested by Chatfield (1989), where T is the number of observations.

3 Empirical Findings

To calculate the coefficient of risk aversion from (8) we use the Campbell (2003) dataset, which combines quarterly data for consumption, interest rates and prices. More in detail, returns are calculated from stock market data sourced from Morgan Stanley Capital International (MSCI), while macroeconomic data on consumption, short-term interest rates and price levels are sourced from the International Financial Statistics (IFS).[7] We present our estimates only for the countries for which at least 100 observations are available in the dataset, namely Australia (1970:1–1998:4), Canada (1970:1–1998:4), France (1973:2–1998:3), Italy (1971:2–1998:1), Japan (1970:2–1998:4), Sweden (1970:1–1999:2), UK (1970:1–1999:1), and the US (1947:2–1998:3 and 1970:1–1998:3). To allow for a direct comparison with the evidence in Campbell (2003), we present two measures of risk aversion. The first one, termed RRA(1), is calculated directly from (8), whereas the second one, denoted by RRA(2), assumes a unitary correlation of excess returns with consumption growth. Although this is a counterfactual exercise, we follow closely Campbell (2003) and we postulate a unitary elasticity between returns and consumption growth to account for the sensitivity of the implied risk aversion on the smoothness of consumption rather than its low correlation with excess returns. We then identify the short-run

[6] For example, the periodogram of $f_{ee}(\omega)$ is given by $I_{ee}(\omega) = g_0 + 2\sum_{k=1}^{T-1} g_k \cos(k\omega)$.

[7] The data are available from http://scholar.harvard.edu/campbell/data. Details on sources and data transformations are given in Campbell (2003).

Table 1 Short-run cross-country estimates of risk aversion

Country	$RRA^{cb}(1)$	$RRA^{cb}(2)$	ρ_{ec}	$\sigma_e = \sqrt{f_{ee}}$	$\sigma_c = \sqrt{f_{cc}}$	c^2_{ec}	RRA(1)	RRA(2)
Australia	58.5	8.4	0.14	6.788	0.715	0.63	33.3	26.5
Canada	59.3	12.0	0.20	6.527	0.514	0.65	89.0	71.6
France	<0	12.3	<0	9.154	0.857	0.70	98.2	82.3
Italy	<0	10.4	<0	10.855	0.516	0.65	21.0	16.9
Japan	82.6	9.3	0.11	7.561	0.844	0.73	53.6	45.8
Sweden	1713.2	26.5	0.02	8.928	0.651	0.62	195.5	154.4
UK	186.0	17.2	0.09	7.727	0.817	0.68	134.7	110.7
US1	240.6	49.3	0.21	6.554	0.380	0.43	449.9	296.5
US2	150.1	41.2	0.27	6.893	0.257	0.54	436.3	319.6

Notes:
(1) $RRA^{cb}(1)$ and $RRA^{cb}(2)$ denote the estimates of risk aversion reported by Campbell (2003). See Sect. 3 for details
(2) ρ_{ec} denotes the correlation coefficient between excess returns and consumption growth as reported by Campbell (2003)
(3) See the text for the definitions of σ_e, σ_c, c^2_{ec}. Both σ_e, σ_c, are reported in annualized percentage points
(4) US1 refers to the sample starting at 1947:2 and US2 at the sample starting at 1970:1

estimates of risk aversion as the averages of fluctuations corresponding from 2 to 8 quarters in the time domain, the medium-run (or business cycle) estimates as the averages of fluctuations from 8 to 32 quarters, whereas the long-run estimates are derived from the averages of oscillations with duration above 32 quarters.

Table 1 presents the short-run spectral estimates of the variabilities of excess log stock return, σ_e, consumption growth, σ^2_c, the squared coherency c^2_{ec}, and the implied coefficients of relative risk aversion, γ_ω. To facilitate the exposition we report next to the country name the values of relative risk aversion from Table 4 in Campbell (2003), termed $RRA^{cb}(1)$ and $RRA^{cb}(2)$ respectively. To gain some insight on the benefits of our proposed methodology, we also report the time domain correlation coefficient, ρ_{ec}, as reported in Campbell (2003). The time domain coefficients of relative risk aversion ($RRA^{cb}(1)$) are in general extremely large ranging from 58.5 (Australia) to 1713.2 (Sweden), while negative coefficients pertain to the cases of France and Italy. In general, reported correlations (ρ_{ec}) are low (below 0.30) and even negative at the cases of France and Italy. Even allowing for a perfect comovement between consumption growth and excess returns, given by $RRA^{cb}(2)$, estimated coefficients still exceed the value of 10 for all countries at hand, with the exception of Australia and Japan. Turning to our spectral estimates, we find an increased comovement between consumption and returns at high frequencies as suggested by the estimated coherency (c^2_{ec}). The respective values range from 0.43 (US for the sample starting 1947:2) to 0.70 (France). However, our results at high frequencies corroborate the evidence by Campbell (2003) suggesting that risk aversion at high frequencies is found to be extremely large (with the possible exception of Italy). More in detail, estimated risk aversion coefficients range from 21 (Italy) to 449.9 (US for the sample starting 1947:2). More importantly, this picture

Table 2 Medium-run cross-country estimates of risk aversion

Country	$\sigma_e = \sqrt{f_{ee}}$	$\sigma_c = \sqrt{f_{cc}}$	c_{ec}^2	RRA(1)	RRA(2)
Australia	9.609	0.414	0.74	44.3	38.1
Canada	8.535	0.756	0.76	45.4	39.6
France	8.007	1.011	0.65	97.3	78.5
Italy	12.638	0.529	0.62	21.9	17.3
Japan	8.593	0.633	0.67	67.6	55.3
Sweden	9.412	0.326	0.59	383.9	294.3
UK	9.682	0.624	0.63	149.5	118.5
US1	7.856	0.392	0.80	270.9	242.7
US2	7.008	0.401	0.68	245.3	202.2

Notes: see Table 1

Table 3 Long-run cross-country estimates of risk aversion

Country	$\sigma_e = \sqrt{f_{ee}}$	$\sigma_c = \sqrt{f_{cc}}$	c_{ec}^2	RRA(1)	RRA(2)
Australia	10.860	3.651	0.68	5.0	4.1
Canada	9.147	3.402	0.63	10.6	8.4
France	13.002	2.613	0.84	22.1	20.3
Italy	13.047	3.512	0.26	5.2	2.6
Japan	7.076	4.796	0.68	10.3	8.5
Sweden	14.688	2.336	0.97	28.5	28.1
UK	12.736	3.704	0.97	16.2	15.9
US1	10.723	3.788	0.87	20.4	19.1
US2	13.121	3.144	0.89	16.2	15.3

Notes: see Table 1

continues to hold under the assumption of a unitary elasticity between excess returns and consumption growth and is in line with the findings typically reported in the literature on the C-CAPM.

Table 2 performs the same exercise for the medium-run or business-cycle frequencies. As the time horizon increases, the variabilities of consumption growth and returns along with their correlation in general increase. The medium-run coherency exceeds 0.59 for all the countries at hand and reaches 0.80 for the US. As a result risk aversion is in general lower, but is still found to be implausibly high exceeding the value of 10, even when a unitary correlation is imposed. The lower estimate is 21.9 and 17.3 for Italy, respectively. Thus we find that the equity premium puzzle persists at business-cycle frequencies.

Next, we turn our attention to the long-run, i.e. low frequencies where we find that the performance of the C-CAPM improves substantially. As shown in Table 3, the coefficients of risk aversion now range between 5.0 (Australia) to 28.5 (Sweden). When a unitary correlation coefficient is imposed, these estimates are slightly reduced for all the countries at hand and range from 4.1 to 28.1. This improvement in the low-frequency estimates of relative risk aversion is driven by the spectral properties of the data at hand. As we move to lower frequencies, the variability of

consumption growth, σ_c^2, increases significantly, reaching even ten times its high-frequency value matching the variability of log excess returns, σ_e^2, and as such the covariance of returns and consumption increases. This property is coupled for most of the countries at hand with a rise in the estimated coherency between consumption and returns.[8]

4 Conclusions

The paper attempts to re-address the empirical issue of implausibly high risk aversion within the context of the C-CAPM by looking at the pattern of risk aversion over the frequency domain. Our results show that as lower frequencies are taken into account, risk aversion falls substantially across countries and, in many cases, is consistent with more reasonable values of the coefficient of risk aversion. This evidence shows some improvement towards understanding the dynamics of the C-CAPM by reconciling its standard single-factor version with lower values of risk aversion and thus the equity premium over the frequency domain appears to be less of a puzzle.

However, we emphasize that a limitation of our paper is that the point estimates of long-run risk aversion remain relatively high and a more in-depth analysis is warranted to align the model with reasonable coefficients of risk aversion across countries. To this end, a number of studies provide interesting insights on the cross-country aspects of the 'equity premium puzzle'. For instance, Bekaert (1995) and Henry (2000) point out that the cost of capital decreases as markets get integrated due to risk sharing that implies a lower required risk premium. Wavelet analysis, which can assess simultaneously the strength of the comovement at different frequencies and how this strength has evolved over time, offers a promising route for further research in this area.

References

Abel AB (1990) Asset prices under habit formation and catching up with the Joneses. Am Econ Rev 80:38–42

Anderson T (1971) The statistical analysis of time series. Wiley, New York

Bansal R, Yaron A (2004) Risks for the long run: a potential resolution of asset pricing puzzles. J Financ 59:1481–1509

Bansal R, Dittmar RF, Lundblad CT (2005) Consumption, dividends, and the cross section of equity returns. J Financ 60:1639–72

[8]It is worth mentioning that the coherency is not maximised at the lowest frequency for all countries. The coherency reaches its maximum in the short-run (2–6 quarters) for Italy and Japan and in the medium-run (2–4 years) for Australia and Canada.

Barberis N, Huang M, Santos R (2001) Prospect theory and asset prices. Q J Econ 116:1–53
Barro RJ (2006) Rare disasters and asset markets in the twentieth century. Q J Econ 121:823–66
Bekaert G (1995) Market integration and investment barriers in emerging equity markets. World Bank Econ Rev 9:75–107
Berkowitz J (2001) Generalized spectral estimation of the consumption-based asset pricing model. J Econom 104:269–288
Brainard WC, Nelson WR, Shapiro MD (1991) The consumption beta explains expected returns at long horizons. mimeo, Yale University
Breeden DT (1979) An intertemporal asset pricing model with stochastic consumption and investment opportunities. J Financ Econ 7:265–296
Campbell JY (1996) Consumption and the stock market: interpreting international experience. NBER Working Papers 5610
Campbell JY (2003) Consumption-based asset pricing. In: Constantinides G, Harris M, Stulz R (eds). Handbook of the economics of finance. Amsterdam, North-Holland
Campbell JY, Cochrane J (1999) Force of habit: a consumption-based explanation of aggregate stock market behavior. J Polit Econ 107:205–251
Chatfield C (1989) The analysis of time series. Chapman and Hall, London
Cochrane J (2005) Financial markets and the real economy. Found Trends Financ 1:1–101
Cogley T (2001) A frequency decomposition of approximation errors in stochastic discount factor models. Int Econ Rev 42:473–503
Constantinides G (1990) Habit formation: a resolution to the equity premium puzzle. J Polit Econ 98:519–543
Crowley P (2007) A guide to wavelets for economists. J Econ Surv 21:207–264
Daniel K, Marshall DA (1997) Equity-premium and risk free-rate puzzles at long horizons. Macroecon Dyn 1:452–484
Danthine J-P, Donaldson JB, Giannikos C, Guirguis H (2004) On the consequences of state dependent preferences for the pricing of financial assets. Financ Res Lett 1:143–153
Engle RF (1976) Interpreting spectral analysis in terms of time-domain models. Ann Econ Soc Meas 5:89–109
Epstein L, Zin SE (1991) Substitution, risk aversion and the temporal behaviour of consumption growth and asset returns: an empirical investigation. J Polit Econ 99:263–286
Fama EF, French KR (1992) The cross-section of expected stock returns. J Financ 47:427–465
Fernandez V (2006) The CAPM and value at risk at different time-scales. Int Rev Financ Anal 15:203–219
Gallegati M, Gallegati M, Ramsey JB, Semmler, W (2011) The US wage Phillips curve across frequencies and over time. Oxford Bull Econ Stat 73:489–508
Gençay R, Whitcher B, Selçuk F (2003) Systematic risk and time scales. Quant Financ 3:108–16
Gençay R, Whitcher B, Selçuk F (2005) Multiscale systematic risk. J Int Money Financ 24:55–70
Ghosh A, Julliard C (2012) Can rare events explain the equity premium puzzle? Rev Financ Stud 25:3037–3076
Gordon S, St-Amour P (2004) Asset returns with state-dependent risk preferences. J Bus Econ Stat 22:241–252
Granger CWJ, Hatanaka M (1964) Spectral analysis of economic time series. Princeton University Press, Princeton
Grossman SJ, Shiller RJ (1981) The determinants of the variability of stock market prices. Am Econ Rev 71:222–227
Grüne L, Semmler W (2008) Asset pricing with loss aversion. J Econ Dyn Control 32:3253–3374
Hannan EJ (1969) Multiple time series. Wiley, New York
Hansen LP, Singleton KJ (1983) Stochastic consumption, risk aversion and the temporal behavior of asset teturns. J Polit Econ 91:249–268
Hansen LP, Heaton JC, Li N (2008) Consumption strikes back? measuring long run risk. J Polit Econ 116:260–302
Henry P (2000) Market integration, economic reform, and emerging market equity prices. J Financ 55:529–564

Kalyvitis S, Panopoulou E (2013) Estimating C-CAPM and the equity premium over the frequency domain. Stud Nonlinear Dyn Econom 17(5):551–572
Kim S, In FH (2003) The relationship between financial variables and real economic activity: evidence from spectral and wavelet analyses. Stud Nonlinear Dyn Econom 7:1–18
Kocherlakota NR (1996) The equity premium: its still a puzzle. J Econ Lit 34:42–71
Koopmans LH (1974) The spectral analysis of time series. Academic, New York
Lettau M, Ludvigson SC (2009) Euler equation errors. Rev Econ Dyn 12:255–283
Lucas R (1978) Asset prices in an exchange economy. Econometrica 46:1429–1445
Mehra R (2003) The equity premium: why is it a puzzle? Finan Anal J 59:54–69
Mehra R, Prescott EC (1985) The equity premium: a puzzle. J Monet Econ 15:145–161
Nakamura E, Steinsson J, Barro RJ, Ursúa J (2013) Crises and recoveries in an empirical model of consumption disasters. Am Econ J Macroecon 5:35–74
Parker JA (2001) The consumption risk of the stock market. Brookings Pap Econ Act 2:279–348
Parker JA (2003) Consumption risk and expected stock returns. Am Econ Rev Pap Proc 93:376–382
Parker JA, Julliard C (2005) Consumption risk and the cross-section of expected returns. J Polit Econ 113:185–222
Priestley MB (1981) Spectral analysis and time series, vol I and II. Academic, New York
Ramsey JB (1999) The contribution of wavelets to the analysis of economic and financial data. Phil Trans R Soc A Math Phys Eng Sci 357:2593–2606
Ramsey JB (2002) Wavelets in economics and finance: past and future. Stud Nonlinear Dyn Econ 6(3). doi:10.2202/1558-3708.1090
Ramsey JB, Lampart C (1998a) Decomposition of economic relationships by time scale using wavelets. Macroecon Dyn 2:49–71
Ramsey JB, Lampart C (1998b) The decomposition of economic relationship by time scale using wavelets: expenditure and income. Stud Nonlinear Dyn Econ 3:23–42
Ramsey JB, Zhang Z (1997) The analysis of foreign exchange data using waveform dictionaries. J Empir Financ 4:341–372
Rua A (2012) Wavelets in economics. Economic Bulletin and Financial Stability Report Articles, Bank of Portugal
Rubinstein M (1976) The valuation of uncertain income streams and the pricing of options. Bell J Econ 7:407–425
Semmler W, Grüne L Örlein C (2009) Dynamic consumption and portfolio decisions with time varying asset returns. J Wealth Manage 12:21–47
Weil P (1989) The equity pemium puzzle and the risk-free rate puzzle. J Monet Econ 24:401–421