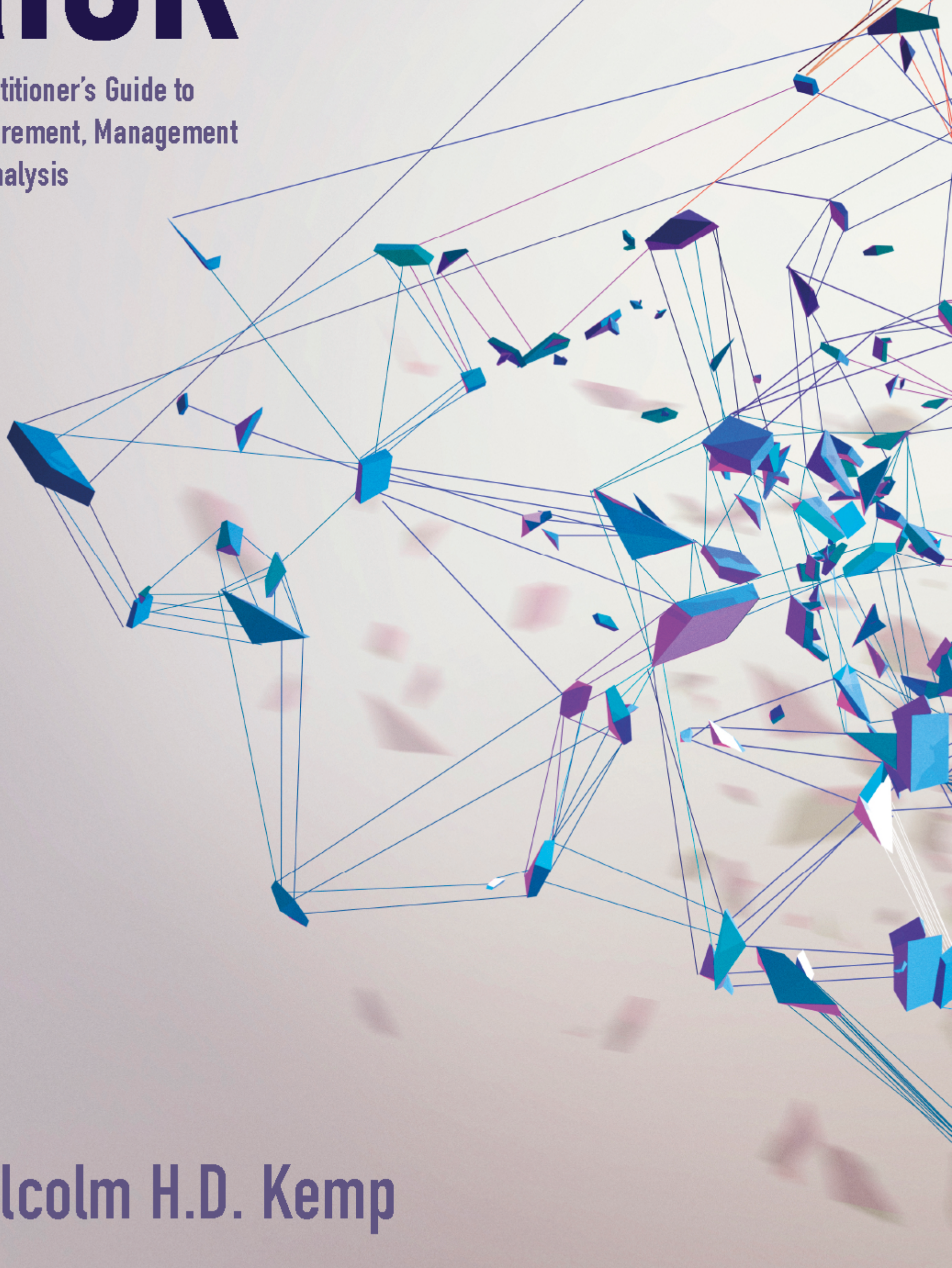
SYSTEMIC RISK

A Practitioner's Guide to Measurement, Management and Analysis



Malcolm H.D. Kemp

Systemic Risk

Malcolm H. D. Kemp

Systemic Risk

A Practitioner's Guide to Measurement, Management and Analysis

palgrave
macmillan

Malcolm H. D. Kemp
Nematrian Limited
London, United Kingdom

ISBN 978-1-137-56586-0 ISBN 978-1-137-56587-7 (eBook)
DOI 10.1057/978-1-137-56587-7

Library of Congress Control Number: 2017939908

Cover illustration: Mina De La O / Getty

Printed on acid-free paper

This Palgrave Macmillan imprint is published by Springer Nature
The registered company is Macmillan Publishers Ltd.
The registered company address is: The Campus, 4 Crinan Street, London, N1 9XW, United Kingdom

Preface

Economists and other commentators often remind us of the importance of a soundly functioning financial system. It is perhaps surprising that financial systems don't fall over more often. We all share a common interest in financial stability. Daily life is much easier if the financial system of which we are a part is behaving robustly. However, shared interests aren't sufficient to eliminate fragilities. Financial stability is a public amenity. Like many other public resources, it is all too easy to understate its worth when times are good, and to believe that the costs of keeping it functioning smoothly should be borne largely by others.

All these facets were visible in the 2007–2009 Credit Crisis that engulfed much of the Western world. It wasn't as severe as the 1930's Great Depression and doesn't (yet?) seem to have created the sort of fertile breeding ground for new political ideologies that led to the traumas of the Second World War. However, in the depths of the Crisis there were certainly worries that it might do so. This influenced some of the unconventional economic responses that were then adopted.

The Crisis was also sufficiently severe to result in authorities applying a largely new term to such situations, i.e. *systemic risk*, or at least to redefine how this term might be used in this context. Lots of new initiatives were started that collectively aimed to make the financial system more robust in the face of systemic risk. New bodies were set up specifically to focus on it. Additional roles and responsibilities were given to central banks, given the perceived strong interplay between systemic risk and monetary policy.

When viewed over the sorts of timescales historians typically analyse, financial collapses are not so unusual. Many regime changes, particularly violent ones, have resulted in financial upheavals. Within the financial community the tendency is to limit the notion of 'systemic risk' to events and exposures that are not this extreme. The focus is usually on stresses that can somehow be pinned on faults within the financial system (or within the economic system of which it is a part) rather than on more existential threats deriving from wider political developments. But this narrower focus is not necessarily how those outside the financial community interpret the term.

For those outside the financial community, *systemic risk* is most obviously something that affects systems. Systems of one sort or another have become commonplace across nearly all aspects of life. Firms and individuals access computer systems every day of the year. Computer systems have the unfortunate habit of falling over from time to time. Almost any sort of system has vulnerabilities. In this respect, financial systems differ only in terms of scale, extent of interaction with difficult to predict human behaviours and magnitude of undesirable consequences when they go wrong.

One reason why applying the term 'systemic risk' to the financial system seems relatively novel is because we haven't suffered too many large-scale systemic risk events until recently. The 2007–2009 Credit Crisis was a once in a working lifetime event (we hope!), akin to the 1930s Great Depression. It still gets mentioned all the time at financial conferences even though eventually it too will fade in the memories of leaders, influencers and policymakers. Systemic events have of course still been happening in the meantime, just not quite so large. Commentators may describe events like the Savings & Loans crisis in the USA in the 1980s or the failure of Long Term Capital Management (LTCM) in the 1990s as systemic risk events.

A consequence of this relative novelty is that there is no fixed view on exactly what systemic risk involves or which types of organisation are most likely to create, transmit or be affected by it. People do not even agree on whether its causes are primarily exogenous (i.e. caused by something outside the system) or endogenous (i.e. caused by vulnerabilities within the system). This lack of consensus is frustrating to those involved in business planning within the financial sector. How can you effectively respond to a fickle and shifting topic? One of the aims of this book is to survey this otherwise complex scene, to help practitioners best respond to such uncertainties.

There is a tendency (more prevalent outside the regulatory community than within it) to assume that systemic risk is largely or wholly to do with the banking system. Most policy measures introduced to date that can be explicitly badged 'macroprudential' (and hence focused largely or wholly on systemic risk) have indeed related to banking. The banking system was at the epicentre of the last Crisis. However, those within the regulatory community tend to adopt a broader focus. Sorting out the part of the financial system that suffered the most during the last Crisis has some undoubted logic. But doing so to the exclusion of *other* elements of the financial system runs the risk of underemphasising issues that may be at the heart of the *next* crisis.

The very broad impact of (financial) systemic risk events and the complexities of the system itself (and what should be deemed inside or outside it) make systemic risk both highly important and fascinating to analyse. We can be almost certain that systemic crises will happen again in the future. My main aim with this book is to help equip those who work (or aspire to work) in the financial system (and beyond) with a better understanding of systemic risk and how it can be measured, managed and analysed. I hope that by doing so readers will be better able to respond appropriately to (and ideally even profit from) current trends that are placing increasing importance on systemic risk.

Acknowledgements

I would like to thank Aimee Dibbens, Nicole Tovstiga and their colleagues at Palgrave Macmillan (and before them, Pete Baker) for encouraging me to write this book. Thanks are also due to Paweł Fiedor and colleagues at the European Systemic Risk Board and to Enrico Biffis, Paul Sweeting, Javier Suarez and Colin Wilson who read parts of this manuscript and provided helpful comments on how it might be improved. A special appreciation goes to my wife and family for supporting me as this book took shape.

Contents

Abbreviations

ABCP	Asset-backed commercial paper
ABS	Asset-backed securities
ADC	Acquisition, development and construction
AIF	Alternative investment fund
AIFM	Alternative investment fund manager
AIFMD	(EU) Alternative Investment Fund Managers Directive
AMA	Advanced measurement approach
AUM	Assets under management
BCBS	(BIS's) Basel Committee on Banking Supervision
BCR	(IAIS's) Basic Capital Requirement
BIS	Bank for International Settlements
BRRD	(EU) Bank Recovery and Resolution Directive (i.e. Directive 2014/59/EU)
CAS	Complex adaptive system
CCP	Central (clearing) counterparty
CDC	Collective defined contribution (pensions)
CDO	Collateralised debt obligation
CDS	Credit default swap
CLO	Collateralised loan obligation
CNAV	Constant (or stable) net asset value
COMFRAME	Common supervisory framework (that is being developed by IAIS)
CoVaR	Value at risk (VaR) of financial system conditional on institution being in distress
CPSS	(BIS) Committee on Payment and Settlement Systems
CPU	Central processor unit
CRD	(EU) Capital Requirements Directive
CRDV01	Credit DV01
CRR	(EU) Capital Requirements Regulation
CTE	Conditional tail expectation

CVaR	Conditional Value-at-Risk
DB	Defined benefit (pensions)
DC	Defined contribution (pensions)
DJIA	Dow Jones Industrial Average (index)
Dodd-Frank	(US) Dodd-Frank Wall Street Reform and Consumer Protection Act
DTCC	Depository Trust & Clearing Corporation
DV01	Dollar value of a one basis point change in interest rates /credit spreads
EAD	Exposure at default
EBA	European Banking Authority
ECAI	External credit assessment institution
ECB	European Central Bank
ECL	Expected credit loss
EIOPA	European Insurance and Occupational Pensions Authority
EL	Expected loss
EMIR	European Market Infrastructure Regulation
ESBies	European Safe Bonds
ESFS	European System of Financial Supervision
ESM	European Stability Mechanism
ESRB	European Systemic Risk Board
ES	Expected shortfall
ETF	Exchange traded fund
FASB	(US) Financial Accounting Standards Board
FC	Financial counterparty
FCA	(UK) Financial Conduct Authority
FCL	Financial corporation engaged in lending
FDIC	(US) Federal Deposit Insurance Corporation
Fed	(US) Federal Reserve Board
FHLBB	(US) Federal Home Loan Bank Board
FinTech	Financial technology
FMI	Financial market infrastructure
FNAV	Floating net asset value
FPC	(UK) Financial Policy Committee
FRTB	Fundamental review of the trading book
FSA	(UK) Financial Services Authority
FSB	Financial Stability Board
FSLIC	(US) Federal Savings and Loan Insurance Corporation
FSOC	(US) Financial Stability Oversight Council
FVA	Fair value accounting
FVC	Financial vehicle corporation (engaged in securitisation transactions)

G-SIB	Global systemically important bank
G-SIFI	Global systemically important financial institution
G-SII	Global systemically important insurer /institution
G20	Group of Twenty international forum for governments and central banks
GAAP	Generally accepted accounting principles
GAR	Guaranteed annuity rate
GDP	Gross domestic product
HCA	Historic /amortised cost accounting
HLA	Higher loss absorbency
IAIS	International Association of Insurance Supervisors
IAS	International Accounting Standard
IASB	International Accounting Standards Board
ICAAP	Internal capital adequacy assessment process
ICS	(IAIS's international) Insurance Capital Standard
ICG	Individual Capital Guidance
IFRS	International Financial Reporting Standard
ILAAP	Internal liquidity adequacy assessment process
IM	Internal model
IMF	International Monetary Fund
IORP	Institution for occupational retirement provision
IOSCO	International Organization of Securities Commissions
IRB	Internal ratings based
IRDV01	Interest rate DV01
IRRBB	Interest rate risk in the banking book
IT	Information technology
LCR	Liquidity coverage ratio
LDC	Less developed country
LEI	Legal entity identifier
LGD	Loss given default
LIBOR	London interbank offered rate
LTCM	Long Term Capital Management
LTV	Loan-to-value
MA	Matching adjustment
MBS	Mortgage backed security
MiFID	(EU) Markets in Financial Instruments Directive
MiFIR	(EU) Markets in Financial Instruments Regulation
MMF	Money market fund
MES	Marginal expected shortfall
MTE	Marginal tracking error
MTVaR	Marginal tail value-at-risk
MVaR	Marginal value-at-risk

NAMA	(Irish) National Asset Management Agency
NAV	Net asset value
NBNI	Non-bank non-insurer
NFC	Non-financial counterparty
NPL	Non-performing loan
NSFR	Net stable funding ratio
NTNI	Non-traditional non-insurance
O-SII	Other systemically important institution
OFI	Other financial institution
OIS	Overnight indexed swap
OFR	(US) Office of Financial Research
PBGC	(US) Pension Benefit Guarantee Corporation
PD	Probability of default
PIT	Point-in-time
PPF	(UK) Pension Protection Fund
PPS	Pension protection scheme
PRA	(UK) Prudential Regulation Authority
RWA	Risk-weighted assets
S&L	Savings and loan institution
SAMC	Spanish Asset Management Company
SA	Standard assessment
SCAP	(US) Supervisory Capital Assessment Program
SEC	(US) Securities and Exchange Commission
SES	Systemic expected shortfall
SF	Standard formula
SFT	Securities financing transaction
SIFI	Systemically important financial institution
SIV	Structured investment vehicle
SMA	Standardised measurement approach
SME	Small and medium enterprise
Solvency II	(EU) Solvency II Directive, i.e. Directive 2009/138/EC (recast)
SPV	Special purpose vehicle
SREP	Supervisory review and evaluation process
SSM	(EU) Single Supervisory Mechanism
TARP	(US) Troubled Asset Relief Program
TBTF	Too big to fail
TCF	Treating customers fairly
TE	(Ex-ante) tracking error
TLAC	Total loss-absorbency capacity
TTC	Through-the-cycle
TVaR	Tail value-at-risk

UCITS	(EU) Undertakings for collective investment in transferable securities
UL	Unexpected loss
UFR	Ultimate forward rate
VA	Volatility adjustment
VaR	Value-at-risk
XBRL	eXtensible Business Reporting Language

List of Figures

List of Tables

List of Boxes

1

Introduction

The invention of money, several thousand years ago, ranks as one of humanity's more important inventions. Niall Ferguson cogently argues in his 2009 book *The Ascent of Money*, see Ferguson (2009), that the extent to which a society adopts modern financial practices (such as those underpinning modern capital markets, banking and insurance) strongly links to its overall competitiveness. Money makes practical the effective division and specialisation of labour and of other factors of economic production. It allows us to borrow or save depending on whether our current productive activities are less than or greater than our immediate consumption needs. Systemic risk (in a financial context) is about the ways in which this backdrop can get disrupted (and what we can do to mitigate this risk). Who amongst us, in their heart of hearts, would be happy to return to stone-age barter?

Systemic risk as applied to finance has many analogies with the analysis of risks that other (non-financial) systems can exhibit. We will introduce some of these analogies as the book progresses. However, this is primarily a book on finance. Unless the context expects otherwise, when we refer to 'systemic risk' we will mean the sort of systemic risk that applies to the financial system. However, we won't limit ourselves to risks that arise solely from within the financial system. To do so would ignore perhaps the most common historic cause of financial system collapse, i.e. war, revolution or (other) violent regime change causing the collapse of the political or societal system within which any specific financial system resides.

Fundamentally, the study of (financial) systemic risk derives from the realisation that financial systems are fragile and can seize up or otherwise

M.H.D. Kemp, *Systemic Risk*, DOI 10.1057/978-1-137-56587-7_1

fail to operate as intended if they suffer a severe enough stress. The focus is on the potential collapse of the entire financial system (or more precisely a major part of it, such as some specific financial market or sector) rather than on the failure of (isolated) individual components that individually do not threaten to bring the rest of the system down. The study of systemic risk is important precisely because money is such an important part of how economies and societies operate.

Implicit in its study is a belief that the financial system has an 'intended' mode of operation. The expectation is that its behaviour has some elements of design rather than being merely the random outcome of other economic factors. This brings systemic risk within the remit of the political process. The general aim is for the financial system to provide a useful contribution to wider well-being (otherwise why would we care if it was stable or not?). But what this means in practice will be influenced by what we (and others) view as 'useful' and by our views about how governments should operate within the economic sphere.

All these factors mean that identifying a *precise* definition of (financial) systemic risk is not straightforward. Definitions influence thought-processes and hence decisions adopted. Mindful of this, politicians and other commentators may work backwards from end decisions they would most like others to reach to select definitions they think are most likely to lead to these end outcomes.

Usually a reasonably generic definition of systemic risk is adopted, giving politicians and policymakers quite a lot of flexibility to guide the outcome as they so wish. An example of a relatively broad definition is one given in the EU Regulation that established the European Systemic Risk Board (ESRB):

> 'Systemic risk' means a risk of disruption in the financial system with the potential to have serious negative consequences for the internal market and the real economy. All types of financial intermediaries, markets and infrastructure may be potentially systemically important to some degree, see European Union (2010).

This is the definition that we will mostly stick to in this book although we will explore in more detail in Chapter 2 why some commentators prefer broader definitions and others narrower ones.

We have structured the remainder of this book as follows

(a) Chapter 2 explores further what we mean by 'systemic risk', the 'financial system' to which we think it applies and to the concepts of financial stability, procyclicality and macroprudential policy, which are at the heart of interpretations of the term ‘systemic risk’. We also introduce two overarching views of what systemic risk might involve, the ‘domino’ view and the ‘tsunami’ view and why they colour how we think about systemic risk more generally.

(b) Chapter 3 explores further the interaction between capital, solvency and systemic risk. It illustrates how most systemic risk events fit neither the domino view nor the tsunami view in isolation, instead requiring the two to happen in combination.

(c) Chapter 4 surveys current developed market financial systems. We describe the activities of different players in the financial services industry and of those who interact with them. We explore how they are typically viewed by macroprudential policymakers (which may not always be the ‘right’ way to view them, but does at least help us to understand where macroprudential policymakers are coming from). Direct players include banks, institutional investors such as insurers and pension funds, asset managers (and the funds that they manage on behalf of others) and other market ‘utilities’ such as exchanges and clearing houses. Others who interact with such players include governments, regulators (and other central authorities), advisors, non-financial corporates and the public.

(d) Chapter 5 explores how to measure systemic risk, both at an individual firm level and at a system-wide level. We draw out the difference between how academic theorists typically propose measuring systemic risk (which tends to have a more ‘domino’ like view of systemic risk) and how policymakers typically seem to analyse systemic vulnerabilities in practice (which tends to adopt a broader perspective).

(e) Chapter 6 discusses the macroprudential policies that have been proposed (and in some cases already implemented) by regulators, politicians and other commentators. We highlight some of the issues generic to such policies as well as some issues specific to policy measures that have already gained some traction.

(f) Chapter 7 explores broader trends in the current technological and societal environment that are influencing finance and the direction of travel of macroprudential policy. We also highlight some of the underlying rationales behind a range of current regulatory initiatives and how these interact with systemic risk.

(g) Finally, Chapter 8 draws together ideas on how firms (and individuals) can best respond to current developments in finance linked to systemic risk. Hopefully these ideas will help readers plan better for changes that are happening because of the increased societal focus on (financial) systemic risk, perhaps even profiting from the direction of travel involved.

Throughout the book the aim is to be both as insightful as possible and as helpful as possible to its readers. There is tension between these two aims. An insightful book will typically want to explore broad themes and be relatively succinct. However, in a field like systemic risk, which is relatively new and complex, adopting a more detailed and systematic exposition can be helpful.

To address this tension, we have sought to make available additional resources on systemic risk via the Nematrian website, see http://www.nematrian.com/SystemicRisk.aspx. These resources include descriptions of mathematical tools and quantitative techniques that can be used to study, analyse and manage systemic risk. They also include a searchable library of references to what some others have written on these topics. Unless otherwise stated, the figures included in this book are copyright Nematrian Limited 2017 and have been reproduced with the kind permission of Nematrian. Copies of many of them are available through the Nematrian website.

If you fail to find what you are looking for in these references then it is worth noting that bodies such as the Financial Stability Board (FSB), European Systemic Risk Board (ESRB), International Monetary Fund (IMF), central banks and other regulators have not been idle in the years since the 2007–09 Credit Crisis. There is now a wealth of publications on systemic risk from such organisations, usually available via their websites. These include regular financial stability reviews and occasional or other working series papers focusing on specific topics of interest to the macro-prudential community.

References

European Union (2010). Regulation (EU) No 1092/2010. *European Union.*

Ferguson, N. (2009). *The Ascent of Money*. Penguin Books, London, UK.

2

Systemic Risk and the Financial System

The working definition of systemic risk that we typically use throughout this book is: *"'Systemic risk' means a risk of disruption in the financial system with the potential to have serious negative consequences for the internal market and the real economy. All types of financial intermediaries, markets and infrastructure may be potentially systemically important to some degree"*, see European Union (2010).

However, it is important to understand early on that embedded within any such definition are lots of implicit assumptions, some of which will significantly colour our interpretation of systemic risk and of what constitutes the financial system, to which (financial) systemic risk is supposed to apply.

In this Chapter, we provide an overview of these implicit perspectives and we introduce different mindsets that commentators tend to adopt when thinking about systemic risk. The mindsets that industry representatives typically adopt are not identical to those that tend to be adopted by policymakers and regulators, because their perspectives differ.

2.1 Reasons for Adopting Broader Definitions of Systemic Risk

The definition of systemic risk given above is relatively broad, but even so, some commentators favour more malleable and potentially even broader definitions for systemic risk. Reasons include:

M.H.D. Kemp, *Systemic Risk*, DOI 10.1057/978-1-137-56587-7_2

(a) *Any* system, financial or otherwise, typically exhibits fragilities. For example, ecosystems can become damaged, computer systems can fall over, exam systems can become discredited, incentive systems can become misaligned, health systems can become overloaded, pension systems can fail to meet the needs of employers and former employees etc. An important strand of research in (financial) systemic risk involves exploring the sorts of (systemic) risk that other types of system face and then applying the insights to the financial system. For this to work, the definition selected needs to be broad enough to permit such an analogy to be developed.

(b) The financial system does not exist in isolation. Instead, it has a broader purpose, which is to support the economy more generally. Choosing whether to address a systemic risk exposure involves a trade-off between the economic consequences of it materialising and the economic consequences of trying to address it. The precise definition adopted can implicitly tilt this trade-off.

(c) Sometimes the definition used may link to the political motivations of the individual or organisation involved. Even the economic system (of which the financial system is just a part) does not exist in isolation. Instead it too needs to be seen in the context of its overriding purpose. But individuals of different political hues may disagree about what this purpose is. Some may view its purpose from a purely utilitarian angle, e.g. exclusively targeting the maximum wealth for the maximum number of (today's) individuals. Others may have different political ideals, including a desire to bequeath a sustainable environment (or political system or economy, etc.) to future generations. Even a cursory historic review of the robustness of past economic and financial systems shows just how important the political dimension can be.

2.2 Reasons for Narrowing the Definition

Conversely, some commentators favour narrower definitions for systemic risk.

For example, not everyone agrees on what forms the 'financial system'. Do occupational (or state?) pension arrangements belong to the financial system? To some it may seem self-evident that they do, because they involve financial commitments. To others, such a linkage may be anathema, since they see the financial system as largely or wholly synonymous with the banking system and view pensions as quite unlike banking. Moreover, pension commitments are often very long-term in nature.

Related to these questions is the question of whether analysis of systemic risk should focus primarily on might happen in the relatively here and now (e.g. the possibility that a bank might suffer a run tomorrow or next month)? For an example of what we mean by a bank 'run' see Section 4.1.2 or e.g. Box 3.2. Or should it also be concerned with developments that might take many years to materialise? Uncertainties usually cloud any longer-term analyses.

Again, taking pensions as an example, in many jurisdictions pension arrangements are closely aligned with employment arrangements, which may be viewed as being more aligned with the 'real' economy than with the financial system per se. Following this line of reasoning, state pension arrangements might be viewed as even less a part of the financial system. People do not generally think of a country's welfare state as being part of its financial system (except perhaps if they work in the country's finance departments and need to figure out how the costs of the welfare state should be met!).

Commentators can also be influenced by the geographical application of the concept of systemic risk. Is systemic risk about the global financial system or just the financial system in a specific geographical location or jurisdiction? Should the focus be on an individual country, an entire region, the entire world or some other agglomeration? EU Directives that refer to systemic risk tend to give primary focus to the 'system' within each individual member state (whilst still recognising that there may be spill-overs across member states or more globally). The USA has states but they have a different constitutional status to EU member states. The focus of applicable US legislation is more at the federal than at the state level. Supranational bodies, such as the Financial Stability Board (FSB), the Basel Committee on Banking Supervision (BCBS), the International Association of Insurance Supervisors (IAIS) and the International Monetary Fund (IMF) tend by their nature to focus more on the supranational financial system.

Where the focus is on individual geographies or jurisdictions then commentators may concentrate on the structure of the system in their own locality, see Fig. 2.1. Pension funds are relatively small in relation to other parts of the financial system in some jurisdictions but are more important in others. Banks tend to be important in most jurisdictions. In Europe banks form a particularly large part of the financial system, a feature that has led some commentators to argue that Europe is 'overbanked', see ESRB (2014c).

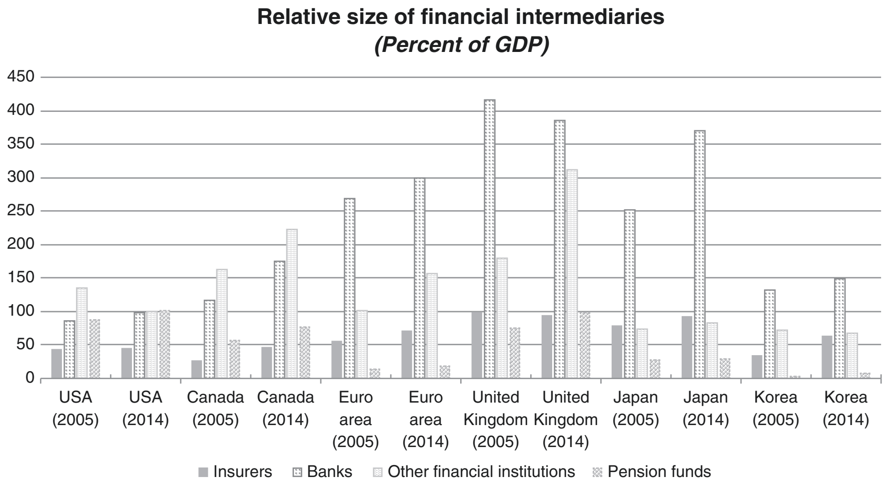


Fig. 2.1 Relative size of financial institutions

Source: Nematrian. Adapted from IMF (2016). There is no single measure that uniquely identifies the size of the financial system, either globally or within a specific country. The above chart gives approximate sizes of some parts of the financial system in some jurisdictions.

2.3 Interconnectedness and Domino Effects

Systemic risk may also be closely associated with the propensity of the financial system to suffer from domino effects. For example, Stephen Schwarcz writing in March 2008 argued that 'A common factor in the various definitions of systemic risk is that a trigger event, such as an economic shock or institutional failure, causes a chain of bad economic consequences – sometimes referred to as a domino effect', see Schwarcz (2008).

The assumption that systemic risk is *solely* about a system's propensity to suffer from domino effects involves several implicit assumptions which can have a significant impact on how we think about systemic risk. Some debates about the scope of systemic risk ultimately derive from these implicit assumptions, so it is worth bringing them out into the open early on in this book.

Conventionally in this context, propensity to suffer from domino effects is referred to by the term *interconnectedness*. A highly interconnected system has more opportunity for individual failures to propagate to others, so is more susceptible to such domino effects.

An analogy can help to explain why interconnectedness is not the sole factor to consider when analysing systemic risk. Consider, for example, a local community such as a seaside town. How might its smooth functioning, its ‘system’ of operation, so to speak, be seriously disrupted?

Focusing on domino effects, i.e. interconnectedness, is like focusing on the potential risk that the town might suffer from an outbreak of a serious contagious disease. The more interconnected the community, e.g. the more that individual participants in the community interact with each other by visiting each other or going to each other’s shops, business premises, schools or other public places, the more likely such an outbreak is to grow into an epidemic (if the disease is virulent enough). Some of the tools that health authorities use to limit the spread of an epidemic, such as closing schools, naturally resonate with a domino perspective. For example, we might view hiving off problem or non-performing loans (NPLs) from a troubled bank into a separate ring-fenced entity as akin to quarantining that part of the bank and limiting the likelihood that the problem will spread to other entities.

Ideas borrowed from the field of epidemiology *do* have relevance to (financial) systemic risk. But to assume that it is the *only* field worth drawing from leaves us vulnerable to other types of systemic risk that are not so strongly linked to interconnectedness.

We can see this by asking whether there are other ways in which a seaside town can ‘fail’. Ephesus was a major seaport in the first and second centuries AD, probably at the time one of the ten largest cities in the Roman Empire. However, it is now a ruin. The primary reason was that its harbour silted up, rendering its commercial heart unviable. Its economy was overwhelmed by a structural problem that arguably had little to do with how interconnected its individual members were.

IMF (2016) includes a variant on this analogy, by referring to how financial stability can be imperilled either by ‘domino’ effects or by ‘tsunami’ effects, see Fig. 2.2. Their analogy was probably influenced by the tsunami triggered by the Tōhoku earthquake on 11 March 2011 that overwhelmed the Japanese seaside town of Fukushima. The tsunami was severe enough to flood the town’s nuclear power plant over the next few days causing a major nuclear disaster. This had many repercussions, including temporary suspension of nuclear power generation across Japan and disruption of global supply chains. Later that year, flooding in Thailand uncovered other fragilities in global supply chains. Propagation of problems due to

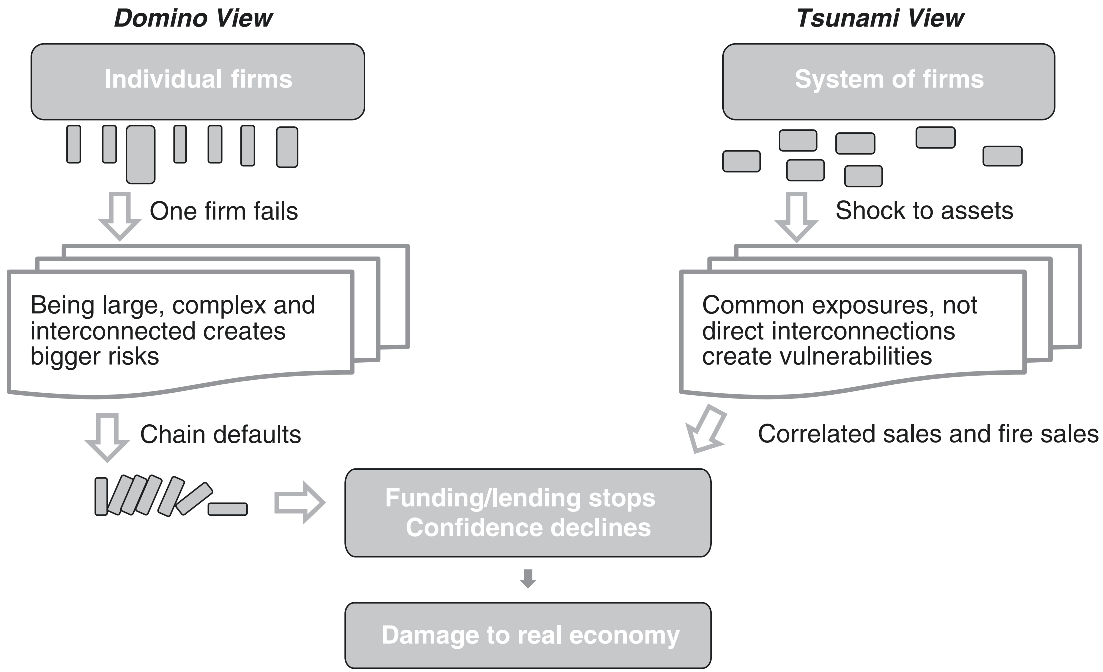


Fig. 2.2 Domino versus tsunami views of interconnectedness

Source: Nematrian. Adapted from IMF (2016).

interconnectivities shouldn't be ignored, but does not by itself provide the whole picture. It does not always help us identify where large scale disruptions might come from.

2.4 Hidden Vulnerabilities and Tsunamis

An underlying assumption of the epidemic analogy is that the starting state of firms (in 'normal' times) in the part of the financial system being considered is essentially healthy. Some external shock then comes along that triggers failure /problems for one firm. If the way in which different firms are interconnected means that this can lead to large adverse impacts on other firms then the initial shock can trigger a cascade of failures /problems elsewhere. In the most extreme case, the cascade creates an explosive chain reaction and the financial system is wrecked. Physicists will recognise that an atomic bomb just needs a critical mass of fissile material to be gathered together. Make the fissile mass sufficiently large (and concentrated) and elements that drive how a chain reaction propagates tip into an unstable state, making an explosion essentially certain.

But what if for some reason the starting 'normal' state of a significant fraction of firms is unhealthy but this is not yet obvious to outside observers? In such a scenario, interconnectedness becomes less relevant to how the system might evolve. Indeed, little interconnectedness might be a hindrance. It could reduce the likelihood that otherwise unhealthy organisations get a leg up from the remainder through some means or other.

The extreme (perhaps illustrated by Ephesus) is if the entire cohort is unsound. Hopefully we won't come across this too often, but history suggests that it can't be ruled out entirely, particularly in times of technological or political upheaval.

Systemic risk events are generally 'tail events', i.e. hopefully very low probability but also high severity events. But we can kid ourselves about just how severe are most of the ones in the (financial) systemic risk space we might come across. Put in the context of the damage that a nuclear war, a major asteroid impact, a global pandemic or Yellowstone erupting, and the 2007–09 Financial Crisis doesn't seem quite so calamitous. Conversely, the entire system doesn't need to be completely wiped out for a systemic event to feel decidedly painful.

The most obvious consequence of an over-focus on (direct) interconnectedness is to place too much emphasis on sectors that are highly interconnected and not enough on sectors that may have hidden sector-wide vulnerabilities. Banking tends to be more interconnected than most other sectors. Of course, that doesn't stop banking sectors also potentially having structural weaknesses. As we shall discuss in the next Chapter, it is perhaps it is when these come to the fore *alongside* interconnectedness that dangers become most acute.

2.5 Systemic Risk and Political Risk

Worrying about whether an entire cohort might be unsound is more usually consigned to the field of 'future gazing'. Systemic events that wipe out entire business cohorts tend to be associated with (perhaps violent) political regime change. Many commentators would view this type of event as 'too extreme' to be within the remit of (just) financial systemic risk, unless the regime change can be fingered as having been mainly caused by financial instability. Traditionally, the term 'systemic risk' has usually been deemed to refer to *risks arising from the propagation of shocks inside the financial system*. Only more recently, has the term been expanded to refer to *any risks to the whole financial system*.

The problem with this line of reasoning is that some causal link is often reasonably plausible to postulate. Historians like to remind us that the French Revolution was significantly facilitated by structural weaknesses in the French economy and financial system in the years leading up to 1789, see Box 2.1.

Box 2.1: The Mississippi Company, the South Sea and other historical bubbles

According to Ferguson (2009), John Law, born in Edinburgh in 1671 and at one time financial controller to the French monarchy, was *'an ambitious scot, a convicted murderer, a compulsive gambler and a flawed financial genius . . . He may also be said to have caused, indirectly, the French Revolution by comprehensively blowing the best chance that the ancien régime monarchy had to reform its finances.'*

After being sentenced to death for duelling he escaped from prison and fled to Amsterdam in c. 1692. Here he came across a number of financial innovations such as arguably the world's first central bank (the Amsterdam Exchange Bank) and one of the first joint-stock companies, the Dutch East India Company, originally founded in 1602 from the merger of around six predecessor companies. John Law appears to have become enamoured with these financial innovations. He also appears to have become convinced in his own mind that they could be improved upon by combining the properties of a monopoly trading company with a public bank that issued notes in the manner of the Bank of England (the Bank of England had been created in 1694). He proposed the idea to the Scottish parliament in 1705 (then separate from the English parliament as this was before the Act of Union in 1707). The central element of his proposal involved a new bank that would issue interest-bearing notes that would supplant coins as currency. He made a similar proposal to the Duke of Savoy in 1711. He is reputed to have told a friend *'I have discovered the secret of the philosopher's stone* . . . *it is to make gold out of paper'.*

He finally succeeded in winning approval for his ideas from the French monarchy, even though they appear to have realised that he was a professional gambler. Ferguson (2009) argues that this was due to the especially desperate nature of France's fiscal problems. It had an enormous public debt (because of Louis XIV's wars) and was teetering on bankruptcy. Law's first proposal for a public note-issuing bank was submitted to the royal council in 1715 but rejected as it involved the bank also acting as the crown's cashier (and hence handling all its tax payments). Law did however manage to get a private bank, Bank Générale, founded in 1716. In 1717 he also managed to obtain a decree that Bank Générale notes should be used for all tax payments. Mindful of the success that the Dutch were having in expanding their empire (via the Dutch East India Company) a new Company of the West (*Compagnie d'Occident*) was set up in 1717 with a monopoly of the commerce of Louisiana, then a French possession. In 1719 this company took over the French East India and China companies to become the Company of the Indies (*Compagnie des Indes*), better known as the Mississippi Company. A couple of months later, Law secured the profits of the royal mint for a nine-year term. Later that year he took control of indirect and direct taxes. He also agreed to lend the Crown sufficient money to pay off the entire royal debt. Colloquially this was known as the Law 'System'.

However, these acquisitions were financed not out of company profits but simply by issuing new shares. For the System to work, it needed to create a financial bubble. Dilution of existing shares due to the issuance of new shares should have caused the price to decline, but instead the shares were talked up by reference to future profits from Louisiana. Prices of Mississippi Company shares peaked in December 1719. Law and his royal patrons then started to resort to a range of artificial measures to prop up its share price, including ones that made its banknotes legal tender and made it illegal for private citizens to hold more than a certain amount of metal coin. Eventually, meltdown occurred, when a violent public outcry highlighted limits to the Royal absolutism that in effect underpinned the System.

Law fled the country in December 1720, having lost his fortune. But in Ferguson's opinion, the losses to France were more than just financial. The boom and subsequent bust set back France's financial development, putting the French off paper money and stock markets for generations. The French monarchy never resolved its fiscal crisis, and royal bankruptcy eventually precipitated revolution. Perhaps the Law System can be thought of as an early example of an organisation (here the French monarchy) 'gambling for resurrection', i.e. betting big in the hope that the gamble comes good. Regulators often worry about this sort of behaviour in a systemic risk context.

At about the same time Britain was also suffering from an extraordinary popular delusion and a madness of crowds, as Mackay (1841) puts it, in the form of the South Sea Bubble. It too exhibited classic symptoms of a financial bubble, such as:

(1) *Economic innovation*: some apparently new economic opportunity comes along
(2) *Euphoria*: there is rapid growth in share prices as future profits become hyped up
(3) *Mania*: apparently easy capital gains attract further investors
(4) *Distress*: smarter investors (or maybe just those with better inside knowledge) work out that expected profits cannot justify continuation of now exorbitant prices
(5) *Discredit*: a more general stampede for the exit occurs, causing the bubble to burst (or merely to an unwinding of a 'crowded trade', if prices haven't hit fully fledged bubble territory)

The South Sea bubble was much less extreme than the Mississippi Company bubble, ruining fewer people. At their peaks, stock prices had risen by a factor of 'only' 9.5 in the case of South Sea stock rather than 19.6 for the Mississippi Company. The prices of other British stocks, such as those of the Bank of England and the English East India Company rose substantially less. The South Sea bubble seems to have created little lasting systemic damage to the English financial system.

Other major stock market bubbles and subsequent crashes also paint contradictory pictures. Some, like the Wall Street crash of 1929 had dramatic impacts, see Box 4.1. Others, like the dot com bubble from c. 1995–2001, led to spectacular share price rises and falls but little wider systemic risk fall out.

One thesis might be that bubbles that are market-wide have a greater propensity to create systemic risk shocks. This would seem to fit the 1973-74 bear market which affected all major stock markets, particularly the UK's. In the two years from 1972 to 1974 the US economy slowed from 7.2% pa real GDP growth to −2.1% and inflation jumped there (and elsewhere, particularly UK, leading to an era of stagflation). Conversely, the October 1987 stockmarket crash saw some of the largest one-day moves ever in several major stockmarkets without appearing to have much systemic impact. Even the dot com boom and bust (which is most classically associated only with certain industries) led to substantial movements in wider stockmarket index levels.

More probably, systemic impact requires some level of hidden vulnerability. The 1973–74 bear market came after the collapse of the Bretton Woods system of currency management that had been established around the end of the Second World War. It was compounded by the outbreak of the 1973 oil crisis, highlighting the vulnerability of Western economies to oil price shocks. The Wall Street Crash was followed by erection of trade barriers worldwide, highlighting a dependence on the benefits of globalisation that politics proceeded to dismantle.

Massive financial instabilities don't have to be linked to violent political change, but when they happen they often create significant political stresses that in due course can be impractical to contain without significant political ramifications. This does seem to have been a fear amongst some policymakers during the 2007–09 Credit Crisis, and influenced their willingness to try out radical monetary policy solutions in the hope that they would address some of the issues involved. Examples of financial instabilities that were either the result of political issues or in due course led to significant political change include some of the hyperinflations and other de facto repudiations of earlier financial commitments that occurred during the twentieth century, see Box 2.2.

Box 2.2 Systemic risks related to hyperinflation

In economic terms, inflation is the sustained increase in aggregate prices. Hyperinflation is very high inflation, perhaps where the monthly inflation rate is a sustained rate of 30–50% per month or higher for at least a few months in a row. There is no specific threshold that all economists agree on, so the aim of such a threshold is to exclude exceptional one-off price adjustments, e.g. due to a very bad harvest in an agrarian economy. Hyperinflation as defined above is a largely twentieth-century phenomenon. The most spectacular examples typically occurred between World War I and World War II or, for some countries such as Hungary, in the immediate aftermath of World War II. Between August 1922 and November 1923 aggregate prices in Germany rose by a factor of 1.02×10^{10}, i.e. on average quadrupled each month. From August 1945 to July 1946 the general level of prices in Hungary rose more than 19,000% per month (in July 1946 more than tripling each day).

Such strains typically result in the collapse of the monetary system of the state in question. They also reallocate wealth, both from the public (which holds money) to the government (which issues money), and between borrowers and lenders. The collapse in the usefulness of money typically has a serious negative impact on a country's economic efficiency as it forces people to revert to barter.

Hyperinflations are caused by extremely rapid growth in the supply of 'paper' money, and usually occur when monetary and fiscal authorities issue sufficiently large amounts of paper money to fund government expenditure. Hyperinflation can therefore be thought of as a form of taxation suffered by whoever needs to use paper money for whatever reason. Up to a certain point, hyperinflations tend to be self-perpetuating, even self-reinforcing, with individuals trying to spend paper money as quickly as possible to minimise the amount of 'inflation tax' they are being forced to pay, and governments trying to issue paper money as quickly as possible to maximise the effective spending power being generated by the 'tax'. Usually they reach such a fever-pitch that something gives, e.g. there is political collapse or a major monetary reform is implemented (the German hyperinflation mentioned above was eventually solved by introduction of a new currency that could be converted on demand into a bond having a certain value of gold).

More recent examples have tended to occur in Latin America and former Eastern bloc nations, and in some African nations. In Latin America, they often led to 'dollarization', i.e. use of dollars in place of the domestic currency.

Political developments do not need to be violent to create systemic risks. For example, Kriwaczek (2010) notes that the financial system that flowered in Hammurabi's Babylon several thousand years ago had many features like modern financial systems, including credit elements. It also had a tendency from time to time to involve a build-up of general indebtedness that could at times grow so large as to threaten the financial or even political stability of the state. A radical solution appears to have been adopted, involving a general 'debt forgiveness', in which all loans (or perhaps just short-term personal loans) were declared null and void. These tended to occur on the accession of a new king but were sometimes promulgated mid-reign. Kriwaczek notes, unsurprisingly, that politically inspired debt-remissions often decimated businesses that involved lending money to others.

2.6 Systemic Risk and Societal Change

As risk managers may attest, an important element to a robust risk management framework is the inclusion of horizon-scanning for 'emerging risks' i.e. risks that are not well quantified or even necessarily well-articulated but nevertheless have the potential to disrupt business models significantly. In this book, we will explore some of these emerging risks. However, we need to accept that such horizon-scanning is necessarily an inexact science. It is impossible to come up with an exhaustive list of such risks, particularly for something as disparate and heterogeneous as a modern financial system.

However, it is possible to envisage lesser 'revolutions' that affect only specific parts of the economy and/or financial system. At the time of writing, worries are surfacing about whether the core business propositions offered by, amongst others, commercial banks, life insurers and central clearers have longer-term futures. Other industries have been substantially disrupted by technological innovation. Will money-orientated businesses suffer a similar fate at the hands of 'FinTech', see Box 2.3?

Box 2.3: FinTech

Financial technology (Fintech) is a catch-all term for information technology (IT) that relates to the finance industry. For many years, firms within the finance industry have been heavy consumers of IT services. Indeed, some large investment banks have sometimes been characterised as specialist types of IT firms, given their level of dependence on IT!

At the time of writing, some commentators are wondering about the possibility that *standalone* FinTech companies (or other new entrants from outside the financial services industry) might reshape the finance industry. For example, PwC (2016) believes that more than 20% of financial services business is at risk of migrating from existing institutions to FinTech firms by 2020. Despite (or sometimes because of) substantial past investment in IT, mainstream financial services players are often viewed as struggling to develop IT-enabled financial services that have compelling customer appeal. The proposition made by FinTech enthusiasts is that it is now difficult to imagine a world without the internet or mobile devices. They have brought a high degree of disruption to virtually every area of business. These enthusiasts argue that the internet and mobile devices are likely to do the same for the financial services industry, since the digital revolution is also transforming the way customers access its products and services.

There are some notable examples of emerging disruption in this area, including WeChat, a successful Chinese mobile-messaging service. A recurring dream of businesses in the internet age is the cashless economy. In mid-2013, very few people bought things using WeChat (known as Weixin in mainland China). In contrast, by mid-2016, Economist (2016) notes that roughly a third of WeChat's users were making regular e-commerce purchases directly through its app. Other social networks such as Facebook are seeking emulate its success in this part of the financial services value chain. If everyone used such tools to make all their purchases, wouldn't it decimate the role of banks in the realm of consumer banking and payment services?

Most FinTech activities are less eye-catching but the general theme is similar. The Fintech firm targets a part of the value-chain that banks or other financial institutions do not currently do well but is valued by some customers. It typically then seeks to create a compelling internet or mobile orientated solution that will appeal to today's digital-savvy customers and steal them from existing players. If the FinTech firm is particularly successful it then expands out its solution to related activities and/or gets snapped up by a bigger fish (such as a tech titan or an existing financial services firm).

Commentators who are more pessimistic about the inroads that FinTech firms may achieve typically highlight the challenges they face, including the heavily regulated nature of the financial services industry. Probably the picture is somewhere in between. Existing players do have some competitive advantages because of e.g. greater familiarity with regulatory issues. However, they also have competitive disadvantages such as legacy systems and less nimbleness when it comes to implementing new ideas in a digitally enabled form. The worry for existing players is that the areas where they have competitive advantage are viewed as of limited value by customers, so the profitable parts of their business will be hollowed out by the upstarts.

From a systemic risk perspective, FinTech, if it is as successful as its enthusiasts propose, is likely to lead to transformed structures within the financial services industry, a breakdown of divisions between the financial sector and the rest of the economy, perhaps higher likelihood of IT system failures and consequential contagion risks and probably less ability to control what is going on (since less of the total value-chain will be within the financial services sector and therefore practically capable of being subject to regulatory oversight by financial services regulators). Respondents to the PwC survey described in PwC (2016) ranked areas most likely to be disrupted by FinTech as shown in Fig. 2.3.

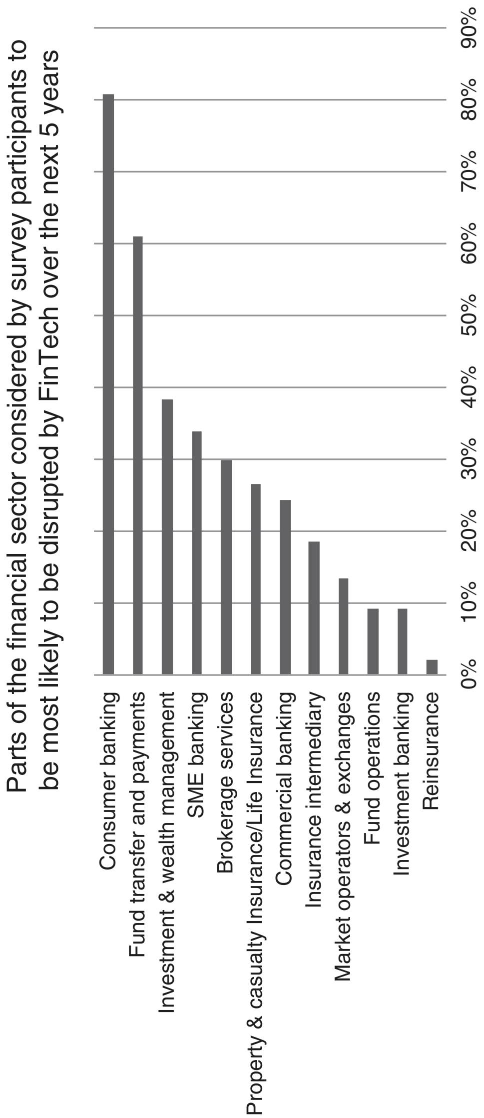


Fig. 2.3 Parts of financial sector subject to FinTech disruption
Source: Nematrian. Adapted from PwC (2016).

Historically, innovation in the financial services sector has often been catalysed by regulatory change. FinTech may be no exception. For example, in August 2016 the UK's Competition and Markets Authority (UK C&MA) completed a review of retail and small business banking in the UK, see Competition & Markets Authority (2016). It identified several areas where it considered existing competition to be deficient and imposed several remedies to be introduced during 2017 and 2018. These include requirements on the largest players to agree suitable standards for open application programming interfaces (APIs), open data and open data sharing. These may facilitate price and service comparisons between players and later allow customers to consolidate their transaction histories etc. across providers (and for example allow automated sweeping from one provider's account to another's, given prior customer consent to such transactions).

Many people seem happy for FinTech to revolutionise the financial sector. Wolf (2015) notes that *'Information technology has disrupted the entertainment, media and retail businesses and, most recently, the supply of hotel rooms and taxis. Is it going to do the same to finance? My first response is: please. My second response is: yes. As Bill Gates has said, 'We always overestimate the change that will occur in the next two years and underestimate the change that will occur in the next 10.' Don't let yourself be lulled into inaction.'* However, those in the regulatory community may not be so enthusiastic. The Riksbank was reported in Spring 2015 to be worried about the pace of developments in payments technology and to believe that technological developments might be proceeding 'too rapidly' (with a risk that evolution of payment systems could jeopardise broader access for those who are less tech savvy).

2.7 Financial Stability

Closely aligned with systemic risk is the concept of financial stability. Indeed, most people involved in systemic risk treat financial stability and systemic risk as essentially two sides of the same coin. The goal is to maintain financial stability. Systemic risk is the risk that the financial system becomes 'unstable'. This then begs the question: what does 'unstable' look like?

Views on this topic can have an important impact on how we think about systemic risk. In the Preface, we note that financial stability can be thought of as a public good or amenity which all of us benefit from and which somehow society needs to foster if it is to prosper.

Returning to the epidemic analogy, the behaviour of an epidemic (in the absence of steps to mitigate it) can be characterised by the average infection rate (i.e. the number of individuals that any given individual who gets infected goes on to infect). If the rate is below unity then the epidemic will eventually die out. Above one and eventually many people will be infected.

The worst-case scenario is if the infection rate per disease carrier is greater than one *and* the disease is highly virulent, leading to death in a high proportion of cases. Panic ensues.

One way of returning such a system to stability is to reduce the infection rate per disease carrier, e.g. by quarantining people exhibiting symptoms of the disease carrier, thus reducing their interaction with others. However, there is a more radical alternative. This is to not have any interaction in the first place. We can guarantee 100% financial stability by dismantling the financial system. In days gone by, the wealthy typically decamped from their city residences whenever there was a hint of plague in the air, leaving the hapless poor to their own devices. Countries can do likewise in the financial sphere. They can for example abandon their own currency if it has become discredited enough, adopting some other more credible currency instead. This seems to be a common outcome when hyperinflation strikes, see Box 2.2.

Even if a soundly functioning financial system is beneficial to the economy (as nearly everyone believes) then seeking financial stability *to the exclusion of all else* probably isn't a rational strategy for a society. We might be able to reduce our risk of catching an infectious disease by never going out, but how would the rest of life then proceed?

The corollary is that maximising the usefulness to society of the financial system is likely to make it susceptible to bouts of financial instability from time to time. Again, we face a (societal) trade-off. The more we seek financial stability, the less effective the financial system may be at helping to achieve other desirable societal aims. Conversely, we all know the havoc that can be wrecked by a large mudslide or earthquake. These sorts of disasters typically arise because the system of which they are a part changes from a stable to an unstable configuration (even if some more proximate event can often be fingered as the immediate cause of the disaster).

Achieving financial stability is hard because the financial system (and the economy of which it is a part) is a complex adaptive system that has self-adjusting and self-reinforcing features, see Box 2.4. These characteristics make it challenging to understand and predict how it will react to different stimuli, whether these stimuli are intentional or by accident.

Box 2.4: Complex adaptive systems

When discussing how the financial system might react to shocks, some commentators focus on features they share with other complex adaptive systems (CASs). The development of thinking behind CASs took hold in the mid-1980s with the formation of the Santa Fe Institute, a New Mexico think tank formed in part by former members of the nearby Los Alamos National Laboratory. It emphasised crossing traditional disciplinary boundaries when studying systems (deliberately drawing from diverse disciplines such as economics, physics, biology, ecology and archaeology). The aim was to identify common theoretical strands that addressed complexity and could provide better understanding of the spontaneous, self-organising dynamics evident in real-life systems. There are many examples of CASs in both the natural and the human world. In the natural world, they include nervous systems (brains), immune systems and developing embryos. In the human world, they include scientific communities, the economy and society more generally.

One important insight is that CASs generally require some external energy source to drive their development. Otherwise they suffer from the consequences of the Second Law of Thermodynamics and eventually lose discernible structure. A corollary is that they generally don't operate in an equilibrium mode. This makes application to systemic risk of traditional (i.e. neo-classical) economic theory potentially suspect. Most of this theory implicitly assumes, up to a point, that the economy is (broadly speaking) in equilibrium at all points of time, shifting only because the equilibrium conditions are changing through time. It has been recognised for many years that such theory does not provide a complete picture of how economies operate (see e.g. Box 2.5 which discusses business cycles). Some commentators however take this one stage further, seeking largely to abandon traditional economic theory, particularly when it comes to interpretation and management of extreme events such as ones of most relevance to systemic risk.

Non-equilibrium modes of thought are not alien to scientific or even economic thinking. Some years ago in mathematical circles the fields of 'catastrophe theory' and 'chaos theory' became fashionable. Catastrophe theory noted that many dynamical systems do not exhibit an equilibrium in the conventional mathematical sense. Instead they cycle irregularly, typically around a limited number of discernible states of the world, occasionally switching from being broadly aligned with one mathematical state to being broadly aligned with another one, particularly if some sufficiently large external shock bounces them between the two. Chaos theory is a related mathematical discipline that analyses in more detail the underlying randomness that such systems often exhibit even in the absence of external stimuli.

In theory, such results are applicable to economic and financial systems. The difficulty is that whilst the theory correctly predicts behaviour of the sort that is consistent with behaviour seen in the real world, it also makes it very difficult to predict in advance exactly how systems might react to external stimuli.

We can still nevertheless identify insights that help us understand systemic risk and how it might arise and/or propagate. For example, self-adjusting systems, including economic systems, often have some features akin to sand dunes. If the dune is steep enough then adding a little more sand to it will result in most of the added sand merely trickling down the side, with the dune getting slightly higher and slightly wider in the process. There is a maximum slope that a (dry) sand dune can exhibit. When it is in this configuration the conflicting forces of gravity and friction are just balanced and we can view the entire sand dune as in some sense only 'marginally' stable (for any individual grain of sand on the outside of the sand dune).

Such systems are self-adjusting from one perspective (i.e. they adjust their height when sand is added to them) but are on the edge of criticality from another perspective (i.e. the overall configuration is at the maximum possible angle above which the dune would collapse). Adjusting the dynamical characteristics (e.g. here adding water, turning the sand into mud) can make the 'system' temporarily stable beyond its otherwise 'natural' state. However, if some other external factor (an earthquake, excessive rain, . . .) changes its dynamical properties still further then what was previously stable can become unstable, and disaster (here in the form of a mudslide) can happen.

Financial sector participants also seem at times to want to push the envelope to its boundary. Indeed, a strong focus on shareholder returns or any other business metric is likely to incentivise participants to go as far as possible in optimising delivery of this metric. The result is that the marginal player (and sometimes a whole class of players) can end up taking on additional debt or additional risk (in pursuit of additional return) at close to the maximum aggregate sustainable rate. Change the dynamical features so that the maximum aggregate sustainable rate falls and we have the recipe for disaster.

2.8 Procyclicality

Debates about financial stability often also include debates about *procyclicality*. An activity is procyclical if it tends to exacerbate behaviour associated with the financial, economic or business cycle.

It is usually taken as self-evident that economies undergo booms and busts, growth spurts and recessions. When business optimism is falling, companies become reluctant to invest in new productive activities that may not deliver a decent return and economic growth becomes subdued (or negative). Eventually, pessimism becomes overplayed and some economic agents start to expect better times, business investment rises and economic growth starts to rebound. This leads to a virtuous circle, as business optimism rises, other firms start investing more, leading to increased economic growth and a further boost to business optimism. However, eventually productive capacity becomes misallocated or other breaks on growth appear, and business optimism starts to fall, with the cycle starting all over again.

Likewise, the financial system has a stylised credit cycle. If debt burdens appear to be becoming overstretched then borrowers may become less keen to add debt and lenders less keen to advance credit. Weaker firms and individuals that have borrowed excessively may then run into trouble, as they struggle to renew any borrowing they need to stay afloat. Eventually, calm returns and the cycle starts again.

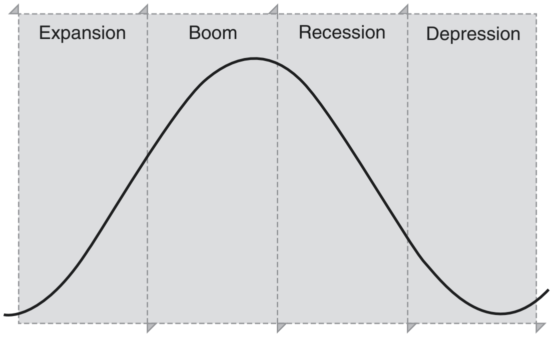


Fig. 2.4 Stylistic representation of the business cycle
Source: Nematrian.

Less clear is whether the actual evidence supports the existence of cycles of definable lengths, stylistically illustrated in Fig. 2.4. The concept of cycles as such is no longer so well entrenched in modern economic and finance theory, see Box 2.5.

Box 2.5: The business cycle and the financial (credit) cycle

The business cycle is most commonly defined as the downward and upward movement of Gross Domestic Product (GDP) around some longer-term growth trend, involving periods of expansion, boom, recession and depression, which then repeat themselves. The first clear exposition that such cycles exist is usually ascribed to Sismondi in 1819, although perhaps his theory can more accurately be classed as involving periodic crises rather than cycles per se.

Stylistically it can be viewed as involving four stages as shown schematically in Fig. 2.4, i.e. expansion, boom, recession and depression.

In theory, the cyclical nature of economies can be analysed using mathematical techniques such as spectral analysis, which can tease out the length of time between consecutive repeats in a business cycle. The output of such an analysis is a spectrum akin to the way in which white light splits into light of different wavelengths and hence frequencies when passed through a prism. If cycles really repeated themselves at predictable time intervals then the resulting spectrum coming out of such an analysis should show sharp peaks at the relevant frequencies. In practice, spectra coming out of such analyses tend to be broadly spread, and it is hard to discern clear peaks. So at least some of the notion of the existence of (regular periodic) business cycles may be an example of what in behavioural science is known as 'confirmation bias'. We recognise that economic activity goes up and down, so we back-fit in our mind anything that looks vaguely cyclical in nature to confirm our presupposition that there are clear dynamical drivers creating cycles of specific frequencies.

In the mid-twentieth century several economists sought to classify business cycles by periodicity including:

(a) The Kitchin cycle (3 to 5 years), focusing on cyclical developments in inventory levels
(b) The Juglar cycle (7 to 11 years), focusing on levels of fixed investment
(c) The Kuznets cycle (15 to 25 years), focusing on infrastructural investment levels
(d) The Kondratiev wave or long technological cycle (45 to 60 years)

Interest in such classifications has waned with the development of modern macroeconomics which tends not to support cycles of fixed periodicities. To the extent that business cycles do exist, it is probably the case that they do not align particularly well with financial cycles. Financial cycles, to the extent that they exist and fit into any of the four cycles referred to above are probably most aligned with levels of fixed investment, if this is financed by credit. However, much debt these days is advanced to consumers so may be more linked to changes in consumer confidence and other factors driving consumer financial behaviour, which don't seem to be well aligned with any of the cyclical drivers referred to above.

2.9 Macroprudential Policy

Closely aligned to systemic risk and financial stability is the notion of macroprudential policy. This can be defined as the set of regulatory and other measures that central authorities can adopt (or are adopting) to mitigate systemic risk. Macroprudential policies can be applied to the whole financial system or just to individual sectors within it (although if the measures are granular enough then they will verge into 'microprudential' policy, i.e. regulations designed to address the financial robustness of an individual institution).

For example, some jurisdictions have adopted policies that seek to prohibit banks from extending too high a proportion of new house loans to people asking for very high loan-to-value (LTV) mortgages. The motivation for this is that financial meltdowns in the past have often been preceded by general adoption of overly lax bank lending standards.

Will introduction of such requirements be universally welcomed? They may not be welcomed by any individuals who are consequently denied mortgages and therefore unable to buy their 'dream' house (unless these individuals take the enlightened view that the debt they would otherwise have taken on might have proven unsustainable). And how objectively do we define 'very high'? It must be a lot higher than zero, as otherwise the unintended consequence of the policy measure would be to deny *everyone* a mortgage. We face another trade-off. Typically, we will want macroprudential policies to intrude as little as possible in market mechanisms but to reduce as much as possible the systemic vulnerability being targeted. This implicitly assumes that we believe that the market mechanisms in question do operate in ways that help society. This was not a view that everyone in retrospect held when surveying some market developments that occurred in the run up to the 2007–09 Credit Crisis.

Added to this mix is the challenge of justifying any specific stance. Central authorities have had many years to get used to setting and justifying monetary and fiscal policies. They have had less experience targeting macroprudential policies. Taking the policy referred to above as an example, exactly how should we interpret the term 'lax' and how can we most robustly identify whether lax lending standards are prevalent at any given point in time?

Even if we can overcome definitional challenges we also need to find levers that affect the macroprudential picture. Most economies nowadays are not command economies in which individual firm decisions are rigidly set from on high. Instead, firms are usually allowed considerable freedom of action. It is generally believed that an approach that gives them this freedom will ultimately lead to more rational economic choices than a centrally planned economy. But what might be desired by central authorities on macroprudential grounds may not match up with what firms themselves desire. Firms may find ways of legitimately circumventing restrictions if they feel they are too onerous.

Identification of which central authorities should have which macroprudential powers can also be a political hot potato. Some of the more important bodies concerned are set out in Box 2.6. Firms in the financial services industry are typically considered to be relatively heavily regulated, i.e. subject to rules on how they must structure their balance sheets and business behaviours if they are to be allowed to continue to operate in a specific field. These rules are set by regulators and monitored by supervisors (although colloquially either term may be applied to either role). Existing regulatory and supervisory bodies are therefore likely to play an important role in macroprudential policy. So too are central banks and/or finance ministries, as they will need to write the cheques if things go wrong.

Macroprudential policy can in principle be contrasted with microprudential policy. Microprudential policy involves regulations and the like that apply to individual firms rather than to the whole system. They generally aim to reduce the risk of that entity running into difficulties. Or if the entity has run into difficulties then they generally aim to promote orderly recovery or resolution of the entity. By 'recovery' we mean the business refocusing in a manner that addresses the underlying issues whilst still leaving the business as a going concern. By 'resolution' we mean the winding up the business in an orderly fashion to avoid it failing in a disorganised manner, see Section 3.2.5.

In practice, the boundary is less clear. An indication of this is that as at 31 December 2015, according to ESRB data (see http://www.esrb.europa.eu/national_policy/systemically/html/index.en.html for latest available information), 160 EU banks had been deemed sufficiently 'systemically important' to warrant imposition on them of a 'systemic risk buffer' in addition to their normal capital requirements under the EU's Capital Requirements Directive (CRD), European Union (2013a), and the Capital Requirements Regulation (CRR), European Union (2013b). The CRD /CRR is the way in which the EU has implemented the latest Basel Capital Adequacy Accords applicable to all globally active banks. More were expected to be added to this number during 2016 as Austria, Bulgaria, Poland and UK decided how to implement the systemic risk buffer. Whilst ostensibly a macroprudential tool targeting systemic risk, the particularly wide adoption of the systemic risk buffer seems to be partly because it can be more flexibly applied at an individual firm level than some other tools available under CRD /CRR. 160 is not a very large number in the context of the total number of EU banks. The financial systems of some individual EU member states may also be disproportionately focused on a small number of banks. But even so, looked at from an EU-wide perspective, the number feels intuitively high and veering more into the area of microprudential policy than might otherwise have been expected.

There is also some overlap between microprudential and macroprudential policy when it comes to firms that are deemed sufficiently large (however 'large' is defined) to be individually systemically important. The identification and regulation of systemically important financial institutions (SIFIs) has far reaching implications both for the firms so designated and for others in their industry.

Another feature of CRD and CRR is the extent of responsibility it gives for macroprudential policy to individual EU member states. The CRD is an EU Directive which member states need to translate into their own legislative frameworks. The CRR is a maximum harmonising regulation that applies directly in all member states. A high proportion of microprudential (i.e. specific firm-level regulatory requirements), particularly ones involving capital requirements are encapsulated in the CRR, so regulatory bodies in individual member states have relatively little discretion to vary them. The Eurozone (which forms a large part of the overall EU) has gone even further, with the introduction of a Single Supervisory Mechanism (SSM) that directly supervises most larger banks across the Eurozone.

In contrast, responsibilities in relation to financial stability focus primarily on the financial system of the member state in question (although there are coordination requirements if there is expected to be spill-overs to other member states, see e.g. Table 4.1). The powers involved tend to be set out in the CRD and are largely reserved to competent authorities *within* individual member states. Central EU authorities such as the European Banking Authority (EBA) tend to become more involved only if authorities in individual member states disagree. National competent authorities therefore retain the greatest formal powers under CRD in macroprudential rather than microprudential matters.

Box 2.6: Bodies with macroprudential mandates

The EU has a European System of Financial Supervision (ESFS). The four bodies forming the ESFS are the European Banking Authority (EBA), the European Securities and Markets Authority (ESMA), the European Insurance and Occupational Pensions Authority (EIOPA) and the European Systemic Risk Board (ESRB). Other central bodies such as the European Central Bank (ECB) also in effect have system-wide responsibilities and dedicate resources to analysing, formulating and applying policy in this area. Each of the three sectoral bodies of the ESFS (EBA, ESMA and EIOPA) have financial stability sections that focus on financial stability matters within their sector. ESRB has no formal regulatory powers but does have the power of persuasion and in effect provides a forum where cross-sectoral aspects of financial stability can be discussed. The three sectoral bodies of the ESFS and the ECB are members of the ESRB, the current ESRB chair is the president of the ECB and the ESRB secretariat is currently based at the ECB's main office in Frankfurt.

For banking, each EU member state is also required by the CRD to designate a national authority responsible for financial stability in that member state. This authority is responsible for certain macroprudential activities, see Section 4.1.3. Often these authorities form part of the country's central bank or are closely associated with it.

A similar picture applies in the USA. Responsibility for most macroprudential policy matters lies with the Federal Reserve Board (Fed) or bodies linked to it. Under the US Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank), a body was set up called the Office of Financial Research (OFR) as an independent bureau within the US Treasury Department to undertake research into financial stability.

Internationally, the body with chief systemic risk responsibilities is the Financial Stability Board (FSB), established by the Group of Twenty international forum for governments and central banks (G20) in April 2009. Other key components of the international financial scene such as the International Monetary Fund (IMF) also nowadays tend to have financial stability directorates or the equivalent and often publish financial stability reviews. This reflects the importance now being given to systemic risk, financial stability and macroprudential policy in modern financial regulatory thought.

The increased focus on systemic risk can be seen in the way in which some countries now organise their (financial sector) regulatory structures. For example, the UK has a Prudential Regulation Authority (PRA, part of the Bank of England) that supervises the capital requirements of banks and insurers. It also has a Financial Conduct Authority (FCA) that supervises firm behaviours. However, arguably at the top of the tree is its Financial Policy Committee, which is able to give guidance to both the PRA and the FCA, see e.g. PRA (2012).

2.10 Key Takeaways

Points we can draw out from the remarks in this Chapter include:

(a) There is currently a lack of a settled definition of what is meant by systemic risk. Probably there always will be, as different definitions reflect different underlying perspectives about what the system is that we should be interested in, what level of aggregation we should focus on, a range of other implicit views about the risk characteristics to which 'our' part of the system relates, what bodies are best placed to monitor and address such risks and how important are these risks. The flip side of this definitional uncertainty is that practically any type of entity is potentially exposed to being considered by some as contributing to, transmitting or otherwise meaningfully interacting with a suitably selected form of systemic risk. Political and other factors seem likely to significantly influence the regulatory environment, which makes it harder but more important to manage the risks involved in an effective manner.

(b) Macroprudential policy seems to be the new frontier and 'where the action is at' for many regulators. To the extent that national regulatory bodies retain flexibilities to drive their own futures, these are (at least in the EU and for banking) becoming more concentrated in the macroprudential area. This could easily in the short-term increase the apparent emphasis that these bodies place on systemic risk or at least on policies that can somehow be viewed as associated with systemic risk even if in practice they include a significant microprudential element.

(c) A key factor influencing how commentators and practitioners tend to view systemic risk is the notion of interconnectedness (particularly 'direct' interconnectedness, involving explicit contractual relationships between different parties). However, (direct) interconnectedness is not the only factor regulators consider when devising macroprudential policies. Lack of interconnectedness is no guarantee of avoiding the systemic risk regulatory net!

References

Competition & Markets Authority (2016). Retail Banking Market Investigation: Final Report. *UK Competition & Markets Authority.*

Economist (2016). WeChat's World. *The Economist*, 6 August 2016.

ESRB (2014c). Is Europe Overbanked? *European Systemic Risk Board.*

European Union (2010). Regulation (EU) No 1092/2010. *European Union.*

European Union (2013a). Capital Requirements Directive, i.e. Directive 2013/36/EU. *European Union.*

European Union (2013b). Capital Requirements Regulation, i.e. Regulation (EU) No 575/2013. *European Union.*

Ferguson, N. (2009). *The Ascent of Money*. Penguin Books, London, UK.

IMF (2016). Global Financial Stability Report, April 2016, Chapter 3: The Insurance Sector – Trends and Systemic Risk Implications. *International Monetary Fund.*

Kriwaczek, P. (2010). *Babylon: Mesopotamia and the Birth of Civilization*. Atlantic Books, London.

Mackay, C. (1841). *Extraordinary Popular Delusions and The Madness of Crowds*. Wordsworth Reference, Ware, Hertfordshire, UK.

PRA (2012). The PRA's Approach to Banking Supervision. *The Bank of England, Prudential Regulation Authority.*

PwC (2016). PwC Global FinTech Report March 2016: Blurred lines: How FinTech is shaping Financial Services. *PwC.*

Schwarcz, S.L. (2008). Systemic Risk. *Duke Law School Legal Studies Research Paper 163.*

Wolf, M. (2015). Good News – Fintech Could Disrupt Finance. *Financial Times.*

3

Overall Features of the Financial System

In this Chapter, we introduce and discuss generic features of systemic risk shared across nearly all parts of the financial system. Most of the specifics of individual sectors are left to the next Chapter.

We introduce the Chapter by exploring further the ‘domino’ and ‘tsunami’ models of systemic risk, to help us understand which parts of the financial system might practically be exposed to or contribute to systemic risk. We argue that large systemic risk events can nearly always be expected to involve a ‘combined’ model, in which elements of both ‘domino’ and ‘tsunami’ perspectives are present, the tsunami element usually being the more dominant if the systemic crisis is large.

We then explore the purpose and nature of regulation in the financial sector (including the concept of resolution planning) and how this interacts with the amount of regulatory capital financial firms are required to hold. Measuring how much they hold involves preparation of balance sheets and other accounting information, so we also explore the impact of relevant accounting principles on these measurements.

Even well capitalised firms can run into difficulties if they do not have sufficient liquid assets to meet customer expectations in a timely manner, so we also explore how liquidity interacts with capital requirements and with systemic risk more generally.

M.H.D. Kemp, *Systemic Risk*, DOI 10.1057/978-1-137-56587-7_3

3.1 What Predisposes the Financial System to Suffer from Systemic Risk?

3.1.1 Introduction

A natural question to ask is what features are common to nearly all parts of the financial system rather than being unique to individual components of it, and which of these features are most important in terms of systemic risk.

This, of course, implicitly assumes that essentially any part of the financial system is potentially caught up in the scope of systemic risk. This is a view that some industry participants would dispute from the start. So we start our review by identifying the types of players that might be viewed as within the possible scope of systemic risk and the sorts of characteristics we might expect them to exhibit if they are to be material contributors to or participants in a systemic risk event.

As explained previously, policymakers typically start with a broad scope in mind. They also typically seem to have shifted through time away from focusing just on (direct) interconnectedness and towards a stance which views systemic risk as potentially coming from wider range of sources. This has added to the number of players that they believe might be 'in scope' of systemic risk.

Policymakers (like other humans) are predisposed to consider recent past experiences when forming views. In the public mind, the 2007–09 Credit Crisis is probably most associated with the default of Lehman Brothers and problems faced by the banking industry. However, at about the same time as Lehman defaulted the US Government stepped in to rescue an insurer, AIG, see Box 3.1, and some money market funds (MMFs), see Box 4.8. Other sorts of investment funds (albeit usually more exotic ones like LTCM, see Box 4.9) have failed within recent memory or have needed formal or informal support from central authorities. Going further into the past, experiences such as the US Savings & Loan debacle colour policymaker perspectives. They also remind policymakers that direct interconnectedness is not necessarily a pre-requisite for large bailout costs, if enough smaller players all individually adopt similar sorts of strategies that then go wrong.

Box 3.1: AIG and other recent insurance failures

The near failure of AIG was a milestone in the 2007–09 Credit Crisis. Its near failure in 2008, at about the same time as the default of Lehman Brothers, was primarily due to losses arising from the sale of credit default swaps (CDS) by one of its non-insurance affiliated entities (its Financial Products division) and its securities lending business. These losses led it to be on the receiving end of very large margin calls, with other market participants requiring it to post additional collateral to allow it to continue to hold derivatives it was using to hedge some of its risks. When it became clear to the authorities that it lacked sufficient readily available assets to be able to meet these margin calls, the authorities bailed it out, fearing that otherwise panic would ensue.

It was bailed out by the US authorities at just about the same time as these authorities let Lehman Brothers default. To the extent that they were only prepared or able at the time to bail out one of the two firms it is noteworthy that they chose to bail out AIG, presumably implying that they thought AIG presented greater systemic risks than Lehmans.

Several monoline insurers specialising in providing credit guarantees also ran into difficulty during the 2007–09 Credit Crisis, as the costs of honouring credit guarantees they had written ballooned. A monoline insurer is technically one that specialises in just one line of insurance but in the run-up to the Crisis the term became particularly associated with insurers that specialised in providing 'credit enhancement'. This involved an issuer issuing a bond within a structure that also included a credit guarantee from such an insurer. The net result was that the overall structure commanded a higher credit rating than was practically available had the issuer issued the bond without credit enhancement. This increased the appeal of the bond to end investors, who in many cases only bought bonds that were deemed sufficiently creditworthy by ratings agencies.

Some other insurers received public support during the Crisis, e.g. Aegon received a EUR 3bn capital injection from the Dutch government, see Financial Times (2008). The capital injection aimed to bolster its capital position in the light of losses and impairments it had suffered on mortgage-backed securities including ones associated with the collapse of Lehman Brothers and Washington Mutual.

The Aegon case can be presented as an example of an insurer that ran into difficulties due to contagion from the banking sector. However, presumably part of the reaction of central authorities in such cases involved a desire to limit the likelihood of further contagion and broader loss of confidence. It presumably therefore reflected a worry that Aegon might be a link in a domino chain rather than at the end the chain.

3.1.2 Too Big to Fail (TBTF)

The term *too big to fail* (TBTF) features prominently in the lexicon of systemic risk. Even with the best will in the world, we can expect firms to run into difficulties from time to time. It is what happens afterwards that is the issue.

From the regulators' perspective, the worst scenario is that the government feels obliged to step in to support the firm in difficulties because doing otherwise is expected to create more problems than supporting the firm. In extremis, the costs of addressing TBTF can become astronomical, even at times threatening the finances of the state itself. Part way through the 2007–09 Credit Crisis some countries such as Ireland guaranteed their banks' debts, in the hope that this would calm market fears. However, the Crisis failed to blow over, indeed it got worse. Some of these countries then ended up needing to be bailed out themselves.

Banks seem to have a propensity to run into trouble in a way that presents difficult challenges for regulatory authorities. It was one such episode, the failure of Continental Illinois National Bank and Trust Company in 1984, that led to the coining of the term 'too big to fail'. Continental Illinois was a large but not very large US bank at the time, see Box 3.2.

During the 2007–09 Credit Crisis, a significant number of banks ended up being partially nationalised or otherwise supported by governments or deposit protection schemes. In effect, a significant number of banks were found to be 'too big to fail'.

Several regulatory changes are in the process of being implemented that seek to reduce the TBTF problem. These include mandatory 'bail in' of some bank investors, the aim being to impose losses on these investors before losses get borne by the state. This has led to the development of the *total loss-absorbency capacity* (TLAC) regime applied to the largest global banks, see Section 3.4.6, which requires firms to identify specific investor classes that will be automatically 'bailed-in' in specified circumstances. In the EU, the Bank Recovery and Resolution Directive (BRRD), see European Union (2014), gives authorities the power to bail in a large fraction of investors and wide discretion over exactly who suffers what loss.

Different commentators have different views on how effective these reforms are likely to be at addressing the TBTF issue. Some believe that TBTF is well on the way to being solved, and by the time TLAC is fully up and running (bank) TBTF will have been consigned to history. Others are less sure.

History in the form of Continental Illinois is rather sanguine. Noteworthy about the supervisory response to its failure was that the relevant central authority, the US Federal Deposit Insurance Corporation (FDIC), agreed to guarantee *all* its deposits (because of worries that to do otherwise would create systemic ripples that would cripple the financial system or at least significant parts of it). This was despite the agreed regulatory response at the time to such failures being only to cover insured deposits up to some pre-specified upper limit. But when push came to shove, the authorities scrapped this rule and effectively insured all deposits.

Politicians have flexibility over which firms should be treated as large enough to be too big to fail. They are also able (within bounds) to rewrite or at least reinterpret the rules if the going gets tough enough. At the heart of the TBTF issue is regulatory and political moral hazard. If a firm is large enough or interconnected enough, and if its failure under existing rules creates a sufficiently challenging situation, then how do you eliminate the risk that the rules will be rewritten in a manner that does not reintroduce the TBTF issue? Part of the challenge in Continental Illinois' case was the possible impact that its failure would have on other US banks weakened by the Less Developed Countries Debt Crisis, see Box 3.2.

Can we legislate against such an outcome? This was indeed one of the responses to the failure of Continental Illinois, as described in FDIC (2000). Speed bumps were introduced that were designed to make it more difficult for (larger) firms to lobby politicians for support. Arguably, these sorts of speed bumps contributed to the lack of willingness to support Lehman Brothers when it ran into difficulties. A line in the sand had metaphorically been drawn and it was held for a while. However, events during the 2007–09 Credit Crisis turned out so adverse that the line in the sand proved impossible to sustain and broader institutional failure followed, see Box 3.3.

The term 'too big to fail' is not limited to banks or even the financial sector. In autumn 2016, the South Korean Government allowed Hanjin Shipping to default after having propped it up by for many years whilst it suffered significant losses. Much of South Korea's exports were shipped abroad by this company. Allowing it to fail could have had significant consequences for the wider economy.

Box 3.2: Continental Illinois, TBTF and the LDC Debt Crisis of the 1980s

Continental Illinois

The US Federal Deposit Insurance Corporation (FDIC) was set up during the financial and economic crises of the 1930s. In 1984, shortly before the S&L Crisis (see Box 3.3), it had to manage the then largest banking failure in US history, involving Continental Illinois National Bank and Trust Company (Continental Illinois), see FDIC (2000). Some of the issues that faced the global banking system in 2007–09 resonate with issues that afflicted Continental Illinois.

Continental Illinois was the nation's seventh-largest bank. It not only forced regulators to recognise that very large institutions could fail but also that satisfactory ways were needed to cope with these failures. The differential treatment it received was to be given a specific name that resonates to our own time, namely 'too big to fail' (TBTF). All of its deposits were covered by the FDIC when for other smaller failures FDIC had tried to adopt an approach involving 'modified payoffs', in which only a proportion of the amount owed to uninsured depositors and other creditors was paid (based on the estimated recovery value of the institution's assets).

The management of Continental Illinois began to implement in the mid-1970s a growth strategy focused on commercial lending. By 1981 it had become the largest commercial and industrial lender in the USA. With the benefit of hindsight there were elements of its financial profile that were harbingers of problems to come (and that echo with the experience of the 2007–09 Credit Crisis). As it grew it experienced a substantial increase in its loans-to-assets ratio. It was heavily exposed to participations from Penn Square Bank in Oklahoma which failed in July 1982 leading to substantial losses. Continental Illinois was also involved with three large corporate bankruptcies and suffered from further negative perceptions arising from the onset of the less developed country (LDC) debt crisis (brought on by Mexico's default in August 1982). It had never had a large retail deposit base and had instead relied primarily on Fed funds and issuance of certificates of deposit, i.e. on wholesale money markets, for its funding. Shortly after Penn Square's failure, Continental Illinois started to run into difficulties funding its domestic (i.e. US) operations from US wholesale money markets. It turned increasing to foreign money markets (principally the Eurodollar market) as a source of funding. It then suffered a bank 'run' in May 1984. On 9 May, Reuters asked Continental Illinois to comment on rumours that the bank was on the road to bankruptcy. Stories circulated that a Japanese bank might be interested in acquiring it. Anxious overseas depositors started to shift their deposits away from Continental Illinois. By Friday 11 May Continental Illinois had had to receive liquidity support from the Fed. A loan package that was established with other banks that weekend was insufficient to stop the run and its domestic depositors also started to withdraw funds. Three bank regulatory agencies provided additional assistance and the FDIC guaranteed all its depositors and other general creditors. Ultimately, the FDIC ended up owning 80% of the bank.

Too-Big-To-Fail (TBTF)

The term 'too big to fail' was coined by the US Comptroller of the Currency, C.V. Connor in a US congressional hearing about Continental Illinois in September 1984, according to Dymski (2011). It figured centrally in the resolution of banking crises later that decade, see Box 3.3, and in the 1990s with the failure of LTCM, see Box 4.9, and some large Japanese and French banks. It became a household term during the 2007–09 Credit Crisis and its aftermath.

The Less Developed Countries (LDC) Debt Crisis of the 1980s

At first sight TBTF seems to be particularly aligned with the domino theory of systemic risk, i.e. the worry that failure of one firm might cause others to fail, leading to still others failing in a cascade fashion. Therefore, so the line of reasoning goes, we are forced to intervene to avoid the failure of the initial domino, to stop others also falling over.

However, if we look carefully at the failure of Continental Illinois we see that the backdrop involved some suspiciously tsunami-like factors affecting lots of US money centre banks at the same time. In particular, all faced challenges because of the Less Developed Countries (LDC) debt crisis, which is described further in another chapter of FDIC (2000). Between 1980 and 1994 more than 1,600 banks insured by the FDIC were closed or received FDIC financial assistance. This was far more than had previously occurred in any other period since the advent of US federal deposit insurance in the 1930s. It is far more than can be easily be explained purely via a domino-type model of bank failure. During most of the 1980s, arguably developments in the US national economy should have been relatively favourable for the US banking industry. The challenge was that during much of the 1980s the financial position of US money-centre banks became increasingly entwined with less favourable developments further afield.

Between end 1978 and end 1982, total LDC debt held by the eight largest money-centre banks expanded from $36 billion to $55 billion. By the end of this period, total LDC portfolios held by these banks averaged more than double the banks' aggregate capital and reserves. Some attempt to curtail this lending was made by bank regulators but these attempts were relatively limited. FDIC (2000) argues that in aggregate it may be said that government policy supported LDC lending activity by the banks.

In August 1982, the government of Mexico announced it could no longer meet interest payments, and by the end of the year 40 nations were in arrears. By the end of 1983, 27 countries were in negotiations to restructure their existing loans. Following the Mexican default, U.S. banking officials did not require that large reserves be immediately set aside for the restructured LDC loans, apparently believing that some large banks might have been deemed insolvent and that an economic and political crisis might have been precipitated. By the end of 1986, loss reserves still averaged only approximately 13% of the total LDC exposure of the money-centre banks. According to former FDIC Chairman Seidman, US bank regulators, given the choice between creating panic in the banking system or going easy on requiring US banks to set aside reserves for Latin American debt, chose the latter. Only starting in 1987, did more aggressive and more realistic recognition of losses start to take place. By the end of 1989, total reserves at the money-centre banks had risen to nearly 50% of total LDC loans.

The LDC experience illustrates the high priority given to maintaining financial market stability in the treatment of large banks. It also arguably represents a case of regulatory forbearance, in which money-centre banks were effectively allowed to delay recognition of losses, thereby avoiding repercussions that might have threatened their solvency. In time, loss reserves and charge-offs were greatly increased, and no money-centre bank failed because of LDC loans. The creation of the Brady Plan in 1989 reflected recognition that banks would not recover the full principal value of existing loans and turned international efforts from debt rescheduling to debt relief. Ultimately, the LDC crisis was resolved by shareholders of the world's largest banks eventually assuming the losses under the Brady Plan.

3.1.3 The 'Domino' Model of Systemic Risk

Commentators often express the view that (financial) systemic risk can be viewed as exclusively a product of some sort of instability arising within the financial system (e.g. undercapitalisation, inappropriate behaviour, liquidity panics etc.) which propagates across the financial system and then leads to undesirable consequences for the wider 'real' economy.

Commentators proposing this point of view typically assume that the 'normal' state of the financial system is not to exhibit such instabilities. Something then comes along and triggers a change in the way financial institutions interact with each, leading to the system shifting from a stable to an unstable state. Once the system passes a threshold that takes it into an unstable mode then a much wider range of outcomes become possible, including some that have substantial adverse consequences for the wider economy.

It is not difficult to identify ways in which systemic risk might arise through this route. Perhaps the most obvious is if a liquidity shock occurs, like the one that formed the epicentre of the 2007–09 Credit Crisis.

As this Crisis worsened, see Box 3.3, banks became unsure how robust were other banks they might be dealing with. They became less willing to provide each other with short-term funding. Some markets that had previously been used by some banks to source funding for their asset holdings largely or wholly ceased to function (e.g. large parts of the commercial paper market). Banks with more liquid balance sheets began to 'hoard' liquidity, starving others of liquidity. Banks most dependent on ready access to liquidity from other banks ran into trouble and in several cases failed. The Crisis was only in effect resolved when central banks flooded the banking sector with liquidity.

The natural mathematical toolset for the domino model of systemic risk is one borrowed from the study of disease epidemics. We assume that the financial system involves a mesh of interconnections. From time to time random events cause some firms to run into difficulties, which create issues for others because of these interconnections.

However, there is a problem with assuming that interconnectedness is the *sole* driver of systemic risk. The level of interconnectedness of the financial system, even just the banking element of it, doesn't seem to be high enough to lead to the sorts of problems actually seen in practice.

Or rather, there seems to be insufficient interconnectedness if we assume that the nature of the interconnections is sufficiently stable through time. Suppose such stability is present. Then immediately prior to any postulated trigger that might start off the domino collapse we should expect nearly all firms to be adequately capitalised. This is because those that aren't should have failed beforehand but in an idiosyncratic rather than systemic manner. Then when the trigger comes along it may take out the first domino in the chain. It may even take out one or two others. But at each consecutive step in the chain the impact of the trigger should diminish as extra (surplus) capital falls within the scope of the toppling domino chain. A chain reaction fails to occur, because the impact of the initial trigger is dampened at each consecutive step of the chain.

So, what we also need is for the system itself to change through time. If elements of the domino chain are no longer essentially static we can envisage circumstances in which this sort of dampening is temporarily diminished and the conditions needed for a chain reaction become temporarily present.

For example, we might hypothesise that changes in banking business models in the extended calm before the 2007–09 Credit Crisis were sufficient to tip what had previously been a stable mode of the financial system into one that when stressed enough would lead to larger scale instabilities. Several policy responses after the 2007–09 Credit Crisis specifically targeted changing bank business models in a manner believed by policymakers to promote financial stability, i.e. to remove these vulnerabilities and to improve the capital bases of the firms involved. These included the Volker rule against proprietary trading introduced for US banks.

IMF (2009) was one of the earliest responses to the 2007–09 Credit Crisis. In it the IMF sought to provide initial guidance on how authorities could assess the systemic importance of financial institutions, markets or instruments. It defined systemic risk along the lines of: '*Systemic risk is the risk of disruption of financial services that is caused by impairment of all or parts of the financial system and has the potential for serious negative consequences for the real economy*'. It argued that fundamental to this definition is the notion of 'negative externalities' from a disruption or failure in a financial institution, market or instrument. By an 'externality' we mean something that impacts other players not the original player in question, typically because the original player is not incentivised to worry about such impacts. Externalities play an important role in economics, particularly where the system in question is seeming not to deliver the solution that society might like.

Following this line of reasoning, IMF (2009) concluded that all types of financial intermediaries, markets and infrastructure could potentially be systemically important to some degree. This then raised the question of which ones were likely to be *most* systemically important. Criteria it highlighted as likely to help in identifying whether any specific institution, market or instrument might be systemically important included:

(a) Size (the volume of financial services provided by the individual component of the financial system);

(b) Substitutability (the extent to which other components of the system can provide the same services in the event of a failure); and
(c) Interconnectedness (linkages with other components of the system).

Given its ground-breaking analysis, IMF (2009) influenced the formal mandates given to some of the bodies set up to address systemic risk. However, drafters of these mandates seem even early on to have wanted to hedge their bets about exactly how important interconnectedness might be in this context. For example, we've already given a definition of how systemic risk is defined by the Regulation that established the ESRB. It comes immediately after two other relevant definitions:

> (a) 'financial institution' means any undertaking that falls within the scope of the legislation referred to in Article 1(2) of Regulation (EU) No 1093/2010, of Regulation (EU) No 1094/2010 and of Regulation (EU) No 1095/2010, as well as any other undertaking or entity in the Union whose main business is of a similar nature;
> (b) 'financial system' means all financial institutions, markets, products and market infrastructures;

So, in principle bodies like the ESRB do not need to limit themselves to elements of the financial system that are interconnected, as long as collectively the elements in question might create an economy- or market-wide disruption large enough to fall within the deemed scope of systemic risk.

In the UK, new legislation on 1 April 2013 established an independent Financial Policy Committee (FPC) at the Bank of England. The FPC was charged with a primary objective of identifying, monitoring and acting to remove or reduce systemic risks with a view to protecting and enhancing the resilience of the UK financial system. It also has a secondary objective to support the economic policy of the UK Government. Again, the definition assumes that it is practical to identify the 'financial system' but again is relatively silent on whether occurrence of systemic risk will necessarily be characterised by a domino-like cascade of failures or by some other mechanism.

Box 3.3: The 2007–09 Credit Crisis

A very considerable amount of material has been written about the 2007–09 Credit Crisis. Timelines describing what happened that were reasonably contemporaneous with the Crisis include Kemp (2009) and Bank of England (2008):

(1) Prior to 2007 there was a period of lax bank lending standards, benign economic and liquidity conditions and a 'search for yield' as investors believed a new paradigm was resulting in mitigation of risk even for bonds and other financial instruments that were offering otherwise attractive yields. This period was also characterised by an increased focus on an 'originate and distribute' model for banking in which banks increasingly made loans which they repackaged and sold on to others, the financing for these activities being supported by the then benign liquidity conditions.
(2) At the end of July and in the first two weeks of August 2007 several quantitatively run investment funds suffered surprisingly large losses due to dislocations to previously ruling relationships between secured and unsecured money market rates. The dislocations forced them to close leveraged positions which happened to be relatively similar across the funds affected and across some other similarly positioned portfolios (an example of a 'crowded trade'). Whilst this received only limited attention outside a small group of investors and investment bankers, the dislocations were eventually to grow to be very sizeable and to have a dominant impact on financial markets. With the benefit of hindsight, this fortnight can therefore be deemed to mark the start of the Crisis.
(3) Valuations of mortgages parcelled up into retail mortgage backed securities and collateralised debt obligations (CDOs) started to scatter in summer 2007, the proximate trigger being a major deterioration in delinquency rates on (US) sub-prime mortgages originated in 2006. Confidence in such valuations continued to decline as 2007 progressed and into 2008 eventually leading to contagious loss of confidence across a broad swathe of instrument types. Banks started to distrust their counterparties, perhaps reasoning that if they themselves couldn't easily value their own books then others couldn't easily value their books either. The corollary was that nearly anyone might be insolvent if scrutinised hard enough. Banks with a business model dependent on continued benign liquidity conditions were thus in trouble.
(4) The resulting loss of confidence led to failures, initially apparently isolated but eventually becoming frequent and problematic. An early casualty in the US (in early 2008) was the near collapse of Bear Stearns, a US investment bank. In the UK, a specific focus of attention early on was Northern Rock, a medium-sized bank with a prime residential mortgage book, which experienced the first material bank 'run' in the UK for many years. Its business model was dependent on access to funding from the wholesale money markets. When these dried up, depositors descended on it, demanding to withdraw their deposits. The bank needed emergency liquidity support from the Bank of England and it was eventually nationalised by the UK Government.
(5) Even at this stage many commentators were expecting the problems to be isolated and to blow over. This was not to be. A critical phase was reached in September and October 2008. In September, the US Government rescued Fannie Mae and Freddie Mac, the two main quasi-government entities important in repackaging of US housing market mortgages. The US government allowed Lehman Brothers to default, but rescued AIG, an insurance company. Almost immediately afterwards a money market fund, the Reserve Primary Fund ran into trouble and needed rescuing.

(6) This led to a further breakdown of the interbank funding markets and broader institutional distress. Governments recapitalised some banks (and some other financial institutions), guaranteed some debts and /or depositors and introduced or increased the size of liquidity schemes supporting banks. This stabilised the financial system but had wider negative impacts on several developed Western economies during the subsequent Eurozone sovereign debt crisis, see Box 4.14.

In the USA, the bailout in (6) was implemented via the Emergency Economic Stabilization Act of 2008. This authorised the US Treasury Secretary to spend up to \$700 bn to purchase distressed assets (particularly MBS) and to supply cash directly to banks (both domestic and foreign) under the Troubled Asset Relief Program (TARP). Dodd-Frank subsequently reduced the authorised size of TARP to \$475 bn. On 19 December 2014, the US Treasury sold its remaining holdings of Ally Financial (previously General Motors Acceptance Corporation, i.e. GMAC), effectively bringing TARP to a close.

3.1.4 The 'Tsunami' Model of Systemic Risk

At its extreme, the domino model of systemic risk involves the failure of one firm triggering the failure of another etc. It is in effect a one-on-one model of failure propagation or interconnectedness. Implicit is the model of a chain reaction, akin to that in say an atom bomb. A stray neutron triggers the fission of an atom of uranium 235 or plutonium, causing another neutron or two to be emitted. Given certain conditions (in the case of an atom bomb that there is enough fissile material in a sufficiently confined space), a cascade results.

As we saw previously, modern regulatory thought is increasingly willing to countenance other types of propagation, particularly the tsunami model of systemic risk. In this model, firms across the financial system (or some part of it) share common exposures. There may be relatively little direct interconnectedness, but if some broader factor kicks in, their common exposures all come home to roost at the same time, again leading to disaster.

The banking sector also seems to be exposed to this model of systemic risk. Some features of the 2007–09 Credit Crisis can be cast in this light, e.g. a relatively common focus on strategies that assumed continuing benign liquidity conditions. A clearer example is the US Savings and Loan Crisis.

Savings and Loan institutions (S&Ls) are akin to banks. As explained in Box 3.4, they became less strongly regulated in the 1980s and 1990s and in some cases this resulted in them taking on excessive amounts of risk relative to their available capital (and their available skills for managing such risks). They were generally relatively small institutions, each one in isolation not obviously large enough to be deemed particularly systemically important. Instead, it was the large fraction of them that failed (roughly one-third) which resulted in the S&L crisis becoming one of the costliest debacles in US banking history.

Box 3.4: The US Savings & Loan Crisis

The US Savings and Loan Crisis (the 'S&L crisis') involved the failure of over 1,000 US savings and loan associations from 1986 to 1995 (i.e. around one-third of the total number of such institutions). A fuller analysis of the S&L Crisis and the extraordinary upsurge in banking defaults that occurred at roughly the same time is set out in e.g. FDIC (2000). The US General Accounting Office estimated the total cost of the S&L crisis to be $160 bn, including $132 bn from federal tax payers. Much of this cost could have been avoided if the government had had the political will to act appropriately in the early 1980s. But, believing that the marketplace would provide its own discipline, FDIC (2000) argues that 'The government used rapid deregulation and forbearance instead of taking steps to protect depositors. The government guarantee of insured deposits nonetheless exposed US taxpayers to the risk of loss – while the profits made possible by deregulation and forbearance would accrue to the owners and managers of the S&Ls'.

Around 30% of these failures involved closure or other sorts of resolution by the Federal Savings and Loan Insurance Corporation (FSLIC) between 1986 and 1989, whist the remainder were closed or otherwise resolved by the Resolution Trust Corporation (RTC) between 1989 and 1995, after FSLIC had in effect itself run out of money.

S&Ls (or 'thrifts') are types of financial institutions that accept savings deposits and make mortgage, car or other sorts of personal loans to individual members of the S&L. Around 75–80% of S&Ls in 1980 were mutually owned. The nearest equivalent in the UK are building societies. US federal regulation of S&Ls developed under a different legislative framework to that applicable to commercial banks. The S&L legislation was driven by a public policy goal of encouraging home ownership, and was overseen by the Federal Home Loan Bank Board (FHLBB), which oversaw FSLIC. FHLBB comprised 12 regional Home Loan Banks supervised by FHLBB.

FHLBB's examination, supervision and enforcement practices are generally considered to have been weaker than those of the corresponding federal banking agencies. FHLBB was a small agency that oversaw an industry that was perceived to be performing a type of public service. Banking agencies generally recruited higher quality candidates, paying them c. 20–30% more than FHLBB could offer. Although FHLBB legally had similar enforcement powers to those of the banking agencies, it used them much less frequently (partly because enforcement was a lengthy process if contested by the institution). The industry was significantly involved in its own supervision, a result of it being perceived to have exhibited few past mismanagement problems, to be carrying out a quasi-public service and to it having a favourable image and protected status with lawmakers. FDIC (2000) quotes one S&L lobbyist as later writing: 'When we [the U.S. League of Savings Institutions] participated in the writing of the supervisory law, hindsight shows that we probably gave the business too much protection against unwarranted supervisory action'.

This did not appear to generate too many issues prior to c. 1980. But in 1979 the Fed doubled interest rates that it charged its member banks (seeking to reduce inflation). The S&Ls had issued long term loans at fixed interest rates that were lower than the rate at which they could borrow. 118 S&Ls failed in the first three years of the 1980s, costing FSLIC c. $3.5 bn to resolve. There were also 493 voluntary mergers and 259 supervisory mergers, but these did not create costs for FSLIC. During the previous 45 years only 143 S&Ls had failed, costing FSLIC $306m. More ominously, as at end 1982 there were still 415 S&Ls, with total assets of $220 bn, that were insolvent based on the book value of their tangible net worth.

According to FDIC (2000), 'many government officials believed that the insolvencies were only "on paper", caused by unprecedented interest-rate levels that would soon be corrected . . . Most political, legislative, and regulatory decisions in the early 1980s were imbued with a spirit of deregulation. The prevailing view was that S&Ls should be granted regulatory forbearance until interest rates returned to normal levels, when thrifts would be able to restructure their portfolios with new asset powers. To forestall actual insolvency, therefore, the FHLBB lowered net worth requirements for federally insured savings and loan associations from 5% of insured accounts to 4% in November 1980 and to 3% in January 1982'.

Phase-in rules for meeting net worth requirements, averaging rules for computing deposit bases and the applicability of some relatively lax accounting principles meant that actual capital requirements could be lower still, which made setting up new institutions very attractive. Accounting practices for supervisory goodwill were also relaxed, to encourage healthy S&Ls take over insolvent institutions, without the FSLIC having to compensate the acquirer for the (entire) negative net worth of the insolvent institution. Some of these changes were in effect written into legislation, namely the Depository Institutions Deregulation and Monetary Control Act of 1980 and the Garn-St Germain Depository Institutions Act of 1982. The aim was to help the S&L industry. However, the effect was to increase significantly the eventual cost of the crisis. This legislation also gave S&Ls new and expanded investment powers and eliminated deposit interest-rate ceilings. Particularly problematic, with the benefit of hindsight, were expanded authority to make acquisition, development and construction (ADC) loans and elimination of previous statutory limits on loan-to-value (LTV) ratios. These allowed S&Ls to make high-risk loans to developers for 100% of a project's appraised value. The S&L industry can thus be viewed as an example of the dangers of 'regulatory capture' and excessive regulatory forbearance, see Section 3.1.6.

Unsurprisingly, the S&L industry showed extremely rapid growth from end 1982 to end 1985, with total S&L assets increasing by 56% to over $1 tn. Risk takers gravitated towards the S&L industry, some fraudulent, many others just greedy. Between 1980 and 1986 nearly 500 new S&Ls came into existence. S&L investment portfolios rapidly shifted away from traditional home mortgage financing (down from 78% in 1981 to 56% in 1986). A large percentage of S&L assets were devoted to ADC loans, which helped fuel boom-to-bust real estate cycles in some parts of the USA. Losses mounted as declines in real estate values deepened. Efforts to recapitalise FSLIC eventually failed and a separate body, the RTC, eventually needed to be set up to address the resulting widespread undercapitalisation.

(a) There is a need for strong and effective supervision of insured depository institutions (particularly if they are given new or expanded powers or are growing rapidly).
(b) Structures need to limit the influence an industry has over its regulator(s) and regulatory agencies need to have sufficient resources and funding to allow adequate hiring, training and retaining of suitably qualified staff.
(c) It is desirable to close promptly insolvent insured financial institutions, to minimise potential losses borne by deposit insurance funds and to ensure a more efficient financial marketplace.
(d) Effective resolution of failing financial institutions requires adequately capitalised deposit insurance funds (FDIC argued that this capitalisation should involve '*real reserves, not just federal guarantees*').

3.1.5 The 'Combined' Model of Systemic Risk

A core thesis of this book is that in most circumstances we should expected major systemic risk events to be characterised by a 'combined' model, one in which domino-like and tsunami-like features are present simultaneously. In this model, we simultaneously need:

(a) Some (often relatively opaque) vulnerability or vulnerabilities to be present across a substantial part of the financial system; and
(b) Some firms that have a level of interconnectedness with other parts of the financial system that is sufficiently high to lead to market reappraisal of plausible vulnerabilities of other firms if one or more of these interconnected firms runs into difficulties.

More precisely, we assume that across the financial system (or across the part of it that we are interested in) there are some firms that are stronger than others. The 'combined' model of systemic risk proposes that for a 'full scale' crisis to arise we need a significant fraction of all firms to be in a (usually hidden) precarious financial position. By 'precarious' we mean a position which if it became widely recognised would likely lead to firm failure.

Such a structure creates the possibility of a crisis. But for one to occur we need some trigger that causes previously prevailing interpretations of the health of firms in the system to be reappraised, bringing to light undercapitalised organisations.

In this model, the three types of criteria mentioned by IMF (2009), i.e. size, substitutability and interconnectedness take on more nuanced interpretations.

Size and interconnectedness become important not primarily in their own right but because failure of a small organisation that operates largely in isolation is less likely to cause a major reappraisal of financial strength of other financial system participants.

It is less clear what importance should be accorded to substitutability. The less substitutable a specific organisation is, the less likely it is to be seen as illustrative of others. Conversely failure of such an organisation remains more likely to lead to problems beyond the financial system.

Another way of interpreting such a model is that it primarily focuses on *indirect* interconnectivity. We don't primarily worry about the direct impact that failure of one firm might have on another firm. Rather, we worry about the impact that difficulties faced by one firm will have on *perceptions* of the financial strength of other firms. The notion of indirect interconnectivity is explored further in ESRB (2016). One important point to note is that entities triggering indirect interconnectivity do not need to be in the same industry as the entities affected by the resulting contagion or even to have any contractual relationship with them. They do not even need to get into serious financial difficulties. All that is needed is that when they are reappraised it leads to a non-trivial reappraisal of the health of others.

The relative importance we place on the domino versus tsunami model also has important implications regarding the tools we will naturally focus on when designing macroprudential regulatory structures. At its most basic, a domino model favours entity-based regulation, targeting individual firms. For example, it favours imposing relatively higher capital requirements on larger firms, since they are more likely to figure in the propagation of domino-like chains of failures. In contrast, a tsunami model favours activity-based regulation, targeting control of activities deemed likely to be particularly important in the context of systemic risk. For example, we might view securities lending and derivative transactions as potentially adding to systemic risk. We might then seek to limit systemic risk by imposing controls on how such activities are carried out (irrespective of the type of organisation carrying out the activity). We explore this linkage further in Section 6.6.

3.1.6 Regulatory Forbearance

Precariousness in the above context can be heavily linked to the extent to which a firm is benefiting from *regulatory forbearance*.

Explicit (entity-specific) regulatory forbearance can be said to involve regulators explicitly allowing a firm that has inadequate levels of capital to continue in business. In extremis, a firm could be 'bust' on nearly all rational measures of its intrinsic financial strength, but if its supervisors do not raise any objections it may be able to carry on much as before.

A supervisor would probably need to be quite brave or foolhardy to extend much explicit regulatory forbearance, although the larger the firm and the more it can plead special extenuating circumstances the harder it can be not to be flexible to some extent.

Implicit (usually sector-wide) regulatory forbearance would typically be subtler. For example, firms might face a material amount of risk of a specific type, but the regulatory regime might not require them to set aside much or any capital to cater for the possibility that the risk might materialise. Working out *exactly* how much capital a financial entity needs to face the risks to which it is exposed is very tricky so it is nearly always difficult to tell whether implicit regulatory forbearance is happening.

Why might supervisors be willing to extend any regulatory forbearance at all to organisations, explicit or otherwise? Supervisors have some tricky balancing to do. Suppose they come down too hard on an organisation, maybe requiring it to raise more capital or imposing restrictions on what it can do. This may exacerbate its difficulties creating greater losses if the organisation's difficulties would otherwise have turned out to be temporary. Suppose, instead, they do nothing. The eventual losses may be larger and the supervisors may have egg on their face, if the organisation fails to turn itself round, see e.g. Box 3.4. All roads are invidious. However, we still give supervisors these responsibilities because their privileged access to confidential information from the organisation means that they are probably better placed than others to take such decisions.

Sharing this balancing act with others has natural resonance for supervisors. This is one reason why modern financial regulatory frameworks typically include elements relating to transparency and market disclosure.

If there are undercapitalised firms in the system (and this undercapitalisation is not very temporary) then these firms may be said to be benefiting from regulatory forbearance. However, it may not be clear to anyone at the time that this is happening. Further deterioration in the financial position of these firms won't necessarily increase the likelihood of immediate failure. It may just lead to increased regulatory forbearance. What we need is for regulators or supervisors to tire of extending regulatory forbearance or for the market to wake up one day realising they have extended too much of it to be sustainable.

Huizinga and Laeven (2010) argue that during the 2007–09 Credit Crisis banks typically overstated the value of distressed assets they held, resulting in them showing a more favourable capital position than would otherwise have been the case. In a sense, therefore, they benefited from some regulatory forbearance. With the benefit of hindsight this may have been an appropriate response for supervisors to have adopted. However, if the Crisis had got worse rather than better it might have led to even bigger challenges further down the line. In any extreme situation, there are almost certain to be arguments that favour quicker responses and arguments that favour slower ones. Valuation of assets and liabilities (particularly distressed ones) is not an exact science, so exercise of some discretion is a necessary part of a supervisor's role. The hope and expectation is that it will be done so wisely, fairly and not unduly favouring a specific sector or individual firm.

A worry that sometimes surfaces in the context of regulatory forbearance is the concept of *regulatory capture*. This involves organisations (or whole sectors) somehow gaining undue influence over regulatory and supervisory decisions. This can result from overly cosy relationships that lead regulators to place too much credence in what they are being told by individual firms. Occasionally it can involve outright bribery. Conversely, it would be perverse to ban any interaction between the regulators and the regulated. Both can gain understanding by listening to the other.

3.1.7 Viewing the 2007–09 Credit Crisis Through a 'Combined' Model Lens

We referred above to the 2007–09 Crisis as a possible example of the 'domino' model. But we could equally view it as an example of the 'combined' model. Prior to the Crisis many banks ran down their equity assets relative to their loan portfolios, took out sizeable positions ultimately linked to sub-prime debt and engaged in business models that were vulnerable to liquidity drying up. In short, there were system-wide weaknesses that went unrecognised. A trigger came along and the precarious state of their financial health became better recognised, leading to broader institutional distress. Whatever policymakers may say in public about the importance of interconnectedness, the actual response to the Crisis has primarily been that banks should become better capitalised and with capital of a higher quality than before. Most policy responses adopted subsequently have tried to address underlying weaknesses perceived in hindsight to be present in the run-up to the Crisis, rather than elements of bank business models that provided the proximate trigger for the Crisis.

3.1.8 Financial Services that are Critical to the Real Economy

Another way to identify which parts of the financial system are most 'in scope' of systemic risk is to identify those sorts of financial services that are most critical to the real economy. This sort of approach is illustrated by French, Vital and Minot (2015). They indicate that, as at the time of writing, the Bank of England's Financial Policy Committee (FPC) viewed as critical to the real economy anything relating to:

(a) Payment services;
(b) Channelling savings into investments; and
(c) Insuring against and dispersing risk

Nearly all 'direct' elements of the financial services industry (and potentially many other parts of the economy) get included on this basis. Money has two core roles, i.e. as a medium of exchange and as a store of value. These come within the remit of (a) and (b) respectively. By 'direct' we mean ones that involve some transfer of value from time to time between the entity and the rest of the economy.

So, depending on the breadth of your definition of systemic risk, most or the whole of the financial sector can potentially be caught within its net. As far as policy makers are concerned, systemic risk is not something that is only relevant to banks. Many other financial institutions are also perceived by them to be within the remit of their deliberations. The sorts of systemic risks involved are not all the same or of the same size, and the ways of tackling the risks involved may not be uniform across institutional types. However, this does not remove the sector from the scope of systemic risk deliberations, it just alters the outworking of these deliberations. These views chime with those implicit in the FSB's consultations on how non-bank non-insurer (NBNI) global systemically important financial institutions (G-SIFIs) might be identified, see Section 6.5 and see also e.g. FSB (2014) and FSB (2015a).

3.1.9 Economic Importance

Another argument supporting reaching beyond (direct) interconnectedness is based on broader economic importance. Usually financial sector firms are either stand-alone entities or part of groups specialising in financial services. However, this is not always the case, even for large firms. For example, GE Capital was designated in 2013 as a systemically important financial institution by the Financial Stability Oversight Council (FSOC), even though at the time it was owned by General Electric, a multinational conglomerate better known for manufacturing activities such as building jet engines.

If a part of the financial services industry is important enough to garner a significant fraction of the total revenue base, salary bill and profitability of the financial services industry then presumably it is doing something to merit these transfers from other parts of the economy. The saying goes 'no reward without risk'. The art of outcompeting others is to maximise reward whilst minimising risk, but it is hard for every firm in an entire industry to do this simultaneously. The size of an industry, almost certainly correlates to some extent with the potential risks the industry might engender at a macro-economic level if it were to hit hard times.

Put like this, we should expect policymakers to view any large part of the financial services industry as potentially having systemic relevance and to flex the meaning given to 'systemic risk' to make sure that the sector is not then ignored. Industries can't simultaneously be both important enough to command a decent slice of the total financial revenue of an economy and unimportant enough not to figure somehow in a reckoning of systemic risk, if systemic risk is understood broadly enough.

In broad terms, Alves et al. (2015) indicate that large EU insurers display relatively low interconnectivity overall, at least compared to large EU banks in isolation. Does that mean that policymakers will view insurers as less systemically important? Only to a certain degree. They are also non-trivial components of the whole financial system.

So far in this Chapter we have focused on systemic risks linked to capital adequacy and liquidity risk. It is worth noting that these are not the only factors that may contribute to financial stability if this term is interpreted more broadly. For example, some Bank of England Financial Stability Reviews have focused on broader issues that deplete public trust in financial institutions and therefore, in its opinion, hinder financial stability. Bank of England (2015) includes discussion of the potential systemic risks arising from misconduct risk and cyber risk. It notes on these topics:

(a) Misconduct risk: *'Misconduct has undercut public trust and hindered progress in rebuilding the banking sector after the crisis. It has also posed risks to systemic stability, with direct economic consequences. The fines and redress costs paid by UK banks, at £30 billion, are equivalent to around all of the capital that they have raised privately since 2009.'*
(b) Cyber risk. *'While in some areas the financial sector is leading efforts to combat cyber crime, the adaptive nature of the threat means that ways of managing this risk must evolve. As well as looking to build defensive resilience to threats, firms must build the capability to recover quickly from cyber attack, given the inevitability that attacks will occur. The evolving nature of the threat means that strong governance at the most senior levels of banks is required to build this capability in defensive resilience and recovery across technology and personnel.'*

3.2 Financial Sector Regulation

3.2.1 Introduction

Financial contracts can be quite complicated for those not expert in finance to understand (and even at times for experts!). It can be difficult to identify in advance the 'quality' of a financial service or product. They can be quite long-term in nature and it is generally only possible to test whether you have actually got what you expected in arrears. Economists describe such a situation as involving information asymmetry between the firm and its customers. The firm usually knows much more about whether the contract it is entering into is appropriate and likely to be honoured than the customer does.

Regulation in the financial services industry is designed to tackle this asymmetry. By imposing requirements on how firms behave and on how well capitalised they need to be, societies seek to level the otherwise uneven playing field between firm and customer and to limit the likelihood of firms taking 'unfair' advantage of their customers.

In broad terms, there are four main generic ways of regulating markets:

(a) *Voluntary codes of conduct.* For example, providers in a specific industry could group together to agree their own rules about how markets in the industry should function. These voluntary codes might for example cover how they will behave, what will be sold and how these products will be marketed.

(b) *Self-regulation*. This operates in a similar manner to voluntary codes of conduct except that the rules are not voluntary; everyone covered by the regulation will need to comply with them.
(c) *Statutory regulation*. The state sets out the rules that need to be followed by everyone operating in a specific market, and it (to be more precise, some supervisory body that it appoints) polices compliance with the rules.
(d) *A mixed regime*. This involves a combination of the above approaches. For example, statutory regulation could apply to pricing whilst voluntary codes of conduct could apply to marketing.

Voluntary codes of conduct and self-regulatory approach do exist within the financial sector, but increasingly organisations operating in this space have found themselves subject to statutory regulation. The financial services industry is typically viewed as heavily regulated. This reflects policymaker views as to the apparently limited effectiveness of other approaches to addressing the information asymmetries referred to above. Regulation also has the unintended consequence of erecting barriers to entry into this sector.

3.2.2 Overarching Political Considerations

How financial regulation ought ideally to be structured is inevitably heavily influenced by our views on the purpose of regulation and on how 'unfairness' might work out in practice.

At the very highest level, we have the issue of the extent to which society should or is seeking to adopt a 'command' (or 'collective') as opposed to a 'market' economy. For those favouring a command approach, regulation may be viewed as just another means of ensuring that the right commands get implemented in practice. Regulations might aim to prohibit or limit activities that those in power view as undesirable.

Classically, this sort of topic might have been viewed through a Cold War perspective, with collective Soviet-style command economies deemed to be pitted against highly capitalist economies in which resources were exclusively apportioned according to market forces. In practice, there are many shades in between (and there were even at the height of the Cold War). In nearly every major developed economy some sectors are nearly exclusively government controlled, e.g. the army and police force, and others are nearly exclusively in the private sector, e.g. retailing. For the latter types of sector, regulation is seen principally as a means of tempering some of the excesses or undesirable social consequences that exclusive focus on market forces might otherwise bring. For example, regulations may impose minimum standards of hygiene on restaurants etc., to limit public health risks.

Quite where financial regulation fits into this spectrum is a matter of debate and probably always will remain so. Traditionally, within the capitalist West, a relatively laissez-faire approach has been deemed appropriate. This was consistent with the Anglo-Saxon capitalism that seemed prior to the 2007–09 Credit Crisis to have been in the ascendancy. However, one consequence of the Crisis was a loss of confidence in these economic norms, and by implication in the ways in which such economies handled their financial systems.

A reminder of this is the perceived incongruity of Hank Paulson, the then US Treasury Secretary, having to prostrate himself before the US Congress in 2008, seeking the authority (and money) to prop up the US financial system in a country perceived as averse to government bailouts of any sort. In any case, even previous Anglo-Saxon laissez-faire approaches often had elements of 'instructing' the economy to work in a specific manner. Successive UK governments had for many years prior to the Crisis promoted London as an international financial centre. This included adopting a relatively light touch regulatory framework because it was thought likely to foster innovation.

One step down, we might subdivide the purpose of regulation in the financial community into two broad strands:

(a) *Conduct regulation*. This relates to the behaviours firms are required to adopt, e.g. stopping firms from selling services to individuals that firms should have known that these individuals did not need. Essentially it is attempting to answer the question: *What are fair sorts of behaviours we want to encourage/discourage financial entities to adopt within their own businesses/structures and in relation to how they interact with their customers?*
(b) *Capital regulation* (otherwise known as *prudential regulation*). This relates to amounts of (surplus) capital that a firm is required to hold. This type of regulation is seeking to answer the question of what level of capital (and of what type) do we want financial entities to hold, individually and in aggregate, to limit the potential loss to customers or drain on the public purse if the entity (or the whole sector) fails?

In some jurisdictions, the subdivision between conduct and capital regulation is formalised by having separate regulatory bodies focusing on each type of regulation. For example, the UK has a Financial Conduct Authority (FCA) that primarily focuses on (a) and a Prudential Regulation

Authority (PRA) that primarily focuses on (b). We include the word 'primarily' because some UK financial firms are subject to minimum capital requirements but are only regulated by the FCA. Also, being knowingly inadequately capitalised is probably a behaviour that we want regulation to discourage so there is no strict dividing line between the two sorts of regulation mentioned above.

3.2.3 The Legal Form of Financial Sector Regulation

Regulation is ultimately imposed by law. However, given the complexity of modern finance it is not usually possible to provide the level of detail needed within primary legislative statutes. Instead, many elements of any regulatory framework may be contained in secondary legislation and/or in rules set out by regulatory bodies that are created by primary legislation.

In the EU, the top of this legal tree generally involves formal Directives, agreed by member states, the EU Commission and the EU Parliament. Directives relating to financial services regulation generally set out the overarching characteristics of the regulation to be applied to specific sectors (e.g. the Solvency II Directive applies to insurers whilst the Capital Requirements Directive to banks and some other types of institution). Even higher up the tree are the treaties that establish the EU itself. These treaties limit what can be included within Directives. For example, social and labour law is generally reserved to member states under these treaties.

Below Directives come Delegated Regulations. For example, the Capital Requirements Directive includes powers to introduce Regulations on various topics specified in the Directive. These are codified in the Capital Requirements Regulation. There are also Directives and Regulations that establish centralised regulatory bodies that form part of the European System of Financial Supervision, i.e. EBA, ESMA, EIOPA and ESRB, see Box 2.6. EU Directives also require member states to set up national supervisory bodies to apply the regulation in practice and to sanction firms not adhering to the relevant regulatory requirements. These national supervisory bodies do not need to be exactly aligned with the centralised regulatory bodies (provided all regulatory supervisory requirements are being met) but unsurprisingly there is reasonably good alignment between the two.

A similar picture applies in the USA. Primary legislation such as the Dodd-Frank Act specifies in broad terms how financial services regulation should operate, but there are then several regulatory and supervisory bodies charged with fleshing out the details and supervising individual firms. There can be tensions between the federal and the state level. This is particularly evident with insurance, which is generally regulated at a state level except for the very largest firms (i.e. the ones deemed systemically important) which at the time of writing were in effect regulated at the federal level (by the Fed).

It should not be assumed that these regulatory requirements and regulatory structures are static. Far from it! Whenever a new crisis strikes, politicians tend to react by modifying existing regulation, often in a piecemeal fashion although sometimes (most typically if the crisis is sufficiently large) in a more big-bang fashion. Regulatory responsibilities also get shuffled around, extended and (occasionally) removed as time progresses. For example, the UK used to have a unitary regulator until 2013 called the Financial Services Authority (FSA) combining both conduct and capital regulation, before these two roles were split between the FCA and the PRA. One outcome of the 2007–09 Credit Crisis has been a tendency to place prudential (i.e. capital) regulation within a jurisdiction's central bank. In extremis, bailing out financial firms typically involves close interaction between relevant regulatory bodies and those who hold the purse strings.

3.2.4 Supervisory Actions

Regulation that has no teeth is unlikely to be effective. This is as true in the systemic risk field as it is in any other area touched by regulation. It is little use identifying factors that might cause systemic risk if there is no practical way of incentivising individuals and organisations to address these factors.

Regulatory sanctions vary depending on the type of regulation being breached. Regulation usually gives powers to supervisors to fine regulated firms or to intervene in their management if they breach conduct regulations. Firms may also be required to recompense customers who suffer loss due to their regulatory breaches. Firms whose conduct is sufficiently poor may be stopped from carrying out specified regulated activities or even occasionally may have their entire licence to operate in the financial sector revoked.

However, if the breach relates to minimum capital requirements then imposing fines on the firm will merely make the position worse. So more usually in such circumstances firms are required to identify ways of obtaining extra capital. An escalating ladder of supervisory actions may get triggered if the firm is unable to address capital shortfalls quickly enough to address supervisory concerns. For example, the firm may be stopped from writing new business or may have other restrictions placed on it. In extremis, the supervisor may impose new management on the firm or forcibly 'resolve' it, i.e. impose actions on it that have the aim of closing it down and as far as possible meeting its liabilities to its existing customers.

The existence of such powers highlights another asymmetry between firm and customer. Firms can and do from time to time decide to wind themselves up in an orderly fashion, if they have the capital resources to do so. For example, they may no longer believe that a specific business activity is an attractive one to pursue. However, the risk if they run into difficulties is that they may become incentivised to 'gamble for resurrection'. This involves taking on outsized bets in the hope that the bets will come good but worsening the eventual position of customers if they go sour. Supervisors need to be able to react quickly if such behaviour appears to be on the cards.

Usually in such circumstances supervisors do not actually take over the day-to-day running of a business. Instead they install experts, e.g. liquidators, receivers or other outside consultants, who have more hands-on experience than they do of maximising the recovery value in such circumstances.

3.2.5 Recovery and Resolution

The notion of 'recovery and resolution' has gained greater prominence recently, given high profile failures within the financial sector. Ideally, supervisors somehow prod an otherwise failing organisation in the direction of recovery back to good health. However, if this is impractical then supervisors may need to step in somehow to ensure that it is resolved effectively and in a way that as far as possible honours liabilities to customers.

Recovery and resolution takes on added importance in the context of systemic risk. If there is a risk that the 'disorderly' failure of one organisation might lead to contagion and failure of others then it becomes particularly desirable to resolve the organisation in as orderly a fashion as possible should the need arise.

There is no exact definition of 'disorderly' in this context although the assumption is that most people would recognise a disorderly failure if one happened. Some element of disorder is almost inevitable if a firm fails unexpectedly.

Most firms in the financial sector are established as companies so can also be bankrupted, although some regulatory frameworks in effect stop this happening if the regulator believes that it would result in undesirable outcomes. The impact that disorderly failure has on contracts the firm has previously entered into may be quite complex, particularly if the firm itself has a complex structure.

One of the best-known examples of a 'disorderly' failure was that of Lehman Brothers in 2008, see Box 3.5. Particularly problematic in its aftermath was uncertainty over who what entitled to what. This issue, and other similar cases that arose as the 2007–09 Credit Crisis unfolded, has led regulators to require many types of financial firms to prepare 'recovery and resolution plans' identifying how they could (where practical) be resolved in an orderly fashion, see e.g. Bank of England (2009).

Colloquially these were originally known as 'living wills' although more recently the concept has been extended to cover the sorts of governance frameworks that need to be in place in advance for such plans to be likely to help in a distressed situation.

Anyone who has had to sort out the estate of a deceased close relative will appreciate the benefits of the relative having a well-drafted will (and e.g. centralised records of a person's financial affairs). Much the same sort of logic arguably also applies in the business world. Arguably, it is wrong to assume that companies won't die. Indeed, they probably on average have a higher mortality rate per annum than humans! From governments' and regulators' perspectives, such documents and frameworks, if soundly drawn up, should simplify the process of winding down a company. The process of drawing them up might also highlight weaknesses in group structures. These weaknesses might add complexity in actual resolution situations but might be relatively simple to rectify in the meantime.

Conversely, firms caught by such requirements have argued that their (possibly complex) structures may serve a useful purpose (rather than being simply a result of unintentional accrual of business complexity through time). They may argue that unravelling these complexities in the meantime may be costly (and possibly counterproductive). Preparing such plans can also be hard work. There are many in the financial community who are sceptical of their usefulness, particularly for sectors (such as insurance) that were largely out of the firing line during the Crisis.

Even in sectors where the practice is better developed (such as banking) there seems to be at present a gulf between regulator expectations and actual depth of analysis being carried out by regulated firms. This arguably reflects divergence of interest between regulated firm and regulators. Shareholders of a regulated firm have little direct interest in making themselves simpler to resolve. In the event of a firm running into such a level of difficulty, most shareholder value will have already evaporated. Moreover, by making themselves simpler to resolve they may be less likely to be deemed TBTF and hence bailed out by governments.

It is therefore unsurprising that banks do not appear to have made as much progress as regulators might have liked in developing credible resolution plans. Conversely, affected firms argue that at least some of this lack of progress results from regulators providing unclear guidance on what they want. The Fed noted the following as typical of plan shortcomings (at the time) in Federal Reserve (2014a) (for US firms) and Federal Reserve (2014b) (for non-US firms):

> 'While the shortcomings of the plans varied across the first-wave firms, the agencies have identified several common features of the plans' shortcomings. These common features include: (i) assumptions that the agencies regard as unrealistic or inadequately supported, such as assumptions about the likely behavior of customers, counterparties, investors, central clearing facilities, and regulators, and (ii) the failure to make, or even to identify, the kinds of changes in firm structure and practices that would be necessary to enhance the prospects for orderly resolution.'

Related to the topic of resolution planning is an increased emphasis being placed in regulatory frameworks on the concept of reverse stress testing, see Box 5.3. Reverse stress testing deliberately requires firms to postulate how their business model might crumble and what might be done to mitigate the causes and consequences.

Box 3.5: The failure of Lehman Brothers

The most notorious failure of the 2007–09 Credit Crisis was Lehman Brothers. Prior to its bankruptcy in September 2008 it was the fourth-largest US investment bank (after Goldman Sachs, Morgan Stanley and Merrill Lynch) and involved in a wide range of activities, including equity and fixed income sales, market-making and proprietary trading, provision of investment research, investment management, prime brokerage, private equity and private banking.

Going into the 2007–09 Credit Crisis, Lehmans had large positions in subprime and other lower-rated tranches of mortgage-backed securitisations, which went sour as market conditions deteriorated. Its market capitalisation declined by c. 73% during the first half of 2008 and it announced a large quarterly loss on 9 June 2008, as credit markets deteriorated and its hedges proved ineffective. It seems to have used some cosmetic accounting fixes to make its finances appear more robust than they were. It announced an even larger quarterly loss on 9 September 2008 causing a further plunge in its market capitalisation. Frantic attempts to sell the business were made over the next few days, including at a meeting on Saturday, 13 September 2008, called by Timothy Geithner, the then president of the Federal Reserve Bank of New York. However, no-one was prepared to buy the company and shortly before 1 am on Monday, 15 September 2008, Lehman Brothers Holdings announced that it would file for Chapter 11 bankruptcy protection, the largest in US history.

Many commentators view Lehman's bankruptcy as itself having played a major role in the unfolding of the Crisis. As explained in Box 3.3, it came at a crunch time when other organisations were also spiralling towards failure or at least a need for major recapitalisation. Given its highly interconnected status a particularly large number of market participants had it as a counterparty (indeed, one of the reasons it failed was because many market participants lost confidence in its likely ability to continue, i.e. it in effect suffered the investment banking analogue of a bank 'run'). To the extent that one organisation's failure might create domino-like failures of other organisations then failure of a large (and therefore almost inevitably highly interconnected) US investment bank such as Lehmans was likely to be problematic. Indeed, Lehman's failure does appear to have been the proximate cause of failures of some other market participants, such as the Reserve Primary Fund, see Box 4.8.

An alternative interpretation is that the failure was more a reflection of the severity of the situation at the time. The Federal Reserve Bank of New York sought to adopt an approach that had worked previously. This involved calling together leading banks to arrange a bailout or otherwise orderly liquidation of the organisation in trouble, as had happened with Bear Stearns, an investment bank that had been sold to JP Morgan earlier in 2008 and before that with the rescue of LTCM in 1998. Over the same weekend as Lehman's failed, Merrill Lynch agreed to be acquired by Bank of America and a few days later the other two large US investment banks (Goldman Sachs and Morgan Stanley) agreed to become traditional bank holding companies, drawing to a close the ascendancy on Wall Street of securities firms, 75 years after they had been separated from deposit-takers under Glass-Steagal, see Box 4.1. During the same weekend as Lehmans failed, the US authorities also had to deal with the impending near collapse of AIG, see Box 3.1. To the extent authorities did not at the time have sufficient financial or political firepower to deal with all the problems swirling around at the time, maybe letting Lehmans fail was the least bad option.

3.2.6 Compensation and Protection Schemes

Ideally, resolution of a financial organisation would result in its wind-down with no (ultimate) losses to customers. However, this is likely to be too optimistic in some (many?) circumstances. Many countries therefore operate compensation or protection schemes which aim to limit the loss customers suffer in the event of failure of regulated financial organisations. An early example was the FDIC, introduced by the USA in the 1930s, see Box 3.2.

It can be very expensive to provide full protection of all customer liabilities. Often protection schemes include limits on coverage being provided and only apply to certain customers (e.g. retail customers). Sometimes such schemes are set up as (or provided by) insurance companies but often they are set up as separate statutory bodies (in which case they will not legally be providing insurance cover even if economically their coverage is akin to insurance).

Financial sector compensation or protection schemes can either be funded in advance (typically from other market participants in the same sector) or in arrears. Either approach should (eventually) address the costs of meeting the required coverage, if the generic business model of the sector supporting the scheme is sufficiently robust relative to the losses sustained. However, the two approaches will in general apportion costs differently. The levies raised by the scheme usually depend on business volumes which will change through time at different rates for different firms. Whether a scheme is funded in advance or in arrears may also influence the likelihood of the government feeling obliged to support it. Prefunded arrangements have funds already set aside, which can be earmarked for support purposes. Conversely, setting money aside in advance may seem unfair to earlier contributors if the funds are not then needed. Some countries, like the UK, have compensation schemes that have multiple compartments (for different parts of the financial sector) but with some scope for costs to jump across compartment boundaries if losses are large enough.

However, if a particularly major systemic risk event arises then these schemes may run out of money or the political will to address the problems purely from contributions from other industry participants may dry up. The latter is more likely if the firms otherwise being asked to stump up for such losses are sufficiently effective in arguing that it is ‘unfair’ on them to be asked to do so. Things can then get less predictable. Fairness is a very important concept in regulatory policy!

A very wide range of possibilities arise in such circumstances. They can be thought of as akin to the disorderly failure of the compensation /protection scheme (or at least of its failure to achieve fully its original purpose). Possibilities include:

(a) The state bails out individual firms and/or the protection scheme
(b) Parties that politicians and/or lawyers can somehow associate with the failure end up ‘bailing in’ or contributing to the cost of meeting these losses. Accountants’ professional indemnity funds (and hence to some extent the insurance industry, if these funds have taken out third party insurance or reinsurance) seem to be one area that lawyers have in the past been effective at targeting.
(c) If the scheme’s resources are exhausted and no one else meets the losses involved then ultimately the losses fall on the relevant customers, perhaps

with the sharing out of these losses being recast by the courts, if they consider the contractual arrangements otherwise applicable to be unreasonable.

Why are some such schemes called 'compensation schemes' whilst others are called 'protection schemes'? 'Compensation' perhaps has the connotation of being something that customers are entitled to, whereas 'protection' perhaps carries more the connotation of doing the customers a favour. This perhaps reflects societal norms about the expected role of the state in financial services. Some financial sectors in some countries do not have such schemes. In such cases customers are more exposed to the principle of caveat emptor, i.e. buyer beware, and are more at risk in the event of failure of a firm they have bought a financial service from. The structure of compensation /protection schemes and the expected response of states if these schemes prove inadequate influences the extent to which systemic risk events may trigger losses to the public purse. Even if the public purse does not pick up losses directly, it may also do so indirectly in stressed scenarios, e.g. due to reduced tax take or costs borne by other parts of the state budget.

3.2.7 Conduct Regulation

Regulations defining what sorts of conduct organisations can adopt are very widespread across the whole economy. For example, we have rules limiting the scope of businesses to include unfair terms in contracts agreed with retail customers and we have employment legislation designed to stop employers taking unfair advantage of employees. Humans have a very strong general sense of 'fairness'.

Some anthropologists see this human characteristic as core to our 'hypersocial' nature. It has allowed us to collaborate much more broadly and successfully than most other creatures and seems an innate part of our make-up. Religious people like me may view it as a God-given attribute of humans that if we lose somehow diminishes our soul or essence. Others may view it merely a consequence of drivers present in human evolution. Without some notion of intrinsic fairness, and by corollary some willingness to believe that we will ourselves be treated broadly 'fairly' by others, much economic activity would grind to a halt. Who would be willing to invest for the future if they lacked any confidence that they would be fairly treated in terms of rewards that might flow from such investing?

Whilst nearly everyone thinks that the rewards from human endeavour should be 'fairly' divided up, different people have different views on exactly what this means in practice. Practical perspectives on what is 'fair' seem to differ somewhat across societies. In Chapter 7 we explore further the impact that shifting societal norms might have on systemic risk issues.

The experience of the 2007–09 Credit Crisis has led to several regulatory initiatives in conduct regulation. Kemp and Varnell (2010) noted that regulators are quick to argue that additional capital is not necessarily the most practical or even the most appropriate way to protect customers against risk. Instead regulators often seek to place a strong emphasis on firm behaviour, including governance practices, organisational structures and corporate culture. Proposals emanating from the Crisis along these lines have included:

(a) *Restricting the size of systemically important organisations*. Direct intervention in relation to this goal has been limited, although it could be argued that some of the changes to the banking sector have indirectly facilitated this outcome, e.g. by making leverage more expensive or banning some types of activity, see (b). Governments also appear more mindful of the risks of having an out-sized banking sector relative to the size of their whole economy.
(b) *Limiting the types of activity that regulated entities can undertake (especially if the regulated entity can access deposits benefiting from implicit or explicit government deposit protection guarantees)*. This idea has gained more traction, e.g. with introduction of restrictions on bank proprietary trading and with the proposed ring-fencing in some countries of retail banking activities.
(c) *Facilitating changes to market structures perceived likely to reduce systemic risks*. The most obvious example is the requirement to trade specific types of instruments through centralised exchanges and central clearing rather than over-the-counter (OTC), see Section 4.8.2.
(d) *Limiting the scope of others to profit from a firm's weakness and hence to increase the cost to the government of bailing it out*. This initially led to bans on short-selling but these mostly proved problematic to implement or unworkable. However, to some extent similar effects have arisen naturally through market forces. Arbitrageurs are facing significantly higher funding costs because of added capital requirements imposed on banks, see e.g. Devasabai (2014).
(e) *Improving resolvability of firms in difficulties*. This has become a major strand in regulatory thought, see Section 3.2.5.

(f) *Improving liquidity risk management processes*. Such changes have proved relatively uncontentious given the extent to which the 2007–09 Credit Crisis can be viewed as a liquidity crisis rather than merely a credit crisis. The banks that failed during the crisis disproportionately relied on wholesale money markets for their funding. It was when these funding sources dried up that they ran into problems. They were then unable to use other assets they possessed to source the liquidity that they needed to continue as going concerns. The liquidity management practices and backstops for large asset management companies have also come under the spotlight, see Section 4.5.

(g) *Improving overall risk management processes and governance disciplines*. This has gained widespread traction. It ties in with the increasing emphasis being placed on 'enterprise' risk management (or 'entity-wide' equivalents for institutions like pension funds that do not necessarily see themselves as 'enterprises' per se). In the UK enhancements to risk disciplines, risk functions and risk management processes mandated by e.g. HM Treasury (2009) have been applied across nearly all areas regulated by the PRA or the FCA rather than just to banks. A recently agreed EU Directive on pension funds, known in the EU as institutions for occupational retirement provision (IORPs) will beef up regulatory expectations for the quality and depth of risk management carried out by such organisations.

3.3 Regulatory Capital and Economic Capital

3.3.1 Introduction

At a very high level capital requirements are imposed on financial firms to address an information asymmetry between firms and their customers. Customers often cannot in practice test the 'quality' of the financial service they are purchasing at point of purchase, because it is generally only fully testable in arrears. This conceptually differs from a situation where a customer is, say, purchasing an apple, which can more easily be inspected at point of purchase. Jurisdictions that do not impose any capital requirements on financial firms are (more) susceptible to unscrupulous individuals setting up sham arrangements. The risk is that individuals part with their money in the expectation of receiving a valuable financial benefit in the future which never then materialises (with the unscrupulous individual making off with the money instead).

Of course, we can overplay this difference. Even apples can have flaws that are not obvious to the eye, e.g. they might be covered by poisonous pesticides or might turn out to be inedible. In more advanced economies, nearly all goods and services bought by the public are subject to *some* state-imposed regulations that aim to maintain minimum quality standards. Even when these standards are limited, firms often have a vested interest in maintaining their own self-imposed quality standards, to enhance their own brand reputation and hence the price that they can charge for their goods and services.

The point about financial services is that they often involve quite complicated financial elements, are often much larger in value than transactions people might make as part of, say, their grocery shopping and they often involve customers placing considerable trust in the financial services provider in question and in it still being around to honour its commitments when they fall due.

Occasionally governments may view financial services firms as honey pots to raid when necessary. In these circumstances, governments may implicitly or deliberately formulate capital requirements more as a means of taxing or gaining support from these firms than as a means of delivering overall financial stability. Examples might include requiring banks to deposit excessively large amounts of money with central banks or finance ministries on which artificially low interest rates are paid (especially if the government concerned is unable to borrow from others at the time).

More commonly, at least in advanced economies, capital requirements in the financial sector are not (deliberately) a disguised form of arbitrary taxation. Instead, they ostensibly aim to be a proportionate way of addressing the information asymmetries referred to above in a manner that is fair between different market participants. This means that policymakers are generally not deliberately seeking to favour one firm over another within a given sector or even across sectors. They should therefore in principle be open to arguments that seek to demonstrate that regulations are inadvertently doing otherwise.

3.3.2 Risk-Sensitive Capital Requirements

Different types of financial institution are subject to different regulatory capital requirements and it is not normally easy to compare regulatory frameworks from this perspective. In general, regulators seek to adopt approaches which exhibit some degree of risk sensitivity, i.e. impose higher capital requirements on firms that are facing higher amounts of risk. Conversely, risk measurement is an imprecise science, so some trade-off between risk sensitivity and simplicity is usually sought, see e.g. Box 3.10.

An early example of a risk-based approach to setting capital requirements was introduced by the Basel Committee on Banking Supervision (BCBS) in its original Basel I accord. It included the concept of risk-weighted assets (RWA). This involves assigning different risk weights to different types of assets (and to off balance sheet items), e.g. typically assigning higher weights to loans to corporates versus loans to governments, because corporates are typically deemed more likely to default than governments. In some cases, banks can set their own risk weights, see Box 3.10. Originally, banks merely needed to identify all their assets, risk weight them and add them up. The minimum capital requirement they were then subject to was a specified fraction of this total and this was compared to the amount of (surplus) assets they had, to identify if they were adequately capitalised. Over time the computation has become more complicated but still at its core is a computation expressed in the form of RWAs.

RWAs do not feature as a concept within Solvency II. This is perhaps because Solvency II explicitly allows for diversification effects between different exposures, so a linear adding up of exposures as is implicit in RWAs is less meaningful. Also, a much greater part of an insurer's overall risk exposure arises from interest rate or other market risks, whereas RWAs are arguably best suited to credit risks of the sort that form a much higher proportion of a bank's total risk 'budget'.

3.3.3 Regulatory Capital Requirements

There are many other specialist texts that describe in detail the (minimum) capital requirements applicable to banks, insurers and other financial services entities. The precise structure that applies under any specific regulatory framework is often very complicated, so it is often helpful to develop a conceptual framework that caters for as many as possible of these complications. We include in Box 3.6 such a conceptual framework. Some key issues that are particularly relevant to systemic risk are:

(a) *The strength of the inherent desire of firms to remain adequately capitalised*
We might expect most financial services firms to have an incentive to remain adequately capitalised, to maximise the value of the relationships they have with their customers. Firms ought to have some incentive to hold some level of surplus capital since they themselves suffer in the event of their own failure.

Regulatory frameworks like CRD and Solvency II do require regulated firms to identify what capital the firm believes it 'intrinsically' needs irrespective of any applicable regulatory requirements and then to remain adequately capitalised relative to this level of desired capital. If these incentives are sufficiently strong then explicit capital requirements imposed by regulation should become less important. Before the 2007–09 Credit Crisis most banks that subsequently failed during the Crisis presumably believed that they were adequately capitalised even if with hindsight this view was wrong. A fundamental issue is that these intrinsic incentives may be too weak to result in an 'adequate' level of capitalisation from the perspective of regulators.

(b) *The inherent financial stability aspects of any capital regulation*

At a very high level, capital requirements exist to increase the likelihood of financial services entities honouring their commitments. These commitments can be expected to be more robust, i.e. more 'stable', if suitable capital requirements are in place. In some sense, *all* capital regulation in the financial sector has a financial stability aspect, if financial stability is interpreted sufficiently broadly.

This type of argument leads some academics to conclude that it is ultimately impossible to differentiate rigorously between macroprudential and microprudential policies. It also leads some to conclude that macroprudential policy is ultimately just a politically convenient way of implementing otherwise less popular microprudential policies.

(c) *The inherent link between capital requirements and risks being faced*

It is natural to aim for capital requirements to link somehow with the actual risks present. It would seem perverse if firms that reduced the likelihood of failing to honour their commitments were typically penalised for doing so by being subject to higher regulatory capital requirements.

Conversely, designing a regulatory capital framework that completely avoids any possible anomaly would be extremely difficult or impossible and is likely to be prone to excessive model risk. In practice, minimum capital requirements implicitly involve a balance between simplicity and complexity and between risk sensitivity and more formulaic elements. They also ultimately involve judgements (by whoever ultimately signs-off on the framework details) on how much capital is needed to face any given risk and on how to handle combinations of risks.

The trade-off between simplicity and complexity influences views on the usefulness of *internal models* versus *standardised formulae*. A standard formula approach to setting regulatory capital requirements involves a formula that is relatively simple to implement. These usually involve computations based on business volumes or other metrics that are relatively easy to source, or involve aggregating the results of applying standardised stress tests to the firm's balance sheet. In contrast an internal model is a (usually more sophisticated) model that the firm itself has identified as appropriate for assessing the capital that should be held for some or all the risks the business faces.

Philosophically, there are two conflicting views on use of more sophisticated models in regulatory capital frameworks. On the one hand, financial services firms are often quite complicated and hence fully understanding the risks that they face is also likely to be a complicated task. Sophisticated models that align with such risks should therefore help with this understanding. On the other hand, sophisticated models offer greater scope to finesse or otherwise arbitrage regulatory capital requirements, i.e. may offer firms too much incentive and scope to select capital requirements that suit themselves, however much they might protest otherwise.

The extent to which firms can (or are encouraged to) use their own models for setting regulatory capital requirements is an indication of the amount of trust regulators have in firms in this respect. Of late this trust seems to have been in relatively short supply, particularly within the banking regulatory community. This is influencing the extent to which internal models are being deemed acceptable for regulatory capital purposes in that sector.

(d) *The treatment of parts of the financial services industry that are not subject to capital regulation*

Whilst many parts of the financial services industry are subject to capital regulation, this is not true of all players, especially in the institutional, i.e. non-retail, arena. For example, some types of entity, colloquially known as 'shadow' banks, see Section 4.6, loosely speaking carry out bank-like activities but are not regulated like banks. They may not subject to any meaningful regulation or minimum capital requirements. This is perceived by regulators to create the risk of regulatory and/or capital arbitrage, see Section 3.3.5.

(e) *Any overarching policy goals we may want regulatory capital requirements to exhibit*

At a very high level, if regulatory capital requirements are lowered (and are a practical constraint on how regulated firms operate) then this may make firms more willing to take on risks. In a banking context, this might mean a willingness to expand lending, which may be deemed desirable for macroeconomic reasons. We might also decide we want regulatory capital regimes to favour diversity of business strategies across a sector. Part of the perceived problem during the 2007–09 Credit Crisis was the wide adoption of business strategies that depended on their viability on benign liquidity conditions.

The strength of the incentives mentioned in (a) may vary by sector. This may explain an otherwise puzzling difference between banks and insurers. As a sweeping generalisation, banks (or at least the people working in them) currently seem to place relatively less emphasis than insurers on their own intrinsic assessments of required capital and relatively more emphasis on regulatory minima. Perhaps insurers generally feel a greater need to be intrinsically adequately capitalised, because of the usually longer-term nature of the contracts they enter into with their customers. A puzzle is that many banking activities also have long time-scales and rely heavily on trust between the bank and its customers. Maybe banks (or bankers as a class of individuals) just don't have such a long-term mindset. Some of the changes in remuneration policies imposed on banks since the 2007–09 Credit Crisis aim to foster a longer-term perspective on such matters. Or maybe it is just that banks are still rebuilding their capital bases following the 2007–09 Credit Crisis, and so (hopefully) will one day reach the happy position where regulatory minima no longer have such a dominant impact on their behaviour.

Box 3.6: A conceptual framework for capital adequacy

1. A conceptual framework that caters for many different aspects of and approaches to capital adequacy is contained in Kemp (2009) and Kemp and Varnell (2010). It is based on the insight that (absent future new business or capital raising) the full (or 'economic' or 'holistic') balance sheet of any financial firm or organisation can be conceptually organised as in Fig. 3.1. In this Figure it is assumed that all assets and liabilities are included, even ones that are off balance sheet items according to relevant generally accepted accounting principles used to draw up traditional financial statements
2. In this representation, 'customer liabilities' correspond to liabilities to depositors (for a bank), policyholders (for an insurance company) or beneficiaries (for a pension fund). There may be some liabilities that rank above customer liabilities (e.g. mortgages secured on specific assets). Usually, however, most non-customer providers of the organisation's capital have a priority ranking below the firm's customers. In the event of default customers will be paid in preference to these lower-ranked capital providers.

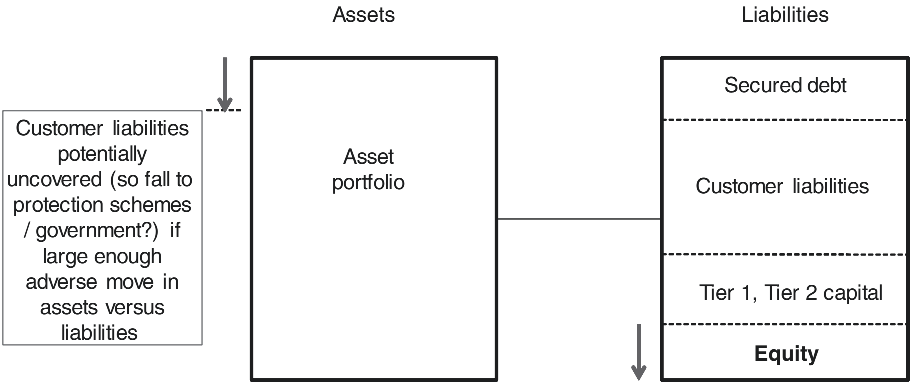


Fig. 3.1 Stylised description of balance sheet of a financial organisation
Source: Nematrian

3. Stand-alone entities are typically only able to replenish capital ranked below customer liabilities by raising new capital from elsewhere or from profit-making activities. An entity's ability to raise fresh capital normally depends heavily on the extent to which it is expected by outsiders to have access to profitable new business flows in the future. An exception is where the 'capital raising' involves calling in capital already committed but not yet paid over to the entity. This, for example, might apply if a subsidiary has run into trouble and its parent has previously implicitly or explicitly agreed to support the subsidiary in such circumstances.
4. All other things being equal, the greater the amount of capital the organisation has ranking below its own customer liabilities the better protected are its customers against the organisation running into difficulties, provided interest and /or dividend rates payable on the relevant capital instruments are sufficiently modest and their expected maturity or repayment dates are sufficiently far into the future. Only after this capital cushion is exhausted would customers start to find their liabilities not being fully honoured. A corollary is that 'solvency' is never absolute. As long as some customer liabilities exist there will always be outcomes we can envisage that are severe enough to exhaust this cushion and lead to customer liabilities not being honoured in full. For example, the organisation might suffer a particularly massive fraud. It might be hit with a particularly large back tax or liability claim. It might suffer reputational damage which exhausts its future earning power. Or it might just make the wrong business decisions and end up making losses which exhaust its capital base.
5. Although it is conventional to specify capital adequacy in terms of the amount of capital an organisation requires, we can instead specify the problem in terms of the yield spread (versus some suitable risk-free rate) that would or should apply to customer liabilities were they to be traded freely in the

market place. Such a conceptual framework highlights many subtleties that arise in theory and in practice with solvency computations, e.g.:

(a) All other things being equal, more capital (if its priority is below customer liabilities) provides greater protection for customers, but lowers the returns to capital providers (unless it leads to greater access to profits from new business).
(b) The required target capital level depends on the extent to which assets match customer liabilities (since the greater the volatility in the difference the greater the likelihood of capital being exhausted).
(c) The merits of capital that helps in a 'gone concern' as opposed to capital that merely helps in a 'going concern' situation become easier to appreciate, thus providing a clearer theoretical justification for treatment of different capital 'tiers', see Section 3.4.5.
(d) Treatment of liquidity risk becomes conceptually easier to visualise. If we invested in paper ranking the same as customers, how would the yield spread we would want be influenced by the liquidity characteristics of either assets or liabilities (or both in tandem)?
(e) A yield spread, being ultimately derived from a weighted average of outcomes over all possible scenarios includes the entire spectrum of outcomes. It therefore includes ones in which the entity defaults. These would otherwise be ignored if the focus is merely on limiting ruin probability to a given level. This has relevance to the question of whether to use expected shortfall (or tail value-at-risk) rather than value-at-risk (VaR) as the main risk measure for capital adequacy purposes, see Box 5.1 and Section 8.4.
(f) Given (e), the framework can conceptually handle who bears any losses (and the sums involved) arising from entity default. It is these losses that arguably are the ones that have the most visible potential to flow through to governments and/or industry-wide protection schemes.
(g) Many issues relating to procyclicality become clearer. In effect the issue becomes how we might want the target yield spread to vary through time (and between sectors) depending on economic circumstances.
(h) The appropriate treatment of 'own credit' risk in solvency computations is clarified, see Box 3.8. In effect, it no longer features in the calculation, since we are now solving for a given target level of own default risk rather than trying to work out how to take account of the actual level present.
(i) By referring to the spread that would otherwise apply on the open market, the approach can be formulated in a market consistent manner (if so wished), which in principle may help to limit scope for regulatory arbitrage. By 'market consistent' we mean aligned with observable market prices, see Section 3.4.2 and Section 7.5.

The framework is also sufficiently rich to illuminate more subtle issues. For example, it can frame a discussion of what, if any, allowance should be incorporated in regulatory capital computations in respect of sovereign default risk (not just of other sovereigns but also of the government of the jurisdiction in which the entity is domiciled). The definition of 'risk-free' against which the spread is measured can, for example, be set before or after allowing for this risk, depending on whether it is thought that customers would expect their liabilities to carry this risk.

The 'waterfall' style of approach implicit in such the framework can also be applied to the financial or economic system (with the asset portfolio now consisting of multiple firms), or to whole sectors within it, rather than just to individual firms in isolation.

3.3.4 The Structure of Regulatory Frameworks

The difference between regulatory minimum capital requirements and intrinsic assessments of how much capital is needed by a firm is embedded in the structure of most modern financial sector regulatory frameworks. These tend to adopt a '3 Pillar' approach as illustrated in Fig. 3.2. The exact contents of each Pillar vary a little between different regulatory frameworks but in broad terms:

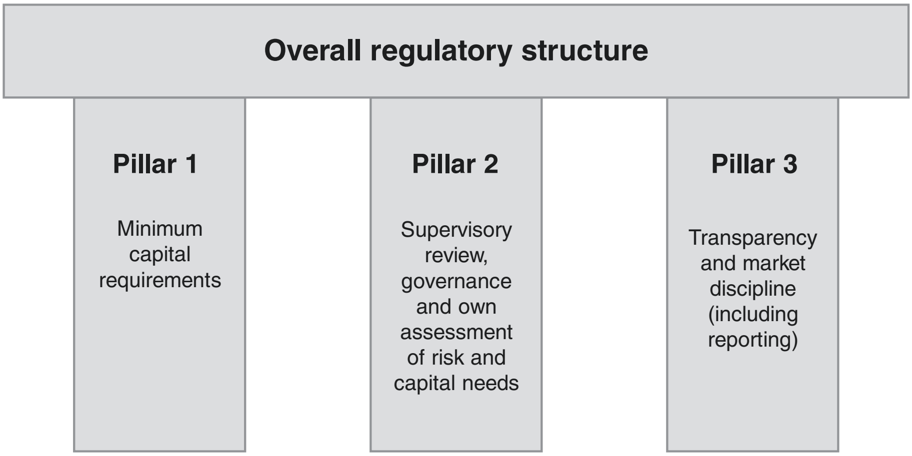


Fig. 3.2 Graphical description of 3 pillar regulatory framework
Source: Nematrian

(1) **Pillar 1** involves minimum (regulatory) capital requirements. These should make it more likely that regulated financial firms will be able to honour their promises.
(2) **Pillar 2** includes mechanisms allowing for supervisory review and for establishing appropriate systems of control including effective governance of an organisation. Pillar 2 is usually also deemed to include a (required) own assessment of the risks the organisation faces and hence a required own assessment of the intrinsic capital (and liquidity) needs of the organisation as well as supervisory review of these assessments and governance systems. This encourages organisations to develop robust risk management disciplines. Implicit is the assumption that firms are likely or ought to have a better understanding than outside parties (including supervisors) of the risks present within their own businesses.
(3) **Pillar 3** focuses on provision of information to others and market discipline. This should improve transparency and market-wide confidence which should help promote a soundly functioning financial system. As a by-product, it may also incentivise firms to develop suitable management information to help them manage their risks more effectively.

Perhaps the greatest area where there is scope for different high-level interpretations of the 3 Pillar structure is with Pillar 2. The underlying legislative documents establishing regulatory frameworks do not typically refer to 'Pillars' as such. Instead the Pillars are merely ways in which others interpret and make sense of the underlying regulatory requirements.

For example, the CRD and CRR refer to a supervisory review and evaluation process (SREP) which according to PRA (2015) is *'a process by which the PRA, taking into account the nature, scale and complexity of a firm's activities, reviews and evaluates the: arrangements, strategies, processes and mechanisms implemented by a firm to comply with its regulatory requirements laid down in PRA rules and the CRR; risks to which the firm is or might be exposed; risks that the firm poses to the financial system; and further risks revealed by stress testing'*. On the basis of the SREP, the PRA (for UK banks) determines whether the arrangements implemented by the firm and the capital held by it provide sound management and adequate coverage of its risks, if necessary imposing Individual Capital Guidance (ICG) on the firm and/or a 'PRA buffer'. The CRD and CRR also require firms to have an internal (i.e. individual) capital adequacy assessment process (ICAAP) and an internal (i.e. individual) liquidity adequacy assessment process (ILAAP).

The SREP (within the banking world) is typically viewed as a Pillar 2 mechanism. In the UK, an explicit part of the SREP, according to PRA (2015), involves the PRA reviewing the firm's ICAAP, as well as other factors. There are two main areas that the PRA considers when assessing a firm's capital adequacy under a SREP: (i) risks to the firm which are either not captured, or not fully captured, under the CRR; and (ii) risks to which the firm may become exposed over a forward-looking planning horizon. The PRA refers to the first area as Pillar 2A and the second as Pillar 2B.

Thus, the PRA in effect includes SREP, ICAAP (and ILAAP) and ICGs (and PRA buffers) within Pillar 2 (or Pillar 2A) for banks. However, the inclusion of ICAAP (and ILAAP) is somewhat roundabout in nature, resulting in some commentators in other jurisdictions interpreting Pillar 2 more narrowly, to include only SREP-type components. Similar conceptual elements also exist within insurance regulation as epitomised by Solvency II. However, essentially everyone categorises Solvency II's Own Risk and Solvency Assessment (ORSA), its analogue to the CRD's ICAAP /ILAAP, as falling within Pillar 2.

The perceived categorisation of ICGs within the 3 Pillar structure can also vary. They fall somewhere between formal additional capital requirements (which in Fig. 3.2 would imply categorising them as within Pillar 1) and private SREP-style interactions between the supervisor and the supervised firm (which in Fig. 3.2 would imply categorising them as within Pillar 2).

3.3.5 Regulatory Arbitrage

The complexities of categorising different elements of regulatory frameworks into simple overarching structures capable of catering for different industries may seem relatively arcane, but it does illustrate an important practical consideration for macroprudential policymakers. If it is hard to interpret regulatory structures in a common manner, it is even harder to ensure that detailed elements of different capital requirements are also consistent across regulatory regimes. This creates the potential for *regulatory arbitrage*.

Regulatory arbitrage is most usually associated with arbitraging capital requirements, although it can also apply to the types of activity that an entity can undertake or other conduct-related matters. Larger financial firms often have several (even many) subsidiaries, each one potentially subject to different regulatory regimes. For example, a holding company may own (or have a part interest in) a bank, an insurer, an asset manager (that may manage assets for the insurer and bank but also manage third party assets), intra-group service companies and potentially many other types of subsidiaries. Some activities may be capable of being carried out ('booked') within different subsidiaries, incentivising the firm to carry them out in the subsidiary where the capital requirements are lowest. Alternatively, it may be possible to substitute one activity for a similar one that faces lower capital requirements (e.g. making the activity more fee based in nature) even within the same sector. If capital is expensive to service then firms will be incentivised to minimise the amount of capital that they hold. We should therefore expect them to manage the services they sell in ways that seek to minimise the capital employed when doing so.

The microprudential and macroprudential consequences of intra-group regulatory 'arbitrage' as described above can be limited by imposing group--wide capital requirements on such groups and by requiring firms to carry out robust group-wide ICAAP /ILAAP /ORSA-style own assessments of their intrinsic capital needs. If the economic substance of an activity is the same then the intrinsic amount of capital the group as a whole should need to face the risks involved should be independent of where within the group the activity is being booked. Moreover, if the own assessment of risk is robust enough, then any changes in risk profiles caused by substituting one type of activity with another type should align with the capital needed to face such activities.

Of perhaps greater macroprudential concern is that activities may move over time between organisations in ways that on average result in a shift from more regulated to less regulated sectors and reduced capital supporting these activities. Left unchecked, this tendency could result in activities (or close substitutes) shifting towards sectors whose regulatory regimes are least effective at ensuring adequate capital is held to face the risks in question. Hidden vulnerabilities, akin to some that presaged the 2007–09 Credit Crisis, could then build up.

The apparent breakdown in trust between regulators and regulated hinted at in Section 3.3.3 appears here too. There seems to be a tendency amongst some regulators to deem nearly anything that has the slightest hint of being akin to banking but does not involve a bank to be an example of 'shadow banking'. However, usually regulators are not so worried if the entity involved in such activities is subject to other types of prudential oversight, e.g. if it is an insurance company.

Some entities deemed to be shadow banks may be funds that are themselves largely outside the regulatory net but are managed by others, e.g. asset managers, who fall within the regulatory net. They may not then in practice be subject to prudential regulation but at least they are caught by some conduct regulations.

The greatest worry is reserved for entities that are largely or wholly outside the regulatory net and largely or completely stand-alone. Regulators may then be worried that intrinsic incentives favouring adequate capitalisation are too weak to result in an appropriate outcome.

Ultimately, the only way of eliminating regulatory arbitrage within and across sectors (whilst still retaining *some* capital requirements to limit the information asymmetry referred to in Section 3.3.1) is to frame regulatory capital requirements in a sector and activity agnostic fashion. This would likely be extremely challenging to achieve, both from a political and a technical perspective. The material in Box 3.6 provides a way of framing the technical challenges involved. More probable is that the already discernible trend towards convergence across the financial services industry will continue over an extended period, perhaps catalysed by broader adoption of SREP /ICAAP /ORSA-style mechanisms across less regulated parts of the financial services industry.

3.4 Accounting

3.4.1 Introduction

Measuring the capital base of an organisation requires the creation of a balance sheet in which assets and liabilities are totted up. The most obvious profession that has a role in this process is the accounting profession. The actuarial profession plays an important role for insurance and some other (longer-term) liabilities, such as pension liabilities. Other specialist professions also get involved, particularly for less liquid assets, e.g. surveyors and other property valuers when valuing real estate. Members of these professions are typically subject to professional qualification requirements, standards and codes of practice designed to ensure that their work is of a suitably robust standard.

Normally these professionals are not themselves explicitly responsible for financial statements and other material demonstrating adherence to minimum capital requirements of the firms they work for or advise. Instead it is the firms themselves (and particularly their boards of directors) who take overall responsibility for the correctness of such statements. In addition, firms are often required to have their financial statements validated or audited. A summary of the main elements of financial statements and associated material published alongside them is set out in Box 3.7.

Accounting and other standards relevant to the measurement of the capital base of an organisation therefore generally apply to the organisation rather than to the professionals advising the organisation (although of course accounting standards may need some interpreting, perhaps with the help of lawyers and the sorts of professionals mentioned above). Collectively, in the accounting sphere, the standards that organisations must typically adhere to are called *generally accepted accounting principles* (GAAP). These can vary somewhat by jurisdiction. Sometimes (local) GAAP requirements are contrasted with (international) accounting standards such as *International Financial Reporting Standards* (IFRS), if firms can choose between the two.

Box 3.7: Main elements of financial statements

In most jurisdictions, the main elements of a firm's financial statements are:

(1) The *revenue account*

The revenue account measures the flow of funds into or out of the organisation over some specified period, e.g. a year. Some simple entities account on a receipts and payments basis, looking just at the cash paid or received over a given period. For larger organisations, accounts are generally prepared on an accruals basis. This takes account of how much of a given income or expense item should be allocated to a specific period of account.

A challenge is how to apportion income and expenses and hence revenue between time periods for multi-period contracts. If we apportion a lot of income but very little expense to early in the life of a contract then it will boost the apparent profitability of the business early on (at the expense of depleting apparent profitability later in the life of the contract). This can give an overly rosy picture of the robustness of the firm's business part way through the contract (particularly if a high volume of such contracts has recently been written).

Not all net income has the same characteristics. Quite apart from any segmentation by business unit to help understand which are contributing most to the business, usually a firm's accounts will also seek to differentiate between:

(a) continuing net revenue that has a reasonable likelihood of recurring if the business continues much as at present, and

(b) one-off net revenue that needs to be recorded to reconcile the revenue account with the balance sheet but is not otherwise as sustainable as the revenue included in (a).

For example, parts of IFRS split the revenue account into two elements, a *profit and loss* (P&L) account and an *other comprehensive income* (OCI) account. The latter includes for some assets changes in how their fair value compares with the value at which they are recognised in the profit and loss account.

(2) The *balance sheet*

Firms can exist essentially without any assets or liabilities and still make a profit or loss (think of an individual trader who walks into a car boot sale or trades on eBay for a day, buying some goods and then almost immediately selling them on to others). However, usually capital must be employed by a firm to make profits.

The balance sheet aims to identify this capital base and how it is formed. It is struck at (i.e. is a valuation as at) a specific point in time (instead of being a flow of value through time, like the revenue account).

Ultimately, when the firm winds up it will liquidate all its remaining assets (and liabilities) and the net balance will show up in the (closing) revenue account. More generally, the revenue account and movements in the balance sheet should exactly tally, although some of the elements of the revenue account may e.g. correspond to unrealised gains and losses on assets and liabilities present in the balance sheet, i.e. may in some sense be present largely or wholly to ensure that the tallying is exact.

(3) Other *reports* and *notes* (and *auditor and director opinions*) included in the financial statements to amplify or clarify elements in the revenue accounts and balance sheets or to provide more general business commentary

These elements of the financial statements don't alter the stated profit or stated net assets as such, but can substantially influence the interpretation placed on these figures by third parties (e.g. investment analysts forming a view on whether the current price of the firm's shares represent a good or poor investment).

Nowadays there are lots of such reports and notes, ranging from an overview of business development and analyses of revenue by geography or business line to statements on remuneration paid to senior employees or directors. Financial statements provide an opportunity for a firm to present itself to third parties and to explain what it is doing, why it is doing it and how successful it is at doing these things.

We can view regulatory capital assessments for financial organisations as a (special) type of additional note (sometimes only made available to the supervisor) that re-expresses the values placed on the organisation's assets and liabilities onto whatever basis the supervisor mandates for this purpose.

3.4.2 Fair valuation and other valuation approaches

An important accounting issue is how to value an asset or a liability. There are several ways in which assets and liabilities can be valued in financial statements, including at:

(a) *fair value*

This loosely speaking is the price at which an asset (or liability) would be bought or sold in the market if there were simultaneously willing buyers and willing sellers of it. If an asset or liability is actively traded on a (not rigged) market then its fair value is typically the same as its market value. Identifying a value in such circumstances is called *marking to market*. Life becomes more complicated if the asset or liability is less liquid and/or it is necessary to infer its price from those of more actively traded instruments with which it shares characteristics. It then involves *marking to model* (as usually some sort of 'model' is needed to identify the price to use).

In some regulatory frameworks (such as the EU's insurance regime, Solvency II) fair valuation is referred to by the term *market consistency*, since the aim is to use valuations that are consistent with market observables. Kemp (2009) explains the interaction further and explores how market consistent valuations can be identified in practice. He defines a market consistent value of an instrument as its '*market value, if it is readily traded on a market at the point in time that the valuation is struck, or a reasoned best estimate of what its market value would have been had such a market then existed, in all other situations*'.

International Financial Reporting Standard (IFRS) 13 *Fair Value Measurement* mandates a fair value hierarchy involving 3 different levels of fair valuation (depending on the source of the inputs used in the computation of the fair value). Accounts drawn up under IFRS 13 need to indicate what proportion of fair values fall into each level. Level 1 inputs involve quoted prices in active markets for identical assets or liabilities. Level 2 inputs involve inputs other than those included in Level 1 that still involve market observables (e.g. quoted prices for similar assets or interest rates, yield curves, credit spreads or implied volatilities observable at commonly quoted intervals). Level 3 inputs involve unobservable inputs.

(b) *historic cost* or *amortised cost*

The historic cost of an asset corresponds to the price at which it was purchased. If the asset is a wasting asset (e.g. computer equipment) then historic cost values might be further reduced to reflect depreciation. Small organisations might value assets such as property (i.e. real estate) at historic cost, perhaps revaluing the property from time to time if the cost is believed to underrepresent its 'true' value. Where assets are likely to be held to maturity and on maturity can be expected to have a specified price (e.g. the par value for a fixed income bond or loan) then a variant called amortised cost is more common. This involves valuing the asset in some manner that progresses smoothly from its historic cost to its end par value through time, i.e. amortising the difference, e.g. linearly through time (and applying certain other adjustments, see Section 3.4.4 and Box 3.9).

(c) *value in use*

This represents the value that the firm can gain from the asset because of its usefulness to the firm, e.g. a steel manufacturer might value its steel mills on a value in use basis.

In recent times, there has been increased focus on use of fair values for reasons that are explained further in Kemp (2009). It provides greater transparency and is typically a 'fairer' way of apportioning value between different parties with different interests, particularly if the asset or liability is likely to be actively traded. In the banking world, however, some types of asset (e.g. loans expected to be held to maturity within a bank's banking book) are usually valued on an amortised cost basis in financial statements. The values in the bank's financial statements also usually form the basis for capital adequacy computations (possibly with some modifications, see below). Amortised cost is also used for some types of assets held by some types of insurers in some jurisdictions, although under Solvency II the focus is on market consistent, i.e. fair, values.

Even if the financial statements themselves are not drawn up using fair values there may be a requirement to publish in the notes to the financial statements details of how the asset values would change if they had been fair valued.

The use of fair values creates a lot of debate in some circles. Some view increased use of fair values as contributing to procyclicality and hence imperilling financial stability by increasing the tendency of firms in trouble to need to engage in fire sales. The thesis is that if profits and losses are drawn up using fair values rather than (smoother) valuation measures such as amortised cost then this will add to the volatility of results, adding pressure on firms to liquidate positions if the firm is short of capital resources, creating further downward pressure on the prices of the affected assets, exacerbating adverse price movements.

Others view the fair valuation approach as unfairly tarnished by such arguments. For example, firms often actually sell more liquid assets first, if they need to raise funds. Declines in the prices of less liquid assets may therefore create less serious feedback loops than might initially be expected. Moreover, not using fair values runs the risk of under-reacting to problems. Also, in stressed times, market participants seem to spend a considerable amount of effort trying to work out the 'true' (fair valued) position of their counterparties, to try to understand whether the counterparty is at risk of experiencing a run. On this basis, it may be simpler and more reasonable to mandate provision of fair value information, to enhance transparency and the broader financial stability it is usually presumed to create. Commentators such as Laux and Leuz (2010) have analysed experience during the 2007–09 Credit Crisis and find little support for claims that fair value accounting exacerbated the Crisis. Alternatives such as the amortised cost approach do not necessarily provide greater financial stability. They may just end up making the problem worse further down the line (with less transparency in the meantime), if the underlying fundamentals are deteriorating.

3.4.3 Impact of accounting standards on capital requirements

It is natural to ask how much influence accounting standards might have on regulatory capital requirements and on systemic risk exposures. This is a surprisingly difficult question to answer since:

(a) Organisations issue different sorts of financial statements each of which may have a different role and may therefore be drawn up in accordance with different rules.

General purpose financial statements tend to be drawn up on a 'going concern' basis, because this is believed to provide the maximum insight to shareholders and users of the accounts with similar interests. However, other stakeholders interested in the financial progression of the organisation may have different requirements and may therefore want information supplied to them to be prepared on a different basis. For example, valuations and revenue assessments for tax purposes are mandated by tax authorities and therefore do not necessarily accord with valuations and revenue assessments applicable to general purpose financial statements.

(b) The usual purpose of regulatory capital computations is to help supervisors understand the organisation's robustness to adverse scenarios. This purpose is not necessarily well served by a 'going concern' accounting assessment of the organisation.

Regulators may therefore require other approaches to be adopted, overriding the figures that otherwise appear in the organisation's main financial statements. Sometimes these figures are only made available to supervisors but more commonly summarised versions of these figures also need to be made available at regular intervals to the wider public

(c) Some important users of financial statements such as financial analysts will often go to considerable effort to re-express information that is made public via financial statements (and via other disclosures, if available to them) onto bases that they consider to be more meaningful for their own purposes.

In stressed times, analysts appear to place strong emphasis on identifying how the firm would look if its assets and liabilities were carried at fair value, irrespective of the actual carrying value used by the firm. This is partly because fair values are more naturally comparable between firms. The likelihood of firm failure and the consequences for different parties if this happens also comes to the fore. Implicitly, analysts may be carrying out analyses akin to those described in Box 3.6. The results of sector-wide stress tests as described in e.g. Box 5.3 can provide reference points that can help them do this.

An example of (a) might be the treatment of depreciation, i.e. the expected decline in value of an asset over its working life. Sometimes tax authorities allow firms to use initially more favourable depreciation allowances in their tax computations (for some assets) than would be considered suitable in their general purpose financial statements. This lowers the firm's taxable profits relative to its 'going concern' profits at outset, and increases them later (as the value of an asset with a finite lifetime will eventually depreciate to zero on either accounting basis). Governments may want to encourage specific sorts of capital investments for wider economic purposes. Reducing the initial tax cost the firm incurs may provide this encouragement.

An example of (b) is the contentious issue of own credit risk, see Box 3.8. General purpose accounting figures can take some account of the fall in value of debt issued by a firm if its creditworthiness declines. The market value of such debt declines in such circumstances, so it is difficult to fault the accounting logic of such an approach. But it seems intuitively wrong (and arguably *is* wrong, if we adopt the conceptual model set out in Box 3.6) to take credit for this effect when deriving the firm's available regulatory capital base. The apparent addition to the capital base it creates only arises because the likelihood of the firm honouring its commitments is declining.

Box 3.8: Own credit risk

Own credit risk is a complexity of modern accounting that arises because the (fair) value of a firm's liabilities depends on the creditworthiness of the firm, i.e. its likelihood of honouring these liabilities.

Within modern finance there are several ways in which firms can in effect have an economic exposure to their own credit risk or broader financial health.

Take, for example, exposure to a firm's own equity. Treasury departments of firms in some jurisdictions are permitted to own some limited fraction of their own equity, perhaps for just limited periods of time, or at least to have indirect economic exposures to themselves. This might be because they are buying up their own stock in lieu of issuing a dividend, and have yet to cancel the shares they have just purchased. Investment banks might also hold their own shares (or those of a quoted parent) as inventory within their market-making businesses, particularly if their parent is sufficiently large to be included in major market indices that form the basis of major equity index futures contracts. They are likely to use such futures from time to time when hedging overall market exposures, and will be adding or subtracting indirect exposure to themselves whenever they do so. A conceptually straightforward way of addressing such exposures is simply to reduce the amount of equity deemed to be outstanding when drawing up a firm's balance sheet.

Own credit risk can be thought of as relating to a similar issue, but referring to the firm's outstanding liabilities to its debt (or loan) holders rather than to its equity investors. The complication is that as the firm's creditworthiness deteriorates, so too do the values of these debt or loan instruments, partially compensating equity holders for the declines in value involved. Put another way, equity investors benefit from what is known as the 'shareholder put'. Ultimately, however far the value of the firm's debts or loans decline, its equity holders only stand to lose at most the value of their equity, if the firm is a limited liability company. Particularly complex are some derivatives, since the value of a derivative instrument in principle depends on the creditworthiness of both parties to the derivative (and on the types of collateral and frequency of posting of the collateral applicable to the derivative contract). An investment bank running a complex trading book will typically be adding or subtracting indirect exposure to its own creditworthiness almost all the time.

From an accounting perspective (and if instruments are being valued at fair value, which it is difficult to fault for at least those exposures that are being actively traded), it is hard not to conclude that instruments expressing exposure to a firm's own credit risk should be valued in a way that takes account of this risk. If the instruments are bought and sold in the market then the price paid or received will depend on the creditworthiness others ascribe to the firm, so why shouldn't the firm's own traders do likewise.

But from a regulatory capital perspective, the results seem perverse, since they involve the firm's position being flattered due to it becoming a poorer credit risk, precisely not the sort of result we might want to protect customers against potential default of the firm.

At one level the way in which to reconcile these two conflicting stances is to accept that general purpose financial statements (often framed particularly with equity holders in mind) and regulatory capital computations have different overall objectives, so do not need to adopt the same valuation approach. At a deeper level, as we note in Box 3.6, the problem goes away if we express the task of regulatory capital assessment to be the identification of the amount of capital needed to achieve a specific level of creditworthiness.

As noted above, derivative exposures present added challenges. The amounts of collateral that a firm may need to post in relation to open derivative positions may depend on its creditworthiness. If the firm has sufficiently large amounts of derivatives outstanding then the task of finding the extra cash to post as margin can create further downward pressure on the firm's creditworthiness, in effect creating a possible collateral 'run' on the firm (which can have the same adverse effect as a bank 'run' has on a bank). Potential liabilities under derivative contracts may at first sight appear to be subordinated to customer liabilities, but the process of collateralisation means that in some circumstances derivative counterparties can end up with a higher priority on the firm's assets than customers. Some of AIG's problems (see Box 3.1) appear to have been the result of these effects.

Own credit risk came to the fore in the 2007–09 Credit Crisis as it became clear that some firms were using the write ups available by recognising own credit uplifts to improve their apparent capital position (versus what would otherwise have been the case). Regulatory capital frameworks in the financial sector have since been adjusted to restrict the ability of firms to do this, even though this results in valuations for regulatory capital computations diverging (further) from those applicable under GAAP or IFRS accounting.

3.4.4 IFRS 9

Several of the above issues can be illustrated by considering features of a new accounting standard that is currently being introduced by the International Accounting Standards Board (IASB), namely International Financial Reporting Standard (IFRS) 9 *Financial Instruments*. This standard is replacing the older International Accounting Standard (IAS) 39 *Financial*

Instruments: Recognition and Measurement. Firms subject to IFRS will need to apply IFRS 9 for annual periods beginning on or after 1 January 2018.

Under IFRS 9 (as in IAS 39), some loans and other financial instruments are carried at amortised cost (HCA) whilst others are carried at their fair value (FVA). Under IFRS 9 the approach used depends on the business model applicable to the portfolio in which the instrument resides, see IASB (2014):

(1) HCA applies to most loans in a business model 'whose objective is to hold assets in order to collect contractual cash flows'.
(2) FVA applies to most loans in a business model 'whose objective is achieved by both collecting contractual cash flows and selling financial assets', with changes to the fair value carried through to the OCI account.
(3) FVA also applies to most loans in business models other than in (1) or (2), but for these loans changes to the fair value are carried through to the Profit and Loss account.

As a very rough rule of thumb (because the definitions involved are different), (1) is akin to a bank's 'banking book' under IAS 39, (2) is akin to a bank's 'available for sale book' under IAS 39 and (3) is akin to a bank's 'trading book' under IAS 39.

The fair valuation approach (FVA) involves use of market values (or some suitable equivalent), see Section 3.4.2. Loosely speaking, amortised cost approaches involve a carrying value that starts at the value at which an instrument is bought and then progresses (usually linearly) to the par value of the instrument at maturity (the par value is the amount that will be returned to the investor at maturity if the instrument hasn't defaulted in the meantime). The amortised cost approach is also often known as the historic cost approach (HCA), since it is based on the original historic cost of the instrument. Sometimes, however, HCA is differentiated from other amortised cost approaches. For equities and other instruments that have no defined maturity or par value then it may not be practical to identify any meaningful amortisation schedule so HCA may then default to the actual unadjusted historic cost.

Superimposed on the basic amortisation schedule in HCA is some allowance for credit losses. Under the older IAS 39, credit losses for loans and other financial instruments that were being carried at amortised cost were only typically recognised when a loss arose (or some other trigger event occurred). This is called an 'incurred loss' model of credit loss recognition. The original rationale for this approach was to limit an entity's ability to create hidden reserves that could be used to flatter earnings during bad times.

However, as the 2007–09 credit crisis unfolded, it became clear that some firms were adopting an opposite type of earnings management, postponing recognition of losses and thereby potentially overstating the firm's available capital base. IAS 39 did not require firms to wait for actual default before recognising some loss, but in practice firms often did so.

Perhaps the most important change being introduced by IFRS 9 is to credit loss provisioning for such instruments. Under IFRS 9, losses will still, of course, be recognised as they occur, to the extent that they have not been recognised previously. But in addition, as soon as a financial instrument is originated or purchased, 12-month expected credit losses need to recognised in the P&L account. A corresponding loss allowance therefore needs to be established, reducing the net balance sheet by a same amount. By 'net balance sheet' we mean the value of the assets the firm holds less the value of its liabilities (ignoring for this purpose equity liabilities). A loan in this position is said to be in Stage 1 of the IFRS 9 impairment process.

Subsequently, if credit risk increases 'significantly' (and the resulting credit quality is not considered to be low credit risk), the loan is said to move to Stage 2 of the IFRS 9 impairment process and lifetime expected losses need to be recognised. For a multi-year instrument this involves projecting expected losses over the remaining lifetime of the instrument, not just focusing on expected losses over the first 12 months. The term 'significantly' is not explicitly defined, so is likely to be influenced by interpretations placed on it by firms and their auditors.

The computation of lifetime expected losses can be carried out on a portfolio basis if the loans are not considered credit-impaired. By 'portfolio basis' we mean not focusing on the characteristics of individual loans but on representative instrument types. Loans and other financial instruments that have deteriorated in credit quality by enough to be deemed credit-impaired are said to be in Stage 3 of the IFRS impairment process. IFRS 9 generally expects such assets to be individually assessed.

IFRS 9 is therefore introducing an 'expected loss' provisioning model, with only the first 12 months' expected loss allowance needing to be recognised (in addition to incurred losses) if the instrument's credit quality hasn't fallen too far since outset but a lifetime expected loss allowance needed if the credit quality has fallen far enough.

In broad terms, this change increases loan loss provisioning throughout the life of the instrument (the increased loan loss provision becoming larger for multi-year instruments if its credit quality falls by enough to be deemed to have deteriorated 'significantly'), thus depleting a firm's net balance sheet.

The broad magnitude of this effect can be estimated by, for example, creating a model that characterises the likelihood of the credit quality of an instrument changing from one quality to another through time, and then projecting through all possible paths the credit quality might take between now and the maturity of the instrument. In some of these paths the instrument will default and hence a loss will become fully ‘incurred’.

The mathematics involved in this model is explained further in Box 3.9. Using plausible assumptions, we find that the average depletion of a firm’s capital basis depends on:

(a) The rating profile of affected loans and other financial instruments for the firm in question
(b) The extent to which any given credit rating needs to have deteriorated for the firm to deem it to have deteriorated ‘significantly’
(c) Some other factors, such as the typical recovery rate on affected loans.

In many cases the depletion will be material relative the original expected loss that might have been incurred on such an instrument, particularly for ones which have deteriorated in credit quality ‘significantly’.

What is less obvious just from a model such as the one set out in Box 3.9 is what impact a change like IFRS 9 might have on regulatory capital computations. The key here is to realise that the needs of regulators may differ from the needs of users of accounts prepared under IFRS 9, and regulators can mandate different accounting treatments if they so wish.

For insurers subject to Solvency II, assets are typically valued at market or market consistent value (typically akin to fair value), so the refinements being introduced by IFRS 9 described above will probably have little if any influence on such firms’ regulatory balance sheets, i.e. on how well or badly capitalised it is relative to its regulatory capital requirements.

For banks that use the Internal Ratings Based (IRB) approach to setting capital requirements (see Box 3.10), implementations of Basel III such as in the EU’s CRR require them to use an expected rather than an incurred loss approach. So, again, we might expect that these aspects of IFRS 9 will have relatively little influence on such banks’ regulatory balance sheets. However, the position is not quite so simple, because although the total balance sheet may not be much affected, the decomposition of the firm’s available regulatory capital into different tiers (some of which are capped in size) can be affected by whether the IFRS 9 provision is higher or lower than the corresponding CRR provision. Banks not using the IRB approach will also be more affected. At the time of writing, the Basel Committee is reviewing its approach to the regulatory treatment of accounting provisions including those applicable to banks not using the IRB approach, see BCBS (2016a).

Even if the change has little impact on the firm's regulatory capital position, it still has an impact on the firm's stated profits (as recorded in its financial statements). This could have systemic risk implications, since firm behaviours might be partly driven by how quickly profits are recognised in their financial statements, especially if the firm has an apparently reasonable level of surplus versus its regulatory capital requirements. Conversely, we have previously noted that investment analysts and the like put a lot of effort into restating a firm's financial position onto bases that they find more meaningful for their purposes. This can diminish the practical impact of figures contained in financial statements, particularly if there is other publicly available information (such as the results of regulator mandated stress tests) that analysts can also work on when analysing the firm.

Another change being introduced by IFRS 9 relates to the 'own credit' issue. As explained in Box 3.8, this involves banks and others (when they have elected to measure their debt at fair value) booking gains when the value of their own debt falls due to a decrease in their own credit worthiness. IFRS 9 stops firms from including such gains in their P&L account. Instead it requires such gains to be booked to their OCI account. However, in the balance sheet such debt presumably still is carried at its fair value, consistent with its economic substance but not conducive to the effective measurement of the (regulatory) capital base the firm might need to face the risks to which it is exposed. As noted previously, recent regulatory capital frameworks such as Basel III and insurers Solvency II usually now disallow such gains or losses in computation of a firm's capital base for regulatory capital purposes, overriding the treatment otherwise applicable under IFRS 9. For firms outside the scope of Basel III or Solvency II but subject to IFRS 9, the issue would still apply, since the balance sheet still includes such gains and losses, even if they flow through the OCI and not the P&L account.

Perhaps the main lesson that we can draw from this discussion is that accounting principles are not unimportant in terms of how healthy a firm might appear to be. However, if regulators and politicians feel strongly enough they can always override general purpose accounting rules with bespoke ones that define how firms should value assets and liabilities for regulatory capital purposes. There are costs incurred when doing so, particularly for the companies concerned, which mean that there is usually a desire when practical to reuse valuations for multiple purposes.

Box 3.9: IFRS 9 loan loss provisioning

One way of estimating the magnitude of the additional provisions IFRS 9 requires versus IAS 39 is to build up a model of how the credit rating of an instrument might change through time. A common method used in portfolio credit risk modelling is to assume that the creditworthiness of an instrument migrates through time according to a *transition matrix* describing the probability of transitioning from one credit rating to another over a specified time period, often one year. For example, instruments might be assigned one of the following credit ratings: AAA, AA, A, BBB, BB, B, CCC/C and D and might be assumed to migrate each year with probabilities in line with the transition matrix shown in Table 3.1. A transition matrix defines the likelihood of an instrument ending the year with a specific rating conditional on it having started with a specific rating.

Often, transition likelihoods in different years are assumed to be independent of each other. Such a model is then known mathematically as a *Markov chain*. Using the transition matrix, and knowing the rating at outset we can work out the likelihood of any potential way in which the credit rating can evolve through time. We might assume that a loss is incurred whenever the credit reaches D (i.e. the 'default' state) and in such circumstances the loss is typically not 100% but 100% x (1 – *Recovery Rate*), where *Recovery Rate* is some specified parameter, e.g. 0.40. For any given initial credit rating and time to maturity we can then work out the average loss a cohort of such instruments would be expected to have incurred up to any given point in time after inception and the average additional loan loss provision set up in relation to the instruments.

In this exercise, we also need to keep track of whether the path the credit rating has taken has reached a level that is deemed to have deteriorated 'significantly' since outset. This is akin to shifting from Stage 1 to Stage 2 of the IFRS 9 impairment process.

Given such a model, and a transition matrix as above (and making some assumptions about how to handle ratings that have become 'N/R', i.e. not rated, e.g. that we scale the remaining likelihoods by a single ratio to get them to add to 100%) we can estimate the difference in the loan loss provisions between IAS 39 and IFRS 9 as in Table 3.2 at different times for different starting credit ratings and maturities etc.

The IFRS 9 loan (credit) loss provision starts higher than the IAS 39 provision because it includes an allowance for expected losses over the first 12 months of the instrument life. Thereafter, the difference typically rises further (for loans that are longer than one year), sometimes quite materially, as in some circumstances lifetime expected credit loss provisions need to be established under IFRS 9. Eventually the difference reduces, falling to zero at maturity, since any realised losses will then have been recognised on either basis.

The maximum difference depends quite significantly on the meaning we ascribe to a credit rating having deteriorated 'significantly', highlighting a judgemental aspect of this accounting standard. Another subtlety introduced by this feature is that if, say, deteriorating from AA to BBB is deemed a significant deterioration but staying at BB isn't then (depending on assumptions adopted) the loss provision ascribed to the loan that has deteriorated to BBB may be higher than the loss provision ascribed to the BB loan even though it is of higher quality.

Table 3.1 Illustrative yearly credit rating transition matrix

%age To From	AAA	AA	A	BBB	BB	B	CCC/C	D	NR
AAA	87.17	8.69	0.54	0.05	0.08	0.03	0.05	–	3.38
AA	0.54	86.29	8.36	0.57	0.06	0.08	0.02	0.02	4.05
A	0.03	1.86	87.26	5.53	0.36	0.15	0.02	0.07	4.71
BBB	0.01	0.12	3.54	85.09	3.88	0.61	0.14	0.22	6.39
BB	0.02	0.04	0.15	5.18	76.12	7.20	0.72	0.86	9.71
B	–	0.03	0.11	0.23	5.42	73.84	4.40	4.28	11.68
CCC/C	–	–	0.16	0.24	0.73	13.69	43.89	26.85	14.43

Source: S&P (2013).

Table 3.2 Cumulative IFRS 9 recognised losses less IAS 39 recognised losses

T (maturity)	10	10	10	10	10	10	4
Initial credit rating	AAA	A	BBB	A	A	A	A
Rating at or beyond which deemed 'deteriorated'	A	BB	B	BBB	B	CCC/C	BBB
Recovery rate	0.4	0.4	0.4	0.4	0.4	0.4	0.4
t (time from inception)							
0	0.00%	0.04%	0.14%	0.04%	0.04%	0.04%	0.04%
2	0.09%	0.23%	0.61%	0.51%	0.15%	0.07%	0.12%
4	0.14%	0.40%	1.00%	0.70%	0.26%	0.10%	–
6	0.15%	0.46%	1.11%	0.66%	0.32%	0.13%	–
8	0.10%	0.34%	0.83%	0.42%	0.27%	0.14%	–
10	–	–	–	–	–	–	–

Source: Nematrian. See Box 3.9 for details of methodology used.

Refinements that can be made to the sort of model described above include making the recovery rates 'stochastic' (i.e. random in some predefined manner), allowing for more sophisticated transition rules, modelling uncertainty in transition matrix values, capturing the impact if any of changes in interest rates applicable on fixed interest instruments etc. Several of these refinements are included in the model described in Abad and Suarez (2016), which is otherwise conceptually similar to the above model but uses fewer rating categories and is therefore more analytically tractable. The above model also assumes that once a loan has suffered a 'significant' deterioration it can't 'un-deteriorate', but in practice this may be possible.

Technically the approach described above only models how on average a well-diversified portfolio might evolve. Less well diversified portfolios exhibit additional variability around expected outcomes, which can be modelled by simulating the behaviour of each individual exposure.

IFRS standards are commonly adopted by EU firms. In the USA, it is more common for firms to prepare their financial statements using Financial Accounting Standards Board (FASB) standards. At the time of writing FASB had also recently set out new rules on loan loss provisioning in situations where firms are accounting for such instruments on an amortised cost basis. These standards require firms to include lifetime expected loan loss provisions in more cases. The provisions banks need to establish for regulatory capital computations depend on whether they use Basel III Internal Ratings Based (IRB) approaches or Standard Assessment (SA) approaches, see Box 3.10.

The main differences in expected credit loss (ECL) models under IASB, FASB and Basel III IRB computations are summarised in Table 3.3, according to BCBS (2016a). The 'model' here refers to how the relevant probability of default (PD), loss given default (LGD) and exposure at default (EAD) are estimated. Estimated PDs can be point-in-time (PIT) or through-the-cycle (TTC) depending on whether

Table 3.3 Differences in expected credit loss (ECL) models

		IASB	FASB	Basel III IRB
PD	Measurement period	12 months (Stage 1) Lifetime (Stage 2)	Lifetime	12 months
	Cycle sensitiveness	Point-in-time, considering forward-looking information, including macroeconomic factors		Economic cycle[1]
LGD /EAD	Measurement	Neutral estimate, considering forward-looking information, including macroeconomic factors		Downturn estimate

Source: BCBS (2016a).

[1] i.e. a 'through the cycle' (TTC) measure, in contrast to the PIT measure mandated by IASB and FASB.

they seek to capture current likelihoods or ones averaged over a financial cycle. PIT estimates are more market consistent but also potentially more procyclical than TTC estimates, see e.g. Benford and Nier (2007) or Kemp (2009). Financial cycles, to the extent that they exist, do not appear to have well-defined timescales, see Box 2.5, making it hard to reach agreement on what exactly is meant by a TTC PD. BCBS (2016a) also describes possible ways in which the ECL model applicable under the Basel SA approach might be changed to be more akin to the model applicable under the Basel IRB approach.

Box 3.10: Internal models, IRB approaches and advanced measurement approaches

Modern regulatory capital frameworks for financial organisations such as Basel III and Solvency II generally include the concept of internal models. These are models that the firm itself develops, capturing its own idiosyncrasies, to determine its (Pillar 1) capital requirements. They contrast with standardised approaches. The exact terms used differ by regulatory framework. The ones used under Basel III and Solvency II are shown in Table 3.4. Generally, internal models require prior approval from the firm's supervisor. The larger the firm, the greater is the supervisory expectation that it will develop an internal model. Larger firms tend to be more complex and to have more resources available to develop such models. Firms using standard assessment methodologies may be required to justify why it is not appropriate for them to develop internal models.

The typical form of such a model (e.g. what it models) also varies by regulatory regime, e.g. an IRB model under Basel III will naturally focus on deriving PDs, LGDs and EADs for different credit-sensitive instruments, see Box 3.9. These appear in the SF methodology under Solvency II but may be less important for many insurers.

Internal models can be very complicated, making it difficult for supervisors to assess the robustness of the model. Models in this field are subject to model risk!

Table 3.4 Internal model approaches versus standard assessment approaches

Regulatory framework	Name given to internal model approach	Name given to standard assessment approach
Basel III (banks)	*Credit risk*: Internal Ratings Based (IRB) *Operational risk*: Advanced Measurement Approach (AMA)	*Credit risk*: Standardised approach (SA) *Operational risk*: Standardised Measurement Approach (SMA)
Solvency II (EU insurers)	Internal model (IM)	Standard formula (SF)

Source: Nematrian.

Not surprisingly, regulatory frameworks expect models to adhere to specified (some in the industry would say extremely onerous) criteria before they will be acceptable to supervisors. These include governance criteria, e.g. clarity over who is responsible for the model and any updating of it, evidence of sign-off and challenge by the firm's governing body and evidence that the model is being used more broadly within the business (the 'use' test). They also include more technical modelling criteria, e.g. robust model design selection and implementation processes, backtesting protocols, etc.

Internal models for Pillar 1 capital requirements are conceptually different to any models that firms may be required to develop for their own intrinsic assessment of their capital needs, e.g. under ICAAP (Basel III) or ORSA (Solvency II) protocols. However, if a firm has a Pillar 1 internal model then typically there will be some link between it and models the firm develops for equivalent risks for ICAAP /ORSA purposes. Both Basel III and Solvency II effectively include the concept of 'partial' internal models that involve models that are more sophisticated than standard assessment approaches for individual risk types, although it may then be necessary for the firm to justify why it is using an internal model for some risk types instead of others.

Regulatory enthusiasm for internal models seems to wax and wane from time to time. Internal models ought to be more accurate and therefore more risk sensitive than standard assessment approaches, since they can better capture firms' idiosyncrasies. However, they may be less comparable across firms (although this can in part be addressed by requiring firms to supply numbers on both bases). Regulators worry that firms may (consciously or unconsciously) select models that flatter the firm's (Pillar 1) capital position. Many of the criteria imposed on internal models before they can be approved by supervisors stem from trying to address this scope to 'game' the regulations.

Regulators also worry about how intrinsically amenable to quantification are the risks and uncertainties that might be covered by internal models, i.e. how 'Knightian' they are (see Section 7.4.2 for an explanation of Knightian uncertainty). The more sceptical they are about the potential for the robust quantification of risks, the more likely they are to favour use of standard assessment approaches.

Examples at the time of writing include BCBS (2016b) (which is imposing minima on market risk capital requirements irrespective of firms' own internal models for such risks), BCBS (2016c) (which is exploring limiting some of the flexibilities firms have in IRBs for credit risk) and BCBS (2016d) (which is exploring eliminating the ability of banks to use internal models for operational risks).

3.4.5 Going Concern Versus Gone Concern

We have seen above that regulatory capital computations may mandate ways of valuing some types of assets and liabilities that differ from the values that might be used in general purpose financial statements. The surplus capital (i.e. assets less liabilities) may also be *tiered*, depending on its likely effectiveness at supporting a 'going concern' versus a 'gone concern' business model. Limits may then be placed on how much of any specific type of capital can be included in such a computation.

In this context, 'going concern' means that the business is still functioning and therefore, typically, can be assumed to have some franchise value. For example, a manufacturer might have machine tools that it can continue to use whilst it is making goods it can sell. These tools then have some value in use for the manufacturer which we might include in a going concern valuation. In contrast, 'gone concern' refers to the situation where the business has failed. The machine tools that had some value in use under a going concern basis may have little or no value if the manufacturer using them has gone bust and everyone else in the industry already has more modern tools capable of doing the same manufacturing but at lower cost.

The 'gone concern' focus of tiering in most Pillar 1 capital computations typically focuses on circumstances akin to the firm being put into liquidation. The capital type that is most appealing to supervisors is common equity, since it has the most ability to absorb losses in nearly all situations. However, it may be more expensive for firms to service, so firms may prefer to issue debt instruments. The less subordinated a debt instrument is, the less effective it is at protecting customers if the firm runs into difficulty. In the extreme case where the debt instrument is 'secured' and has higher priority than customer liabilities (see Fig. 3.1) then it will be of little or no use for such purposes. Issuing debt ranked this highly may increase rather than reduce the effective spread applicable to customer liabilities in the conceptual framework for capital adequacy described in Box 3.6.

3.4.6 TLAC and MREL

Liquidation is the usual outcome of a 'gone concern' situation for smaller firms, but the larger the firm the more likely it is that it will go into some sort of resolution process (particularly if it is a bank), see Section 3.2.5.

For the largest and globally most important banks, i.e. the bank G-SIFIs (also called G-SIBs or in some circumstances G-SIIs), the FSB is introducing an additional capital standard, TLAC. TLAC in effect requires firms to specify which capital instruments are subordinated to liabilities to depositors and gives powers to regulators to 'bail in' these liabilities (i.e. to get them to contribute to capital shortfalls) in situations where the firm needs resolving. Chennells and Wingfield (2015) describe further the concept of *bailing in* a bank and some of the potential pitfalls.

There seems to be some divergence between EU authorities and US authorities on how the concept of bailing in should best be implemented. The TLAC approach, which accommodates the US perspective and is set out in FSB (2015b), aims to be relatively prescriptive over which capital instruments can be bailed-in. In contrast, EU rules as per the BRRD allow relevant resolution authorities discretion to bail in nearly all capital instruments. The relevant authority is typically the EU's Single Resolution Mechanism for larger banks within the Eurozone, but typically national supervisory authorities in other member states or for smaller banks. The EU approach applicable for firms other than global systemically important banks (for which TLAC applies) is called the *Minimum Requirement of Own Funds and Eligible Liabilities* (MREL).

The TLAC term sheet as per FSB (2015b) makes clear that the TLAC requirements apply alongside any minimum capital requirements set out in Basel III rather than superseding them. TLAC requirements apply to each individual 'resolution entity' where a resolution entity is defined as a 'parent company, an intermediate or ultimate holding company, or an operating subsidiary'. A G-SIFI bank may have one or more resolution entities.

3.5 Tranching

3.5.1 Introduction

Another term for the multi-layered liability side of the balance sheet illustrated in Box 3.6 is that it is *tranched*. Tranched capital structures involve an entity having liabilities that have different levels of subordination. They are perhaps most easily illustrated by reference to *collateralised debt obligations* (CDOs). Some investors in CDOs or similar structures suffered large losses during the 2007–09 Credit Crisis and so in their own right they have relevance from a systemic risk perspective.

Tranched structures have also been proposed as ways of tackling some of the systemic risk consequences of the close linkage between sovereigns and their own banking systems. Some commentators have proposed the creation of *European Safe Bonds* (ESBies), see e.g. Brunnermeier et al. (2016). These would be formed by the senior tranche of a portfolio of EU government bonds. More recently, these structures have been renamed *sovereign backed securities*. It would be an interesting twist for the same underlying structure to be simultaneously the cause of some of the challenges evident in the 2007–09 Credit Crisis and a possible solution to some of those evident in the subsequent Eurozone sovereign debt crisis.

A key point to bear in mind with tranching is that it involves a rearrangement of who suffers what if there are (credit) losses within a portfolio. The most basic way in which losses can be shared involves all investors suffering a proportionate share of the losses (e.g. if I own 30% of the portfolio then I suffer 30% of the losses incurred by the whole portfolio). This is typically how a mutual fund or other unitised vehicle operates. It is akin to a portfolio with just one tranche. But with two or more tranches the tranches share in losses differently.

The implication is that the losses in tranched structures are *disproportionately* shared. Some of the issues that have arisen with such structures involve investors not fully understanding this. Unless carefully explained and understood their behaviour may appear to be at variance with the desire for 'fairness' that is hard-wired into most human brains.

3.5.2 Traditional CDOs

There are two main types of CDO, i.e. physical CDOs and synthetic CDOs.

The original sorts of CDOs were physical ones. They involved the establishment of a *special purpose vehicle* (SPV) that held one set of debt instruments and funded the purchase of these positions by itself issuing several different tranches of debt, see Fig. 3.3. We immediately note the visual similarity between this diagram and the schematic diagram of the balance sheet of any financial organisation in Fig. 3.1. The different tranches have different priority levels and therefore command different credit ratings and

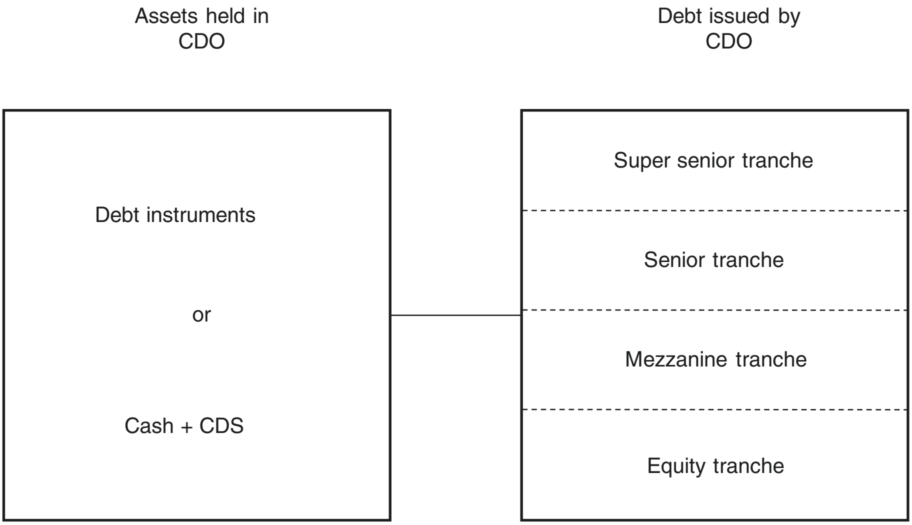


Fig. 3.3 Structure of a collateralised debt obligation (CDO)
Source: Nematrian

credit spreads. A SPV is simply a legal structure that a firm establishes with a specific purpose in mind (in this case to hold debt securities and to fund their purchase by issuing other sorts of debt).

By 'priority level' we mean that in the event of losses arising on the portfolio of debt instruments held by the CDO, losses up to a certain amount would be met wholly by the lowest tranche until it was exhausted. Above this level, losses would be met by the next tranche. Working up, eventually if the whole portfolio defaulted (and there were no recoveries on the debt instruments held), some of the losses would be borne by the highest tranche.

Conventionally, the lowest tranche, i.e. the one that takes the first loss, is called the equity tranche, the next one the mezzanine tranche, etc. up to the highest which might be called the senior or super-senior depending on how many tranches there are. The loss above which a given tranche starts to suffer losses is called its subordination level or its attachment point. The loss above which a given tranche has been wiped out and therefore does not suffer any further losses is called its detachment point, see Fig. 3.4. It is hopefully very unlikely that most of the debt instruments will default. So, the likelihood of super-senior tranches not receiving payment in full should be low, and hence they should be highly creditworthy and should command only a modest credit spread.

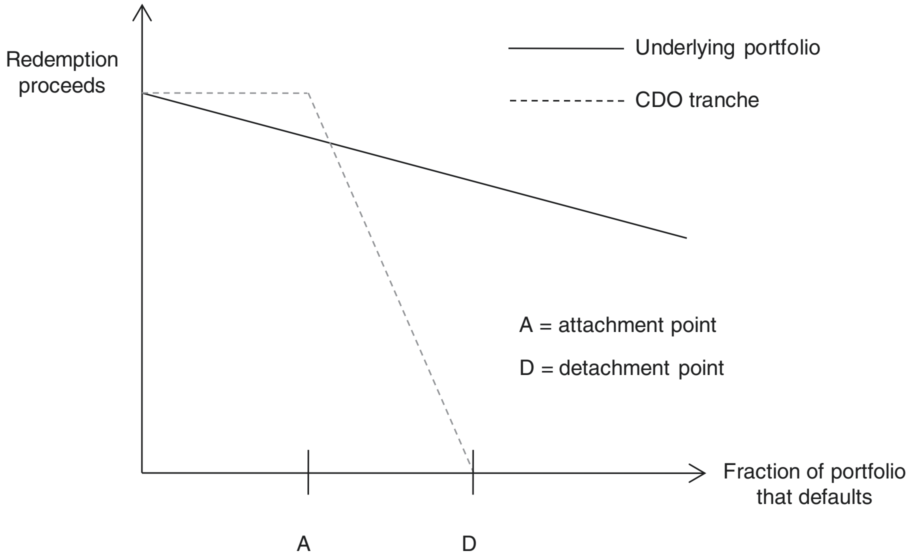


Fig. 3.4 Redemption proceeds of different CDO tranches
Source: Nematrian

3.5.3 Initial Rationale Behind Issuance of CDOs

The initial reason banks started to issue CDOs or collateralised loan obligations (CLOs) was to eliminate debts or loans from their balance sheets. They could do so by creating SPVs and selling the debts or loans to the SPV. The SPV needs to raise sufficient funds to be able to purchase the debts or loans from the bank. The SPV would have its own capital structure, issuing various tranches of debt. Different entities would subscribe to the different tranches of the SPV's debt, the spreads being demanded being dependent on where in the priority ladder the relevant tranche lay.

The underlying economic rationale for CDOs (and tranching more generally) is that different market participants may find different parts of the credit risk spectrum particularly relevant to their own needs. For example, different investors will have different risk profiles, perhaps because they are subject to different regulatory requirements. By repackaging risks so that each tranche can be sold to the sort of investor to which it is most suited, the theory is that the sum of the parts can in some sense be 'worth' more than the whole. This is also the ultimate economic rationale behind the development of other risk transference or risk sharing mechanisms such as derivatives and insurance markets. It is perhaps worth noting that if regulatory capital was always set in the manner implicit in the conceptual framework for capital adequacy set out in Box 3.6 then the apparent arbitrage involved here would fall away, since the credit risk involved would be treated consistently across all market participants.

Originally, the debt instruments held within CDOs were typically passively managed or subject to very limited substitution rights, i.e. defined rules for replacing, say, a bond that had defaulted with another non-defaulted bond, to avoid the CDO having defaulted paper on its books. By the time the 2007–09 Credit Crisis struck, it had become more common for CDOs to be actively managed. Good active management benefits the investors in the CDO (just as it benefits investors in any other sort of actively managed investment product).

3.5.4 Synthetic CDOs

Traditional tranched CDOs suffer from the significant disadvantage that the SPV needs to sell *all* its tranches to raise the funds it needs to buy its debt portfolio. Selling the equity tranche to third parties was often particularly difficult. To circumvent this challenge, investment banks developed the concept of the *single tranche* CDO in which the investment bank synthetically acquired all bar a given tranche by selling to (and/or buying from) the SPV some credit protection that replicated what would have happened had there been the remaining tranches and these had been sold to third parties.

These portfolio-level transactions are examples of *basket* credit default swaps (i.e. credit derivatives dependent on a whole basket of credit names), rather than the more standardised *single-name* credit default swaps (that depend merely on the behaviour of a single credit).

The investment bank would then typically try to hedge the risks it had incurred via these tranche CDS contracts. A good way for it to hedge at least some of these risks is for it to buy suitable amounts of single-name CDS protection on each of the individual credit risk exposures contained within the underlying portfolio. It therefore became commonplace for single tranche CDOs to be structured so that their credit exposures were implemented using credit default swaps rather than physical bonds. Such structures are also called *synthetic CDOs*.

Prior to the 2007–09 Credit Crisis, issuance of single tranche CDOs exploded as they appeared to offer high credit quality combined with high credit spreads. When investors seem to be buying instruments that have these two arguably inconsistent characteristics simultaneously (and without perhaps carrying out as much due diligence as they should on such instruments) then they are said to be *searching for yield*.

Alongside other vulnerabilities, markets in the run up to the 2007–09 Credit Crisis seemed to be characterised by a strong search for yield, not just in terms of CDOs but also across nearly all credit asset classes. There was talk of a new paradigm, with changes to the structure of markets having supposedly led to a greater and more effective sharing of risk across the system. One of the warning signs macroprudential authorities now look for is evidence of renewed search for yield.

3.5.5 Risk Analysis of CDOs

How did investors in the run up to the 2007–09 Credit Crisis so rashly ignore the inconsistency between apparently high credit quality and high credit yields? Many commentators blame an undue reliance on the credit rating assigned to different instruments by a third-party credit ratings agency. Such rating agencies are called external credit assessment institutions (ECAIs) in modern EU regulatory frameworks.

The ratings agencies used Monte Carlo simulation and other techniques to identify how likely they thought a given tranche was to suffer a default (and its likely recovery rate). Usually attachment and detachment points were set so that the tranche in question achieved a certain credit rating at outset, with some margin built in to provide some protection against a downgrade in somewhat adverse circumstances. Detailed methodologies varied by rating agency. For example, in 2008, one agency apparently concentrated just on the subordination level (i.e. attachment point), whilst another one apparently also allowed for the expected loss if the attachment point was reached (which depends on where the detachment point is placed).

One of the major issues that arose with such instruments as the 2007–09 Credit Crisis progressed was the extent to which these ratings diverged from what appeared to be the 'true' (or at least market implied) creditworthiness of the tranche in question. Also, there was not always parity between a rating given to a tranche and an apparently identical rating given to a more conventional bond, providing further scope for confusion for those who were relying too blindly on such ratings.

An important consequence is a strong desire by regulators to reduce the dependency on external credit ratings of regulatory regimes they police. This has proved very tricky to implement in practice, so regulators have in some cases reverted to mapping credit ratings from ratings agencies onto standardised *credit quality scores* (CQS). These at least allow regulators to 'downgrade' the deemed reliability of a given rating agency's ratings if this is considered appropriate.

3.5.6 Tranching and Leverage

One important point to note is that CDO tranches potentially involve leverage. There are several angles to this. That some leverage may exist can be surmised by noting that if, say, a tranche is \$100m in size and has an attachment point at, say, 6% and a detachment point at, say, 10%, then this \$100m must in some sense have some underlying 'reference portfolio' of \$2.5bn, i.e. \$100÷(0.10–0.06)m. Conversely, there should also be some dampening effect due to the subordination characteristics of any given tranche. This makes it unclear how much leverage any given tranche might exhibit in practice.

One way of identifying a practical measure of leverage is to consider the sensitivity of the price of a tranche to (small) movements in credit spreads and to compare this sensitivity with that of a portfolio containing the same credit exposures as those underlying the whole CDO. Commonly such sensitivities are measured using 'credit DV01' (CRDV01), i.e. the change in value created by (say) a 1 basis point change in credit spreads. The interest rate risk analogue to CRDV01 is 'interest rate DV01' (IRDV01), which is the change in value created by a 1 basis point change in interest rates.

The ratio between the spread sensitivity of the individual tranche and that of the whole underlying portfolio is called the *tranche delta*. Suitably weighted, the average tranche delta of all tranches in a physical CDO should equal one, since the sum of the payoffs to all tranches combined mirrors the payoff to the whole underlying portfolio. Tranche deltas are sensitive to several factors, but JP Morgan estimated that the weighted average deltas across all CDOs issued in 2004 were roughly as shown in Table 3.5. Some tranches had tranche deltas less than one but they were not the ones most typically purchased by investors at the time. Instead, investors typically purchased tranches which could be expected to be more sensitive to spread movements than might otherwise have been expected by the investor.

Many ways of exploring capital adequacy struggle to cater with equivalent effects applicable to the firm's own balance sheet. These effects imply that one dollar of additional capital of one type is not necessarily equivalent to

Table 3.5 Approximate average delta of different CDO tranches during 2004

Tranche type	Subordination	% of issuance	Average tranche delta
Junior	0 to 3%	12	14.3
Mezzanine	3 to 7%	33	9.6
Senior	7 to 10%	24	4.3
Super senior	>10%	31	0.4
Total/Average		100	6.0

Source: JP Morgan, Kemp (2005).

one dollar of additional capital of a different type, but the usual accounting approach is to assume equivalence until some limit is reached when any further capital of a specific type is then disallowed. One of the appeals of the conceptual framework for capital adequacy introduced in Box 3.6 is that any such differences would be more appropriately captured, as the implicit credit spread applicable to customer liabilities is sensitive to such differences.

3.5.7 Market Risk Appetite and Liquidity Risk

Leverage is not the only feature relevant to analysing the risk characteristics of different tranches. The riskiness of senior tranches is strongly influenced by the likelihood of many defaults occurring nearly simultaneously, i.e. how correlated are the pattern of defaults in the portfolio.

It is possible to derive market implied correlations from prices of such tranches. These correlations can get very high indeed in stressed circumstances, see e.g. Heidorn and Kahlert (2010). They note that implied (base) correlations for iTraxx Europe senior-high tranches (i.e. ones with 12–22% attachment-detachment points) were above 80% for much of 2008 and peaked at 96.90%! This implies an extremely high likelihood that multiple defaults would occur over the lifetime of the instrument.

Implied correlations applicable to European *equity* markets showed a noticeable rise (relative to earlier levels) at the same time, but peaked nowhere near as high as the ones for European *credit* markets referred to above. Unless you believed in a doomsday scenario, this level of correlation appears implausibly extreme, i.e. appears to indicate that the relevant credit markets had become out of line with reality. But in the depths of the 2007–08 Credit Crisis doomsday scenarios were not self-evidently wrong. If the Crisis had led to a political upheaval which ultimately morphed into something akin to the Russian Revolution of 1917 then any correlation below 100% could have proved optimistic. Market implied parameters embed information about market risk appetite but who truthfully can be sure when fears swing to extremes that they are necessarily unfounded?

Another way of interpreting the extraordinarily high implied correlations applicable to super-senior tranches during the 2007–09 Credit Crisis is to view them as representative of an extraordinary decline in their liquidity. They could only be sold, if at all, at prices that involved a substantial discount to their perceived 'underlying' value, at least if this underlying value was struck using assumptions regarding correlations between defaults more in line with experience in less stressed circumstances. Prices struck this pessimistically necessarily require extreme parameter values to be replicated by the pricing models used to back out relevant market implied information. We have already noted that the Crisis can be thought of as ultimately a liquidity crisis characterised by a particularly savage decline in liquidity being made available to weaker market participants. Market liquidity is subject to some feedback loops that make it subject to particularly extreme outcomes in some circumstances, see Fig. 3.5.

Market risk appetite is another difficult to predict behavioural feature of the financial system. We might naively expect systemic risk events to create one-off shocks which wear off reasonably rapidly, as markets adjust to a new equilibrium applicable post the shock. Price discovery in many financial markets is very now rapid given modern communication technologies and the speed at which it is possible to trade on such information in many modern financial markets.

The reality is rather different. Systemic risk events often involve a breakdown in trust and this can take a more extended time to return, see e.g. Anand et al. (2011). In consequence, financial systems seem to exhibit 'hysteresis', stylistically illustrated in Fig. 3.6. They may only slowly revert to 'normality' (if normality is judged by the state of the financial system between systemic risk events). Hysteresis is a well-known physical phenomenon involving a switch between different states. Many useful physical processes make use of this phenomenon, e.g. the switching between polarities in a computer magnetic disk drive used to store data. As we have previously noted, the financial system is an example of a complex adaptive system and these sorts of systems often shift between discernible states. Liquidity conditions are very heavily influenced by market participants' assessments of the likely behaviour of other market participants, so can be expected to be particularly sensitive to such features.

Modelling such behaviour is challenging as it requires models that include more than one state or regime in which the (financial) world might be in at any given point in time (as well as a component identifying the likelihood of switching between these states). Generically these are called regime switching models.

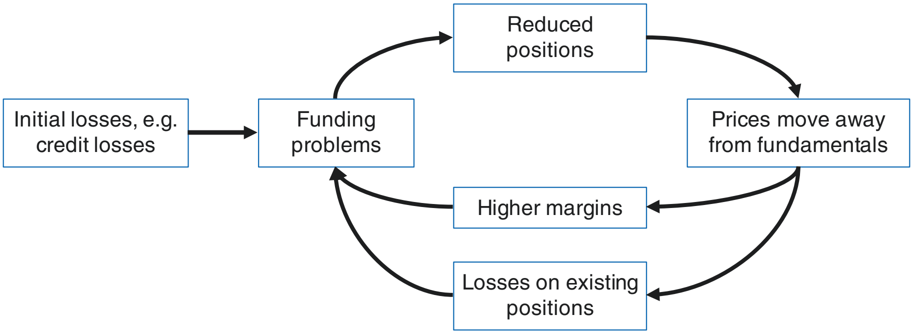


Fig. 3.5 Liquidity feedback loops

Source: Nematrian. Adapted from Brunnermeier (2009)

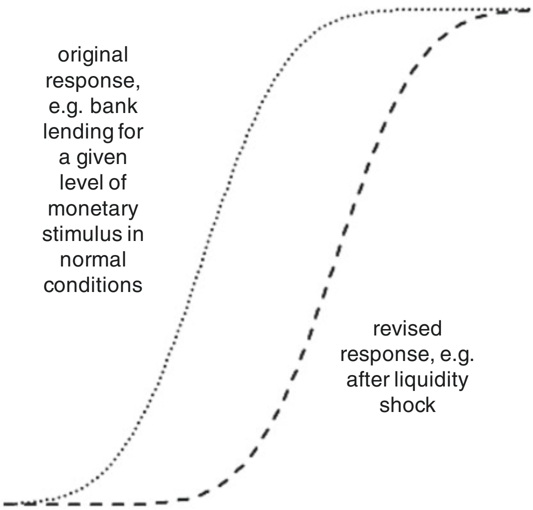


Fig. 3.6 Financial system hysteresis

Source: Nematrian

3.6 Rational and Irrational Behaviours

A question that follows on from comments in the last Section is whether individuals, markets and other elements of the financial system behave rationally or irrationally when systemic risk events come along. Franklin D Roosevelt is famously quoted in his US Presidential inauguration speech in 1933 as having said: '*the only thing we have to fear is fear itself*'. If only we looked at matters sensibly ('rationally'?) would we not see the irrationality of our collective behaviour and would we not respond collectively to address the issues concerned?

This question underpins a larger debate that appears to be taking place at present within the field of economics. This is the question about the extent to which classical or neo-classical economics, with its typical focus on equilibrium style approaches to how the economy operates, is a suitable description nowadays of how the economic world operates in practice.

Increasingly, researchers appear to be seeking insights from the field of *behavioural economics* or the closely allied discipline of *behavioural finance*. Both disciplines posit that there are certain common human traits that significantly influence our behaviours. Some are set out in Box 3.11.

Few people doubt that these influence financial markets, sometimes significantly. In particular, many booms and busts (the 'animal spirits' as J.M. Keynes would put it) are inflated by the collective exercise of such factors.

And yet, there is also little doubt that in times of stress people can be surprisingly rational and cold-blooded in their dealings with others. In the depths of the 2007–09 Credit Crisis, market participants expended a lot of effort trying to estimate the 'fair value' position of precariously positioned counterparties, and stopped doing business with them if they looked to be in danger of folding. The American economy did not properly recover after President Roosevelt's inauguration in 1933 until full employment returned with the Second World War in the 1940s. Being told that fear was the only thing the American public needed to worry about (i.e. that if only the public were more rational about the economic challenges facing the USA then everything would be OK), seems a little way from the full picture, even if it made for a good sound-bite at the time.

Box 3.11: Behavioural finance and its relationship to 'classical' economics

In recent years, a range of economists and behavioural specialists have questioned the theoretical basis assumed to underlie most economic theory. They have argued that the notion that humans behave in the theoretical manner of 'homo economicus' (that is supposed to govern the rational behaviour of economic agents when interacting economically with others) is fundamentally flawed.

Instead they highlight a range of biases humans seem to exhibit in practice, most or all of which can be identified in laboratory-style experiments on humans (disproportionately drawn, it may be said, from university student populations, who may not form a fully representative sample of the wider human populace). Proponents of this branch of economics or psychology include e.g. Kahneman (2011).

In an investment context, such biases include, see e.g. Montier (2007):

- *Overconfidence and overoptimism*: 'I know better because I know more'
- *Size bias*: assuming 'big' equals 'important'
- *Confirmatory bias*: seeing what you want to see
- *Self-attribution bias*: 'heads was skill, tails was bad luck'
- *Hindsight bias*: 'I knew it all along'
- *Anchoring*: an irrelevant value/factor influences where we pitch the answer
- *Representativeness bias*: judging by appearances not likelihoods
- *Recency effect*: preferentially remembering the recent past
- *Framing*: how we are asked the question can influence the answer we give
- *The status quo and endowment effect*: once you have something you may start to put a higher value on it than others

It is hard to disagree with the existence of these biases. Anyone willing to adopt a little introspection will soon recognise many of these biases in their own lives and economic actions. Some of these biases even arguably appear in the field of systemic risk analysis. For example, we tend preferentially to remember the recent past (e.g. the 2007–-09 Credit Crisis) over more distant events (which is one reason in this book we have attempted to include details of earlier historical systemic risk events). Perhaps focusing too much on TBTF is an example of a size bias (i.e. assuming too readily that 'big' equates to 'important'). The strong current focus on improved transparency and data provision as a means of tackling systemic risk (as noted in e.g. Section 5.5) may eventually turn out to be an example of the fallacy of 'I know better because I know more'.

Most of the behaviours described above can be viewed as akin to a preference towards 'convenience'. Indeed, writers in this field such as Kahneman (2011) talk about our conscious (rational) versus our unconscious (behaviourally driven) brain, with it being all too convenient to revert to the latter because it offers heuristics that work tolerably well much of the time. There is business value in 'convenience'. Convenience here encompasses not just our willingness as customers to go to say a convenience grocery store rather than to make a longer trip to a bigger supermarket where prices are likely to be lower. It also encompasses broader network effects such as in IT, where value accrues because it is an effort for individuals in isolation (or en masse) to move from a product they are familiar with to a different one. Money shares many of these attributes, so may be particularly exposed to behavioural effects.

Perhaps, though, the most challenging question to pose is why, if our behaviours are so riddled with behavioural biases, do they not appear to make *more* difference in the real world?

From an investment market perspective, the existence of behavioural biases such as those set out above ought to offer plenty of attractive opportunities to make money for active investment managers able to resist these biases more than the generality of investors. And yet, investment markets for the most part appear to be reasonably 'efficient' (using this word in the same way as is used in the efficient markets hypothesis). Whilst some investors do seem to be able to outperform their peers, the number who do so very consistently is very few, and some of these apparent instances seem to involve a form of 'self-attribution bias', or at least effective marketing to those who seem to be swayed by such a bias.

Most probably, this reflects the powerful moderating impact markets have on such behaviours. In economic affairs, we are usually willing to tolerate some irrationality on our own behalf, because it is too much of an effort to be super-rational all the time. But if the price we are paying is too far from an easily observable market price then nearly everyone eventually becomes persuaded that it is worth shopping around. We may also compartmentalise issues (and parts of our own financial affairs) because it is a convenient and usually prudent mindset to adopt, whilst knowing that compartmentalisation ultimately gets broken if the needs are great enough.

Where this breaks down from a financial stability perspective is if the price is primarily driven by what we think other people may think an asset is worth. Then prices can diverge from some more rational view of the value of the asset, and a bubble (or, the opposite, i.e. bust) can form. Even then, though, it is still tricky to persuade people to buy at a price too far away from the market price; it is the market price itself that exhibits irrational exuberance or irrational pessimism.

3.7 Key Takeaways

This Chapter has explored features influencing systemic risk that are shared across the whole of the financial system. Key points noted include:

(a) It is common amongst practitioners to assume that contribution to systemic risk is largely about the degree of interconnectedness a firm exhibits with other parts of the financial system. This is only a partial picture at best. Systemic risk events commonly reflect underappreciated vulnerabilities that one or more parts of the system possess. Interconnectedness seems primarily to be the trigger that translates these vulnerabilities into financial crises, rather than (usually) the root cause of the crisis itself.

(b) Financial sector regulation primarily seeks to address an information asymmetry that exists between firms and their customers. One important regulatory development since the 2007–09 Credit Crisis has been the concept of enhanced resolution planning, i.e. firms needing to plan how they might be closed in an orderly fashion. This development arguably seeks to address an information asymmetry between firms and their regulators and is most acute for firms that are deemed to be or may prove to be too big to fail.

(c) Regulatory frameworks contain many elements, including regulatory capital minima, incentives for firms to maintain adequate levels of economic capital for the risks that they face, governance and supervisory review disciplines and methodologies that seek to achieve transparency and hence market discipline. Most frameworks tend to be relatively microprudential in nature, i.e. focusing on the risks faced by individual firms. As we shall see in later Chapters, many also now contain overrides that seek to increase their macroprudential elements. However, industry participants are not always keen on such overrides, particularly if they do not see their part of the financial system as practically contributing much to systemic risk.

(d) A particularly important part of regulation for many market participants relates to regulatory capital computations. Whilst it is not the way that regulatory capital computations are normally formulated, we can conceptually think of regulatory capital as being held to deliver some suitable level of credibility to the promises organisations give their customers or other (direct) stakeholders, such as depositors for banks, policyholders for insurers and beneficiaries for pension funds. These computations can involve internal model or standard assessment approaches. Internal models are potentially more risk-sensitive than standard assessment approaches, but are typically also more complex and less comparable across firms. Different regulatory frameworks tend to come down differently on exactly how much they favour one style of approach over the other and to change through time as policymaker opinions change.

(e) Regulatory capital computations require valuations of assets and liabilities. Third parties might therefore expect results published in financial statements to be a major driver behind such computations. In practice, however, regulators can and do override valuations using generally accepted accounting principles as shown in these statements with valuations that are more suitable for demonstrating whether an organisation is adequately solvent. The general trend is towards greater use of fair, i.e. market consistent, valuations that take greater note of market valuations. Some commentators believe that this potentially makes regulatory capital computations more procyclical, but there is little real evidence that this is the case. In stressed times market participants seem to put a great deal of effort into understanding how less robust players might appear if other valuation approaches are restated to be more on a fair valuation like basis. Lack of ready availability of this information and the consequential lack of transparency can itself foster financial instability.

(f) An important aspect of an organisation's balance sheets is the notion of tranching, i.e. the subdividing of an organisation's liabilities by the priority accorded to them in the event of wind-up or failure. Some of its consequences are hidden in traditional sorts of financial statements that e.g. do not capture some of the implicit liabilities that would fall to governments or government organised deposit insurance arrangements in such circumstances. Some of the vehicles that proved most problematic during the 2007–09 Credit Crisis explicitly focused on tranching. Conversely, the financial technology involved in tranching has also been proposed as a solution to some other financial stability problems.

(g) Commentators debating systemic risk naturally come with their own biases, reminding us that humans do not always adopt rational stances in their interactions with the financial system. Few people doubt that broader behavioural biases play important roles in how financial markets and hence finance systems evolve, as they can at times add fuel to bubbles and busts and hence to the magnitudes of systemic risk events.

References

Abad, J. and Suarez, J. (2016). A Recursive Model for the Assessment of the Cyclical Implications of IFRS 9. *CEMFI*, mimeo.

Alves, I., Brinkhoff, J., Georgiev, S., Héam, J-C, Moldovan, I. and Scotto di Carlo, M. (2015). Network Analysis of the EU Insurance Sector. *ESRB Occasional Paper No 7.*

Anand, K., Gaib, P. and Marsilic, M. (2011). Rollover Risk, Network Structure and Systemic Financial Crises. *SSRN.*

Bank of England (2008). Financial Stability Report, October 2008. *Bank of England.*

Bank of England (2009). Recovery and Resolution Plans – Remarks by Andrew Bailey. *Bank of England News Release*, 17 November 2009.

Bank of England (2015). Financial Stability Report, July 2015. *Bank of England.*

BCBS (2016a). Discussion Paper: Regulatory Treatment of Accounting Provisions. *Basel Committee on Banking Supervision.*

BCBS (2016b). Minimum Capital Requirements for Market Risk. *Basel Committee on Banking Supervision.*

BCBS (2016c). Consultative Document: Reducing Variation in Credit Risk-Weighted Assets – Constraints on the Use of Internal Model Approaches. *Basel Committee on Banking Supervision.*

BCBS (2016d). Consultative Document: Standardised Measurement Approach for Operational Risk. *Basel Committee on Banking Supervision.*

Benford, J. and Nier, E. (2007). Monitoring Cyclicality of Basel II Capital Requirements. *Bank of England*, Financial Stability Paper No. 3, December.

Brunnermeier, M.K. (2009). Deciphering the Liquidity and Credit Crunch 2007–09. *Journal of Economic Perspectives*, 23, No. 1, Winter, pp. 77–100.

Brunnermeier, M.K., Langfield, S., Pagano, M., Reis, R., Van Nieuwerburgh, S. and Vayanos, D. (2016). ESBies: Safety in the Tranches. *ESRB Working Paper Series*, No 21.

Chennells, L. and Wingfield, V. (2015). Bank Failure and Bail In: An Introduction. *Bank of England Quarterly Bulletin* 2015 *Q3.*

Devasabai, L. (2014). No Arbitrage. *Risk*, September.

Dymski, G.A. (2011). Genie out Of the Bottle: The Evolution of Too-Big-to-Fail Policy and Banking Strategy in the US. *Post-Keynesian Economics Study Group.*

ESRB (2016). Indirect Contagion: The Policy Problem. *European Systemic Risk Board*, Occasional Paper 9.

European Union (2014). Bank Recovery and Resolution Directive, i.e. Directive 2014/59/EU. *European Union.*

FDIC (2000). An Examination of the Banking Crises of the 1980s and Early 1990s. *Federal Deposit Insurance Corporation.*

Federal Reserve (2014a). Guidance for 2013 Section 165(d) Annual Resolution Plan Submissions by Domestic Covered Companies that Submitted Initial Resolution Plans in 2012. *Federal Deposit Insurance Corporation, Board of Governors of the Federal Reserve System.*

Federal Reserve (2014b). Guidance for 2013 Section 165(d) Annual Resolution Plan Submissions by Foreign-Based Covered Companies that Submitted Initial Resolution Plans in 2012. *Federal Deposit Insurance Corporation, Board of Governors of the Federal Reserve System.*

Financial Times (2008). Aegon gets EUR 3bn in government capital. *Financial Times*, 28 October.

French, A., Vital, M. and Minot, D. (2015). Insurance and Financial Stability. *Bank of England Quarterly Bulleting* 2015 *Q3*

FSB (2014). Consultative Document: Assessment Methodologies for Identifying Non-Bank Non-Insurer Global Systemically Important Financial Institutions: Proposed High-Level Framework and Specific Methodologies. *Financial Stability Board.*

FSB (2015a). Consultative Document (2nd): Assessment Methodologies for Identifying Non-Bank Non-Insurer Global Systemically Important Financial Institutions: Proposed High-Level Framework and Specific Methodologies. *Financial Stability Board.*

FSB (2015b). Principles on Loss-absorbing and Recapitalisation Capacity of G-SIBs in Resolution, Total Loss-absorbing Capacity (TLAC) Term Sheet. *Financial Stability Board*

Heidorn, T. and Kahlert, D. (2010). Implied Correlations of iTraxx Tranches during the Financial Crisis. *Frankfurt School – Working Paper Series*

HM Treasury (2009). A Review of Corporate Governance in UK Banks and Other Financial Industry Entities – Final Recommendations. *UK HM Treasury*, November.

Huizinga, H. and Laeven, L. (2010). Bank Valuation and Regulatory Forbearance During a Financial Crisis. *SSRN.*

IASB (2014). IFRS 9 Financial Instruments: Project Summary. *International Accounting Standards Board.*

IMF (2009). Guidance to Assess the Systemic Importance of Financial Institutions, Markets and Instruments: Initial Considerations: Report to the G-20 Finance Ministers and Central Bank Governors. *International Monetary Fund, Bank for International Settlements and Financial Stability Board*

Kahneman, D. (2011). *Thinking, Fast and Slow*. Penguin Books, London, UK.

Kemp, M.H.D. (2005). Risk Management in a Fair Valuation World. *British Actuarial Journal*, 11, No 4, pp. 595–712

Kemp, M.H.D. (2009). *Market Consistency: Model Calibration in Imperfect Markets*. John Wiley & Sons, Chichester, UK.

Kemp, M.H.D. and Varnell, E. (2010). Regulatory Frameworks – Lessons Learned and Potential Implications of the Credit Crisis. *Presentation to Institute and Faculty of Actuaries Risk and Investment Risk Conference, June 2010*

Laux, C., and Leuz, C. (2010). Did Fair-Value Accounting Contribute to the Financial Crisis? *Journal of Economic Perspectives*, 24, 93–118

Montier, J. (2007). *Behavioural Investing: A Practitioner's Guide to Applying Behavioural Finance*. John Wiley & Sons, Chichester, UK.

PRA (2015). The Internal Capital Adequacy Assessment Process (ICAAP) and the Supervisory Review and Evaluation Process (SREP). *The Bank of England, Prudential Regulation Authority.*

S&P (2013). 2012 Annual Global Corporate Default Study and Rating Transitions. *Standard & Poor's.*

4

Individual Elements of the Financial System

The previous Chapter explored some topics that span the whole financial system. In this Chapter, we explore in more detail the systemic risk characteristics of individual parts of the financial system. Separate Sections cover banks, insurers, pension funds, investment funds, asset managers, shadow banks and providers of market infrastructure such as exchanges and clearing houses. We also explore how these organisations interact with other players in the wider financial system, including regulators and supervisors, governments, sovereign wealth funds (and other 'long-term' investors), the real economy and the public more generally.

For reasons explained previously, we adopt a deliberately broad definition of systemic risk. In this Chapter, we focus mainly on the *current* structure of the financial system. Other Chapters explore how the financial system might evolve in the future. The financial system, particularly the more commercial parts that we might group together as the 'financial services industry', has seen many changes over the last few decades. We may expect change to continue in the future, not just in how the industry is regulated but also in the products it offers and the tools and techniques it uses to risk manage these products. Most practical ways of measuring risk implicitly assume that the present (and past) is a reasonable guide to vulnerabilities that might appear in the (near) future. Implicit in adopting a similar philosophy for systemic risk is a worry that the financial system might suffer a shock that is sufficiently painful for it to freeze up. Knowing the current shape of the system should help us identify how it might enter such a state.

M.H.D. Kemp, *Systemic Risk*, DOI 10.1057/978-1-137-56587-7_4

4.1 Banks

4.1.1 Introduction

The 'combined' model of systemic risk introduced in the previous Chapter suggests that when exploring organisations across the financial system we should focus on ones that exhibit a lot of interconnection as well as having (possibly hidden) vulnerabilities. On both counts a natural place to start is the banking sector.

Most commentators assume that banks are the most likely players to give rise to systemic risk issues. Policies seeking to address systemic risk are more developed in the banking sector than elsewhere. As French, Vital and Minot (2015) point out, a high proportion of banks have some involvement with payment services, so a healthy banking system is typically deemed to be a crucial contributor to a robust real economy. The banking sector in its different guises also forms, on various measures, the largest part of the financial system, see Fig. 2.1.

There are many different types of bank. Traditionally, banks were split between *commercial banks* and *investment banks*. In the USA, this split was originally institutionalised by the Glass-Steagal Act(s) that came out of the banking crises of the 1930s Great Depression, see Box 4.1. It prohibited firms from doing both activities at the same time. Commercial banking in this context can be characterised by deposit taking from and extending loans to individuals, corporates and other organisations. By contrast, investment banking focuses on activities relating to capital markets, e.g. supporting firms issuing securities.

Commercial banking can itself be split between *retail banking* (focusing on e.g. lending to individuals or small enterprises) and *corporate banking* (focusing on equivalent activities but for larger corporates or other similar organisations).

By the time the 2007–09 Credit Crisis struck the last vestiges of Glass-Steagal had been dismantled. There is now resurgent pressure to limit what banks subject to deposit protection arrangements can do. These include the Volker ban on proprietary trading introduced by Dodd-Frank and ring-fencing of retail banks in some jurisdictions.

It is not always obvious what is or is not a bank. Policymakers usually define them as bodies that perform financial intermediation (particularly any financial intermediation that involves credit intermediation or maturity transformation) and are subject to capital regulation. A more precise definition applicable in the EU is given in the CRR. It uses the term 'credit institution' for them and defines such an institution as an '*undertaking the business of which is to take deposits or other repayable funds from the public and to grant credits for its own account*'.

Box 4.1: The Wall Street Crash, Glass-Steagal, the FDIC and the Great Depression

The Wall Street Crash of 1929 is usually deemed to be the most devastating stockmarket crash in history. It and its aftermath, the Great Depression, were massive shocks to the global financial system. A number of very important US regulatory changes followed, including separation of commercial from investment banking under the Glass-Steagal Act(s) and the development of deposit insurance under the remit of the FDIC.

As with the less severe 2007–09 Credit Crisis, the years leading up to the Wall Street Crash were characterised by business optimism. New technologies such as radio were all the rage. Had they used the terminology at the time, the talk would have been of new paradigms and the like. The Roaring Twenties which preceded the Crash saw substantial industrial growth fuelled in part by extensive migration of rural Americans to cities in the hope of a more prosperous life. The upward climb in the US stockmarket was fuelled by this optimism and by speculation, some of it involving buying stocks on *margin* (i.e. using borrowed money or leverage). In March 1929, after the Fed warned of excessive speculation a mini-crash occurred, but this was halted when a leading bank announced it would continue to provide credit to speculators. Stock resumed their rise in June, even though the economy had turned sluggish. By the time the market peaked in September 1929 it had risen 10-fold over the preceding 9 years.

Selling intensified in mid-October as did market volatility. On Thursday 24 October 1929, the market lost 11% at opening but recovered to be only modestly down by close of day following intervention by several leading Wall Street bankers. However, the following Monday, 28 October 1929, with more investors facing margin calls and liquidating positions to meet these calls, the market suffered a then record loss of 13% as measured by the Dow Jones Industrial Average (DJIA). The following day it lost a further 12% even as leading financiers sought to prop up the market. There was a significant bounce back on 30 October, but for most of the next 3 years the stockmarket fell, eventually reaching a bottom in July 1932 (as measured by the DJIA) almost 90% below its September 1929 peak.

Writers such as Galbraith (1955) have analysed in detail the Wall Street Crash, its causes and its consequences. Some view the Wall Street Crash as the main trigger for the Great Depression. Others view the Crash as more a symptom of rather than a cause of the Great Depression (the US stockmarket held up reasonably well during 1930).

Most commentators seem to agree that the Fed should have adopted a more accommodative monetary policy in the immediate aftermath of the crash, as a tightening of monetary policy played a significant part in the large numbers of banks that failed during this time.

Keynesian (i.e. demand-driven) economists also argue that a large-scale loss of confidence led to a reduction in consumption and investment spending, deflation and eventually inadequate aggregate demand to keep the economy growing. As the Depression went on, the US tried a range of public works, subsidies and other ways of stimulating the economy, but it never properly pulled out of recession until the Second World War.

Most historians and economists also lay some blame on the growth in protectionism that followed in the immediate aftermath of the Crash, epitomised by the US Smoot-Hawley Tariff Act (which was enacted in June 1930). This seriously reduced international trade, and led to retaliatory tariff increases from other countries. International trade was not particularly important for the US but was much more important for many other countries. This mechanism therefore resulted in an exporting of the US's difficulties to others. Intertwined with growing worldwide protectionist trends was a focus on the gold standard. Every major country left the gold standard during the Great Depression. The speed of doing so seems to have been positively correlated with the speed of subsequent economic recovery.

From a broader historical perspective, the Great Depression was so severe that it led to considerable political dislocation in many countries, fostering the rise of fascism in some countries and the horrors of the Second World War which were to follow.

The Great Depression forms one of the seminal studies underpinning modern macroprudential thought. Some of the policy responses adopted after the 2007–09 Credit Crisis were specifically motivated by a desire to avoid problems thought to have been present during the Great Depression.

4.1.2 Banking and Leverage

Nearly all modern banking involves leverage, in the sense that banks take in deposits and other sources of funding from one group of customers and use the monies (funding) received to make loans to another group of customers. This leaves a bank with a net asset base that is typically much smaller than either its assets or its liabilities in isolation.

This leverage is perhaps best illustrated by the notion of 'fractional' banking, which is the process by which banks manufacture 'money'. With fractional banking, banks hold some of their assets in the form of 'true' cash, i.e. legal tender (or instruments issued by the government that are guaranteed to be exchangeable into legal tender at short notice). However, the amount of such cash they hold at any given point in time is substantially smaller than the amount of 'money' held in at-call (i.e. 'current') accounts by their depositors. Depositors can de facto use funds held in such accounts as money in day-to-day transactions, because the bank promises to allow them to withdraw the funds or transfer them elsewhere whenever they like.

Lecturers typically introduce banking concepts to students by referring to a retail bank that obtains most or all its depositors from individual members of the public (perhaps primarily using the bank's current account service as a substitute for otherwise holding cash in the form of notes or coins). The retail bank is then assumed to advance mortgages and other sorts of loans to other individuals who have some reason for borrowing money from the bank.

Such a bank is undertaking financial intermediation (more specifically, credit intermediation) because in principle the borrower and ultimate lender could arrange the loan directly between themselves rather than using the bank as their intermediary. However, such direct lending (also called *peer to peer lending*) presents challenges for both lender and borrower. The lender needs to carry out suitable credit assessments on the borrower and there is the difficulty of linking up borrower and lender. Banks currently benefit from competitive advantages in these fields and most bank-like lending is done by banks rather than peer-to-peer.

A *bank run* develops when demand to withdraw from at-call (i.e. current) accounts overwhelms the bank's capacity to honour such withdrawals. More precisely, a 'run' is associated with a position where many depositors of a specific bank become worried that the bank will be unable to honour its promises and try to get preferential treatment by withdrawing their funds as quickly as possible. If enough depositors believe the bank has got into difficulties (or even just believe that a sufficient fraction of *other* depositors believe that this is the case) then it becomes rational for depositors to withdraw their funds as quickly as possible leading to the bank collapsing. A significant number of US banks suffered bank runs in the USA in the aftermath of the 1929 Wall Street Crash.

In practice, banks have many different business models. For example, some banks concentrate on the advancing of loans to individuals or small businesses but do not have ready access to deposits coming from a large depositor base. They still need money to fund the advances they wish to make, but will need other sources than the *retail funding* available from retail customers' deposits. Collectively these alternative sources are called *wholesale funding*. They often involve use of instruments, such as mortgage backed securities (MBS) that are traded in the *money markets*. Wholesale funding can also involve advances made by one bank to another bank, i.e. make use of the so-called *interbank market*.

In the stylised model of retail banking described above, a bank's assets are typically of longer maturity (duration) and less liquid than its liabilities, i.e. it undertakes *maturity transformation* and *liquidity transformation*. In theory, a bank might do the reverse, but the whole banking sector on average carries out such transformations because of the biases and needs of the typical consumer of banking services.

Nowadays, banks (and many other market participants) make considerable use of derivative contracts. A derivative is an instrument whose behaviour derives from the behaviour of some underlying index or other economic exposure. For example, going long an equity market index futures contract provides a market participant with a similar payoff to one that involves buying the market index outright.

Usually, parties entering into a derivative transaction do not need to put up the whole amount of exposure involved, i.e. they too include elements of leverage. Parties do, however, need to be able to 'fund' the resulting exposures, in much the same way as they need to fund any other market exposures. If the derivative is exchange traded then this generally involves posting initial margin (usually in the form of cash) with a clearing house (or clearing broker, if the party does not have a direct relationship with the clearing house) and then posting variation margin if the value of the derivative contract moves against the market participant. If the derivative is over-the-counter, i.e. not exchange traded, then this generally involves posting collateral of an agreed type with the counterparty at outset and whenever the instrument value moves by more than a certain amount. In either case, less liquid assets are less helpful as sources of funding, and may need to be sold (or used as collateral in other contexts) to generate sufficient liquid funds to be able to hold on to the position without it being forcibly closed by the counterparty, possibly at material loss to the market participant concerned. The collateral (adjusted by any capital gain or loss arising on the transaction) is returned to the original party posting the collateral at the end of the transaction. Leverage introduced through derivatives contracts is usually called *synthetic leverage*.

4.1.3 Regulatory Capital Frameworks

Banking is a very global business, with some of the largest banks active in nearly all countries simultaneously. The largest investment banks typically trade assets across all main financial markets whenever such trading is possible (which is sometimes around the clock except on public holidays).

Policymakers recognised early on the advantages of adopting harmonised global approaches to capital regulation of globally active banks. The main body that has coordinated these activities is the Basel Committee on Banking Supervision (BCBS), a part of the Bank for International Settlements (BIS, sometimes called the 'banker for central banks'). It has developed the various Basel Capital Accords.

The latest of these is called Basel III. These encapsulate many of the reforms to the previous Basel II Accord that were proposed following the 2007–09 Credit Crisis. Nearly all these reforms can be summarised as involving more capital and more of the 'right' sort of capital. Increasing an entity's capital base should reduce the likelihood of default and hence customer loss. It should therefore reduce the potential cost to governments of stepping in to carry the burden of these losses. For example, BCBS (2009), the Basel Committee's Consultative Document on '*Strengthening the resilience of the banking sector*', which ultimately led to Basel III, had the following main strands:

(a) Improve the quality, consistency and transparency of firms' capital bases;
(b) Strengthen the risk coverage of the capital framework;
(c) Introduce leverage ratio limits to supplement the previous largely capital risk-based framework, i.e. the *liquidity coverage ratio* (LCR) and the *net stable funding ratio* (NSFR), see Section 8.4;
(d) Introduce counter-cyclical capital buffers (including contingent capital arrangements[1])
(e) Introduce enhanced liquidity management standards.

At the time of writing, BCBS is also implementing the *Fundamental Review of the Trading Book* (FRTB), see BCBS (2016b). This is introducing new requirements for the trading book component of a bank's balance sheet. BCBS is also introducing changes to operational risk, see BCBS (2016d), and to treatment of interest rate risk in the banking book (IRRBB), see BCBS (2016e).

Most of the globally agreed standards introduced by Basel III have been implemented in EU legislation via the EU's *Capital Requirements Directive* (CRD), see European Union (2013a) and the associated *Capital Requirements*

[1] Contingent capital is capital that doesn't currently reside on the firm's balance sheet but can be accessed if necessary. Bail in mechanisms mentioned in Section 3.4.6 provide one source of such capital.

Regulation (CRR), see European Union (2013b). This Directive isn't the only EU regulatory framework applicable to banks. Others include the BRRD, see Section 3.2.2, and ones described in Section 8.3. However, the CRD is the EU Directive most associated with capital adequacy requirements for banks (and investment firms).

The CRD and the CRR contain many references to systemic risk, financial stability or macroprudential matters, see Table 4.1 (CRD) and Table 4.2 (CRR). In these Boxes 'G-SII' means *global systemically important institution* 'O-SII' means *other systemically important institution*. As the CRD and CRR focus on banking these in practice mean banking SIFIs. To assist readers who are not well-versed in EU legislative documents we set out below a few terms that appear in such documents that may not be obvious to all readers:

(a) *Recital*. EU legislative documents generally start with a series of recitals that state the purpose of the legislation and guide some of its interpretation. The remainder of the document is formed by Articles (or for longer documents Annexes referred to in these Articles).
(b) *Member state*. Individual countries in the EU are called member states. The EU single market applies to some other countries in which case they too are treated as 'member states' in relevant EU legislation.
(c) *Host member state and home member state*. Under the EU single market firms that are based in one EU member state are typically free to provide goods and services to customers on other member states. This freedom is also called *passporting*. The member state in which the firm is domiciled /registered is called the 'home' member. The member state where the services are being provided is called the 'host' member state.
(d) *Institution*. For the purposes of the CRD and CRR, 'institutions' are the organisations being regulated, i.e. mainly banks, although the CRD and CRR also apply to some types of investment firm.

The CRD and CRR both date from June 2013 although some of their components were phased in over time. Some resolution bodies set up by some member states to facilitate sorting out their banking systems following the 2007–09 Credit Crisis are mentioned in the CRR, including the Irish National Asset Management Agency (NAMA) and the Spanish Asset Management Company (SAMC).

The CRR includes references to many other types of 'system' (e.g. systems of governance, ratings systems, systems to manage concentrations of

Table 4.1 References to systemic risk, financial stability or macroprudential matters in the CRD

Component	What EU Capital Requirements Directive specifies
Recital (26), Article 52.3	Host member state regulators should be able to carry out on-the-spot checks of an institution if relevant for reasons of stability of the host member state financial system
Recital (29)	Allows exchange of information between supervisors in different member states to help strengthen stability of financial system
Recital (30)	Recognises that some behaviours by institutions can affect stability and integrity of the financial system
Recital (34)	Priority for resolution planning should be given to systemically important institutions
Recital (47)	Consolidated supervision aims to protect depositors and investors and to ensure stability of the financial system
Recital (50), Article 7	Competent (national) authorities should also consider effect of their decisions on stability of the financial system in other member states
Recital (51)	Shadow banking has systemic risk relevance (sometimes beneficial, sometimes harmful)
Recital (53)	Weak corporate governance has led to some systemic problems in some member states and globally
Recital (80), (81)	It is appropriate to require credit institutions and relevant investment firms to hold a capital conservation buffer and a countercyclical capital buffer (that takes account of, inter alia, growth in credit levels and changes in ratio of credit to GDP) when there is a build-up in system-wide risk
Recital (85), (87), Article 104.3	Member States should be able to require certain institutions to hold a systemic risk buffer (for non-cyclical systemic or macro-prudential risks not covered by CRR), and to recognise systemic risk buffer rates set by other member states
Recital (86)	It is appropriate for the ESRB to develop principles tailored to the Union economy and be responsible for monitoring their application
Recital (90)	Authorities are expected to impose higher own funds requirements on G-SIIs to compensate for the higher risk that G-SIIs represent for financial system and the greater impact their failure might have on taxpayers
Article 3.1(10)	'Systemic risk' means a risk of disruption in the financial system with the potential to have serious negative consequences for the financial system and the real economy
Article 3.1(30)	'Systemically important institution' means an institution, EU parent institution, EU parent financial holding company or EU parent mixed financial holding company, the failure or malfunction of which could lead to systemic risk
Article 50.1	

(*continued*)

Table 4.1 (continued)

Component	What EU Capital Requirements Directive specifies
	Competent authorities should collaborate closely, including regarding liquidity, solvency, deposit guarantee, the limiting of large exposures, other factors that may influence the systemic risk posed by the institution, . . .
Article 50.4	The host member state competent authority can take appropriate measures . . . to protect the stability of the financial system
Article 51.1, 51.3, Article 158	The host member state competent authorities may request that a branch of an institution other than an investment firm be considered significant, reasons given to include . . . the likely impact of a suspension or closure of the operations of the institution on systemic liquidity . . . or size and importance of branch . . . within context of the banking or financial system of the host member state. The Decisions of the home member state competent authority on e.g. the systemic nature of different institutions should take account of the potential impact on the stability of the financial system of the member state concerned.
Article 56, 57, 58	Parties involved in the exchange of information should include '*(b) authorities or bodies charged with responsibility for maintaining the stability of the financial system in Member States through the use of macroprudential rules; (c) reorganisation bodies or authorities aiming at protecting the stability of the financial system*'
Article 70	Administrative penalties (imposed on institutions) should take account all relevant circumstances including '*(h) any potential systemic consequences of the breach*'
Article 74.4	Competent authorities shall ensure that recovery plans are put in place, but requirements can be reduced if, after consulting the national macroprudential authority, competent authorities consider that the institution is, in effect, less systemically important
Article 86.3	Competent authorities to monitor developments in liquidity profiles and to take effective action where they may lead to individual institutional or systemic instability
Article 89	Reference to country-by country reporting to the Commission for all G-SIIs
Article 97, 98	Competent authorities to review arrangements, strategies, processes and mechanisms institutions use to comply with CRD and CRR and evaluate . . . '*(b) risks that any institution poses to the financial system . . . Competent authorities to establish frequency and intensity of supervisory review and evaluation having regard to . . . systemic importance . . . where a review shows an institution may pose systemic risk . . . the competent authorities [should] inform EBA without delay*'.

Table 4.1 (continued)

Component	What EU Capital Requirements Directive specifies
	When assessing liquidity management, the competent authorities in one Member State shall '*consider the potential impact of their decisions on the stability of the financial system in all other Member States concerned*'.
Article 99	Supervisory examination programmes shall include the following institutions ' . . . *(b) institutions that pose systemic risk to the financial system*'
Article 103	Where institutions have similar risk profiles or pose similar risks to the financial system then competent authorities may apply the supervisory review and evaluation process (SREP) in a similar or identical manner
Article 105	Competent authorities to assess whether any imposition of a specific liquidity requirement is necessary to capture liquidity risks to which an institution is or might be exposed taking account of . . . '*(d) systemic liquidity risk that threatens the integrity of the financial markets of the Member State concerned*'
Article 114	Covers information requirements in an emergency situation which potentially jeopardises the market liquidity and stability of the financial system in any Member State where entities of a group have been authorised or it has significant branches
Article 116, 117	Decisions of (and information sharing by) consolidating supervisors should take account of . . . the potential impact on the stability of the financial system in the member states concerned
Article 128	'Systemic risk buffer' means the own funds that an institution is or may be required to maintain in accordance with Article 133. The 'combined buffer requirement' is the capital conservation buffer as per Article 129 extended by the G-SII buffer in accordance with Article 131(4) and O-SII buffer in accordance with Article 131(5) and by any institution-specific systemic risk buffer.
Article 129, 130	Small and medium-sized institutions can be exempted from capital conservation buffer and institution-specific countercyclical capital buffer requirements if such an exemption does not threaten financial stability
Article 131.1	Member states need to designate an authority in charge of identifying G-SIIs and O-SIIs
Article 131.2	O-SII identification to be based on: size, interconnectedness with the financial system, substitutability of services or infrastructure provided by the group, complexity, cross-border activity

(*continued*)

Table 4.1 (continued)

Component	What EU Capital Requirements Directive specifies
Article 131.3	O-SII identification to be based on: size, importance for economy of Union or relevant Member State, significance of cross-border activity and interconnectedness with the financial system. EBA to publish guidelines
Article 131.6	O-SII buffer should not entail disproportionate adverse effects on the whole or parts of the financial system of other Member states or of the Union as a whole or create an obstacle to the functioning of the internal market. O-SII buffer to be reviewed at least annually.
Article 131.9	Specification on subcategorisation of G-SIIs
Article 131.13–15	Specification of how some of the risk buffers work
Article 132	Commission to submit a report on possibility of extending framework for G-SIIs to other types of systemically important institutions.
Article 133, 134	Requirement to consider setting a systemic risk buffer (up to 5% of risk-weighted assets) for the financial sector or one or more subsets of that sector '*in order to prevent and mitigate long term non-cyclical systemic or macro-prudential risks not covered by [CRR]*'. Some requirements need to be met (more onerous if the buffer is to be > 3% of risk-weighted assets), e.g. if applies to other jurisdictions then a justification needs to be supplied for why none of the existing measures in CRD or CRR would be sufficient to address the identified macroprudential or systemic risk
Article 135	ESRB to provide guidance on setting countercyclical capital buffer rates
Article 136	Designated authorities to take account of ESRB guidance and other variables that the designated authority considers relevant for addressing cyclical systemic risk
Article 154	Competent authorities in one member state in general exercise of their duties to consider impact of their decisions on the stability of the financial system in all other member states

exposures, individual firms' risk management systems, etc.). These of course can also be subject to risk of failure, but we do not typically refer to these as examples of 'systemic risk' (except if their consequences are so large that they have implications for wider financial stability).

As we have already noted, there is a tendency amongst some commentators to equate a country's banking system with its financial system. The CRD does not support this stance. Its Article 51 differentiates between a member state's banking system and its (wider) financial system.

Table 4.2 References to systemic risk, financial stability or macroprudential matters in the CRR

Component	What EU Capital Requirements Regulation specifies
Recital (1)	Refers to G-20 Declaration of 2 April 2009 on Strengthening of the Financial System and its call for changes to regulation of banking systems
Recital (14) – (20), (31)	Refers to de Larosière report and need for establishment of a single rulebook and a European framework for macroprudential supervision. Also refers to a number of tools to prevent and mitigate macroprudential and systemic risks built into CRR and CRD including systemic risk buffers, refers to desirability of coordination through EBA and ESRB and gives member states powers to apply tools in relation to liquidity requirements and leverage ratios until harmonised.
Recital (22)	Review of macroprudential rules to justify how appropriate they are and how internationally agreed standards for systemically important institutions interacts with CRD or CRR
Recital (32)	Comprehensive reform needed to encourage economically useful banking activities, which means competent authorities should be allowed to impose higher capital requirements on systemically important institutions
Recital (87)	Higher own funds requirements for bilateral derivative contracts to reflect higher risk that such contracts pose to the financial system
Recital (100)	Institutions should hold a diversified buffer of liquid assets, as a concentration of assets and overreliance on market liquidity creates systemic risk
Recital (103)	Weaknesses in corporate governance contributed to systemic problems
Article 3	Definition of systemic risk
Article 11, 49, 395	Consolidation requirements to be without prejudice to effective supervision of whole or parts of the financial system
Article 99	Consolidated reporting requirements introduced to the extent necessary to gain a comprehensive view of the risk profile of an institution's activities and a view on the systemic risks posed by institutions to the financial sector or real economy
Article 441	Need to disclose values of indicators used for determining whether an institution should be deemed a G-SII
Article 458	Member states shall designate an authority to handle macroprudential or systemic risk (at a member state level) and it has some powers to vary a range of requirements (unless overridden by the Commission) including: (i) level of own funds, (ii) large exposure limits, (iii) public disclosure requirements, (iv) capital conservation buffer, (v) liquidity requirements, (vi) risk weights, (vii) intra financial sector exposures

(*continued*)

Table 4.2 (continued)

Component	What EU Capital Requirements Regulation specifies
Article 459	Commission empowered to adopt delegated acts that impose for a limited time stricter prudential requirements where necessary to address changes in intensity of microprudential and macroprudential risks
Article 513	A requirement on the Commission to review macroprudential rules contained in CRD and CRR are effective, efficient and transparent etc.

The global nature of the Basel Accords means that other major countries' primary legislation on bank capital adequacy have similar overall features to those in EU legislation, see e.g. Box 4.2 which describes the main elements of the US Dodd-Frank Act relevant to systemic risk

Box 4.2: The Dodd-Frank Act

The Dodd–Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank) is a US federal law which was passed in response to the 2007–09 Credit Crisis. It has brought the most significant changes to US financial regulation since the Great Depression. Reasonably similar rules have been introduced in the EU and elsewhere, but often via a wider range of legal instruments, e.g. in the EU similar market-wide requirements are being introduced via EMIR, MiFID II /MiFIR, AIFMD, CRD and wider financial stability requirements via e.g. creation of ESRB.

Dodd-Frank has proved controversial, with some critics arguing that it is not enough to prevent another financial crisis (and more bailouts) and other critics arguing that it has gone too far. A summary of its main components is set out in Table 4.3. Its stated aim is explicitly focused on financial stability:

> '*To promote the financial stability of the United States by improving accountability and transparency in the financial system, to end "too big to fail", to protect the American taxpayer by ending bailouts, to protect consumers from abusive financial services practices, and for other purposes*'

It increases the oversight of specific institutions deemed systemically important (both banks and non-banks) and brings more investment advisors, hedge funds and private equity funds more within the scope of financial services regulation. It has created some new institutions such as the Financial Stability Oversight Council (FSOC), the Office of Financial Research (OFR) and the Bureau of Consumer Financial Protection. The latter highlights the strong perceived link in politicians' minds between the Crisis and poor behaviour and incentivisation of the financial services industry in the run-up to the Crisis.

Table 4.3 Titles (sections) within the Dodd-Frank act

	Title	Comments (not exhaustive)
I	Financial Stability	Overarching macroprudential bias of the Act
II	Orderly Liquidation Authority	c.f. importance now placed on resolvability
III	Transfer of Powers to the Comptroller, the FDIC, and the Federal Reserve	Seeks to streamline banking regulation
IV	Regulation of Advisers to Hedge Funds and Others	Catches entities who did not previously need to register as advisors
V	Insurance	Reflects view that insurance can be systemically important following experience of AIG
VI	Improvements to Regulation	Limits bank proprietary trading
VII	Wall Street Transparency and Accountability	More regulation of derivatives markets, e.g. OTC swaps
VIII	Payment, Clearing and Settlement Supervision	Seeks to mitigate systemic risk in these areas
IX	Investor Protections and Improvements to the Regulation of Securities	Addresses e.g. powers and structure of SEC, regulation of credit rating organisations and client-broker interactions
X	Bureau of Consumer Financial Protection	Bureau will regulate consumer financial products and services
XI	Federal Reserve System Provisions Introduces some changes to how Fed operates	Introduces some changes to how Fed operates
XII	Improving Access to Mainstream Financial Institutions	Incentivises financial system participation for lower-income people
XIII	Pay It Back Act	Identified how Troubled Asset Relief Program (TARP) etc. would be unwound
XIV	Mortgage Reform and Anti-Predatory Lending Act	E.g. imposes obligations on mortgage originators versus borrowers
XV	Miscellaneous Provisions	Miscellaneous
XVI	Section 1256 Contracts	Refers to tax treatment of some futures contracts

Source: Nematrian.

4.2 Insurers

4.2.1 Introduction

Insurers insure their customers against risks. The risks that are potentially insurable are very wide and touch on many aspects of daily life. In some instances, you are required to take out insurance. For example, many jurisdictions require drivers to take out third-party motor insurance, to reimburse others if the driver causes an accident. There are clear public interest benefits in requiring individuals and firms to have certain types of insurance. Conversely, if for whatever reason the supply of such insurance becomes disrupted, this can cause material disruptions in the non-financial world, creating the possibility of contagion from within the financial system to the world beyond.

Insurers also play important roles in channelling savings into investments. Some types of insurance particularly focus on this role. For example, 'pure' unit-linked life insurance involves insurance policies where the proceeds of the policy depend on how assets associated with the policy perform. From an economic perspective these types of insurance products have strong similarities with investing in an investment fund (e.g. a mutual fund in the USA or a UCITS vehicle in the EU), see Section 4.4. The life insurance variant may offer tax advantages or other features not typically available from an investment in an investment fund, perhaps in return for some limitation on when the investor can withdraw funds from the policy. Or it may just be easier to sell to the relevant customers.

Other types of life insurance may include equivalent savings elements but with the addition of guarantees, to limit the downside investment risk to which the insured may be exposed. Such guarantees may become costly if interest rates fall, introducing a systematic exposure to this economic factor that may be shared by many insurers at the same time.

Even those insurers not focusing on savings-related products will still generally need to build up an investment portfolio to be reasonably confident of being able to meet future claim payments to their policyholders as these payments fall due.

The broad subdivision described above is also referred to as the distinction between protection-based insurance and savings-based insurance.

Protection-based insurance primarily seeks to protect the policyholder from some adverse risk that might happen. Nearly all non-life insurance (called property & casualty insurance in the USA), such as motor or household insurance, has this goal in mind. We buy household fire insurance because we are worried about the adverse financial consequences to us of our house burning down. Some life insurance is also primarily protection-based, e.g. insurance that for a modest premium will pay out a large sum if we die relatively soon after we take out the insurance. Again, it is addressing the worry that we might die in the prime of our life, creating subsequent financial hardship for our dependents.

In contrast, savings-based insurance primarily involves the accumulation of assets for savings purposes. This type of insurance competes with other ways individuals (and other organisations) might save for the future, e.g. by buying investment funds, putting money into deposit accounts, buying houses or other assets etc. Many types of life insurance are primarily savings-based in nature, even though originally life insurance primarily targeted protection needs.

This subdivision is rarely absolute. A major reason why many people save for the future is to provide a nest egg against the unexpected, so the motivation for saving often includes some element of 'protection'. This potentially gives insurers, particularly life insurers, a competitive advantage over most other providers of savings products since life insurance contracts can include both savings and protection elements simultaneously.

More usually, the broadest subdivision adopted by the insurance industry is between life insurance and non-life insurance. The legal, accounting and tax characteristics of the two differ significantly. In many jurisdictions firms (unless they are particularly old) can only be either a life insurer or a non-life insurer not both (although holding companies can typically own both types, so both can appear under the same brand). There is also a presupposition that losses you might protect against using non-life insurance are capped (to the total economic loss you might suffer if the risk in question happened). So, you can't insure a house for \$10m if it is only worth \$1m (or if you do the insurer would only be legally liable to pay out \$1m if it burned down). However, you can insure your own life for \$10m even if your current net assets are only \$1m (or even if they are zero or negative). If you then died (in a way that was within the scope of the policy), the insurer would be liable for the full \$10m irrespective of your net wealth at the time. In practical terms, this subdivision aligns substantially but not completely with the broader savings versus protection subdivision described above.

4.2.2 Can Insurers Generate or Amplify Systemic Risk?

In Table 4.4 we set out a comparison between the business models banks and insurers typically follow. The ones typically followed by insurers seem a far cry from the excesses in the banking sector most associated with the problems arising in the 2007–09 Credit Crisis. Insurers do not typically provide payment services. You cannot usually settle bills directly via your insurance policy.

Table 4.4 Comparison of typical banking and insurance business models

	Banks	Insurers
Monetary role industry mainly fulfils	A means of payment in exchange for goods and services	A store of value, permitting deferred consumption and smoothing
Other roles	Financial services	Risk pooling
Comparative advantage	Screen and finance short-term projects	(As investors) invest long-term and gain from illiquidity premium
Core business activities	Largely asset-driven, often supported by leveraged balance sheets	Mainly liability-driven, less leveraged and often less exposed to 'runs'
Exposure to systemic risk from any one firm?	Higher	Lower
Risk that safety net costs fall on government?	Higher (more 'essential' to current economic activity)	Lower

Moreover, commentators generally don't think that insurers were typically undercapitalised going into the 2007–09 Credit Crisis. Insurers didn't face the same pressure as banks in the immediate aftermath of the Crisis to increase their capital bases. This hasn't stopped regulatory change in relation to insurer capital adequacy from arriving in e.g. the EU in the form of Solvency II. However, we can view Solvency II as more part of the wider European 'journey' (i.e. promotion of the single market . . .) than a mechanism designed to increase insurers' capital adequacy per se.

Impavido et al. (2011) expand on Table 4.4 to draw out other differences. These differences include the typically different sources of capital of banks and insurers, the amount of capital they typically have (relative to their total assets and liabilities), accounting approaches used to assess their profitability and net assets and the extent to which they are typically constrained by explicit (Pillar 1) regulatory capital requirements. However, these differences can be overplayed since:

(a) There are some savings products each offers that are similar from an economic (rather than a legal) perspective. For example, both can offer investment related products such as (insurer-based) unit-linked savings products and (bank-based) structured notes with similar pay-offs. Banks also commonly offer term deposits but insurers can also offer term-certain annuities with similar economic characteristics.
(b) Some forms of protection products can also in effect be provided by either type of firm. For example, investment guarantees and options

written by investment banks can have similar economic characteristics to variable annuity options written by insurers. Trade finance offered by banks may look similar in substance to surety bonds offered by non-life insurers.

(c) Both types of firm are active in investment markets. Both may write or buy credit default swaps.

(d) Both may be subsidiaries of each other (or of holding companies spanning both sectors). This means that each is exposed to risk of contagion from the other sector.

Many in the insurance industry, e.g. Insurance Europe (2014), argue that insurers are quite different from banks and operate fundamentally different business models, partly based on the sorts of business analysis carried out in Table 4.4. To cater for cases like AIG that seem to contradict this logic, they typically limit their arguments to insurers that stick to traditional core insurance activities. AIG is an insurance holding company that suffered a near failure during the 2007–09 Credit Crisis, see Box 3.1. It is then perceived to have collapsed because it strayed from systemically low risk (insurance) activities into systemically high risk (banking) activities such as promoting and guaranteeing CDOs.

However, insurers can have central roles in other apparently critical functions provided by the financial system, including effective channelling of savings into investments and insuring against and dispersing risk. Moreover, in aggregate, insurers form a sizeable fraction of many economies' financial services industries. For example, French, Vital and Minot (2015) indicate that UK insurers had investment holdings valued at about 40% of the value of the assets held by UK banks as at end 2014. See also Fig. 2.1. This doesn't sound like an industry component likely to be unimportant in a financial system context!

Much of the debate about whether insurers can be systemically important ends up being a debate about how important (direct) interconnectedness is to systemic risk. This was one reason why we covered this topic in detail in a previous Chapter.

Insurers *are* typically less directly interconnected with other insurers (and banks) than banks are with other banks. But as we have seen previously, systemic risk isn't exclusively about (direct) interconnectedness. The most problematic situations may be when a sector is both (a) reasonably well interconnected and (b) exposed to common (perhaps unrecognised) vulnerabilities. It doesn't need to exhibit lots of interconnections within itself. Instead the S&L Crisis suggests that it is sufficient for there to be a concentration of interconnections to important players outside the specific sector in question. What we should arguably be focusing on is whether there are any underappreciated vulnerabilities within the insurance industry which could lead to widespread failures if they become uppermost in investors' and customers' perceptions.

4.2.3 Pandemics and Other Natural Disasters

It is not too difficult to identify possible ways in which large parts of both the life and non-life insurance industries could be hammered all at once, particularly if you are an aficionado of Hollywood disaster movies.

For example, a major pandemic, nuclear war or any number of other hopefully highly improbable but catastrophic events might result in very large numbers of individuals all claiming on their life insurance at the same time. Likewise, non-life insurers are in aggregate exposed to catastrophe risks. Natural disasters, such as earthquakes and hurricanes, do happen from time to time and if they strike economically important areas can cause large insured losses.

Any particularly large disaster of the sort referred to above would most likely also have other major impacts on the broader economy. The Black Death that struck Europe in medieval times had a major impact on labour markets and ultimately on political frameworks because it created a scarcity of workers.

The life insurance industry in aggregate has some 'natural hedges' that might mitigate its exposure to such risks. These include the annuity business that it writes. If lots of people died due to a pandemic then presumably so too would some annuitants, resulting in unexpected profits for the industry. Annuities involve payments to individuals until they die, so if a block of annuitants has a higher than expected mortality (also described as a lower than expected longevity) then this should be financially beneficial to the insurer. Also, many of the policies the life insurance industry writes involve little if any mortality risk (e.g. many types of unit-linked life insurance, if it is being sold primarily for savings purposes).

Turning to the non-life insurance /property & casualty insurance sector, insurers writing material amounts of natural catastrophe risk typically reinsure much of their more extreme exposures with reinsurers. The reinsurance industry is particularly global in nature, so most of these reinsurers end up with a portfolio of such risks that is quite diversified geographically.

Of course, a large enough natural catastrophe would wipe out the entire natural catastrophe insurance sector if it hit the 'wrong' location (or several of them simultaneously). This type of business tends to be concentrated in the developed world, especially USA, because this is where most of the world's economic capacity is based and because insured coverage is typically higher there.

But is it plausible for such losses to occur, or at least to be within the range of losses that might be deemed practically to fall within the scope of financial systemic risk? The issue here is that the more extreme the loss, if it is a natural catastrophe, the less likely it is that blame will accrue to the financial system. We are unlikely to blame the financial system for being poorly capitalised if humanity is all but wiped out by a large asteroid strike. The private sector already accepts that some types of catastrophe coverage are beyond its capacity, e.g. many risks relating to nuclear war.

Another feature of the non-life insurance industry (and particularly the reinsurance industry) is the so-called underwriting cycle. Immediately after a large loss, insurers and reinsurers are typically able to raise their rates, creating some spreading of losses across time (in addition to the spreading across geography created by reinsurance being a global business).

More to the point, such exposures are relatively well understood. Everyone knows that sometime there will be a major earthquake hitting Tokyo or San Francisco. We may debate the extent to which the insurance industry's capital base would be depleted by such an episode, but it is hardly new news that such exposures exist.

4.2.4 Underappreciated Risks

More systemically problematic are likely to be risks that are currently underappreciated. An example from the past would be losses that insurers incurred from asbestos. Asbestos used to be thought of as a wonder material particularly for buildings (e.g. because of its particularly good fireproofing characteristics). It was therefore widely incorporated in buildings of a certain age. It is now recognised to be a health hazard.

Apparently, many non-life insurers initially thought that the terms of product liability policies they had sold to building contractors and the like would largely exclude liability to rectify asbestos hazards or losses individuals suffered from being exposed to these hazards. They thought that the relevant policy terms would limit insured losses only to those where there was a clear, direct and time-limited link between the activity concerned or property insured and the eventual loss. However, the courts decided otherwise. Some insurers were left nursing very large losses. Asbestos losses nearly brought down Lloyds of London, one of the world's oldest reinsurance markets.

Similar existential worries have surfaced in relation to other societal developments that have scope to create long-term or large-scale losses. For example, Carney (2015) and PRA (2015) expressed the worry that financial system resilience to climate change issues could be impaired due to possible large liability claims that fossil fuel energy suppliers might face if someone managed to link future losses caused by climate change to the past behaviours of such firms, see Section 4.12. Other potentially major societal changes, such as the introduction of autonomous cars, present similar risks.

4.2.5 Underappreciated Concentrations of Risk

Returning to the Lloyds market, another challenge it faced in the 1980s was the so-called 'LMX spiral'. This involved different firms in the Lloyds market effectively reinsuring with each other through chains of reinsurance arrangements, using excess of loss contracts. Excess of loss contracts only pay out if losses exceed a certain amount. An insurer might have thought that it had reinsured away a given risk, only for the losses to reappear in its P&L account, if the loss was large enough to spiral back to it through these reinsurance chains.

Lessons like these highlight the potential systemic risks arising when contractual linkages are not transparent or involve long chains of financial intermediation, a lesson that the banking industry was also to learn to its cost during the 2007–09 Credit Crisis.

4.2.6 Historical Anecdotes

Given these conflicting lines of thought, policymakers may reach for their history books when deciding how systemically relevant is insurance. Insurers generally weathered the 2007–09 Credit Crisis relatively well (excluding AIG, credit monolines and a handful of other insurers), so the predisposition is to believe that they are less likely to be systemically important than banks. But policymakers can argue, with some justification, that this could be painting an overly rosy picture.

Moving away from situations where credit risk was the driver of failure, French, Vital and Minot (2015) also refer to the failure of HIH Insurance

Group in March 2001. Its failure severely disrupted the provision of Australian builders' warranty insurance for a period of almost a year in Australia. This type of insurance was mandatory for such builders. HIH had significantly under-priced risks in this market and by the time it failed held 90% of the relevant market. Competitors exited the market and potential new entrants were discouraged from entering it at the prevailing price level. Most insurance markets are not as concentrated as this, but HIH does highlight the potential issues that can arise if a significant insurer substantially under-prices an economically important risk category.

The Australian economy (particularly the component relating to Australian buildings) is a small part of a larger whole. Systemic issues associated with the HIH failure were an order of magnitude smaller than those faced by the banking industry or even just AIG during the 2007 – 09 Credit Crisis. At the time of its failure HIH had the equivalent of nearly £3 billion of assets. Contrast this with the up to US$85 billion secured credit facility the Fed set up to prevent AIG's collapse. AIG's Financial Products division had entered into credit default swaps to insure c. $441 billion of securities originally rated AAA. But does this make HIH unimportant in the context of systemic risk? As we have noted previously, the most appropriate geographical scope or scale over which to consider systemic risk is not clear.

Around the time that HIH failed, some life insurers were struggling with equity market declines in the wake of the dot com boom and subsequent bust. Giraldi et al. (2000) describe the plight of some European life insurers who were in financial jeopardy due to minimum guarantees offered to policyholders, reminding us that offering of guarantees by life insurers is not riskless.

Some of the complexities involved with guarantees are illustrated by the experience of some UK life insurers at around the same time. Many had written *with-profits contracts* (also called *participating contracts*) that involved policyholders receiving pay-outs that shared in market upside but included guaranteed floors in pay-outs if markets performed badly.

The actual amount policyholders typically received depended in part on the 'equity backing ratio', i.e. the proportion in equities of the assets backing these liabilities. As equities plummeted, the guaranteed floors became more onerous to honour. Some of the firms involved responded by reducing the proportions their with-profits funds invested in equities, mimicking the sort of dynamic hedging that might be used to hedge an option, see Box 4.3.

In some cases, the proportion was reduced to close to zero, thereby limiting the probable cost to the insurer but also largely eliminating scope for the policyholder to benefit from any subsequent equity market rally. This scenario is colloquially called a 'cash-out' event, since the contract remains on the books of the insurer but the policyholder has no realistic likelihood of sharing much in future profits arising from favourable returns on equities (as the fund no longer holds much of this asset class). Many of these policies were sold with an implicit expectation that they would share in a reasonable level of equity-market upside. Typically, the insurers did not 'fail' as such, as they were generally within their rights to make these sorts of adjustments to the equity backing ratios of their with-profits contracts. However, the experience led to loss of faith in the notion of with-profits life insurance amongst UK life insurance customers (which has not returned), and sales volumes plummeted.

One notable UK with-profits life insurer did in effect 'fail' at about that time, Equitable Life, but it faced some additional challenges, see Box 4.4.

Box 4.3 Dynamic hedging and portfolio insurance

Dynamic hedging is the process of dynamically altering a portfolio in a manner that aims as far as possible to hedge a non-linear position such as an option. The value (price) of, say, a call option varies non-linearly as the value (price) of its underlying (e.g. the S&P 500 index for an S&P 500 index option) varies. Dynamic hedging of such an (written) option position involves going long x units of the underlying for a given option position and going short cash (i.e. borrowing) by the amount needed for the hedge portfolio to have the same overall value as the option. x is the option delta, i.e. the slope of the tangent to the curve describing the value of the option, $V(S)$, as a function of the value of the underlying, S, in e.g. Fig. 4.1.

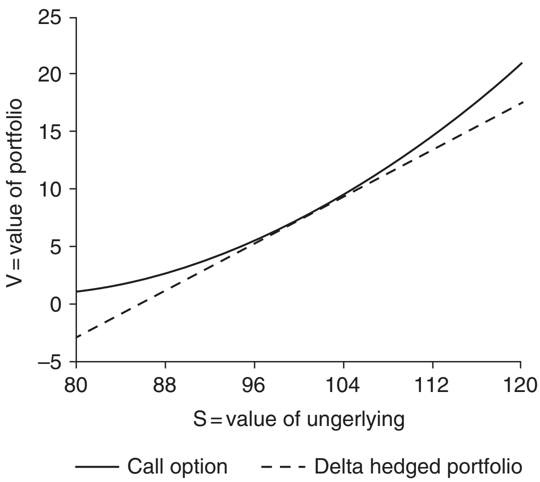


Fig. 4.1 Delta hedging of a call option

Source: Nematrian

All other things being equal, small movements in the price of the underlying result in the option value and the value of the delta hedge portfolio moving in tandem, since both have the same slope. However, the slope of $V(S)$ changes depending on S. To maintain the correct amount of hedging it is necessary to adjust the number of units of the underlying held by the hedge portfolio as S changes (hence the 'dynamic' within dynamic hedging). Dynamic hedging that aims to replicate the payoff of a call option involves buying as the market rises and selling as it falls.

Of course, not all other things are equal. For example, the option value and its delta depend on the implied volatility of the option. However, movements in the implied volatility in isolation do not alter the value of the hedge portfolio.

Dynamic hedging is extensively used by market makers to hedge net option exposures they may have in their trading books. Other institutions may also use dynamic hedging techniques (either explicitly or implicitly if they end up following similar 'buy as it rises, sell as it falls' strategies for whatever reason). For example, with-profits funds may reduce their equity content as the market falls, to make it more likely that they can honour guarantees they are providing.

Prior to the October 1987 Crash, a significant number of institutional investors utilised investment managers who were offering 'portfolio insurance'. This involved overlaying their existing portfolios with dynamic hedging strategies aiming to replicate option-like behaviour (capturing market upside if markets rose a long way, but with capped downside if they fell a long way). Amounts under management were sufficiently large to lead to self-sustaining falls in markets once markets had fallen sufficiently far, as the portfolio insurance algorithms tried to sell more and more of major equity market indices such as the S&P 500 and the FTSE 100 indices. Some commentators have blamed such effects for the October 1987 Crash. Arguably, the dynamic characteristics of the underlying algorithms used by such managers created an enormous crowded trade in the entire equity market.

Box 4.4: The UK Equitable Life

Founded in 1762, Equitable Life is perhaps the world's oldest mutual life insurance company. In its early days, it pioneered the use of age-dependent premium rates that depended on expected (age-dependent) mortality rates. The policyholders who collectively owned the insurer were the with-profit policyholders. As is typical for UK with-profits policies, policy proceeds from such policies involved a fixed sum to which were added bonuses based on the performance of the assets supporting the with-profits policies. Some of these bonuses were accrued each year (so-called reversionary bonuses) and some involved terminal or final bonuses added to policy proceeds only when the policyholder retired (for pension contracts).

From about the mid-1950s to the late-1980s, Equitable Life's with-profits pension policies also typically included guaranteed annuity rate (GAR) options. These allowed policyholders to convert their policy proceeds (including bonuses) into annuities using the better of a current annuity rate (based on then prevailing interest and mortality rates) and a GAR based on a fixed interest rate. If long-term interest rates fall then the value of this sort of guarantee rises.

From the 1980s onwards Equitable Life appears to have recognised the risk that its GARs could prove expensive to honour. However, it did not hedge or reinsure against this GAR risk, believing that it could neutralise the potential effect of its GARs by exercising discretion over how much terminal bonus it gave to different policyholders. In 1994 it started reducing the terminal bonuses it gave to policies with GARs but not to those without GARs. This differential treatment led to a significant number of complaints and Equitable Life sought a declaratory judgement from the courts confirming that it could exercise discretion in this manner. Lower courts came up with conflicting views but eventually in mid-2000 the UK's highest court ruled that it was unlawful for it to exercise discretion in this manner. In late 2000 it closed to new business, put itself into run-off and in effect wrote down policy pay-outs to below those that had had previously been promised, after failing to find a buyer who would take on its liabilities.

Its (effective) failure was therefore only partly a consequence of falling interest rates. A major contributory factor was a failure to hedge the risks involved when the cost of doing so would have been lower, which in turn derived from a failure to interpret correctly what would eventually be deemed to be the fair way to treat customers with different sorts of liabilities.

4.2.7 What Causes Insurers to Fail?

S&P (2013a) carried out an analysis of insurance company failures since the 1980s and explored the factors that seemed to have contributed most to these failures. The authors concluded that prevalent factors, often reinforcing each other included:

(a) Poor liquidity management;
(b) Under-pricing and under-reserving;
(c) An unduly high tolerance for investment risk;
(d) Management and governance issues;
(e) Difficulties related to rapid growth and/or expansion into non-core activities; and
(f) Sovereign-related risks.

Most of these seem at first sight to be company-specific factors, which might suggest that that when insurers fail they typically do so in an isolated manner, without too much systemic spill-over. However, looking in more detail at these failures there do seem to be some commonly shared causes of a more systemic nature that have contributed to several of these failures. These include:

(a) *Challenges relating to the identification of the 'true' underlying financial position of the insurer*

Some insurers have very long-term liabilities which they match with long-term assets. It is intrinsically challenging to assess how robust an organisation is if its assets and liabilities are both long-term and potentially relatively illiquid. Regulatory and accounting frameworks do not always make it easy for outsiders to form a robust view on such matters (or even sometimes for insiders to do so either). From time to time this seems to result in the hiding of excessive risk taking until it is too late for the firm to respond. Any such instances tend to affect multiple companies at the same time.

Several of the failures of Japanese insurers in the 1990s (following the collapse of the Japanese equity market and Japanese bond yields) seem to have been strongly exacerbated by such factors. In the UK in the 1990s some life insurers who had previously included guaranteed annuity rate (GAR) options in policy terms thinking that they were largely worthless discovered to their cost that they had under-reserved for such options when they became more onerous to honour. GAR options allow policyholders to convert accumulated policy funds into annuities at predetermined terms irrespective of the then prevailing cost of buying an annuity. Their values can be quite sensitive to levels of bond yields.

Banks undertaking particularly extensive maturity transformation can also have very long term asset or liability cash flows, so can face similar challenges.

(b) *The broader desire of life insurers in some markets to provide savings products with guarantees*

These products appeal to customers but carry the risk that such guarantees can become onerous to honour in adverse market circumstances.

Some of the drive towards more 'market consistent' based assessments of insurer assets and liabilities epitomised by e.g. Solvency II reflects prior experience regarding how difficult it can be to gain a true understanding of an insurer's balance sheet if other valuation approaches are adopted.

4.2.8 Non-Traditional and Non-Insurance (NTNI) Activities

Many global insurance regulators (and most industry participants) seem to believe that in most cases insurers are unlikely to contribute materially to systemic risk. IAIS (2015) was the first in a series of consultation documents published by the IAIS on the Higher Loss Absorbency (HLA) element of capital requirements for Global Systemically Important Insurers (G-SIIs). It adopted the same working definition of systemic risk as that proposed by IMF (2009) and then summarised the IAIS's position (at that time) on the role that insurers might play in relation to systemic risk. It didn't deny the relevance of insurers to systemic risk but tended to downplay their contribution. It argued that:

- *There is little evidence of traditional insurance either generating or amplifying systemic risk.*
- *NT [non-traditional] and NI [non-insurance] activities within insurance firms or groups may generate or amplify systemic risk.*
- *The insurance sector is susceptible to systemic risk generated in (or transmitted through) other parts of the financial sector.*

Possible weaknesses in these arguments include:

(a) The definition of 'traditional insurance business' is somewhat circular and could perhaps be rephrased as 'insurance business that does not share the systemic risk characteristics of banking'. The systemic risk characteristics of banks are in this context seen as very much tied in with their interconnectedness, so again the logic that is being propounded is that insurers are not typically systemically risky because they are not typically very interconnected.

(b) There are two sides to an insurer's (or indeed any other institution's) balance sheet, i.e. the asset side and the liability side. The definitions of non-traditional (NT) and non-insurance (NI) used by the IAIS, collectively shortened to 'NTNI', only really relate to the liability side of the balance sheet. Maybe insurers can have the equivalent of NTNI on their asset side, e.g. advancing loans and behaving like banks without this sort of activity being obvious from their liabilities.

Those on the industry side of the fence sometimes profess difficulty in understanding exactly how NTNI should be interpreted and why it should be of relevance. Their worry appears to be that the concept may have been introduced primarily for political reasons rather than on solid theoretical grounds.

Who is right? Whilst both perspectives have some merit, it seems to me that both are also missing something. We've discussed earlier that major systemic risk events tend to involve sector-wide weaknesses that apply to many different entities. NTNI focuses on connectedness, of which there does generally appear to need to be some to provide a catalyst for a systemic risk event to materialise. But as we have previously argued, our primary attention should probably be on whether there are sector-wide vulnerabilities which an over-focus on NTNI may miss.

What we can say with a fair degree of certainty is that having any firms in a financial sector deemed globally systemically important has wide ranging impacts on the structure of capital regulation in that sector. The decision to deem some insurers to be G-SIFIs has led to an inexorable chain of logic that is likely in due course to rewrite global insurance capital regulation, making it more harmonised across jurisdictions, see Section 4.2.9 and Box 6.2.

4.2.9 Identifying Systemically Important Insurers

Global insurance regulators (and the FSB) have concluded that some insurers are systemically important financial institutions. How NTNI should be defined then becomes particularly important. As in the banking sector, whether an insurer is deemed systemically important involves several steps that combine quantitative and qualitative assessment. There are also several checks and balances, given the potentially contentious nature of the decision. IAIS (2015b) contains proposals to alter the methodology used to identify which insurers should be G-SIFIs. However, NTNI is still to be given 45% weight when categorising the extent to which any given insurer is 'systemic', see Table 4.5. Also noteworthy is the extent to which interconnectedness is viewed as primarily linked to 'direct' interconnectedness, apart from the contribution from Level 3 asset exposure (i.e. the amount the insurer has invested in hard to value assets, see Section 3.4.2).

In the banking world, it is sometimes argued that being classified as a G-SIFI is not particularly undesirable. Such firms are subject to more onerous regulation and to higher capital requirements (e.g. TLAC), but conversely may benefit from a more explicit recognition that they are probably too big to fail. For insurers deemed G-SIFIs this benefit is tenuous and rarely put forward by commentators as recompense for the likely higher regulatory burdens and capital requirements. This has possibly incentivised

Table 4.5 Factors proposed for identifying globally systemically important insurers

Category	Overall category weighting (%)	Individual indicator and contribution
Size	5	Total assets (2.5%), Total revenues (2.5%)
Global Activity	5	Revenues derived outside of home country (2.5%), Number of countries (2.5%)
Interconnectedness	40	Intra-financial assets (6.7%), Intra-financial liabilities (6.7%), Reinsurance* (6.7%), Derivatives (6.7%), Turnover (6.7%), Level 3 assets (6.7%)
NTNI	45	Non-policy holder liabilities and noninsurance revenues (7.5%), Derivatives trading* (7.5%), Short term funding (7.5%), Financial guarantees* (7.5%), Minimum guarantee on variable insurance products (7.5%), Liability liquidity (7.5%)
Substitutability	5	Premiums for specific business lines (5%)

Source: Nematrian and IAIS (2015b). Most of the indicators are expressed relative to other insurers, but some noted with an asterisk use absolute reference values.

insurers large enough to be deemed G-SIFIs to explore further their business models. Some systemically important insurers appear to have gone to some lengths to get themselves removed from the G-SIFI list. For example, Jenkins (2016) describes how Italy's Generali managed to get itself removed from the G-SIFI list (and replaced by Aegon) by downsizing itself.

4.2.10 Interest Rate Vulnerabilities

Interest rate risk relates to a firm's exposure to industry wide interest rate levels, e.g. yields available on government bonds or other reference rates such as LIBOR swap rates. It is not generally deemed to relate to divergences between these reference rates and the yields available if you lend to (or invest in the paper of) a risky borrower. The difference is the credit spread for that borrower.

Interest rate exposures that firms have do not, therefore, in the main correlate with how interconnected firms might be. This is a type of risk that could in principle affect large parts of the any part of the financial system even if none of them transacted with each other. The 'connection' would be indirect, driven by a common exposure shared across the relevant cohort rather than by how individual cohort members are linked to each other.

Arguably, the low interest rate environment prevailing in many developed economies may have created such a common indirect exposure, particularly for insurers (and pension funds, see Section 4.3.9). Take for example the

Eurozone. In certain EU member states within the Eurozone, such as in Germany, many common types of life insurance policy include long-term guarantees. Amounts that policyholders will have returned to them when their policy matures may be subject to a lower limit. The lower interest rates are, the more onerous it is to honour these guarantees. Interest rates in the Eurozone (and yields on longer term benchmark bonds denominated in euros) are at or close to multi-decade lows. This is likely to have strained the finances of some insurers. Similar concerns have been raised concerning some US insurers, see e.g. Gray (2016).

Competitive pressures may limit the practical ability of firms to reduce these guarantees for new business. Usually existing policyholders will have a contractual right to these guarantees that cannot be broken without the firm defaulting. Even if some firms in a sector are minded to reduce their exposures to guarantees it may still not alter the aggregate profile of the whole sector if new entrants take up the slack.

The Japanese experience is discouraging. Japanese insurers have had to face low interest rates for significantly longer than their compatriots in Europe or USA. Several failed in the late 1990s after Yen interest rates had declined significantly. Those in Europe hoping for some respite from current low interest rates may also be dismayed by how long they have persisted in Japan.

4.2.11 Regulatory Factors Influencing Observed Interest Rate Vulnerabilities

Several factors make it complicated to assess how exposed the life insurance industry (and specific players within it) are to low interest rates. Whether a firm will appear stressed depends on how its financial position is measured.

The EU has recently introduced a new insurance regulatory framework called Solvency II, see Box 4.5. Solvency II seeks in broad terms to adopt a 'market consistent' valuation approach. This seeks to place values on assets and liabilities that are aligned with their 'fair' value as measured by the price at which they might trade in an open market between willing buyers and willing sellers.

That's the theory. The practice is a little more complicated, since no open market exists for many of the long-term illiquid liabilities that insurers owe to policyholders). Solvency II is more market consistent than the patchwork of rules that previously applied in different EU jurisdictions. However, there are still elements of Solvency II that some commentators consider only to be loosely market consistent or appear to have been introduced deliberately to diverge from a full market consistent basis as part of the political process that preceded its eventual adoption, see Box 4.5.

Box 4.5: Solvency II and its long-term guarantees (LTG) measures

Solvency II is the EU-wide regulatory framework for insurers introduced on 1 January 2016, see European Union (2014b) and European Union (2014c). It has three Pillars as per Fig. 3.2. It is more harmonised across the EU than the previous patchwork of rules known colloquially as 'Solvency I'. In most cases assets and liabilities are valued in a market consistent fashion. For most assets and liabilities that can be replicated by assets traded on deep, liquid and transparent markets, the valuation adopted is based on their market value. For less liquid insurance liabilities (also called 'technical provisions') the valuation is based on a 'best estimate' estimate of the liabilities in question, added to which is a 'risk margin' that is designed to reflect the additional (market) costs the insurer might expect to incur were it to try to settle these liabilities onto another party. Risk capital requirements are derived in the main by stressing the net balance sheet (if the firm is using a standard formula approach) or via an internal model (if the firm has agreed one with its supervisor).

As part of the political compromises involved in everyone agreeing to go ahead with Solvency II, some elements were included that are designed to cater for insurance contracts that include long-term guarantees. These are called the LTG measures.

Some of the LTG measures have an indefinite life, including the matching adjustment (MA) and the volatility adjustment (VA). Others involve transitional arrangements with a finite lifetime.

The MA and VA in effect target long term life insurance liabilities which customers cannot liquidate over the short term. If a life insurer can ring fence a portfolio of assets that will generate cash flows sufficiently like those needed to pay such liabilities as they fall due, then the matching adjustment allows these assets and liabilities to be valued as a combined structure. This allows the insurer to pay less attention to general market movements over the lifetime of the combination.

One argument put forward for the MA is that it potentially facilitates socially beneficial life insurer investment in long term less liquid assets (such as some types of physical infrastructure, e.g. bridges, roads, railways and airports). It does this by creating an initial solvency uplift and a more stable trajectory thereafter (if we assume that some of the yield premium available from such assets derives from their illiquid nature and therefore eventually flows through to the insurer). However, if for some reason the ring fencing becomes unsustainable (e.g. the firm becomes a forced seller because its own business model has for some reason fallen apart) then the short term gain to the firm can correspond with longer term pain. From a systemic risk perspective, the LTG measures can probably be expected to mitigate outcomes if the stress is not too large but may amplify them if the underlying stress is more extreme.

The VA provides a similar type of capital uplift (although usually lesser in terms of quantum of uplift) that does not involve the same level of ring-fencing of assets as the MA. Both ordinarily require prior approval from the supervisor.

Some of the transitional arrangements in the LTG package allow firms to phase in the change from Solvency I to Solvency II over an extended period, although the impact of these transitional arrangements will generally need to be publicly disclosed. We therefore return to debates about how relevant material is if analysts have access to additional material that might be more relevant to their analyses.

Not included in the LTG measures but of relevance to likely vulnerabilities of insurers to a low interest rate scenario is a specific element of Solvency II called the *ultimate forward rate* (UFR).

Given limitations in observable market instruments it is not practical to derive a robust market consistent valuation for liabilities that are particularly long dated. However, in some markets, such liabilities form a material fraction of all life insurance liabilities. Regulators under Solvency II solved this conundrum by developing the UFR, a figure selected by the regulators to represent the discount rate applicable to liabilities with terms far beyond what can be observed in the market. The liability valuations will also be sensitive to how slowly the yield curve extrapolates towards this UFR from the so-called last liquid point, i.e. the point along the yield curve designated to be where a more active market in financial instruments ceases to apply.

By its very nature, the UFR is impossible to estimate precisely. Nevertheless, there is a risk that the UFR rate originally adopted under Solvency II (a fixed 4.2% pa) is too high. At the time of writing, charts of the level of the Solvency II discount curve derived from a UFR of 4.2% looked implausibly favourable (particularly if it is considered likely that interest rates will stay low for a protracted period). IAIS has adopted in its field testing for its proposed global Insurance Capital Standard (ICS) a lower UFR of 3.5% for the Eurozone.

At the time of writing (June 2016), EIOPA were consulting on the methodology used to set the UFR, see EIOPA (2016a). They have proposed replacing the fixed UFR (of 4.2% pa) with one that responds relatively slowly to new market data (i.e. gets updated slowly as real returns change). This type of approach is one that the Dutch regulator has adopted when supervising pension funds within its jurisdiction. The Dutch UFR (for pension funds) started its trajectory at a lower level than the Solvency II one.

4.2.12 Resolution of Insurance Companies

If an insurer does get into trouble then, as with banks, an important part of the toolkit available to regulators is the ability to ‘resolve’ companies if necessary.

Insurer resolution regimes vary considerably across jurisdictions. For example, each US state has separate bankruptcy codes and other legislative structures that interact with any resolution process as explained in e.g. NOLGHA (2011). Compared to banks, insurer resolutions tend to be long-term affairs, given the typical nature of insurance liabilities.

Different rules apply in the EU. Some policymakers there have been pressing for rules that are more harmonised across member states, in response to worries that some insurers in some member states could find the current low interest rate environment a challenge, particularly if it persists for a long period.

4.3 Pension Funds

4.3.1 Introduction

Pension systems play very important roles in the economies and social frameworks of most developed economies. They provide a means for individuals to defer consumption of current income to meet financial needs in retirement, often many years into the future. They may also provide welfare safety nets for poorer members of society in conjunction with other social security mechanisms.

Given their size and importance in some jurisdictions, see Fig. 2.1, it is reasonable to expect macroprudential regulatory bodies to explore potential systemic risks and vulnerabilities within pension systems. However, the long timescales over which pension arrangements operate mean that any such vulnerabilities may take much longer to crystallise than for most other institutions. Indeed, their long-term focus may offset vulnerabilities present elsewhere in the financial system, to the extent that the pension system does not itself have underappreciated vulnerabilities. The usual view of policymakers concerning pension funds is epitomised by an Annex to FSB (2016) which includes the sentence: '*Pension funds generally have long-term investment horizons and make a positive contribution to financial stability*'.

There is considerable debate over how much of the pension system practically lies within the 'financial' system and hence within the deemed scope of 'financial' systemic risk. Some systemic risks and vulnerabilities that exist within pension systems may be difficult or impossible to tackle using macroprudential tools currently available elsewhere in the financial system and may need broader societal engagement to be tackled effectively.

Conversely, the assumption that the pension system should be viewed as 'separate' to the rest of the financial system is possibly an example of 'compartmentalisation', a behavioural bias that in other contexts can prove unsustainable in extreme enough circumstances, see Box 3.11.

4.3.2 Overall Structure of Pension Systems

There are considerable differences in how pension systems are structured and operate in different countries. Even within otherwise relatively homogeneous parts of the globe, such as the EU, substantial differences exist. These include differences in the mix between state and private pension provision, between funded and unfunded arrangements, between defined benefit (DB) and defined contribution (DC) arrangements and in terms of the prevalence of occupational versus other sorts of pensions.

Nearly everyone who reaches a certain age (or is expecting to do so in the future, i.e. nearly everyone!) is touched by pensions in some shape or form. Differences between jurisdictions are indicative of the heterogeneity of societies themselves and in their attitudes towards welfare provision, social security and public versus private provision of services etc.

It is conventional to subdivide a country's pension system into three parts or 'pillars' (which are quite different in nature to the three pillars present in modern regulatory frameworks). In this context, Pillar 1 refers to state organised arrangements, Pillar 2 refers to occupational arrangements and Pillar 3 refers to any supplementary arrangements individuals may have that are identifiably pension-orientated in nature. Wealthier individuals save for retirement in all sorts of ways, e.g. by investing in a house that they expect to sell in due course to help fund their retirement (and which they may rent out in the meantime). This means that there is a grey area between Pillar 3 pension arrangements and other longer-term savings that individuals may make during their working lives.

Each of these Pillars comes in many different forms:

(a) <u>Pillar 1 (state organised)</u>

Pillar 1 pension arrangements may include a flat-rate 'universal' basic state pension payable by the state (or some state-sponsored body, such as a specified social security programme) to everyone over a certain age. They may also include elements that depend on the level of social security contributions a participant has paid during his or her working life.

Almost all developed economies now have some form of social security system. Usually this includes a state pension paid to those who have retired, although there may also be other supplementary income elements (perhaps means-tested) that complement such pension payments for those who have few other sources of income in retirement. Pillar 1 arrangements tend to be the most universal form of pension provision in a country, but often provide only a relatively modest fraction of total income for those in retirement, at least for those wealthy enough not to be totally dependent on social security. Often Pillar 1 arrangements are largely or wholly unfunded, with payment of the pension ultimately being a liability of the state in question, met from (future) tax revenues and/or from (future) contributions of (other) individuals covered by the relevant social security programme.

(b) Pillar 2 (occupational)

Pillar 2 pension arrangements can be organised by an individual's employer or centrally on behalf of a class of employers. By the time an individual retires, he or she can belong to several such arrangements depending on his or her work history.

Pillar 2 arrangements come in two main forms:

- *defined benefit* (DB). Individuals receive specified benefits usually linked to the amount of time that they worked for a given employer and their salary history when doing so; and
- *defined contribution* (DC). Individuals receive whatever benefits can be purchased by the accumulated pot of contributions made by them or on their behalf.

Some countries also have collective defined contribution (CDC) arrangements, in which there is e.g. some specified investment risk sharing between participants but collectively the output for members of the whole arrangement depends on the investment returns achieved on the totality of contributions to the arrangement.

The benefits promised by DB pension arrangements come in many different forms. Pension arrangements in some countries typically focus on an income in retirement. In other countries, they may target a lump sum payable at retirement, perhaps with the individual then buying an annuity with the proceeds if he or she desires. Most include some implicit or explicit element of inflation protection, to protect the real value of any income being provided in retirement. For the minority of individuals who stay with the same employer throughout their working life, some arrangements focus on salary at or close to retirement whilst others focus on salary averaged across the member's working life. Nowadays it is rarer for individuals to work for the same employer throughout their working life so it becomes more important to consider what happens to any defined benefits between the data of leaving service and the date when the benefit is eventually taken.

DC arrangements are essentially always funded, in the sense that a specific pool of investments will have been set aside to help honour the promises as they fall due. DB arrangements are often but not always at least partially funded although the extent to which the accumulated funds are explicitly ring-fenced from the employer varies by country (and by whether the employer is in the public rather than private sector). A common way of providing private sector occupational pension provision in Germany and Austria is to establish book reserves that remain within the employer's balance sheet. In the USA, UK and Ireland it is more common for any accumulated pool of assets to be segregated from the employer's balance sheet in the form of a separately invested pension fund. In countries where Pillar 1 arrangements are unfunded it is also common for Pillar 2 public sector pension arrangements to be unfunded for some but not necessarily all state employees.

(c) Pillar 3 (supplementary)

Pillar 3 pension arrangements are nearly always DC in nature and can take a range of legal forms, often involving some form of individually arranged insurance coverage. They may benefit from specific tax privileges to encourage individuals to save for retirement (reducing the likelihood that the individuals concerned will ultimately prove a burden to the state in retirement). There may then be penalties if the individual seeks to draw on his or her pension savings before retirement. In some countries, a high proportion of the total assets managed by life insurers relate to pensions in some shape or form. If savings have no specific pension-related advantages then they are usually deemed to fall outside Pillar 3 although the individual concerned may still consider them to be effectively for retirement purposes.

The exact types of pension arrangement that are most common varies considerably by jurisdiction, as do any associated tax privileges and regulatory frameworks etc. To illustrate these variations, we analyse one pension system in more detail, the UK's pension system, in Box 4.6. Often a variety of approaches can be seen in the same jurisdiction, perhaps because of repeated shifts in the relevant government's policy on pensions. The boundaries between the three Pillars mentioned above (and between them and just saving for a rainy day) can be blurred. For most non-pension experts, pensions arrangements and choices can seem complex and difficult to follow. Added to this, for younger people pensions may seem too far into the future to worry about, but for older people, too immediate to do much about.

Box 4.6: The UK pension 'system'

The UK has a state pension arrangement. As in many other countries facing an ageing population the age at which this pension can be taken is gradually increasing and the proportion of an average employee's income that the state pension arrangement can be expected to replace is declining (although the time-scales involved are reasonably glacial as far as most individuals are concerned).

The UK used to have a strong defined benefit occupational pensions base, typified by final salary pension funds. These involved employers (and to some extent employees) setting aside contributions into separately funded pension arrangements, i.e. occupational pension schemes. In return, employees received the promise of future pension benefits linked to their final salary (plus some inflation uplifting between date of leaving and date of retirement and thereafter). Often the benefit was calculated using a formula like (N/60) × final salary where N is the number of years that the employee worked for the relevant employer.

More recently, the high and potentially volatile cost of such arrangements have come to the fore in the minds of company finance directors. Many of these schemes have been closed to new entrants and quite often also to future accrual for existing members. New employees (and existing employees, if applicable) have then been moved into DC pension arrangements instead.

In EU-speak, these pension schemes are examples of Institutions for Occupational Retirement Provision (IORPs). Most occupational pension provision in the UK has traditionally taken place via IORPs. Typically, an employer wanting to offer such provision established one or more trust based IORPs and its employees then joined these IORPs. The UK has both defined benefit (DB) and defined contribution (DC) IORPs. Most involve a single employer (perhaps though multiple subsidiaries within the same corporate group). Multi-employer, e.g. sector-wide, private sector schemes do exist but they are relatively uncommon. Hybrid schemes in which benefits have both DC and DB elements have become increasingly common as employers have established within the same IORP separate DC sections for future benefit accrual alongside previously established DB sections.

Individuals can also take out DC personal pension arrangements with life insurers. Some employers help their employees collectively save for retirement by setting up group personal pension arrangements which legally involve life insurance contracts between the individual members and specific insurers. According to ABI data, see ABI (2013), in 2013 there were in the UK more assets in either insurer-administered occupational pensions or insurer-administered individual pensions than in all other types of (life) insurance contract combined, highlighting the importance of third pillar pension arrangements to the UK life insurance industry. It is not necessary to use an insurance company to provide these sorts of pensions although insurance company policies are a common approach in the UK. The quantity of assets involved is, however, still less than, say, the amount of money the UK public has invested in private property. Many individuals in the UK also save for their longer-term future (taking a broad view of savings), in ways that do not involve any specific pension arrangements.

UK private sector DB IORPs are required to join to the UK's Pension Protection Fund (PPF), which is the UK's industry-wide pension protection scheme (PPS). Figures quoted below are sourced from PPF (2014), which covers over 99% of such IORPs. As at 31 March 2014, there were around 6,100 such IORPS in the UK, covering around 11.1 million memberships. A 'membership' is one individual's participation in one scheme. One individual can have multiple memberships (e.g. if he or she has changed employer) so the total number of memberships exceeds the total number of individuals covered. So, although 11.1 million sounds a lot, it covers only a minority of the UK working population. UK private sector DB IORPs exhibit a significant bias towards manufacturing. Their glory days coincided with a time when the UK's manufacturing base was larger than it is now. Only around 13% of these IORPs were open to new members in March 2014. The percentage closed to future accrual of benefits even for existing members has been rising since 2006 and had reached c. 32% by March 2014. Many memberships now relate to pensioners or deferred pensioners (i.e. individuals who have left the employment to which the benefits relate but have yet to start receiving a pension). In March 2014, there were around 1.8 million active members in schemes covered by the PPF, a figure that had fallen by c. 50% since 2006.

As UK DB IORP schemes have matured the proportion of their assets invested in equities has fallen, from 61% in 2006 to 35% in 2014. Over the same period, gilts and fixed interest assets have risen from 28% to 45%. Around 41% of the proportion invested in gilts and fixed interest assets is invested in inflation-linked securities. This reflects the high proportion of pension liabilities of these IORPs that are inflation-linked in nature. Commonly, some level of inflation-linking applies to promised benefits both before and after retirement in the UK.

4.3.3 The Social Nature of Pensions

Ultimately (at least for a closed economy), pensioners are only able to consume goods and services that derive from the productive endeavours of those who are currently working. This is true whatever Pillar of the pension system we are considering and whether the arrangement is funded or unfunded. In some sense, the return we accord to savings is linked to how we apportion current productive output between different members of society, especially between those currently active in the workforce and those who were previously working but have now retired. Pension systems thus need a significant element of social cohesion to function effectively given the timescales over which they operate.

Universal pension systems are comparatively modern inventions.[2] Germany was the first nation in the world to adopt an old-age social insurance program in 1889, designed by Germany's then Chancellor, Otto von Bismarck. His

[2] Pension arrangements for specific sections of society appear to have a longer history. Economist (2016a) notes that Imperial Rome effectively had a pension system for members of its army. Emperor Augustus came to power with the help of a private army and was keen to retain the loyalty of its soldiers. The arrangement involved giving soldiers who served for 16 years (later 20 years) the equivalent in cash or land of 12 times their annual salary. Servicing this pension promise was very expensive; military wages and pensions apparently absorbed half of Imperial Rome's tax revenues.

motivations included promoting the well-being of workers, keeping the German economy operating at maximum efficiency and staving-off calls for more radical socialist alternatives. Occupational pensions followed in many countries in the 20th century, originally to reward long serving employees and incentivise others to stay for a long time too, but then more generally becoming seen as another benefit accruing to the individual whilst in employment. As countries became wealthier, interest in individual or supplementary pension arrangements also increased.

The close link between pensions and the broader economy has several consequences relevant for (financial) systemic risk, including:

(a) The 'pension' system may be viewed by some as having little in common with the 'financial' system (which proponents of this view will typically interpret as being essentially the same as the 'banking' system). A corollary is that such commentators may also question whether macroprudential bodies as typically structured should have meaningful oversight over the pensions sector.
(b) Legislative frameworks governing pension arrangements tend to be much more nationally orientated than those applicable to other parts of the financial system. For example, in the EU, the design of pension arrangements is largely within the remit of individual EU member states, although prudential aspects of funded arrangements are to some extent within the scope of relevant EU-wide regulatory bodies (i.e. EIOPA). There are multinational bodies that coordinate regulatory activities in the pensions space but they do not currently appear to have as much influence as e.g. the BCBS dealing with banking supervision or the IAIS for insurance.
(c) Different countries have developed different ways of providing additional benefit security to members of (occupational) pension arrangements, see Section 4.3.6.
(d) If pension arrangements in a country are collectively short of funds then it is often assumed by those within the industry that there will be a collective response to addressing these shortfalls.

4.3.4 The Economic Nature of Pension Liabilities

One systemic risk that exists with pension systems is that the assumed social nature of pensions and the willingness of different parts of society to work together to address any collective shortfalls as per 4.3.3(d) may prove illusory.

Take for example the UK DB pension fund system. At the time of writing it appears to be materially 'underfunded', by which we mean that there are insufficient assets already set aside in UK pension funds to provide for benefits already accrued to date in them, see Box 4.7. As is also explained in Box 4.7, if a UK DB pension fund is wound up then any deficits between what is currently available in the fund and what it would cost to get a third party (an insurer) to provide the benefits is met by the sponsoring employer. If the sponsor cannot afford to do so and itself defaults then the UK's Pension Protection Fund takes over responsibility for some fraction of previously accrued benefits up to a specified level.

Currently, many sponsors cannot fully address shortfalls quickly within funds they have sponsored. In some cases, the shortfalls are large relative to annual cash flow sponsors might realistically be able to set aside from current profits to inject into their schemes. Some commentators have proposed radical ways of addressing these burdens including:

(1) Reducing outright the quantum of benefits members are entitled to, to a level more commensurate with the funds available within a given pension fund.
(2) Reducing the level of indexation that benefits receive once in payment (or in in 'deferment', i.e. between the date a member leaves the scheme and the date the member starts receiving benefits from it). Lower levels of indexation are less costly, and would thus reduce the level of under-funding. This is called 'conditional indexation'. Many larger Dutch pension funds were set up with such benefit structures, which provide a safety valve allowing them to reduce target benefit levels when their funding levels are low.

Both strategies can be thought of as an example of applying Section 4.3.3(d) in practice. What commentators proposing either of these strategies seem to gloss over is that both in effect also involve the scheme and/or employer defaulting on pension promises that have already been given to members.

Looking forwards, it is hard not to be gloomy about the health of some parts of the (UK) pension system, since increased life expectancy coupled with low interest rates forms a perfect storm for some funds. Authors such as Harrison and Blake (2015) writing about the DB part of the UK occupational pension system express this sentiment but also seem to assume that a realistic solution is for members to accept having their benefits cut.

It is tricky to see why members (and hence politicians on their behalf) should be keen to countenance such strategies, even if it provides some respite for the corporates concerned. During the 2007–09 Credit Crisis (and the subsequent Eurozone sovereign debt crisis) individuals who had money in bank accounts took a dim view of being told that a haircut was to be applied to assets in their bank accounts. Bailing in of individuals who had not previously expected to suffer in this manner proved fraught with political difficulties in the banking sector. Why should bailing in members of under-funded pension schemes be different?

There is also a broader angle that this raises for financial stability. We mentioned earlier that the usual view of policymakers is that pension funds have long-term investment horizons and probably therefore make a positive contribution to financial stability. But how true is this?

Going back to when (UK) pension arrangements were younger, inflation was higher than at present and the level of guaranteed benefits was lower. DB pension funds did not necessarily guarantee any level of inflation indexation above 0% pa and if individuals left the scheme before retirement they were not necessarily entitled to anything. This meant that these funds were long-term largely unconstrained investors able to adopt contrarian stances if they thought fit. They could largely ignore the precise nature of their guaranteed liabilities because their guaranteed liabilities formed such a small proportion of the total amounts they expected to pay out.

Instead their investment strategies could focus on identifying strategies that maximised the chance of maintaining the real (inflation-linked) value of benefits. Surpluses generated by these strategies were then often used to provide discretionary pension increases, up to whatever level was affordable at the time. This translated into them typically holding a high proportion of their assets in equities and other return seeking assets. It also meant that they provided a pool of capital potentially capable of supporting other parts of the economy (including its financial system) in tricky times.

Long-term investors such as these ought indeed to be able to make a positive contribution to financial stability, if others for whatever reason get spooked. Some argue that a common pension fund response to the 1973–74 bear market mentioned in Box 2.1 was to see it as a buying opportunity, helping to restrict the negative impact of the bear market on the broader economy.

But that was the position many years ago. Since then, inflation rates have fallen and guaranteed benefits have risen. Pension funds still have long-term liabilities and hence long term investment horizons. But they no longer have the same level of investment freedom and ability largely to ignore their liabilities, especially now many of them are in deficit. This is likely to have significantly reduced their likely ability to contribute to financial stability in a crisis.

Consider a situation where a bank raises long-term finance which it then uses to advance a long-term loan. Does such a balance sheet add to financial stability? Liquidity risk is limited, but the bank is taking on credit risk, so such a balance sheet is not necessarily dampening some possible types of financial instability. It was not primarily the *long-term-ness* of pension funds that may have allowed them to contribute to financial stability in the past. Instead, it was the *investment freedom* they had, because they were not tied down by specifically identifiable liabilities that they had to meet, but only looser best endeavours aspirations.

The shorter-term countercyclical rebalancing typically undertaken by pension funds because of their common use of fixed weight benchmarks for their overall asset allocation probably does provide some contribution to financial stability. However, authors such as those who contributed to Bank of England (2014a) think that pension funds have in recent years exhibited greater procyclical behaviours than previously. Their analysis particularly focused on the substantial shifts out of equities that (UK DB) pension funds have carried out in recent years as they have adopted investment strategies more closely matching the economic nature of their liabilities.

Some of the issues discussed above came to the fore in 2016 with the publication of a UK government consultation paper setting out various options for potentially helping British Steel Pension Scheme (BSPS) as part of a wider package of government support for UK Steel, steel workers and affected localities. Tata Steel, the owner of UK Steel had announced that they were potentially pulling out of the UK with possible loss of a significant number of jobs unless a buyer for the business could be found. A sticking point in the potential sale to a third party was that the third party would likely have to take some responsibility for the (large) shortfall within BSPS. The consultation aired possible ways of addressing the problem, including some that would have altered benefits previously accrued or removed them from within the scope of the third party or the PPF.

Pension experts with long memories can point to the UK government having previously changed the nature of pension promises. Prior to 2003, if a UK DB pension fund defaulted then the sponsoring company was not liable to make good any shortfall. By the 1990s, many funds had built up healthy surpluses, but in the early 2000s there were several high-profile cases where companies 'walked away' from their past pension commitments, one notable example being Allied Steel and Wire in 2002. The ability to do so was eliminated on 11 June 2003. The subsequent Pensions Act 2004 included the establishment of the UK's Pensions Regulator and the UK's Pension Protection Fund. If the government has done so once then why can't it do so again.

The problem with this logic is that 2003 was quite a long time ago. If current pension promise structures are inappropriate then maybe the time to have pointed this out was shortly after 2003 rather than now. Moreover, bailing in members reduces rather than improves the promises previously granted to individuals. Again, we see that underpinning a systemic risk issue is a notion of what is 'fair'. Notions of fairness do change (especially over the long timescales applicable to many pension promises). Historically, however, they have tended to change in favour of individuals (as in 2003) rather than in favour of businesses. Our notions of 'fairness' also include notions about who we should be 'most fair' towards.

More usually nowadays, commentators view DB pension obligations in the UK (and other countries having similar pension systems) as fairly well settled and hence economically like other forms of corporate debt such as bank borrowing. A company's obligation to pay £1,000 in ten years' time to a creditor is then seen (on an economic basis) to be little different from its obligation to pay £1,000 (via the pension scheme) to a pensioner in ten years' time (apart from the latter also being dependent on inflation and mortality in the meantime).

4.3.5 Benefit Security (And Adjustment) Mechanisms

If we accept that a pension promise is now broadly akin to any other debt an employer might have, the main issue as far as pension scheme members are concerned becomes how robust are the promises the member has received. Different countries have developed different ways of attempting to make promises in their jurisdiction more robust. Collectively these are known as *benefit security mechanisms*, see Box 4.7.

With a DC arrangement, members generally carry the investment risk rather than the vehicle providing the pension. The main benefit security mechanism is the setting aside of assets into a separately funded vehicle which represents the members' accumulated contributions. Assuming there is no outright theft of the assets, DC schemes should have similar systemic risk profiles to other similar investment funds, see Section 4.4.

With a DB arrangement, investment risk is (in the main) carried by the vehicle through which the pension is provided (or by the sponsoring employer) rather than the member. The member is in effect promised a certain quantum of pension (and other associated benefits). It is up to the vehicle (and potentially also the employer) to arrange how this promise should be honoured. Some security is provided if the scheme is funded.

However, other benefit security mechanisms may kick in if for some reason the vehicle and employer combined fail to honour the promise.

Box 4.7: Pension fund benefit security mechanisms

Benefit security mechanisms for DB pension schemes vary significantly across countries. The main ones applicable to UK private sector DB pension funds are:

(a) *Funds already built up within the pension fund to support future benefit payments relating to benefits already accrued*.

There are various ways in which the level of support this offers might be quantified. PPF (2014) includes a quantification on a *full buy out basis*, i.e. a (market consistent) estimate of how much it might cost to buy out or transfer the accrued liabilities to an insurer. As at 31 March 2014 the market value of assets of such pension funds (c. £1137 billion) were 67% of aggregate liabilities on a full buy out basis (c. £1690 billion).

(b) *Additional contributions that the sponsor agrees to make in the future to rectify shortfalls*.

If a pension fund is in deficit then the sponsor and the fund trustees typically need to agree on a recovery plan. This will typically involve the sponsor paying additional contributions into the fund at a rate that is expected to rectify the deficit over an agreed time frame (usually several years). The deficit to be addressed is usually measured relative to the pension fund's funding projections. These often take some account of additional returns (above those implicit in the full buy out basis) expected to accrue to the fund from the investment strategy it is adopting, often with some smoothing applied so that the resulting additional contributions are tolerably stable. The additional returns potentially apply to the whole of the lifetime of the liabilities (not just the length of the recovery plan) so pension funds typically reveal a materially better liability coverage ratio on such a basis than on the full buy-out basis referred to above. Funds that have a shortfall as per (a) will therefore not necessarily receive any additional sponsor contributions in the near term via this mechanism (although eventually additional contributions would need to be forthcoming if actual returns are insufficient to make good the shortfall by the time all benefits have been paid).

(c) *Shortfalls becoming debts on the employer if the scheme is wound up or otherwise defaults*.

Typically in such situations in the UK, shortfalls as per (a) are required to be taken over by the sponsor. The sponsor is typically the employer who originally established the fund. However, it may just be an associated company depending on how the corporate group has been restructured in the meantime. This, in combination with (b), means that eventually additional contributions should be paid into the scheme sufficient to provide for all promised benefits. Or rather, it does so if the sponsor does not also default.

(d) *Existence of an industry-wide pension protection scheme*.

The UK has an industry-wide pension protection scheme (the PPF) which UK DB pension funds are required to join. The PPF takes over responsibility for honouring a substantial fraction (but less than 100%) of a member's promised benefits if the sponsor defaults and the IORP is then unable to honour these PPF-covered benefits. For some service periods the level of inflation-linking guaranteed by the PPF can be less than is typically promised by the fund and a 10% reduction is applied to members under normal pension age. The equivalent of the liabilities referred to in (a) but relating to just these PPF-covered benefits is called a 's179 basis' valuation. As at 31 March 2014 these totalled c. £1177 billion. The aggregate funding at that date on a s179 basis was therefore 97% rather than the lower figure of 67% applicable on a full buy-out basis. The PPF is set up as a statutory body (so is not e.g. regulated by EIOPA or any associated national insurance supervisor). To fund itself, the PPF has power, within some limits, to raise levies on individual DB pension funds. Its levies are typically structured to have risk sensitive elements linked to the estimated probability of default of the sponsor and the loss given default (to the PPF) that might then arise.

In Holland, there is much less of a role for an individual's employer in occupational pension provision. Instead pensions are largely provided by industry-wide pension arrangements and need to be self-standing. This means that most of the mechanisms available in the UK context are not relevant in the Dutch context. Instead, the most important benefit security mechanism is conditional indexation, i.e. the ability not to increase pensions in payment by as much as expected if there are insufficient assets to support full indexation. Should there be insufficient assets to provide any indexation (perhaps more likely in a low interest rate environment) then the underlying benefit levels may be cut back. This might be deemed a special case of conditional indexation. However, it would more normally be viewed as a 'last resort' measure that would only be countenanced in extremis as it might be considered tantamount to 'failure' of the pension scheme.

Some other countries have protection arrangements structured as specialist insurance companies and therefore subject to insurance regulation. These are often mandatory for companies to contribute to, if they wish to provide ongoing occupational pension provision.

The main applicable benefit security mechanism for a 'pure' DC pension fund is the build-up of assets within the fund. Assuming these assets are not stolen or otherwise misused, the liabilities payable should match the assets available. Such an arrangement can be viewed as just another form of savings, akin to investing in a UCITS or a 'pure' unit-linked life insurance arrangement. Many DC pension funds include insurance contracts or investment funds. They may therefore benefit from compensation schemes if any that apply to such arrangements.

Countries that do have pension protection schemes (PPS) or the equivalent are potentially exposed to the risk that the PPS itself runs out of funds. This would probably count as a systemic risk event in most people's eyes. The risk of such a failure becomes more likely if the PPS fails to charge an appropriate premium for the risks it is covering. In the UK, these relate primarily to the risk that a given sponsor will default at the same time as the assets available within the pension fund are insufficient to provide for PPF covered benefits. Some elements of the design of the UK PPF were selected to address issues perceived to exist with similar arrangements in other countries. For example, the PPF sets its own levies within suitable bounds. This choice of approach appears to have been designed to address risks perceived to exist with e.g. the US's Pension Benefit Guarantee Corporation (PBGC) where from the perspective of UK-based pensions professionals there was perceived to be greater political involvement in rate setting and therefore a greater risk that politicians would refuse to allow the PBGC to charge a realistic rate for the credit insurance involved (because of the extra costs that doing so would impose on corporations).

Using banking jargon, setting aside funds into a separate special purpose vehicle (here the pension fund) to increase the likelihood that beneficiaries of the SPV will receive what has been promised to them would be called *collateralisation*. The close analogy between setting aside funds in a pension fund and doing so in other collateralised structures means that we can apply credit risk analysis methodologies to quantify the level of security being offered by the presence of these assets (and other benefit security mechanisms applicable to the pension benefit), see e.g. Kemp (2011a).

This 'collateralisation' point of view also highlights another feature shared by many DB pension funds. They are 'not for profit' in the sense that their primary goal is to meet promised payments to beneficiaries as they fall due. In nearly all cases their 'customers' come from their association with their sponsoring employer, so they may not possess much if any financial goodwill of the sort that a firm able to generate future profits from a customer base might have. This means that they cannot realistically expect to raise capital other than by seeking additional sponsor or member contributions (they generally have no profit-making business activities themselves against which anyone else might advance such capital).

4.3.6 Regulation and Supervision of Pension Arrangements

In the EU, unfunded pension arrangements (and funded state-based pension provision) fall within the remit of social affairs rather than finance. This means that under current EU rules on subsidiarity many of the regulatory frameworks involved fall within the remit of the member state itself rather than within the remit of central EU supervisory bodies.

However, when the arrangement falls more definably within the financial system, either because it involves insurance or other modes of investment or because the arrangement is classified as an IORP, it falls within the scope of the EU System of Financial Supervision, with the sectoral supervisor at the EU level being principally EIOPA. For example, most Dutch and UK DB pension funds fall within the IORP rules. However, many German occupational pension arrangements, if they do not involve pre-funding of benefit promises into vehicles legally separate from the sponsor, fall outside them.

The EU has recently agreed a new directive for IORPS, the so-called IORP II Directive, which will change their regulatory framework, see European Union (2014d). No specific minimum capital requirements are mandated by the IORP II Directive (except, it can be argued, for cross-border pension funds, but they are currently rare). These remain the responsibility of individual member states to set in line with social expectations defined at the EU member state level (allowing for any applicable benefit security mechanisms). However, the IORP II Directive will increase the expected level of 'own' analysis of risk that IORPs will be expected to carry out, see EIOPA (2016b). This will bring the analyses they carry out more into line with the internal capital adequacy assessment process (ICAAP) and own risk and solvency assessment (ORSA) rules applicable under Basel III for banks and investment managers or under Solvency II for insurers.

4.3.7 Systemic Vulnerabilities

There are several ways in which pension systems can exhibit systemic vulnerabilities.

At its most basic, the financial cost of providing future pension benefits depends on:

(a) how likely it is that the individual might live long enough to receive the promised benefit;
(b) how long the individual might live thereafter (if the benefit is in the form of regular income in retirement);
(c) how the promised benefit changes through time (e.g. the extent of inflation-proofing it receives);
(d) the present value of payments in the future, taking account of the time value of money; and
(e) how likely it is that the benefit promised will be honoured (presumably we aim for this to be nearly always and would deem 'failure' to be a situation where a suitably sizeable fraction of benefits promised ends up not being honoured)

Pension systems can become untenable if the old-age dependency ratio becomes too extreme and past benefit promises are too generous. In a closed economy, the consumption of goods and services by retirees needs to be met by redirecting some of the productive activities of those who are still working.

If the level of redirection becomes too large then those still working are likely to rebel and the intergenerational social contract implicit in the system is likely to fall apart. Collapse of the system in this sort of manner becomes more likely if the current cohort working also lose faith in the ability of the next cohort to support them in *their* retirement. Perhaps some of the economic changes that occurred in Russia and some other countries following the collapse of the Soviet regime can be viewed in this light. The economic adjustment that took place was so extreme that previously agreed social 'contracts' between generations fell apart and some Russian pensioners in effect had previous pension promises given to them virtually wiped out.

Many Western economies are facing an ageing demographic profile. Addressing this issue is behind many countries targeting increases to state retirement ages, which in some cases also then flow through to target retirement ages in occupational and supplementary parts of the pension system. Some commentators refer to these economies as facing a pension 'crisis'. Compared to the sorts of systemic challenges that afflicted the banking system during the 2007–09 Credit Crisis, the pensions crisis is playing out very slowly. But does this make it less systemically challenging? A 2016 Economist article, Economist (2016a) concludes with the following gloomy outlook:

> This is a slow-motion crisis in which the casualties – the weakest companies and cities – appear intermittently rather than all at once. Although the commitment to pay retired public sector workers is in effect a debt, it does not show up in the official figures. Nine countries – Austria, Britain, Denmark, France, Germany, Italy, Poland, Portugal and Spain – have public-sector pension liabilities of more than 300% of GDP, according to Citigroup.

Pension systems can also become challenged if the present value of payments in the future become too expensive to honour, because of changes to the time value of money. The time value of money for this purpose is clearly linked to (although possibly not identical with) rates that define the extent to which individuals are prepared to forego current consumption for deferred consumption, i.e. interest rates (and the yield curve more generally). Costs of providing pensions generally rise as interest rates fall. Pension systems are therefore potentially vulnerable to low interest rates, in much the same way as some guaranteed life insurance business models are, see e.g. Financial Times (2016a) or Financial Times (2016b).

This vulnerability applies to the whole pension system, even government sponsored elements. For example, a government that has unfunded state-wide pension arrangements (or unfunded arrangements for its own employees) conceptually needs to account for the costs of the benefits being accrued under the arrangement. Eventually the costs will need to be met from somewhere. Unfunded pension liabilities promised to future pensioners are economically akin to issuing government debt (if it is reasonable to assume that the level of 'promise' involved is the same). The practical ramifications are subtler once we consider the full range of possible ways the future might evolve. It may be easier for governments to default on hard to decipher and sometimes not explicitly defined pension promises (or to tax them more) than it is for governments to default on more explicitly defined debt instruments. We look at sovereign risk further in Section 4.9 but would note here that when sovereigns default they often also try (or are encouraged by their creditors) to unpick expensive social contracts they may have previously agreed, including past pension promises they have given to their own employees or to their entire populace.

More usually, policymaker focus is on the private sector. Interest rates encapsulate the trade-off between consumption now and consumption in the future. Yield curves knit together the picture for consumption at different times in the future. Cash liabilities of the sort that banks get involved with lie at one end of the yield curve (the short end), whereas pensions lie largely at the other end (the long end).

Governments and central banks are not disinterested bystanders in this context. Essentially all interest rate bearing instruments involve a liability to one party and a corresponding asset to another party. Altering interest rates therefore involves a reapportionment of resources from one party to another. Monetary policy setting typically focuses more on the impact that interest rate policy might have on aggregate demand and less on these (usually presumed to be second order) redistributive effects. But in today's low interest rate environment some economists are querying whether monetary policy is as effective as it used to be at influencing aggregate demand, and second order effects are taking on added prominence.[3]

[3] Some economists also worry about other second order effects arising from low interest rates, including redistributive effects between wealthier and poorer individuals. Low interest rates, quantitative easing and other unconventional monetary policies have bid up the values of assets and hence made those who currently hold them (disproportionately those who are already wealthy) wealthier, presumably at the expense of those who currently don't hold them (disproportionately those who are currently less wealthy). A low interest rate environment is also a headache for banks, as it squeezes the interest margin they can earn between the rates they charge borrowers and the rates they pay savers.

Taking the UK as an example, PPF (2014) indicates that in March 2014 a 0.1 percentage point (10 basis points) reduction or increase in gilt yields would increase or reduce their s179 liabilities (see Box 4.7 for meaning of this phrase) by around 2 per cent. Pension promises are particularly long-dated. This results in private-sector DB pension funds typically having a structural mismatch between assets and liabilities and hence a structural interest rate exposure, since it is not usually practical for them to buy enough assets of sufficiently long duration to match their pension liabilities. So, some of this change would also flow through to coverage levels.

As well as being worried about scope for funding shortfalls (and possible knock-on implications for stability of any applicable central PPSs), policy-makers also appear to be worried about potential significant (longer-term) interaction effects with the wider economy created by low interest rates. Any consequential reduction in pension benefits presumably can ultimately be expected to impair consumer confidence. In some countries, the extra costs of honouring promises in a low interest rate environment fall primarily onto the sponsoring employer. If this burden is not managed adequately, it could affect the real economy by constraining sponsors' cash flows. Counteracting this, the long timescales over which such effects might come to pass are not aligned with the sorts of shorter timescales that macroprudential policy-makers typically worry about.

Macroprudential policymakers also seem interested in how a low interest rate environment might alter the investment behaviour of pension funds. For example, it might lead pension funds (and/or any investment funds in which they might invest) to adopt riskier investment strategies or to increase non-traditional investment activities such as credit intermediation.

4.4 Investment Funds

4.4.1 Introduction

Another major component of the financial system is formed by investment funds. These come in many different forms and hold many different types of asset.

Classically within the USA these sorts of funds are associated with mutual funds and in the EU with *undertakings for collective investment in transferable securities* (UCITS funds). Growth of both has been spectacular over the last decade or two. This has been helped in part by favourable movements in equity (and bond) prices since the nadir of the 2007–09 Credit Crisis in early 2009 (partly reflecting the impact of quantitative easing and other monetary measures that have favoured the low interest rate environment mentioned previously, see Box 6.1). Commentators also view some of this growth as a symptom of a pull-back by banks from several business areas, which asset managers have been willing to enter.

For example, Hanouna et al. (2015) note that assets in US mutual funds (excluding MMFs and exchange traded funds, 'ETFs') have grown from $4.4 trillion in 2000 to $12.7 trillion in 2014. Funds that invest primarily in US equities remain the largest category, but as a percentage have fallen from 65.2% to 44.5% over this period. Alternative strategy funds have grown faster than any other, growing from £365 million in 2005 to $334 billion in 2014.

US mutual funds and UCITS funds are generally open-ended funds. These sorts of funds create and cancel units /shares whenever investors wish to invest or disinvest from them. An investor might be required to give at least a few minutes' notice that they want to invest in the fund. They would pass cash to the fund and in return the fund would issue new units equal in value to the cash being invested. The fund would invest this new cash (alongside cash previously received from investors and previously invested) as the fund's investment manager sees fit (subject to any relevant constraints, e.g. a fund that said it was going to invest in US equities would not be allowed to be invested primarily in Thai equities). Withdrawing investors would do the same in reverse with the fund then sending some money to the investor once the sale of the investor's units settles.

US mutual funds and UCITS style funds are not the only types of open-ended investment fund that exist. In the EU, fund managers offering non-UCITS funds are caught within the scope of the Alternative Investment Manager Fund Managers Directive (AIFMD) which imposes conduct requirements on how these funds are operated.

Investment funds can also be closed ended, see Box 4.10. Most private equity funds are closed ended. With such funds, the manager does not create and redeem units on demand. Instead a fixed number of units /shares are created and the only way that investors can liquidate their investment is to sell on their investment to another party. Some such funds are legally corporations and their shares are actively traded on stock markets (e.g. UK 'investment trusts'). Others, like private equity funds, are much more difficult to sell, although sometimes secondary markets do operate in these less liquid market areas.

4.4.2 Liquidity Transformation

Open-ended funds potentially undertake liquidity transformation if they are investing in less liquid asset classes or become so large that it would not be practical to liquidate a material fraction of their entire holdings at short notice. Yet investors can in nearly all circumstances still withdraw money from them whenever they like (or to be more precise at some specified time point during a business day).

The investment managers who manage these funds and are usually their distributors are typically very aware that 'in *nearly all* circumstances' is not the same as 'in *all* circumstances'. Investment funds typically have a range of measures that they can adopt if they cannot raise sufficient funds quickly enough to meet requests from investors wishing to withdraw from them. These include:

(a) Funds will typically hold some of their total assets in cash or other very liquid instruments (although this can prove less effective than expected if the cash instruments themselves become less liquid, as happened with funds that had holdings in Lehman paper immediately after Lehman Brothers defaulted).

(b) Funds may have (limited) powers to borrow by pledging as security other assets within the fund. Usually these powers are limited to small scale borrowings, so wouldn't be effective if a large proportion of the investor base wanted to leave simultaneously.

(c) Funds may have the right to impose additional charges if a sizeable fraction of investors withdraw at the same time. Charging structures for investment funds can be tricky to analyse in detail but often these sorts of charges are levied on an investor- and deal- specific basis and accrue to the remainder of the unitholders rather than to the investment manager. They are then known as 'dilution levies'. This is because they limit the dilution or diminution in value that continuing unitholders would otherwise suffer because incoming and outgoing investors wouldn't otherwise pay for the costs of investing or disinvesting new or withdrawn cash. The dilution levy rate can be adjusted to cater for temporarily increased transaction cost levels, but this type of mechanism would not be effective at handling the situation where a lot of investors wanted to leave at the same time and the assets held are fundamentally illiquid, i.e. not able to be transacted at all for some time.

(d) In a similar vein to (c), funds may have the right to honour withdrawal requests with an in-specie slice of the fund (i.e. a portfolio consisting of a pro-rata holding in each of the underlying holdings). This approach is not practical to impose on unitholders with only modest sized investments in the fund, but is more practical for larger institutional investors.

(e) The fund may also be able to defer redemption requests outright, sometimes for an extended period. Most fund managers include such powers in their fund ranges in the event of their own IT systems failing over or in the event of the relevant underlying markets being closed (e.g. many US equity funds imposed deferrals when the US equity market was closed in the wake of Hurricane Sandy in 2012). What is more problematic is when fund managers invoke such powers merely because they are overwhelmed with redemption requests. For some types of investment fund, powers to defer are quite limited. For others, e.g. hedge funds that invest in relatively liquid instruments, imposing such deferrals may be legally possible but may sound the practical death knell of the fund soon afterwards. For others that are recognised as investing in illiquid assets, e.g. property, clients may accept that it is more reasonable for the investment manager to impose such deferrals even if they are usually not very happy that it has happened.

Some open-ended funds, such as 'target date' funds have maturity dates when they expect to wind themselves up. They are used by individuals who think that they are likely to need cash at some specified point in time in the future, e.g. when they are planning to retire. These funds may progressively move into more liquid asset classes (e.g. cash) as the target date approaches. This means that they may specifically build up liquidity when it is most likely to be desired by investors.

A more radical solution to liquidity issues is epitomised by the recent introduction in the UK of *authorised contractual schemes*. This is a type of co-ownership in which individual investors, for example, pay income tax as if they owned a representative slice of each individual asset held by the scheme. Such schemes can be structured to behave as if clients have at all times ownership of a segregated portfolio of assets akin to a specie slice of a fund, but with the assets of different clients being managed as a single portfolio. The concept relies on back office systems that are sophisticated enough to handle the accounting (and tax) challenges involved, so is an example of a fund management product that has only become practical because of advances in IT capabilities. The important insight this type of structure highlights is that to a considerable extent investment funds are merely ways of collectively following a given investment strategy. For systemic risk purposes the key driver is often not the fund itself but the behaviour of the investors who collectively guide what the fund does by choosing to invest in the fund.

However, it can be argued that even if most funds manage their liquidity risks effectively it only needs a handful of funds (particularly larger and more systemic ones) not to do so and problems can arise. This seems to be rationale for the SEC introducing rules requiring (most) open-ended US funds including ETFs to have liquidity risk management controls in place, see SEC (2016).

The above comments are generic to most types of investment fund. We now consider specific types of funds that have for a variety of reasons been the subject of greater macroprudential scrutiny (although in some cases the relevance of this scrutiny seems doubtful).

4.4.3 Money Market Funds

One type of investment fund that is seen by many as having systemic risk implications are *money market funds* (MMF). Many of these funds are ones that investors use instead of bank accounts to park short term funds. Some needed bailing out during the 2007–09 Credit Crisis, see Box 4.8.

Like other investment funds, MMFs will have a *net asset value* (NAV) which is the value of the fund per unit (net of any liabilities, e.g. accrued but not yet paid fees to its manager). Unlike a tranched structure, most investment funds have units that rank pari passu (i.e. rank equally) with each other. Hence the NAV of each unit is simply the total net value of the investment fund divided by the number of units in issue.

Money market funds come in two main types, *constant (or stable) NAV* funds (CNAV funds) and *floating NAV* (FNAV) funds. With a CNAV fund the number of units an investor holds changes (in principle each day) so that the NAV is constant. With a FNAV fund the number of units for the investor works in much the same way as for other investment funds, i.e. it stays the same unless the investor withdraws units or invests more (or is deemed to have invested more, because income generated by the fund may be reinvested on his or her behalf), and the NAV goes up or down depending on the performance of the assets in the fund.

From a systemic risk perspective, CNAV funds are viewed as particularly troublesome. ‘Changing’ day by day has historically meant ‘increasing through time, if necessary including some smoothing so that behaviour is as similar as possible to a bank account on which interest is accruing’. Often contractual terms in a CNAV fund effectively imposed a floor of zero on the movement that the CNAV unit price could exhibit. A CNAV whose NAV price falls in value is therefore potentially problematic. This event is known as ‘breaking the buck’. Introduction of such a floor effectively means that a CNAV fund is susceptible to bank-like runs, as investors may seek to withdraw their funds as soon as possible if they think the CNAV fund is in danger of breaking its buck, adding to its liquidity woes.

The International Organization of Securities Commissions (IOSCO), US Securities and Exchange Commission (SEC) and ESRB have come up with a range of proposals for MMFs, according to FSB (2013). These have involved imposing capital requirements on CNAV funds and/or requiring them to convert into FNAV funds. The FSB preferred requiring all CNAV funds to be converted into floating NAV funds. Where this was not possible it wanted CNAV funds to be subject to rules that are ‘*functionally equivalent to the capital, liquidity, and other prudential requirements on banks that protect against runs on their deposits*’. In contrast, the SEC seemed happy to allow ‘government’ and ‘retail’ MMFs to continue as CNAV vehicles. It defined a retail MMF as one that that ‘*does not permit a shareholder to redeem more than $1 million in a single business day*’ and a government MMF as one that ‘*invests at least 80% of its total assets in cash, government securities, and/or repurchase agreements that are collateralized by government securities*’, according to FSB (2013).

Distribution and management of CNAV funds looks and feels very much like undertaking bank-like activities without specifically involving a bank. It is therefore an example of shadow banking, see Section 4.6. The breaking the buck issue becomes more challenging in a low (or even negative) interest rate environment, because it is harder to identify cash-like assets that will return a sufficiently positive return to provide for manager fees as well as the effective (somewhat smoothed) floor of zero on how the CNAV can move day-by-day.

These challenges to a significant extent mirror those faced by banks in a low interest rate environment. They too may struggle to provide zero or positive interest rates on cash deposits. Banks address this type of issue, when interest rates are low but not particularly negative, by introducing fees on such accounts (i.e. substituting the interest margin they would otherwise receive on the product with fee revenue). A worry with both MMFs and banks in such situations is that they will also try to inflate their interest margins (for banks) or their yield on investments (for MMFs) by searching for yield, i.e. investing in riskier assets than might otherwise be warranted on fundamental risk-return logic, see Section 4.4.7. Riskier investments might in this context involve investment in structured finance vehicles and other tranched structures, see Section 3.5.

When the general level of interest rates for a specific currency becomes negative (as it has done in several European countries in recent years) then banks need to be more creative and may need to impose negative interest rates on customers' deposits. Given the difficulties in explaining this type of behaviour to small retail clients (who may have modest sums which they can instead easily hold in cash, i.e. notes and coin), the tendency so far has been for banks to levy negative interest rates principally only on larger corporate deposits, who are more aware of the practical difficulties arising when interest rates go negative.

MMFs, even quasi-CNAV ones, can in principle do likewise, e.g. by rewriting the method of operation of their unit pricing so that a constant negative drag is imposed on the unit price in the absence of additional returns from investments. Many (but not all) types of investment fund, particularly when they invest in supposedly more liquid asset classes, reserve the right to change any of their contract terms at relatively short notice (e.g. a 3-month notice period), even to the extent of being able to wind themselves up if the manager so wishes. Some MMFs tend to be more used by institutional investors who may be more willing to withdraw their funds at short notice but are also more prepared to accept different modes of operation, if competitive pressures drive such changes. In most cases (at least in Europe), money in MMFs is parked there short-term, awaiting investment opportunities in other asset classes.

FNAV MMFs are viewed as less likely to be systemically important, even though they may hold similar instruments to CNAV MMFs. Their NAVs can rise or fall as the market value of the investments they hold rises or falls. Of course, some clients may have been led to expect that investment in them would behave like cash, and therefore exhibit a level of stability that their actual investments do not deliver. In such circumstances, investment managers have at times needed to contribute to losses suffered by such funds, if some element of culpability is present or is argued to have existed by their customers.

ECB (2015) analysed the state of the euro area MMF industry in mid-2015 and noted that the then prevailing low or even negative interest rate environment presented challenges. However, its analysis suggested that these funds had not in general changed their business model or sought to radically change their investment stance in response to this environment. This suggests that MMFs are more likely to address the challenges of a low or negative interest rate environment by altering contractual terms governing their unit pricing structure than by adopting an aggressive search for yield.

Box 4:8: The Reserve Primary Fund

The US money market mutual fund sector was another part of the financial system that needed propping up by governments at the peak of the 2007–09 Credit Crisis. The Reserve Primary Fund was a CNAV fund which aimed never to lose money. However, it 'broke the buck' (i.e. lowered its share price below its $1 floor) in September 2008 because of its exposure to Lehman Brothers which had just defaulted. The resulting investor anxiety almost caused a run on other similar funds. The run was only averted by the US Treasury announcing a program later that week to insure such funds against them breaking the buck.

One manager's money market mutual fund rarely invests in any other manager's money market fund. So, this is a market where there was (and still is) essentially no 'direct' interconnection between different funds. The averted run affected the entire sector, even those funds that didn't have exposure to Lehman paper. Not only were the funds not significantly interconnected with each other, they were in many instances not even 'directly' interconnected with the proximate reason for why the Reserve Primary Fund ran into difficulties. Only an indirect sort of interconnectivity was present. The need to support this part of the system even though it lacked many direct interconnections is one reason why policymakers are no longer so wedded towards the view that systemic importance is necessarily linked to (direct) interconnectedness.

4.4.4 Hedge Funds

Hedge funds are another type of investment fund that commentators focus on in the context of systemic risk. Hedge funds come in a bewildering variety of types. Even the term 'hedge fund' can arguably be a misnomer for some of them. Originally hedge funds were funds that aimed to take some long positions in markets and to hedge them with some positions expected to perform favourably when markets declined. Gradually the term expanded to cover nearly any type of fund that didn't fit into other more clearly discernible fund types. Some adopt relatively leveraged positions. Others are long-term investors or 'activists' who seek to change the ways firms in which they are invested are managed to boost returns to shareholders. Most are relatively aggressively managed (i.e. may take large positions and /or rapidly change these positions), compared to more traditional investment vehicles. Some adopt highly quantitative approaches to selecting investments. A few are very high frequency traders (and therefore in effect behave like market makers at least from a market microstructure perspective, see Section 8.7).

Most hedge funds have few if any constraints imposed on them, other than those implied by their marketing material or in effect imposed on them by other market participants (e.g. there will be practical limits on how leveraged a fund can become since some other market player needs to advance it funds to allow it to become leveraged). Occasionally their position-taking can go spectacularly wrong. If at the time they are large or systemically important then this can create systemic risk, as illustrated by the case of Long Term Capital Management (LTCM), see Box 4.9.

It is not easy to impose regulation on hedge funds. Many are domiciled in off-shore locations such as the Cayman Islands. Their domicile is usually selected because it is considered by the hedge fund manager to have a suitably robust legal system and business infrastructure. However, funds can usually redomicile elsewhere without too much difficulty. Regulation can more easily be imposed on the managers of the hedge funds. They are often located in financial centres such as London, New York or Zurich. However, they too can typically move elsewhere if they are strongly enough minded to do so (although this may limit their ability to market their services to new investors).

Originally, hedge funds appealed primarily to high net worth individuals, perhaps particularly ones who were footloose for tax reasons. Increasingly, however, hedge funds have been purchased by institutional investors such as pension funds or endowments (e.g. charitable or educational foundations). Regulation can therefore in principle be imposed via their investors by e.g. restricting the sorts of hedge funds that such investors can invest in. However, this flies in the face of other trends in institutional investments that deem it unreasonable to stop institutional investors investing in nearly anything they like, provided they adhere to the so-called *prudent person principle*. Loosely speaking, this principle involves the investor being allowed to select any overall investment strategy they like and then implement it, provided the investor carries out the level of initial due diligence and ongoing monitoring that a prudent person could be expected to undertake if it was his or her own money at stake.

Given these practical challenges, most macroprudential activity in the hedge fund space has focused on monitoring, primarily to gain a better understanding of the impact that hedge funds are having on markets (and hopefully capturing early instances where there might be wider systemic concerns). ESMA (2015) included a section that sought to identify ways of monitoring systemic risk in the hedge fund industry. It assessed the extent to which specific hedge funds appeared to be 'destabilising' (i.e. adopting strategies that appear to be driving the hedge fund sector away from some perceived equilibrium), or 'stabilising' (i.e. pushing the sector back towards this equilibrium). They identified 'destabilising' and 'stabilising' by reference to autocorrelations between an individual fund's return in a specific period with its own return and those of other hedge funds in subsequent periods. It proposed measures based on the fraction of funds that appear to be significantly 'destabilising' (or alternatively 'stabilising') and showed that these measures displayed temporary spikes around the times of some events that appeared to have systemic relevance.

Box 4.9: Long Term Capital Management Fund (LTCM)

LTCM was a hedge fund founded in 1994 by John Meriwether, former vice-chairman and head of bond trading at Salomon Brothers. On its Board of Directors were Nobel prize winners, Myron Scholes and Robert Merton. Despite the evident talent of these and other individuals involved with it, it still imploded in 1998, losing c. $4.6bn in less than 4 months following the 1997 Asian financial crisis and the 1998 Russian financial crisis.

As explained in e.g. Lowenstein (2001), LTCM adopted several investment strategies. One of its core strategies involved a type of fixed income arbitrage strategy in which LTCM shorted the 'on the run' US Treasury bond and went long corresponding 'off the run' US Treasury bonds. The most recently issued US Treasury bond is the 'on the run' one and tends to be more liquid than other (off the run) ones of similar duration that have been issued in the past. The on the run bond therefore tends (all other things being equal) to be slightly more expensive than a suitable basket of off the run bonds, even though it carries the same underlying credit risk exposure (to the US Treasury). Over time (as the US Treasury issued new on the run bonds), this price differential was expected to unwind, or so LTCM thought, creating profits. To benefit in a meaningful way from such a pricing differential (which is typically very modest), LTCM needed to be highly leveraged, i.e. to have (gross) long and short positions much larger than the net capital base derived from the two in tandem. LTCM also followed several other (otherwise apparently relatively independent) strategies that in effect involved shorting more liquid assets and going long assets that were otherwise similar but were less liquid.

Unfortunately for LTCM, some external factors, including an unexpected default by the Russian government on their domestic local currency bonds, led to a flight to quality that bid up the prices of particularly liquid assets relative to less liquid assets. This reduced the capital base of LTCM by so much that it was forced to liquidate some positions at unfavourable times, compounding its losses. Given the size of its gross positions, and fearing a chain reaction were it to fail, the Federal Reserve Bank of New York organised a multi-billion-dollar bailout by its major creditors (mainly Wall Street banks).

Systemic risk issues raised by LTCM include:

(a) The flight to quality led to a flight to liquidity. Like many banks that failed during the 2007–09 Credit Crisis, LTCM was banking on (relatively) benign liquidity conditions continuing (or at least not reversing sufficiently to cause LTCM to become a forced seller at an inopportune time). LTCM demonstrates that investment firms as well as banks can in principle create systemic risks.

Conversely, little longer-term lasting damage to the financial system appears to have been caused by LTCM. The main losers were LTCM's own partners (who had largely excluded other investors from the fund by the time it imploded).

(b) Essentially all (active) investment management involves taking positions, which the investment manager hopes involves buying 'cheap' and selling 'dear'. This requires the manager to form a view as to what is cheap or dear and for the market to come to recognise that this view is correct within a timescale that is not so long that the manager is forced to close the position in the meantime. The wider systemic risk issue is that nearly all participants in investment market take some investment views (even if it is not always clear to them that they are doing so) and if these views fail to come good soon enough (or worse, some unexpected external event causes the opposite to happen) then financial trouble can ensue.

(c) LTCM's liquidity risk manifested itself through derivatives and the need to post collateral to allow it to continue to support its positions. Derivative collateral calls have played a significant role in other systemic risk events, e.g. AIG, see Box 3.1.

4.4.5 Bond Funds (And Funds in Other Traditional Asset Classes)

An area I found particularly bemusing when I first became involved in systemic risk was the apparent focus of some in the macroprudential community on the systemic riskiness of bond funds. Some of the metrics proposed for measuring the risks involved seemed frankly nonsensical to me, given my background as an investment manager. For example, a lot of attention seemed to be placed on the ratio between bond exposures and cash, with the assumption that a high ratio (i.e. a low exposure to cash) was somehow a portent of major concerns.

Gradually I have come to understand better where this line of reasoning is coming from, but I still believe that it confuses decisions being taken by the fund with decisions being taken by the investors who choose to invest in the fund.

From an investment management perspective, bond funds are sold to investors who want to invest in bonds. If the investors are large enough then they will typically hold their bonds in segregated portfolios, i.e. they will legally retain ownership of the individual bonds. But for smaller investors this is less practical and more usually they will club together with other investors in pooled vehicles to gain economies of scale. From an investment manager's perspective, the choice between segregated and pooled is essentially down to the investor (and the fees that the investor might pay).

Once a client has decided to invest in bonds then the usual process (if the client has a segregated account) is for the client to give the investment manager a benchmark and an investment universe. The manager is assessed against this benchmark. The benchmark in effect identifies the neutral position the manager might be expected to adopt if he or she has no investment views about the relative attractiveness of different assets in the investment universe. For a (government) bond fund the investment universe might involve government bonds plus cash (to cater for small amounts of cash awaiting investment or disinvestment). The benchmark might be a market index concentrating exclusively on governments bonds. For a corporate bond fund, the investment universe and benchmark would be adapted to address the desire of the client to invest in corporate bonds rather than government bonds.

For pooled funds the process is conceptually similar. However, as individual clients cannot then have investment universes and benchmarks bespoke to just themselves, the investment manager attempts to craft a sufficient range of funds with different universes and benchmarks to address tolerably closely any client's preferred investment strategy. Different types of pooled fund, e.g. UCITS, US mutual funds or even unit-linked insurance funds, all follow the same basic pattern (except for any specific regulatory requirements to which they need to adhere, to retain such a regulatory classification). The specific type of pooled fund used is driven primarily by client requirements (to the extent that the manager has a platform and geographical reach large enough to be able to offer fund types spanning different regulatory regimes).

A special case of such an approach is a passive vehicle that tracks a specified market index. For example, a (physical) S&P 500 index fund can be expected all the time to be invested essentially 100% in the US equity stocks that form the S&P index (and in the weights ascribed to them in that index). Most commentators accept that such a fund has little obvious relevance to systemic risk (except to the extent that the behaviour of the aggregate equity market might have some systemic risk aspects, see Box 4.3).

What this means is that we can expect nearly all traditionally managed portfolios to be largely or wholly invested in the market that they are supposed to be targeting. For a bond fund, a high ratio between amount invested in bonds and amount held in cash is to be expected. Indeed, for a bond index fund we would expect it to be close to infinity, since the clients are paying the fund manager to be fully invested all the time.

So why do those in the macroprudential community worry about such a ratio? The insight is that banks and others carry out maturity transformation, i.e. borrow short and lend long. Bond funds are effectively doing the same, by issuing units /shares to investors and then investing the proceeds in longer dated assets. So, from the perspective of the macroprudential community metrics such as the ratio between amount invested in bonds and amount held in cash is meaningful as it is quantifying the extent of maturity transformation these funds may be involved with instead of banks.

The disconnect is that it is not the funds themselves who have 'decided' to carry out maturity transformation. Instead it is the clients who have 'decided' to do so by choosing to invest in bonds. If they had done so using segregated portfolios then the same amount of maturity transformation would also be taking place, but it wouldn't show up within the investment funds sector. At its heart this disconnect is about the nature of decision-making in the financial world and about the substitutability of implementation approaches. These may show up in different parts of the financial system depending on how any specific implementation approach is classified.

4.4.6 CDOs, CLOs and Other Tranched Structures

The logic in the previous Section breaks down when the characteristics of investing in a pooled vehicle diverges from those applicable if the same assets were held in a segregated portfolio. The most obvious example of this is one we have already explored in detail in Section 3.5, i.e. where the fund is tranched and different classes of unitholders are entitled to different returns. Although the aggregate performance of all tranches combined should be in line with the performance of the total portfolio, this is not the case for any individual tranche.

CDOs, CLOs and certain types of structured finance vehicles such as mortgage conduits used by banks originate and distribute mortgages had a torrid time during the 2007–09 Credit Crisis. Many investors who thought that they were buying relatively creditworthy and relatively liquid instruments that happened to offer favourable yields found that their investments in these types of securities became much less liquid than they expected. In many cases, they suffered substantial losses. Some investors subsequently sued the originators of such structures arguing that they had been missold to. A consequence is that some originators have suffered large losses and fines.

CDOs, CLOs and mortgage-backed securities are not the only types of tranched structures that have in the past provided fertile ground for lawyers seeking to prove that financial organisations had behaved culpably. In 2002 another type of tranched structure, the split capital trust, also demonstrated that providing differential performance to different investors in the same overall structure could be problematic, see Box 4.10.

Regulatory capital regimes are now more cautious about tranched instruments. And yet, they have not disappeared. Indeed, there have been calls to make it easier to issue 'high quality' structured finance, such as some forms of 'covered' bonds, see Box 4.11, because they are perceived capable of providing a useful financing tool that may contribute to renewed economic growth.

We've noted previously that nearly all types of corporate structure in nearly all walks of life (including finance) contain tranche or tranche-like elements. So, it is not surprising that there is a desire to avoid killing the goose that lays the golden egg merely because a few previous eggs it has laid have proved dud. The problem is how to formulate regulatory frameworks that cater for the huge range of tranched structures that now exist in modern finance, facilitating the high-quality ones but discouraging the more damaging ones. Whether this problem is practically solvable is difficult to say. Tranched structures do not naturally behave in the manner that we would associate as 'fair' to everyone (since our hard-wired notions of 'fairness' generally expect equal treatment for everyone).

Box 4.10: Closed-ended investment vehicles and split capital trusts

An investment trust is a form of exchange traded pooled investment vehicle which is *closed-ended*, i.e. which has a set number of shares (units) in issue. If investors want to realise their investment in such a vehicle they generally need to sell their shares to another party. Investment trusts contrast with *open-ended* vehicles such as (most) unit trusts, open-ended investment companies (OEICs), UCITS and US mutual funds. With an open-ended vehicle, additional units in the vehicle are created (or redeemed) by the organisation managing the vehicle when an investor wants to buy (or sell) units in it (and so the investor technically buys or sells units from the vehicle itself rather than from third parties, with new units being manufactured /removed on demand).

Sometimes both mechanisms operate, e.g.:

(a) ETFs are designed to be bought or sold on exchanges throughout the trading day (rather than typically just once at a set time during the day for most an open-ended vehicles). However, index versions of such funds have an additional mechanism (often occurring once a day) that allows the investors to give to them (or receive from them) an index portfolio in return for new (or redeemed) units. Arbitrage becomes viable via this mechanism if the unit price diverges too far from the value of the underlying holdings (their 'net asset value'). This should keep the unit price and the NAV closely aligned throughout the rest of the day too.

(b) Some hedge funds operate with monthly (or even longer) dealing cycles which can restrict the ability of investors to exit from investment in them in as timely a fashion as the investors might like. Secondary markets can develop, if enough unitholders need to realise their investments urgently (and there are other investors willing to buy their units). Such secondary markets can also form for investments in open-ended funds that are 'gated', i.e. where the manager has deferred redemption of units beyond the usual dealing cycle.
(c) Where the share price of an investment trust diverges too far from the net asset value of the investment trust then it can become appealing for investors to split the investment trust up into its constituent parts, e.g. by carrying out a take-over.

In most cases, unitholders /shareholders in mainstream open-ended or closed-ended vehicles rank pari passu (i.e. equally) with each other. However, sometimes vehicles can be issued (and sold to the public) in which this is not the case. Sometimes different units differ merely in terms of certain voting rights but sometimes they also have more fundamental differences. An example of the latter were split capital trusts issued in the UK in the years up to 2002. These all had at least two different classes of shares, typically with one class receiving all the dividends received by the trust, whilst the other benefited from capital growth.

However, once you have multiple classes of investors (i.e. multiple tranches) you have scope for at least one of the classes to become aggrieved if they consider themselves to be unfairly penalised in adverse circumstances. This was at the heart of the split capital scandal that in the early 2000s engulfed several UK asset managers active in this product space. The regulator fined some of them (and required them to compensate affected investors) arguing that these investors had been the victims of mis-selling.

Box 4.11: Covered bonds

Covered bonds are debt securities issued by banks that are backed by cash flows from mortgages or other loans. They have some similarities with asset-backed securities (ABS) such as MBS. However, in contrast to normal ABS which are issued by special purpose vehicles separate from the bank, covered bonds generally remain part of the issuing bank's consolidated balance sheet. Investors therefore have recourse to both the bank and the collateral backing the covered bond should the creditworthiness of the bond deteriorate. This is called 'dual recourse'.

Dual recourse adds to the appeal of covered bonds for investors. Issuance is substantial in some markets particularly some EU member states such as Germany, Spain, Denmark and France.

Covered bonds come with two twists as far as systemic risk is concerned.

The first is that a covered bondholder is in effect getting priority access to the cash flows backing the bond, potentially moving such bondholders above (mainstream) depositors in priority rankings were the bank to run into trouble. Issuance of covered bonds may help to raise funds, but not necessarily in a way that helps secure the interests of the bank's (other) unsecured customers. The conceptual framework for capital adequacy referred to in Box 3.6 captures this effect. It measures the appeal of issuance of such liabilities (in terms of their contribution to the how secure are depositors' liabilities) by reference to the impact that such issuance has on the credit spread applicable to depositors' liabilities. Only if the cost of funding the higher-ranking liability is sufficiently attractive does such issuance further the aim of securing depositors' liabilities.

The second is that the distinction between covered bonds and ABS may not be as rigid as is suggested above. Some banks found in the 2007–09 Credit Crisis that assets and liabilities they previously thought were segregated from their own balance sheets ended up returning onto their balance sheets through guarantees and other mechanisms, see e.g. Kemp (2009a), i.e. dual recourse was more applicable than they had previously thought. Conversely, there appear to be carve outs for covered bonds in the BRRD that do not apply to many types of ABS, suggesting that in an actual resolution situation the two would still be potentially treated differently. At the time of writing the EU Commission seems keen to enhance the supply of high-quality covered bonds, believing that this is likely to lead to beneficial economic outcomes.

4.4.7 Searching for Yield

We have previously referred to the concept of 'search for yield'. There seems to be general agreement within the regulatory community that investment funds (and others) can create financial instabilities if they undertake excessive search for yield. This can lead them either individually or collectively to enter over-concentrated or excessively large positions (otherwise known as 'crowded trades' to investment managers). This is especially the case if the liquidity terms their clients think they have access to then become divorced from what is practically capable of being honoured by the fund(s) in question. Experience prior to the 2007–09 Credit Crisis included yield spreads between different types of credit investment reaching low levels. At the time, some investors invested heavily in structures like SIVs and CDOs that were then apparently offering very attractive yields (relative to other credit-sensitive asset types) without perhaps properly analysing the risks embedded within such structures.

Exactly what individuals mean when they refer to a 'search for yield' is not always obvious. Possible interpretations include:

(1) *A search for predictable yield, or to be more precise a predictable income stream, when viewed across very broad asset categories*. Risk-averse investors supplying products that include guarantees will typically prefer to invest

in more predictable asset types, to reduce the mismatch between their assets and their liabilities. The more risk-averse they are, the more onerous the guarantees that they offer or the less appealing are the returns on offer from assets offering less predictable income streams, the more these investors are likely to want to move into asset classes offering more predictable income streams. For investors investing in a range of asset types such as equities, properties and bonds, this type of 'search for yield' may show up as a shift from e.g. equities towards bonds (of suitable duration). When inflation rates were high this typically resulted in the overall (running) yield of the portfolio increasing.

(2) *A search for higher returns*. More usually (in the context of financial stability) the term is interpreted to mean the opposite of (1), i.e. to involve a tendency of less risk-averse investors to be unduly happy to invest in riskier assets, in extremis 'gambling for redemption'. Essentially any asset type can be targeted in this way, hence the desirability of monitoring possible formation of asset bubbles across different investment market. For example, the 2015 ESMA Report on Trends, Risks and Vulnerabilities, ESMA (2015), refers to '*The solid EU equity market performance in [the first half of 2015] raised concerns related to excessive asset valuation, as search for yield continued to be sustained by historically low interest rates*'.

Within the more general class of 'searching for higher returns' encapsulated in (2) a particular example is *searching for higher yields within bond portfolios by taking greater credit risk*. When (institutional) investors are referring to investment strategies that are yield related they are most commonly thinking about strategies within the fixed income bond (or inflation-linked bond) arena and are most probably thinking of this type of risk. This is the sort of 'search for yield' that was most obviously exhibited in the immediate run up to the 2007–09 Credit Crisis, see above.

Another angle with 'search for yield' relates to who is taking decisions, see Section 4.4.5. Does a 'search for yield' require an *active* allocation towards a riskier asset class, and if so what do we mean by 'active'? Over the last few years the proportion of EU bond funds invested in AAA paper has declined whilst the proportion invested in BBB paper has increased. Most bond funds are benchmarked against relevant market indices, and the main reason for the shift in average rating appears to be the change in index composition, as existing bonds have been downgraded and new issuance has been less well rated. Would a lack of a search for yield be characterised by a flat line if we plotted the average creditworthiness of bonds held by such funds

(i. e. by investors adjusting what they hold to reflect perceived changes in credit-worthiness), by a line that only changed through time due to changes in proportions of available investments assigned different credit ratings, or by some other characterisation?

4.4.8 Alternative Asset Classes

In Section 4.4.5 we discussed investment funds that invest in traditional asset classes such as equities and mainstream bonds. These asset classes form the bulk (by value) of the readily investable capital market, see Fig. 4.2. The capital market split shown there is derived from the stock of all available

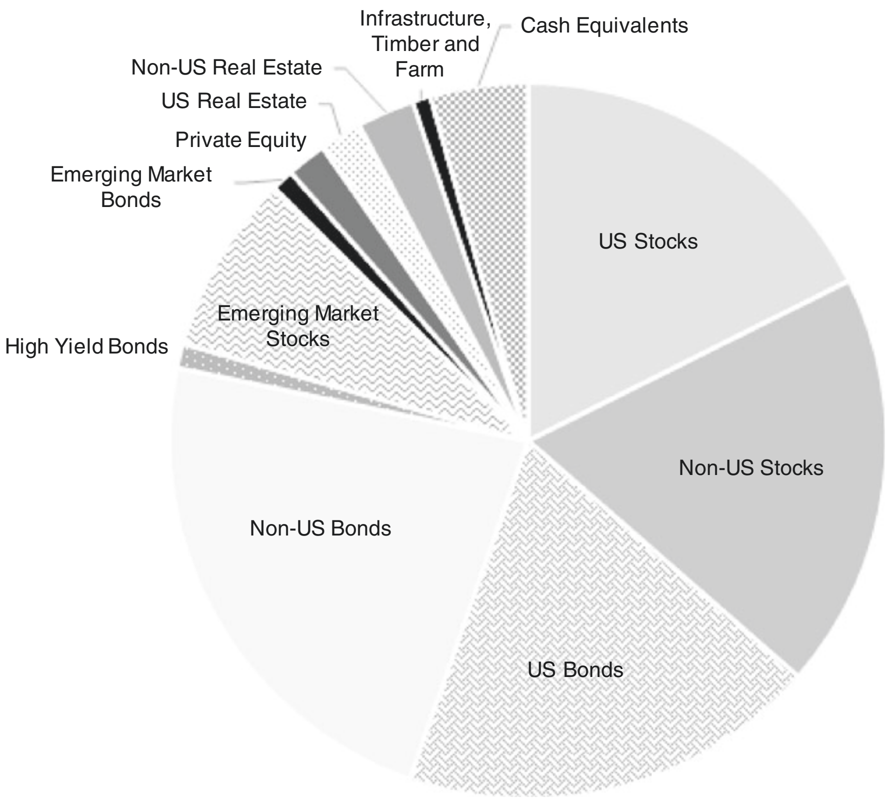


Fig. 4.2 Total investable capital market (31 December 2011)
Source: Nematrian. Adapted from Gibson (2013)

capital market instruments. It is not therefore fully representative of how investment funds might be invested in aggregate, since many of the available instruments will be held in segregated portfolios. The split is also not representative of the fees that investment funds might receive from different asset classes, since fees for more niche asset types tend to be higher than for more traditional asset classes (particularly versus passively managed funds targeting traditional asset classes).

The more niche the asset class the less liquid it tends to be and therefore probably the greater the theoretical scope there is for the asset class to behave in an undesirable fashion from a systemic risk perspective (given the importance of liquidity risk to systemic risk, see Section 3.5.7). Conversely, more niche asset classes tend also to be the smaller ones, providing a practical ceiling on their likely contribution to systemic risk.

A glaring counter-example to identifying 'niche' with 'less liquid' is property, i.e. real estate. It is very illiquid relative to traditional asset classes such as equities and (most) bonds. However, it forms a non-trivial part of the investable capital market, even when looked at through the lens of Fig. 4.2, which focuses on readily traded property vehicles such as real estate investment trusts (REITs) and the like. Include *all* the physical property (such as land and buildings) owned by individuals and corporates and the proportion of the world's capital stock formed by property would be appear much higher. Go back to before the industrial revolution and nearly all societies were primarily agrarian with nearly all wealth being held in this form.

In some intrinsic sense, we should therefore expect physical property to have a particularly important role in systemic risk. This does indeed seem to be the case, since rises and falls in property markets are strongly linked to changes in aggregate levels of bank lending (for mortgages), which in turn are linked to the likelihood of bank capital bases becoming overextended and hence banking crises occurring.

Less clear is the potential role that property investment funds might play in such a landscape. Property funds have from time to time run into liquidity problems, but historically these have been far rarer (and less systemically important) than the sorts of problems that banks have more commonly got themselves into by adopting overly lax mortgage lending policies. Of course, in some cases it is to such funds that the banks have extended such mortgages. But then who is to blame? Is it the investment fund that has provided the scope for individual investors to invest collectively in this asset class or the bank foolish enough to lend it funds so that the fund can overleverage itself?

There is also the possibility that we are adopting double-think. In Section 4.2.6 we referred to the possible systemic importance of the failure of HIH to a specific part of the Australian insurance market. Maybe a major blow up in some (non-property) niche asset class could still be deemed systemically relevant, if the niche happens to be important enough for a specific segment of the financial system.

This risk is perhaps most likely if the alternative asset class is particularly closely aligned with some specific part of the wider, real, economy. For example, some commentators believe that investing in commodities forms an attractive alternative asset class. Where the supply and demand of a commodity are closely aligned (and adjusting either takes a long time) it does not take much new money in the form of vehicles targeting that commodity to have a significant influence on the price of the commodity. But again, is it the vehicle that is creating the possible systemic risk exposure or the investors who have decided to use the vehicle to gain a desired exposure to that asset type? Speculators have always been active in commodity markets and do not need to invest via an investment fund to carry out their speculative activities.

4.4.9 Clarifying Responsibilities

One more general regulatory response to systemic risk seen across the financial sector is to seek enhancements to governance disciplines, including transparency over who is responsible for what. This trend perhaps explains some of the changes introduced for UCITS and other EU investment funds in recent years, see Box 4.12. Perhaps these sorts of changes would have happened anyway, but impetus for them was spurred on by losses some investors suffered due to the failure of Lehman Brothers.

Box 4.12: UCITS and AIFMD

The Alternative Investment Fund Managers Directive (AIFMD), see e.g. A&L Goodbody (2013), and the Undertakings for Collective Investment in Transferable Securities Directive V (UCITS V), see e.g. A&L Goodbody (2014), are directives governing investment products or investment managers in the EU.

AIFMD was formally adopted in 2013. It introduced regulation of alternative investment fund managers (AIFMs) who manage one or more alternative investment funds (AIFs) in the EU and/or market them in the EU. The regulations are akin to some that already apply to UCITS fund managers. AIFMD will impose authorisation and organisational requirements on AIFMs. It includes requirements in relation to transparency, remuneration, depositaries, valuation of assets and leverage. AIFMD also contains detailed provisions on conduct of business requirements, conflicts of interest provisions and risk management and liquidity management provisions.

AIFMD will facilitate EU 'passporting' (i.e. simpler marketing across EU jurisdictions) of AIFs to non-retail investors. This passport will not extend to retail customers, reflecting the perceived potentially sophisticated nature of AIFs and the potential lack of sophistication of retail investors.

UCITS V aligns several aspects of previous UCITS rules with newer regulation applicable to AIFMDs, see above. It amends:

(a) *Depositary and custodian responsibilities and liability*. This issue was triggered by lack of clarity over who was responsible for what when Lehman Brothers International Europe defaulted and by the Madoff case. Lehman was a sub-custodian of some UCITS-like funds. UCITS V requires a UCITS fund to have a single depository and clarifies what responsibilities are placed on this depository including its liability in the event of loss of a financial instrument held in custody. Depositories will need to make good any such loss, with only very limited ability to avoid doing so due to external events beyond its reasonable control. Only credit institutions and investment firms can be depositaries. They will therefore need capital as per applicable regulatory frameworks in relation to the risks involved in being such a depositary.
(b) *Remuneration policies*. New requirements have been imposed on the remuneration policies of firms managing UCITS funds, including e.g. the requirement that '*the remuneration policy is consistent with and promotes sound and effective risk management and does not encourage risk-taking which is inconsistent with the risk profiles, rules or instruments of incorporation of the UCITS*'.

UCITS V also introduces a more formalised whistleblowing regime and rules harmonising sanctions for breaches of UCITS obligations.

4.5 Asset Managers

4.5.1 Introduction

The topic of whether asset managers (or the funds that they manage) can be or are systemically important is a vexing one. Part of the issue is semantics. For the purposes of this Section, we include within 'the funds that they manage' their entire assets under management (AUM) whether in segregated portfolios or pooled vehicles (only the latter were covered in the previous Section). Asset managers generally look after other people's money, so they don't generally see themselves as being systemically important. If anybody is to be deemed systemically important in the context of asset management relationships then asset managers would generally point the finger at their clients rather than themselves.

However, policymakers do not see quite such a rosy picture. They usually focus on the extent to which asset managers (and/or the funds that they manage) might contribute to the propagation of systemic risk. Asset managers might rarely by themselves be instigators of a systemic risk event (although occasionally their funds might be, e.g. LTCM). However, they might amplify it if they react in specific ways to specific market events, perhaps even turning a minor displacement into a major disaster. Policymakers can point to evidence that suggests that asset managers, like banks, potentially react in procyclical ways, see e.g. Timmer (2016).

This can then trigger a debate about whether all types of procyclicality are relevant to systemic risk. This often involves forming a view on whether booms and busts in e.g. mainstream (liquid) equity markets are more likely to create financial instabilities than booms and busts in (less liquid) credit markets or property markets. The issue becomes most relevant when we ask which types of asset managers (for a given size) might contribute most to systemic risk.

Asset managers appear to control sizeable blocks of assets. Global AUM of the asset management industry have risen from \$50 tn in 2004 to \$76 tn in 2014 or 40% of global financial system assets, notes FSB (2016). Of the \$76 tn, approximately \$37 tn were invested in regulated open-ended funds and \$3.0 tn in hedge funds. A handful of individual asset managers individually each have over \$1tn of AUM.

Conversely, policymakers such as FSB do generally seem to believe that the trend towards more market-based financial intermediation arising from the growth in the fraction of assets held by asset managers is likely to bring efficiencies and to be helpful to the broader economy. These developments are typically seen as adding to rather than subtracting from financial stability.

4.5.2 Asset Management Business Models

Of the major types of financial services firms, traditional asset managers perhaps have a business model least like either traditional banking or insurance. Asset managers generally act as *agents* rather than *principals* in investment transactions. They themselves typically have very modest capital bases relative to their AUM. They often charge ad valorem fees on these assets, i.e. some specified percentage of the market value of the assets in question. Some more aggressive asset managers including most hedge funds also charge performance related fees, only payable on investment performance achieved above a specific threshold. Sometimes managers are also remunerated via one-off entry charges when an investor first places some money with them or via corresponding exit charges.

'Aggression' in this context means willingness to take stances quite a way away from the generality of investors within the field in which the asset manager is investing. Measuring the extent of a manager's aggressiveness is possible but sometimes non-trivial as managers may have an incentive to talk up how aggressive they are, especially if they may be paid extra for doing so.

Other ways of characterising asset managers include the asset category in which the manager invests, e.g. equities (and if so where in the world or in what sector), fixed income (i.e. bonds), real estate (i.e. property), money market, venture capital etc. Some managers may also manage multi-asset portfolios which invest in a range of asset classes.

Asset managers manage assets largely on behalf of others, e.g. insurers, pension funds, banks, other institutional investors or individuals. Their own products may compete against products offered by their clients. The main direct risks they face are typically operational in nature (e.g. the risk of investing their clients' money in a manner that is outside the brief given to them by their clients). Indirectly, they may have significant market risk exposures (particularly relative to the size of their balance sheets) due to the exposure of their future fee revenue streams to changes in the market values of their AUM. They may be owned by financial conglomerates and/or dependent for much of their business from banking or insurance associates.

Hedge fund managers (and some other specialist asset managers) typically have business models that are recognisably like those of more traditional asset managers (but often with a greater proportion of revenue coming from performance related fees). However, they may have investment turnover levels more closely aligned to those of investment bank proprietary trading desks (to the extent that investment banks still carry out such activities, since restrictions have been imposed on their ability in the USA to undertake such activities under the Dodd-Frank Volker rule). Hedge funds may have extensively recruited staff from or have been founded by individuals from such trading desks. They are also increasingly seen as potential providers of market liquidity as banks retreat from this activity.

4.5.3 Can Asset Managers Contribute to Systemic Risk?

In the context of systemic risk, the FSB has identified four main structural vulnerabilities associated with asset management that it considered need addressing via policy responses, see FSB (2016). These are:

(a) *Liquidity mismatches between investments held by funds and the redemption terms granted to investors in open-ended funds*. Recommendations focused on open-ended funds both private and public (including ETFs but not MMFs, as the FSB has been addressing MMFs separately) and included:

- Improved information collection and improved disclosure of liquidity risk profiles to investors
- Ongoing requirements for funds' investment strategies to be consistent with terms and conditions governing unit redemptions on an ongoing basis, both in normal and in stressed market conditions
- Regulations to be changed where necessary to allow adoption of tools such as swing pricing, redemption fees and other 'anti-dilution' measures of the sort referred to in Section 4.4.2, to limit incentives for investors to behave in ways akin to a bank 'run', and further direction on open-ended funds' usage of extraordinary risk management tools (e.g. gating)
- Introduction of stress testing requirements, both at a fund level and at a system-wide level

(b) *Leverage within investment funds*. Recommendations focused on all fund types that might use leverage (either through borrowings or derivative activities) and included:

- Proposals to develop simple and consistent measures of leverage as well more risk-based measures
- Improved information gathering (particularly of funds not subject to leverage limits or which pose significant leverage-related risks to the financial system)

(c) *Operational risks and challenges relating to transferring investment mandates in stressed conditions*. Recommendations focused on large or complex asset managers and/or ones that provide critical services. The recommendations involved introduction of requirements or guidance for such managers to have '*comprehensive and robust risk management frameworks and* *practices, especially with regards to business continuity plans and transition plans, to enable orderly transfer of their clients' accounts and investment mandates in stressed conditions*'.

(d) *Securities lending activities*, see Section 4.7. Recommendations focused on monitoring agent lending activities (i.e. where the asset manager lends securities not on its own behalf but on behalf of its clients), particularly provision of indemnities to clients. Where '*these monitoring efforts detect the development of material risks or regulatory arbitrage that may adversely affect financial stability, authorities should verify and confirm asset managers adequately cover potential credit losses from the indemnification provided to their clients.*'

Others have also tried to explore whether asset managers (rather than the funds that they manage) can contribute to systemic risk. Most in the industry are sceptical about the contribution that mainstream types of manager might make to systemic risk. However, this has not stopped e.g. IMF (2015a) exploring whether asset management involving plain-vanilla investment products (e.g. long-only equity portfolios) can create systemic risk. It argues that they can to some extent. However, it is not clear if they are differentiating between funds and the asset managers who manage them when forming this opinion. Neither do they explore how important are the risks identified or whether any large asset managers (or funds) should be classified as systemically important because of them. IMF (2015a) does not believe that that larger funds necessarily contribute more to systemic risk. Instead, in its view, investment focus seems to be a more important driver than size alone.

Others are not so sure that size is unimportant, as far as the asset managers themselves are concerned, since size may be correlated with the types of activity the manager undertakes. For example, Cetorelli (2015) notes that:

(1) the FSOC (in the USA) and the FSB (globally) have initiatives underway seeking to identify SIFIs in this sector (complementing similar activities already underway for banks and insurers). The focus of FSOC has typically been on products or activities whilst the FSB focus was originally on funds or firms (but has now expanded to cover activities).

(2) A stylised view of asset management is that systemic considerations can arise due to liquidity, investment concentration and possible fire sales within securities portfolios, fuelled by leverage (physical or synthetic via derivatives), counterparty exposures and 'run'-ability. The asset manager is not the owner of assets (and its own balance sheet activities are not

directly related to its asset management activities), so it is hard if this stylised view is correct to identify any obvious way in which asset managers themselves should present systemic risks.

(3) However, Cetorelli argues that there are two important changes taking place in financial intermediation. The first involves a shift from banking to shadow banking involving growth in length of intermediation chains. The second is an organisational transformation of banks into 'hybrid intermediaries', with the banks increasingly acting as specialty lenders, underwriters, broker-dealers and engaging in insurance, asset management etc. She thinks that the second change is partly a result of the first change (she thinks that in time a 'solution' to the first change will emerge out of the second, by integrating within the same holding company the long intermediation chains that the first is creating).

(4) There is nothing stopping non-bank financial firms from turning into hybrid intermediaries. So, another possible source of asset manager related systemic risk derives from the extent to which asset managers may be carrying out financial intermediation. Such activities are likely to be present only in material amounts within the largest asset managers, suggesting that it is with these players that the greatest systemic risk issues may arise.

4.5.4 Specialist Asset Managers Who Manage Non-Performing Loans Etc.

Usually differentiated for the purposes of systemic risk from other asset managers are asset managers who specialise in managing non-performing loans (NPLs) or other 'bad' banking assets, typically for what is known as a 'bad bank'. The establishment of a bad bank can allow the original bank transferring assets into the bad bank to be recapitalised, nationalised, liquidated or otherwise resolved. The transfer allows the original bank to have a clean (i.e. 'good') balance sheet. Banks can also set up internal bad banks, to isolate problem exposures into a division that specialises in the management of such assets.

Sometimes a separate (external) bad bank is set up for just one bank. At other times, it may be a more general purpose vehicle established to address problems at many banks at the same time, e.g. NAMA (Ireland) or SAMC (Spain).

Typically, if a bad bank is a separate entity, it will buy the NPLs or other problem assets from the original bank at market value or at some suitable approximation to it. This ensures that the economic losses already incurred by the original bank are borne by its shareholders etc. rather than by the bad bank. A bad bank can therefore be thought of as providing a specialised form of secondary market for such assets. As the assets are likely to be illiquid, identifying an appropriate market value for them can be tricky.

Original banks who are forced sellers may have little ability to get a good price for the assets being transferred. Moreover, doing so may crystallise economic losses that the bank may not previously have needed to recognise (if it had in effect been benefitting from some sort of regulatory forbearance, see Section 3.1.6). Conversely, the capital framework in a specific jurisdiction may incentivise firms where possible to sell such loans in a secondary market if one exists. It could, for example, require firms to write down the assets to zero or to a low value if the asset has been non-performing in some specific way for more than some specific period. The experience of past banking crises suggests that the faster a country forces its banking system to recognise the economic losses that have triggered the crisis, the faster the banking system typically recovers.

As with other types of illiquid asset, the interplay of these sorts of incentives sometimes results in a reasonably active secondary market developing but at other times hinders its development. Active markets only in practice develop if there are both active buyers and active sellers for the assets in question.

If an active secondary market does develop then much of the role of an asset manager responsible for a bad bank portfolio may involve deciding when and how to buy or sell such assets in this secondary market. This mirrors the role of traditional investment managers in other more liquid asset categories. If an active secondary market hasn't developed (and is unlikely to do so in the future) then most of the role of such an asset manager involves maximising the recovery value received from the asset. This might for example involve negotiating with borrowers with the aim of getting them to pay back more of the face value of the loan and more quickly etc.

There are other types of asset manager who may also spend a lot of their time 'sweating' illiquid assets to maximise value obtained from their investments. For example, private equity managers will typically interact closely with the companies they have invested in, including in some cases taking on directorships and other more hands-on roles (even if they are still also typically interested in the value eventually achieved on 'exit', e.g. sale to a third party or public listing). Outside the financial industry, the term 'asset manager' might be associated with someone who looks after some specific asset, e.g. some manufacturing plant or office, and who tries to ensure that the value in use achieved by its owner is as high as possible.

4.6 Shadow Banks

4.6.1 Introduction

I've mentioned above the apparent incongruity to an asset manager of assessing the systemic riskiness of a bond fund by reference to the ratio of the bonds it holds to the amount of cash it holds. *Shadow banking* is another area where language often used in the macroprudential community can appear to diverge from reality as seen by other market participants.

An example of this is the approach sometimes used within the macroprudential community to identifying the size of the shadow banking market. Often a 'broad' definition is adopted, which is then winnowed down to something more akin to how others interpret the term. This means that macroprudential policymakers can seem to start by deeming nearly every possible corner of the financial system that is not a bank to be a 'shadow' bank, using a broad definition capturing any investment fund, here called an 'other financial institution' (OFI), and even sometimes any insurer. Bond funds seem to figure prominently in such analyses even after some winnowing down has taken place, even though most in the asset management community struggle to associate bond funds with shadow banking.

This is not to deny the systemic relevance of shadow banking. To the extent that we might categorise the MMF sector anywhere within the financial system (other than lumped together with other investment funds) we might view it as involving shadow banking, since it offers products somewhat akin to those offered by banks, but it is not regulated in the same manner as banks are. Some of funds in this industry failed during the 2007–09 Credit Crisis, see Box 4.8.

Likewise, some of the mortgage conduits and other structured investment/finance vehicles covered in Section 4.4.6 seem to fall within an intuitively reasonable definition of shadow banking, which might be worded along the following lines: '*activities that are recognisably akin to traditional (commercial) banking activities but not carried out within a banking-like regulatory framework*'.

Much the same sort of definition is used in European Commission (2014), who defined shadow banking as '*a system of credit intermediation that involves entities and activities outside the regular banking system*'. At least it is if you view the main type of activity carried out by banks as being credit intermediation. The Commission noted that shadow banks are not regulated like banks, though their operations are like those of banks, as they:

(a) Take in funds similar to deposits;
(b) Lend over long periods and take in deposits that are available immediately;
(c) Take on the risk of the borrower not being able to repay; and
(d) Use borrowed money, directly or indirectly, to buy other assets.

According to European Commission (2014), shadow banks may include:

- Ad hoc entities such as securitisation vehicles or conduits
- Money market funds
- Investment funds that provide credit or are leveraged, such as certain hedge funds or private equity funds
- Financial entities that provide credit or credit guarantees, which are not regulated like banks or certain insurance or reinsurance undertakings that issue or guarantee credit products.

Included in the last bullet point would be financial corporations engaged in lending (FCLs). These include financial leasing companies, FCLs that grant credit or loans (including consumer credit, credit-card credit and hire purchase and companies that offer factoring services.

Commonly also included in the definition of shadow banking are activities such as securitisation, securities lending and repurchase agreement transactions which constitute an important source of finance for financial entities. We cover these sorts of activities in Section 4.7.

4.6.2 Monitoring Shadow Banks

Metrics that are used in the macroprudential community to monitor shadow banking activities are described in ESRB (2016a) and ESRB (2016b). The tendency is to focus both on entity based analysis and activity-based analysis. Many commentators, e.g. Impavido et al. (2011) and Haldane (2014) have noted the potential for risks to migrate from highly regulated sectors such as banking and insurance to less highly regulated sectors. They have also noted the potential for shadow banking to create, amplify or transmit systemic risk. The issue is summarised in FSB (2013):

> The 'shadow banking system' can broadly be described as 'credit intermediation involving entities and activities (fully or partially) outside the regular banking system' or non-bank credit intermediation in short. Such intermediation, appropriately conducted, provides a valuable alternative to bank funding that supports real economic activity. But experience from the crisis demonstrates the capacity for some non-bank entities and transactions to operate on a large scale in ways that create bank-like risks to financial stability (longer-term credit extension based on short-term funding and leverage). Such risk creation may take place at an entity level but it can also form part of a complex chain of transactions, in which leverage and maturity transformation occur in stages, and in ways that create multiple forms of feedback into the regular banking system.
>
> Like banks, a leveraged and maturity-transforming shadow banking system can be vulnerable to 'runs' and generate contagion risk, thereby amplifying systemic risk. Such activity, if unattended, can also heighten procyclicality by accelerating credit supply and asset price increases during surges in confidence, while making precipitate falls in asset prices and credit more likely by creating credit channels vulnerable to sudden loss of confidence . . . But whereas banks are subject to a well-developed system of prudential regulation and other safeguards, the shadow banking system is typically subject to less stringent, or no, oversight arrangements.

Addressing the potential systemic risks of shadow banking is viewed as a priority area for the FSB. The FSB's policy work to prevent the re-emergence of systemic risks from shadow banking has focused on the following areas, according to FSB (2013):

(a) *Mitigating risks in banks' interactions with shadow banking entities*. Topics here include scope of consolidation, treatment of large exposures and bank investments in the equity of such funds.
(b) *Reducing the susceptibility of MMFs to 'runs'*. This has included a range of proposals for MMFs, see Section 4.4.3.
(c) *Improving transparency and aligning incentives in securitisation*. The FSB believes that the complex structuring and multi-step distribution chains involved in much securitisation prevalent in the run-up to the 2007–09 Credit Crisis generated misaligned incentives. This encouraged a rapid and largely undetected build-up of leverage and maturity mismatches.
(d) *Dampening procyclicality and other financial stability risks in securities financing transactions*. This has included a range of standards on data collection and aggregation, rehypothecation, collateral valuation and management. It has also included policy recommendations relating to

central clearing and changes in the bankruptcy law treatment of securities financing transactions (SFTs). FSB (2014b) set out:

- Minimum standards on haircuts, i.e. margins, limiting the amount of financing that can be provided against a given security; and
- A framework of numerical haircut floors intended to prevent the erosion of margins below minimum levels when non-banks obtain leverage using SFTs backed by non-government securities.

(e) *Assessing and mitigating systemic risks posed by other shadow banking entities and activities*. The FSB has recognised that shadow banking entities and activities take a variety of forms and evolve over time. Its policy framework has therefore sought to include assessments based on economic functions (or activities) and corresponding information-sharing processes capable of responding to this evolution.

According to Tirole (2015), activities usually considered to be shadow bank-like in nature involve credit, maturity, and liquidity transformation taking place without direct and explicit access to public sources of liquidity or credit backstops. Prior to the development of modern central banks, there were usually no such public sources of liquidity or credit but neither were there typically any explicit bank capital regulatory frameworks, so there was probably little obvious that differentiated a 'bank' from a 'shadow bank'.

Entities potentially caught within such a definition include:

(a) Financial vehicle corporations (FVCs) /Structured Investment Vehicles (SIVs) (the two terms appear to mean much the same thing)
(b) Security and derivative dealers (SDDs)
(c) Financial corporations engaged in lending (FCLs)
(d) Specialised financial institutions
(e) Central clearing counterparties (CCPs)
(f) Investment funds including MMFs, hedge funds and bond funds

Activities deemed to be like shadow banking include:

(a) Securities financing transactions, including repo market activity and securities lending (see Section 4.7)
(b) Some types of derivatives activities, particularly ones that involve synthetic leverage and credit enhancement

Of course, devising regulatory frameworks that try to address systemic risk in one area of shadow banking may merely result in the activity moving elsewhere. For example, Johnson (2014) notes that money market fund investors appear to have considered switching to private unregulated CNAV vehicles if European regulators made it impractical for mainstream cash funds to retain CNAV characteristics. Regulatory policy inherently struggles with situations where regulators want one outcome but many customers want another.

It seems likely that shadow banking will continue to be a focus of policymakers' attentions for some time, as the sector seems to be growing relative to the (non-shadow) banking sector. However, there are few simple answers to question of how avoid possible regulatory arbitrage between the two.

4.7 Securities Financing

4.7.1 Introduction

So far, we have primarily concentrated our analysis on *vehicles* or other clearly defined *entities* within the financial services industry. There are several types of *activities* that cross sectoral boundaries that also come in for scrutiny from policymakers. One such area is securities financing, particularly if it involves *rehypothecation*, see Section 4.7.4. ESRB (2014b) explores the potential relevance of securities financing to systemic risk. It provides a summary of the EU securities financing markets, including the extent to which banks and non-banks rely on securities financing transactions (SFTs), how fungible the collateral received is, how interconnected the market appears to be and the extent to which players in the market facilitate credit growth, undertake maturity transformation or take on liquidity risk.

The two main types of SFTs that arise in practice involve securities lending (otherwise known as stock lending if it involves the lending of equities), see Section 4.7.2, and use of repurchase agreements (commonly shortened to 'repos'), see Section 4.7.3.

The European Commission (2014) proposed that:

(a) All SFTs be required to be reported to a central database (allowing supervisors to understand better the links between banking and shadow banking entities);
(b) There should be greater transparency via improved provision to investors of information on the practices of investment funds engaged in SFTs; and

(c) There should be greater transparency on rehypothecation, see 4.8.4, including minimum conditions to be met by the parties involved, written agreements and prior consent. The Commission's ideas in this area were like those of the FSB, see e.g. FSB (2013).

Conversely, it should be remembered that SFTs contribute to availability of collateral and, as Baranova et al. (2016) note, collateral availability has an important impact on liquidity. Measures that deplete collateral availability might have undesirable effects on liquidity, the scarcity of which was a particularly important driver of the severity of the 2007–09 Credit Crisis.

4.7.2 Securities Lending

Typically, securities lending refers to the lending of securities by one party to another, so that the second party can honour a contract it has entered into that requires it to deliver physical securities to someone. Another name for securities lending if it involves stocks (equities) is stocklending. For example, suppose an investor wishes to short a stock (i.e. to be in a position where he or she will benefit if the market price of the stock falls). This can be achieved by borrowing the security, physically selling the security, waiting for the security price to fall and, once it has done so, buying it back in the market and returning it to the security lender. Market makers may also temporarily borrow specific securities to settle transactions. For example, if an upstream trade that they had expected would provide them with a specific security fails (e.g. because there is some mix up in the process for settling the trade) then they may want to borrow securities from somewhere to avoid creating a knock-on failure further downstream on other trades.

Loans will be governed by a Securities Lending Agreement which generally requires the borrower to provide collateral to the lender, usually in the form of cash or non-cash securities which have a value equal to or greater than the loaned securities plus an agreed margin or 'haircut'. The exact types of non-cash securities that the borrower can post as collateral will be specified in the Agreement. Usually lenders seek to limit acceptable forms of collateral to types that are relatively liquid (since the purpose of the collateral is to protect the lender in the situation where the borrower is unable to return the borrowed security as planned). When liquidity is in short supply market participants can in principle run out of collateral that they may be required to post under securities lending transactions. This is a form of 'run' that can in principle apply to nearly any market participant active in securities lending.

Loans can be 'term' loans for specific periods of time or 'overnight' loans, in which case the borrower needs to return the security the following (business) day. As payment for the loan, the parties negotiate a fee, usually an annualised percentage of the value of the loaned securities. The borrower also needs to provide the lender with any cash income generated from the security, e.g. dividends to which the loaned security becomes entitled during the duration of the loan.

From the lender's perspective, the loan involves the lender retaining the economic interest in the security being loaned (as eventually it will be returned as previously, together with any dividends earned etc.) plus receipt of a fee for lending the security to someone else. In securities lending, the borrower is usually liable to return the loaned securities when specified (in return for receiving back its collateral), so the lender will usually only be out of pocket if both the borrower defaults and the collateral held is then inadequate to make good any losses due to unfavourable market movements that the lender then suffers until the position can be made good for the lender. Lenders may include mutual funds, insurance companies, pension funds and other large investment portfolios.

Legal ownership of the security typically changes hands when the security is lent out, which may cause complications in terms of exercise of voting rights if needed. Equity securities lending tends to peak in some jurisdictions around the time of dividend payments. Local investors in a security may suffer less tax than an overseas investor. The tax savings available by temporarily having the security in the hands of a local investor when the dividend is paid can then be split between both parties.

A further complication arises with how any cash 'generated' by the stock-lending activity might be invested. The existence of the cash may not be particularly visible to senior management within the organisation carrying out the stock lending. This can potentially create incentives to 'search for yield' if stocklending is being implicitly viewed as a revenue generating activity. Investment of this cash in other than a risk-free investment is little different to investing using leverage or margin.

4.7.3 Repo Agreements

Closely allied to stocklending are repo agreements. In a (two-party) repurchase agreement (or 'repo'), one party sells to the other a security at some price and simultaneously agrees to buy the security back later for another price. Overnight repos involve the reversal taking place the day after the original sale. A term repo extends the time-period for a fixed time length, which can be up to e.g. three months. An open repo has in effect no initial end date but one or both parties have the option to terminate the transaction within pre-agreed time frames.

A reverse repurchase agreement ('reverse repo') is the same as a normal repo, just seen from the opposite side of the transaction. In practice, the party initially selling the security (i.e. the party entering into the repo) receives cash when doing so, whilst the party initially buying the security (i.e. the party entering into the reverse repo) is handing over cash.

The economic substance of a repo can therefore be thought of as involving the repo-er obtaining a secured loan from the reverse repo-er, even though the nominal form of contract involves a sale and subsequent repurchase of a security.

The maximum security to the lender is typically provided by government debt, so a substantial proportion of repo activity involves such instruments. If there is a supply /demand imbalance for a specific government bond being used in the market then the repo rate for a 'specified delivery' repo involving that bond will diverge from those for the generality of such repos. This is known as the bond in question going 'special', see e.g. Fisher (2002). The 'general collateral' repo (GC repo) rate is the rate for secured borrowing of funds using the generality of such bonds, and is an important interest rate measure in some markets. Tri-party repos are essentially a basket form of transaction, allowing for a wider range of instruments to be used in the secured lending process. In a tri-party repo a third party clearing agent or bank is positioned in between the 'seller' and the 'buyer' in the repo. This agent controls the security pool subject to the lending agreement and processes payment flows between seller and buyer.

4.7.4 Rehypothecation

Rehypothecation involves the process of the re-using collateral posted by a debtor to back the creditor's own trades and borrowing. For example, a bank might have a (hedge fund) client to which it was providing 'prime brokerage' services. These might include lending the hedge fund some securities, in return for which the hedge fund posts back some other securities. If the collateral can be rehypothecated then the securities posted back to the bank can be loaned to other clients.

In the extreme, and with a large prime brokerage business, a bank could source a significant amount of loans 'internally' from other clients of the same prime brokerage business, particularly if some hedge funds wanted to be long and others short the same stock, with the bank taking a fee for doing so.

Before the collapse of Lehman Brothers, the IMF calculated that US banks were receiving over $4 trillion of funding by rehypothecation, with only $1 trillion of original collateral being used, implying that collateral was being rehypothecated on average several times over, see e.g. Singh and Aitken (2010).

After the Lehman Brothers collapse, hedge funds became warier of allowing their collateral to be rehypothecated and in some cases prohibited it entirely (in effect requiring the prime broker to hold any collateral posted in a separate segregated account and prohibiting the collateral from being transferred anywhere else except back to the borrower).

Rehypothecation is generally considered have several systemic risk issues including:

(a) When rehypothecation takes place, the ownership of the financial instruments is replaced with a contractual claim to the return of equivalent financial instruments. In practice, this is akin to an unsecured obligation.
(b) Rehypothecation allows the same financial instruments to create multiple obligations that interconnect different market participants. These obligations amount to a multiple of the value of the rehypothecated financial instruments, creating concerns for financial stability because of amplified leverage and procyclicality.
(c) Rehypothecation forms complex chains of transactions hidden from market participants and regulators. This increases the possibility of a run on a financial company if there are concerns about its creditworthiness.

4.8 Central Counterparties and Other Market Infrastructure Elements

4.8.1 Introduction

Exchanges and clearing houses or central counterparties (CCPs) facilitate market transactions carried out between other market participants. We might initially view them as playing quite different roles to banks, insurers or asset managers /investment funds. In the EU, the primary legislation regulating these entities is the Markets in Financial Instruments Directive (MiFID), rather than the CRD, see Box 4.13.

However, the shift towards central clearing of derivatives, see Section 6.7, has highlighted similarities between such organisations' business models and those of some parts of the firms that use their services. One lesson we can draw from modern business trends is that established business relationships can be destroyed by adoption of disruptive new technologies and business processes, see Chapter 7. Exchanges used to be mainly specific to individual jurisdictions. However, most have now shifted to for-profit business models and have become increasingly global in nature, just like most of the larger firms using them. A core role of exchanges is to facilitate access to market liquidity. They are not, however, the only players who perform this function and so can be disintermediated by (or can disintermediate) others.

At their most basic, we can think of exchanges as locations where buyers and sellers come together and trade financial instruments or other exposures. An important economic role that exchanges provide is *price discovery*, i.e. the process of identifying a fair price that balances supply and demand. The usual mental image of such an exchange is one involving 'open outcry' pits like the ones that stockmarkets and commodity markets used to use. Nowadays nearly all trading on such exchanges is carried out electronically, and open outcry is rarely used.

Box 4.13: MiFID

The Markets in Financial Instruments Directive II (MiFID II) and the associated Markets in Financial Instruments Regulation (MiFIR) cover EU regulation of:

(a) Investment intermediaries providing services to clients in relation to shares, bonds, units in collective investment schemes and derivatives (collectively 'financial instruments'); and
(b) The organised trading of financial instruments

An introduction to MiFID II and MiFIR is given in FCA (2014b).

The primary objectives of the initial MiFID Directive (MiFID I) were to increase competition, improve investor protection and implement EU passporting. The MiFID II package introduces a range of further measures which seek to address consequences of MiFID I and issues raised by the 2007–09 Credit Crisis. It aims to deliver a safer, sounder, more transparent and more responsible financial system and to ensure a more integrated, efficient and competitive EU financial market. It includes requirements such as:

(a) *Extension of MiFID rules to additional products and services*. MiFID II extends MiFID-like provisions to a wider range of retail financial products including structured deposits issued by banks and financial instruments issued by investment firms.

(b) *Harmonisation of requirements applying to different types of trading venue*. MiFID II aims to ensure that all organised trading is conducted on regulated trading venues, applying identical pre- and post- trade requirements on every type of venue.

(c) *Amending MIFID exemptions*. MiFID II defines more precisely what activities are exempt from MiFID requirements. These typically relate to some own account activities not carried out for third parties.

(d) *Upgrades to market structure frameworks*. Various changes are being introduced to take account of recent market developments. For example, to capture 'dark pool' operators and other similar trading systems (e.g. inter-broker-dealing systems), a new category of trading venue called an *organised trading facility* (OTF) has been introduced for non-equity instruments (e.g. bonds, derivatives, structured products). See Section 8.7 for further details on dark pools. OTFs are like the *swap execution facilities* (SEFs) introduced in the US by Dodd-Frank. Derivatives that are subject to central clearing requirements, see Box 8.2, will need to be traded on eligible platforms such as OTFs, *multi-lateral trading facilities* (MTFs) or *regulated markets* (RMs) instead of over-the-counter (OTC).

(e) *Corporate governance*. Some improvements have been introduced e.g. requiring management boards to have sufficient knowledge and skills to comprehend the risks associated with the firm's activities.

(f) *Investor protection framework*. Requirements relating to provision of investment advice and portfolio management have been enhanced.

(g) *New requirements on trading venues*, to publish e.g. annual execution quality data.

(h) An *improved small and medium enterprise (SME) regime*, to assist SMEs in obtaining financing.

4.8.2 Central Clearing

Each party to a financial transaction is potentially exposed to the risk that the other party will fail to honour the transaction. Often exchanges have member firms or clearing participants (also known as 'clearing firms' or 'clearing brokers') that outside parties use to trade on their behalf within the exchange. Clearing houses stand between clearing firms. Their purpose is to reduce the risk (to other clearing firms) of one or more clearing firms failing to honour their trade commitments.

Originally, clearing houses were closely associated with a specific exchange. Exchanges themselves tended to focus on specific types of instrument. However, ongoing merger activity between exchanges has resulted in some clearing houses ending up handling lots of instrument types (including potentially both physical securities, such as equities and bonds, and derivatives, both financial derivatives and commodity derivatives). The clearing process (for e.g. an exchange traded derivative) involves:

(a) Two clearing firms, say A and B, entering into a transaction between themselves
(b) This transaction is 'given up' to the clearing house within a specified short time window after the trade has been agreed, possibly almost instantaneously depending on how the exchange operates)
(c) The original transaction is 'novated' by the clearing house into two separate contracts, one for each of the original parties. Each of A and B ends up with a separate transaction in which their counterparty is the clearing house, so that A and B no longer need to worry about the risk of the other defaulting, and only need to worry about the risk of the clearing house defaulting.

An exchange-based clearing house in effect centralises the risk of settlement failures relating to transactions executed on that exchange onto itself. It needs to be properly managed and well-capitalised. Many such clearing house guarantee funds are capitalised in part with collateral supplied by their clearing firms. If a settlement failure occurs then the clearing firm involved may be deemed to have defaulted, and default procedures may be triggered which may include liquidation of the defaulting firm's positions and collateral. The clearing house may also draw on its guarantee fund to settle trades on behalf of the failed clearing firm.

Central counterparties (CCPs), also known as central clearing counterparties, carry out much the same sorts of roles as clearing houses but (additionally) for over-the-counter (OTC) markets.

4.8.3 Systemic Risks Expressed by Clearing Houses and CCPs

An obvious risk is that a CCP in some sense centralises onto itself possible credit risks otherwise scattered between market participants. Clearing firms and CCPs typically require as part of the novation process the posting of collateral to themselves by each participant to the original transaction. Some of the risks that might arise if a CCP failed are described in Duffie (2015). They include:

(a) *Contagion to clearing firms and others*. A large enough loss would exhaust the CCP's capital and other guarantee funds resulting in contagion to potentially all its clearing firms and others.

(b) *Loss of continuity of critical clearing services*. Its failure would likely cause parts of the market to seize up.
(c) *Disorderly unwind of the CCP's assets*. A CCP failure would likely involve a fire sale assets because some of its positions would likely include some that would be temporarily impaired or rendered less liquid due to its own failure. It has 'wrong way' risk with itself!

Duffie (2015) also wonders whether an orderly resolution of a CCP might be impractical. Partly this is because a CCP failure would likely create systemic uncertainties and worries. But partly it is because typical (derivative) transaction terms that a CCP might novate include early termination rights which counterparties might seek to impose on the CCP if the CCP ran into difficulties, potentially disrupting its own resolution.

Some of these risks might be reduced if market participants have the effective ability to select between multiple CCPs for the same trade. However, this then introduces challenges relating to interoperability, i.e. central counterparties needing in some circumstances to give up trades to each other.

One way of gaining a flavour of the complications that might arise if a CCP were to fail is to note that prior to the introduction of central clearing for standardised derivatives, investment banks were perhaps the closest analogue we had to such entities. Many investment banks had long and short exposures to similar or identical instruments with a variety of third parties as counterparties and sought collateral from them to mitigate counterparty risk. So, failure of a large CCP can be expected to create the same sorts of uncertainties and dislocations as failure of a large investment bank handling similar instrument types. Lehman Brothers was one such investment bank!

This then raises the question of whether CCPs should be subject to external regulatory capital requirements or other bank-like regulatory requirements. Otherwise there is a risk that they undercapitalise themselves (or rely too heavily on guarantee funds from others, which if called upon in a systemic risk event might cause wider contagion). Even if it is not appropriate to set specific Pillar 1 capital requirements for such entities, it can be argued that it would be desirable for them to be subject to some form of ICAAP /ORSA requirement, forcing them to identify the amount (and type) of capital that they intrinsically need to face the risks to which they are exposed.

CCPs such as the Depository Trust & Clearing Corporation (DTCC) are aware of these debates. For example, DTCC (2015) notes:

(a) DTCC is an important element of financial market infrastructure and as a result has been deemed systemically important, reflecting its interconnectedness. Regulators /policymakers are important stakeholders for such players.
(b) It believes that any similar organisation aspiring to play a central role in market activity can also expect to be deemed systemically important, including any big bank that effectively provides some of the facets of a central clearer.
(c) Network-related studies undertaken in the wake of the 2008 global financial crisis suggest that financial networks tend to be robust yet fragile, absorbing shocks up to a certain tipping point, beyond which they spread risks rather than contain them. It is therefore unclear under which circumstances interconnectedness promotes or impairs financial stability. Policymakers' initiatives to address interconnectedness risks have tended to focus mainly on increasing the resilience of the most interconnected (and therefore, in DTCC's opinion, the most systemically important) financial institutions. Similar measures have been introduced to enhance the resilience of critical financial market infrastructures.
(d) Identifying and assessing interconnectedness risks promotes a broader and deeper understanding of the threats to an organisation. The paper proposed guidelines for such activities.

4.9 Governments /Sovereigns

4.9.1 Introduction

No sooner, for Eurozone policymakers, had some semblance of normality returned after the 2007–09 Credit Crisis and another systemic risk issue came to the fore. This was the Eurozone sovereign debt crisis, see Box 4.14. Several Eurozone member states (e.g. Greece, Portugal, Ireland and Cyprus) found themselves for a variety of reasons unable to refinance their government debt or able to bail out over-indebted banks under their national supervision. To do so they required assistance from third parties such as the ECB, other Eurozone countries or the IMF. Details varied by country. Some had faced property bubbles and had already bailed out their banking systems, adding to their own debts.

Box 4.14: The Eurozone sovereign debt crisis

Following on from the 2007–09 Credit Crisis, the Eurozone sovereign debt crisis has so far involved the following:

(a) The euro was established by the EU under the Maastricht Treaty in 1992 but only formally came into existence on 1 January 1999. Most of the major EU member states joined the Eurozone, although noteworthy absentees included UK, Denmark and Sweden. Common euro notes and coins were introduced across the Eurozone on 1 January 2002. Greece joined the Eurozone in 2001, Slovenia in 2007 and Malta and Cyprus in 2008. Several Eastern European countries joined the EU on 1 May 2004, and in 2009 some joined the Exchange Rate Mechanism as a precursor to joining the Eurozone. In late 2008 EU leaders agreed a large stimulus plan to help boost growth in the EU following the 2007–09 Credit Crisis.

(b) In November 2009 concerns about some EU member states' debts started to grow following the Dubai sovereign debt crisis (which itself was in part triggered by the 2007–09 Credit Crisis), with specific focus on Greece's debts.

(c) In early 2010 an EU report condemned 'severe irregularities' in Greek accounting procedures resulting in a substantial revision upwards in its budget deficit, to several times the maximum allowed by EU rules. Greece unveiled a series of austerity measures aimed at curbing the deficit. Concerns started to rise about all heavily indebted Eurozone members, particularly Portugal, Ireland, Greece and Spain. The euro continued to fall against the US dollar and the UK pound.

(d) The Eurozone and IMF agreed a safety net to help Greece and subsequently some emergency loans. Greek borrowing costs reached further record highs and the EU announced that the Greek deficit was even worse than thought. On 2 May 2010, the Eurozone members and IMF agreed a EUR 110 bn bailout package to rescue Greece. In November, they also agreed a EUR 85 bn bailout package for Ireland.

(e) A permanent bailout fund, called the European Stability Mechanism (ESM), worth about EUR 500 bn was established in February 2011. Portugal received a bailout in May. Yields on Spanish and Italian government bonds rose sharply and on 7 August the European Central Bank (ECB) said it would buy their bonds to try to bring down their borrowing costs. The European Commission predicted that economic growth in the Eurozone would come '*to a virtual standstill in second half of 2011*'.

(f) In October 2011, Greece toyed with default but eventually Eurozone finance ministers released a tranche of Greek bailout loans potentially saving the country from default. Weeks of negotiation took place in early 2012, leading to a second bailout package and finally a Greek default on its debt (at the time the largest default in history by a government). As part of the process, private holders of Greek government bonds (banks, insurers and investment funds) were required 'voluntarily' to accept a bond swap with a sizeable nominal write-off. In the meantime, a 'fiscal pact' was agreed by the EU. The UK abstained, as did the Czech Republic, but the other 25 member states signed up to new rules making it harder to break budget deficit targets.

(g) Some calm then returned to the Eurozone, with the ECB announcing on 6 September 2012 free unlimited support for all Eurozone countries involved in a sovereign state bailout/precautionary programme from the ESM (and its precursor, the European Financial Stability Facility). This calm was punctured in 2013 by a near collapse of Cyprus's largest banks amid heavy exposure to Greek debt. Originally an unprecedented one-off levy of 6.7% of deposits up to EUR 100,000 was proposed (more for larger deposits), but this was eventually abandoned.
(h) Return to economic growth enabled Ireland and Portugal to exit their bailout programmes in July 2014. Spain never officially received a bailout as its rescue package from the ESM related to bank recapitalisations rather than financial support for the state itself.
(i) However, a worsening economic outlook in Greece in late 2014 led to rising political unrest and to the Greek government refusing to respect the terms of its earlier bailout agreement. Months of apparently fruitless negotiations (during which Greece defaulted on loans from the IMF) eventually led to a third bailout in July 2015 requiring further reforms. These included changes to pensions and market-based reforms to the economy designed to make it more competitive.

From time to time worries resurface about the structural stability of the Eurozone. Some commentators argue that it lacks the features of a fiscal union that they consider ultimately underpin a stable long-term monetary union.

The strains associated with this crisis can perhaps be gauged by how Greece's economy has fared over the last decade. House of Commons (2015) indicates that Greece's economy was over a quarter smaller in late 2015 than it was in 2008. There is a considerable diversity of views about how much of this fall has been a consequence of the stringent bailout conditions Greece has faced.

4.9.2 Systemic Risk Implications

It isn't my purpose here to consider political ramifications of the Eurozone sovereign debt crisis. Instead I want to highlight some points relevant to systemic risk that can be derived from the experience. These are:

(a) Sovereigns are not risk-free. Indeed, history suggests that the Eurozone sovereign debt crisis has not been as bad as it can get, see Fig. 4.3.
(b) There is often a close alignment between the overall financial health of a country's banking sector and the overall financial health of the country itself. This linkage is called the *sovereign-banking sector nexus*. Reinhart and Rogoff (2009) highlight the high proportion of times that past sovereign debt crises have gone hand in hand with a banking crisis within the relevant state.

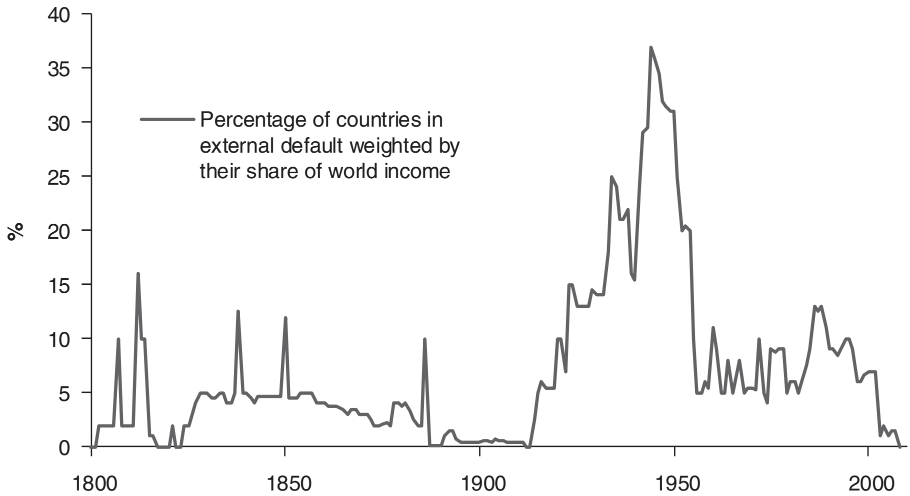


Fig. 4.3 Percentage of countries in external default weighted by share of world income

Source: Nematrian. Adapted from Reinhart and Rogoff (2009)

At its most basic, the Eurozone sovereign debt crisis can be viewed as another example of this type of interplay. Countries cannot afford to have their financial systems (and particularly their banking systems) fall over. Try as they might to avoid doing so, when push comes to shove they often seem to end up bailing out their banking sectors. If the bailouts needed are sufficiently large then the state itself can run into difficulties. Conversely, if a state runs into trouble then there are lots of ways in which it can lean on its own banking sector to support it. Most of these will result in the banks themselves running into serious problems if the state itself then defaults.

(c) Sovereign defaults or near defaults are almost inevitably messy affairs surrounded by considerable uncertainty of outcomes. Some of the Eurozone member states most caught up in the crisis were able in time to exit their bailout programmes with seemingly relatively little long-term financial and economic damage. Others, such as Greece and Cyprus, still (at the time of writing) seem to be struggling, having restructured their debt and in some cases forced debtholders and depositors to accept some share of the losses involved.

(d) There has been considerable angst elsewhere in the EU about the risks posed to other member states' financial systems by their exposures to troubled sovereigns and the local banking sectors of these sovereigns. To a substantial extent these risks have now been removed from the financial systems of other EU member states but only by the risks being pooled together at a central EU level, e.g. on the balance sheet of the ECB.

This has only so far been possible given political willingness to do so from more creditworthy Eurozone sovereigns. However, these more creditworthy sovereigns could one day say 'enough is enough'. Unwinding positions on the ECB's balance sheet in such circumstances could prove very problematic.

(e) On the theme of interconnectedness, essentially everyone in a country is interconnected with the government of that country. The level of interconnectedness may typically be greater for the banking sector, given the nexus referred to above. But it applies to the rest of a state's financial sector too (to the extent that it has one). Indeed, it applies to the entire economy. As John Donne noted: '*No man is an island entire of itself; every man is a piece of the continent, a part of the main*'. Societies stand or fall by the way in which they get the interests of different individuals to work towards a greater whole. If a systemic risk issue threatens the financial viability of the state itself then essentially everyone becomes involved one way or another. Almost inevitably, a large enough systemic risk event crosses into this territory.

Is it possible to eliminate this dependency? And even if it is, is it appropriate to try? Britain arguably became a great power in the 1700s and 1800s because it had a better developed financial system than its competitors, i.e. it took advantage of this dependency for political gain, see e.g. Ferguson (2009). The same is probably also true of the USA in the 1900s.

4.9.3 Potential Remedies

Macroprudential policymakers generally tread carefully when it comes to sovereign risk because of the politics involved (even given its clear linkage to systemic risk and financial stability as noted above). In public policy areas where policymakers fear to tread, a tried and tested approach is to float ideas via think tanks or the equivalent. An apparent example of this is EPSC (2015), a paper published by the European Commission's own in-house think tank. This paper explores ways in which EU banks could be made less dependent on the ongoing creditworthiness of sovereign states within which they are domiciled. It concluded that:

(a) Sovereign-bank links were at the heart of the Eurozone sovereign debt crisis and remain a severe weakness, as banks are still typically heavily overweight their own sovereign.
(b) Zero-risk weighting (for banks) of their holdings in EU sovereign debt creates vulnerabilities (as does the exemption for sovereigns from existing large exposure requirements included in the CRD). Zero-risk weights for such exposures are not global regulatory constraints. The Basel Accords do not prescribe zero risk weights for bank exposures to sovereigns. Rather, it allows them as an option.
(c) Diversification of portfolios leads to sizeable reduction of risk.

Often in such documents, the possible policy responses are aired to create debate. EPSC (2015) thinks that amending the large exposures regime (by including sovereigns in existing large exposure limits applicable to other asset classes) is the simplest approach. Others have suggested:

(1) applying non-zero risk weights (although how such weights might be identified in practice is less clear); or
(2) promoting the development of ESBies (now renamed sovereign backed securities), which would be formed by the senior tranche of a tranched structure investing in a portfolio of EU government bonds, see Section 3.5.1, and then restricting instruments that are zero risk-weighted to these instruments.

Most of these policies become less relevant in a unified state such as the USA. It is not that (local or external) sovereign risk disappears, but many of the political complexities introduced by separate sovereign states each using the same currency are diminished.

4.10 Sovereign Wealth Funds and Other Long-Term Unconstrained Investors

We've concentrated in the previous Section on the liability side of a sovereign's balance sheet. Some sovereigns also have substantial asset portfolios. Alongside pension funds, an Annex to FSB (2016) explores the potential systemic risk characteristics of sovereign wealth funds (SWFs). SWFs are special purpose investment funds or arrangements that are owned by a government. They are often created for broader macroeconomic purposes, may have considerable investment freedom and may not have specific liabilities as such that they are obliged to honour. In times of recent financial stress, some SWFs were a common port of call for executives of troubled financial organisations who were hoping to obtain support from stable long-term investors without needing to be bailed out by their own governments.

The aggregate AUM of SWFs is much smaller than for open-ended investment funds or pension funds but the SWF market is quite concentrated, with the largest 15 funds accounting for about 85% of total AUM of SWFs, according to FSB (2016). Some SWFs are very large. For example, at the time of writing, the Norwegian Government Pension Fund Global (commonly referred to as the 'Oil Fund' and previously known as the Petroleum Fund of Norway) was perhaps the largest owner of equities in Europe. Its purpose is to invest parts of the surplus generated by Norwegian oil taxes and license fees, given the longer-term decline in revenues the country expects from its oil fields. Although its name might suggest that it is like the pension funds referred to in Section 4.3, it is not a pension fund in the conventional sense. Its assets have come from oil profits not pension contributions, and it does not have specific (pension) liabilities as such, merely commitments over time to disburse assets for the greater good of the country and its population.

In a formal legal sense, some funded public sector pension funds that would more recognisably fit the mould described in Section 4.3 also have similar investment freedoms to the Norwegian Oil Fund. However, in practice, their assets are often earmarked to meet more explicitly quantifiable pension liabilities to specific individuals. As noted in Section 4.3, the long-term investment horizons of such funds should potentially contribute positively to financial stability, because these horizons allow such funds to take contrarian stances when everyone else is panicking. Conversely, as noted in Section 4.3.4, this possible positive contribution to financial stability seems to be less about the long-term-ness of the pension financing vehicle in question and more about the flexibilities it may have over how to interpret its liabilities.

From this perspective, some SWFs (and other funds with only loosely defined liabilities, such as some long-term charitable and educational foundations) should potentially be capable of being positive contributors to financial stability, provided they are prepared to adopt the sorts of contrarian investment stances implicit in acting as counterweights to systemic stresses.

FSB (2016) seems relatively cautious about whether in practice SWFs are likely to contribute significantly to broader financial stability. It notes that SWF management practices in relation to liquidity risk and leverage may vary and that some SWFs could be exposed to potentially significant withdrawals depending on the fiscal strength of the government concerned and, possibly, the quality of their own governance practices.

This begs a broader question. We questioned in Section 4.3 whether pension funds were nowadays likely to be white knights who would ride to the rescue if financial instability beckoned. This was because their broader investment freedoms had in many cases become more constrained over time due to their liabilities becoming more explicit (and more expensive in today's low interest and higher life expectancy world). Policymakers also seem to be expressing caution on the ability of SWFs to do so either, even though in some cases their liabilities are more like the best endeavours aspirational pension liabilities of the past. Are there *no* investor types who are likely to offer more effective counterweights to financial instabilities? After all, the last century has seen the largest growth in cumulative asset bases of societies in human history. Surely this has created some flexibility?

There are some organisations that have plenty of spare cash. Financial Times (2016c) reports that US non-financial companies held c. \$1.7 tn of cash on their balance sheets as at the end of 2015, with just five US tech titans (Apple, Microsoft, Alphabet, Cisco and Oracle) holding more than half a trillion dollars between them. Corporations typically have very flexible powers over what they do and how they use their balance sheets. Some corporations, like Berkshire Hathaway, do like to stress the long-term nature of their investment and business strategies and do seem to feature reasonably frequently as white knight investors when financial firms seek assistance to tide themselves over (hopefully temporary) challenges.

However, Financial Times (2016c) also reports that there were non-insignificant cash holdings even before the 2007–09 Credit Crisis. And as the accumulated economic capital of societies has grown so too have the balance sheets of financial institutions. There is typically a positive correlation between the wealth of a nation and the amount of insurance its populace buys. As wealth grows, so too do the savings that add up to that wealth, and the likely needs of individuals for other sorts of financial services.

So, it seems likely to be optimistic to assume that there are natural contrarian players willing to take the opposite side whenever financial instabilities strike. If the price is right then some do appear to be willing to do so. However, if the tsunami model of systemic risk is right then systemic risks will largely crystallise only when underappreciated vulnerabilities come to light. This is exactly the time when such investors are likely to want to charge the most for lending a hand to the financial system. As we noted in the Preface, financial stability is a public amenity and it is therefore all too easy to believe that the costs of keeping it functioning smoothly should be borne largely by others. Believing that others will do so and at the same time undercharge us for the privilege just because their asset base makes them capable of doing so looks suspiciously like a case of falling into this trap.

4.11 Credit Rating Agencies Etc.

The financial sector does not consist solely of financial institutions such as banks, insurers, pension funds and investment funds. Viewed more broadly, it consists of many other organisations including accounting firms, actuarial firms and other consultancies offering all sorts of services to the institutions themselves. Many of the regulatory burdens imposed on firms are not practical for the firms to address themselves; they need expert advice on how to implement regulatory change, enhance business models, apply IT etc.

One type of organisation that incurred some flack during the 2007–09 Credit Crisis were credit rating agencies, called ECAIs in the EU. A natural way of imposing limits and other controls on a firm's credit risks is by reference to the deemed creditworthiness of the credit exposures. An obvious source of views on this are organisations who are paid to express such views. By analysing lots of different firms, credit rating agencies are also capable of gaining economies of scale and of delivering enhanced comparability across firms. However, during the Crisis, many investors lost confidence in the reliability of the credit ratings that these agencies were publishing, particularly ones ascribed to structured products such as CDOs, CLOs and mortgage conduits.

Many regulatory frameworks refer to ratings produced by ECAIs. It has proved very difficult to eliminate such references, despite efforts by the regulatory community to do so. An alternative approach that the EU has been pursuing is to regulate the business activities of ECAIs, presumably in the hope that this will lead to more robust ratings (and mitigation of the conflicts of interest to which ratings agencies are perceived to be subject).

These conflicts of interest arise because the agencies are typically paid by the firms they are rating rather than the ones who make direct use of the ratings they supply. Some commentators have suggested that it would be desirable if ECAIs were instead paid by investors rather than by the issuers of the instruments being rated. However, investor-paid ECAIs only play a marginal role in today's credit markets. Bongaerts (2015) seeks to explore why this is so, using a heterogeneous competition model approach. He concludes that investor-paid ECAIs suffer from several types of free-riding and are generally not competitive enough to compete effectively against issuer-paid ECAIs.

Investor-paid credit analysis does of course get carried out, in the form of proprietary in-house research carried out by asset managers exclusively for the benefit of their own clients. However, this type of research is not particularly appropriate for regulatory capital purposes. It is deliberately seeking out credit exposures believed to have good (or poor) creditworthiness relative that implied by credit spreads, with the aim of then going long (or short) the relevant instrument.

4.12 The Physical Ecosphere

So far, we've implicitly adopted the view that the financial sector is largely self-contained, but that if it suffers a large enough loss then the loss can cascade over into and affect the 'real' economy. But what about the possibility that the chain of causality can work in reverse? Suppose the initial problem is an issue in the 'real' economy, which propagates from there into the financial sector and thereafter creates issues across both.

This type of logic is perhaps most noticeable with *systemic environmental risk*. This is the notion that failure to address ecological issues such as global warming may lead to future sharp economic shocks that may in turn create systemic problems within the financial sector and then beyond it.

It is not difficult to detect political angles within the logic put forward in this area. Some of the leading proponents are politicians or others closely associated with the Green movement. They may be running this type of argument primarily because they want to see progress on climate change rather than because they believe that the financial system can't cope with such change.

Disentangling politics from policy debate is difficult. This is illustrated by UNEP (2015) which is part of the broader United Nations Environment Programme's Inquiry into the Design of a Sustainable Financial System. It notes the emphasis policymakers have placed on macroprudential tools to address flaws in the financial system that led to the 2007–09 Credit Crisis, i.e. to address systemic risk. It describes ways in which these tools might be co-opted to support a wider sustainable development agenda.

More assertive is Carbon Tracker Initiative (2014). This paper argues that carbon rich firms (such as oil and natural gas firms) should be required to provide information in their financial statements on how much of their stated reserves are likely to be burnt and how much would be 'stranded' (i.e. need to be kept in the ground and hence deplete the value to shareholders of the company) if global CO2 emissions targets are to be adhered to. The authors refer to systemic risk issues and macroprudential policies when doing so. However, this is primarily by analogy, as their main thrust seems to be that voluntary codes etc. (which were found wanting during the 2007–09 Credit Crisis) can also be expected to be ineffective ways of addressing the potential market failure they believe is behind this so-called ‘carbon bubble’. Similar sentiments are expressed by Green European Foundation (2014).

Just because politics seems to be influencing the way in this type of argument is presented does not make it inherently unsound. More to the point, if you are a risk manager seeking to understand the potential impact of systemic risk you only need to accept that there is a risk that things will play out as envisaged by such commentators. You don’t need to accept that this is the most likely outcome. It should also be borne in mind that some of the underlying elements of the Green movement’s agenda now have very wide social traction, as evidenced by e.g. the Paris Climate Change Paris Accord.

Types of issues that then become relevant include ones explored by Schoenmaker et al. (2015). They argue that ‘*the health of the ecology and that of the economy of a region or country have always been intimately linked*’ and hence consider ecological imbalances to have the scope to create systemic risks. They also note the strong past link between housing markets and systemic risk. This suggests to them that an (abrupt) change in the services of a major asset class with a long maturity can lead to a major downward adjustment in its price and consequential risk of spill-over into the wider economy, especially if it is often debt financed. They then explore whether the carbon bubble exhibits some of these characteristics, because of the typical business models of carbon rich firms.

This debate gained prominence in the UK in 2015 when the Governor of the Bank of England gave a major speech in which he explored the possible systemic risk implications of stranded assets, see Carney (2015). Subsequent Bank of England analysis, see PRA (2015), suggested that the specific link between the carbon bubble and financial stability concerns linked to the financial health of insurers and pension funds that had been most focused on by earlier commentators might have been overstated. However, it also explored a different angle that might end up being of even greater importance to some insurers. This is the risk of spill-over from product liability risk. If CO2 is in due course deemed by courts to be a waste product with undesirable impacts then someone may be able to sue for loss due to the creation of this waste product (much as they did with a previous wonder-product that went sour, namely asbestos, see Section 4.2.4). Given how dependent our current economy is on fossil fuels, very large product liability losses might then result.

4.13 Non-Financial Firms and the Rest of the Real Economy

In principle, firms do not need to be financial to create (financial) systemic risk. We have previously noted examples like Hanjin Shipping, given its dominant role in Korean exports. Some of the tech titans probably also nowadays have systemic relevance, even just for the financial services industry, given modern business and finance reliance on software tools including some supplied by these companies.

However, most non-financial firms are deemed unlikely to create (financial) systemic risk. Rather, they are expected (alongside members of the public) mainly to be on the receiving end of (financial) systemic risk.

In practice, many companies nowadays carry out activities that to some extent substitute for services provided by the financial sector. If drawn broadly enough these activities might include manufacturers', retailers' and airlines' loyalty points, reward cards or frequent flyer programmes and therefore any firm that has such a programme (although such activities are usually modest in relation to these firms' overall business activities). Some telecoms companies, particularly in the developing world, have also used their mobile networks to provide payments systems.

Some non-financial firms, such as aircraft leasing companies or rental organisations, may engage in a higher proportion of financing-related activities. Macroprudential authorities may try to deem these firms to be shadow banks, see Section 4.6. However, these firms can be very heterogeneous.

And, of course, individual citizens (even ones who are not extremely wealthy) can also contribute to some extent to (financial) systemic risk. When a particularly large financial bubble strikes, the irrational exuberance that has fuelled it tends to be broadly based and not just exclusively in the minds of market professionals. It is disingenuous to blame the whole of the 2007–09 Credit Crisis on financiers. Plenty of others in society were willing to build up higher than sustainable levels of debt to further their own hoped for financial futures.

4.14 Key Takeaways

This Chapter has explored features of specific parts of the financial system of relevance to financial stability. The focus has been broad and wide-ranging because this is the typical stance adopted by policymakers, even if it is not always welcomed by industry participants who may view their own part of the financial system as less systemically important than the remainder. Key points noted include:

(a) The greatest focus has been on the banking sector, as this was the sector most associated with failures during the 2007–09 Credit Crisis. There has been a considerable amount of regulatory change since then that has sought to address systemic risks exhibited by the banking sector. Further changes are in the pipeline.
(b) A few noteworthy insurers also failed during the Crisis. Some insurers now appear to be challenged by low interest rates that are partly a consequence of unconventional monetary policies adopted since the Crisis. This has led policymakers to explore ways in which insurers can contribute to financial instability. Longer term, these are likely to lead to significant changes in how capital requirements are formulated for insurers, at least for larger ones.
(c) Pension funds are often viewed as net contributors to financial stability because of their long-term investing perspective (as are some sovereign wealth funds). However, this implicitly presumes that they are themselves broadly-speaking financially healthy, a view that some commentators disagree with. Whilst their longer-term trajectory may be substantially decoupled from the rest of the financial system this will not necessarily stop parts of the pension 'system' suffering from its own instabilities.
(d) Policymakers worry about the potential for banking activity to pass to the shadow banking sector, where it is less extensively regulated. It is tricky to define shadow banking precisely, to quantify exactly how large it is, to regulate it or to identify exactly what impact such a trend might have on the future nature of the financial system.
(e) Some types of investment fund have historically suffered systemic risk events. Policymakers have mostly focused on the extent to which such entities (and the managers who manage them) carry out liquidity transformation or include leverage (synthetic or otherwise) within their strategies. More broadly, asset managers appear to be expanding into business activities that were previously primarily carried out by commercial or investment banks. Some of the systemic risks previously borne by the banking sector may therefore be migrating to the asset management sector.
(f) Significant changes to how markets operate have been introduced by regulation following the Crisis. These include mandatory central clearing of simpler derivatives. Whether these will reduce systemic risks or merely concentrate them into new hands is not yet clear.

(g) Historically, the biggest systemic risk events have tended to involve breakdowns in society or major political changes, sometimes violent. Any fuller consideration of systemic risk therefore needs to include sovereign risk in its various forms.

Even more broadly, some commentators worry about more existential risks and their potential impact on financial stability. Some of this appears to be in part a usage of the language of financial stability to support a specific political stance. As the financial system ultimately exists within (and to serve) the broader society, it is not in practice possible (or even ultimately desirable) to decouple financial stability from politics.

References

ABI (2013). Funds held in Life and Pension products in 2012. *Association of British Insurers.*

A&L Goodbody. (2013). In Focus: Summary of the Alternative Investment Fund Managers Directive – AIFMD. *A&L Goodbody.*

A&L Goodbody (2014). In Focus: UCITS V. *A&L Goodbody.*

Bank of England (2014a). Procyclicality and Structural Trends in Investment Allocation by Insurance Companies and Pension Funds. *Bank of England Procyclicality Working Group.*

Baranova, Y., Liu, Z. and Noss, J. (2016). The Role of Collateral in Supporting Liquidity. *Bank of England Staff Working Paper*, No. 609.

BCBS (2009). Consultative Document: Strengthening the Resilience of the Banking Sector. *Basel Committee on Banking Supervision.*

BCBS (2016b). Minimum Capital Requirements for Market Risk. *Basel Committee on Banking Supervision.*

BCBS (2016d). Consultative Document: Standardised Measurement Approach for Operational Risk. *Basel Committee on Banking Supervision.*

BCBS (2016e). Standards: Interest Rate Risk in the Banking Book. *Basel Committee on Banking Supervision.*

Bongaerts, D. (2015). The Economics of Investor-Paid Credit Rating Agencies. *DNB Research Seminar.*

Carbon Tracker Initiative (2014). Unburnable Carbon – Are the World's Financial Markets Carrying a Carbon Bubble? *Carbon Tracker Initiative.*

Carney, M. (2015). Breaking the Tragedy of the Horizon – Climate Change and Financial Stability. *Speech given by Mark Carney, Governor of the Bank of England, Chairman of the Financial Stability Board, Lloyd's of London*, 29 September 2015.

Cetorelli, N. (2015). Asset Managers and Systemic Risk. A Financial Intermediation Perspective. *Banque de France-ACPR Conference: Financial Regulation – Stability versus Uniformity: A Focus on Non-Bank Actors.*

DTCC (2015). Understanding Interconnectedness Risks. *DTCC.*

Duffie, D. (2015). Failure Resolution of Central Counterparties. *Banque de France-ACPR Conference: Financial Regulation – Stability versus Uniformity: A Focus on Non-Bank Actors.*

ECB (2015). The Impact of Negative Short-Term Rates on the Money Market Fund Industry. *ECB Economic Bulletin* 5/2015.

Economist (2016a). Fade to Grey. *The Economist*, 24 September 2016.

EIOPA (2016a). Consultation Paper on the Methodology to Derive the UFR and its Implementation. *EIOPA.*

EIOPA (2016b). A Common Framework for Risk Assessment and Transparency for IORPs. *EIOPA.*

EPSC (2015). Further Risk Reduction in the Banking Union. *European Political Strategy Centre.*

ESMA (2015). Trends Risks Vulnerabilities. *European Securities and Markets Authority.*

ESRB (2014b). Securities Financing Transactions and the (Re)use of Collateral in Europe. *European Systemic Risk Board, Occasional Paper 6.*

ESRB (2016a). EU Shadow Banking Monitor No. 1 /July 2016. *European Systemic Risk Board.*

ESRB (2016b). Assessing Shadow Banking – Non-Bank Financial Intermediation in Europe. *European Systemic Risk Board*, Occasional Paper 10.

European Commission (2014). Memo on Reporting and Transparency of Securities Financing Transactions – Frequently Asked Questions. *European Commission.*

European Union (2013a). Capital Requirements Directive, i.e. Directive 2013/36/EU. *European Union.*

European Union (2013b). Capital Requirements Regulation, i.e. Regulation (EU) No 575/2013. *European Union.*

European Union (2014b). Solvency II Directive as Recast in 2014 by Omnibus II Directive, i.e. Directive 2009/138/EC. *European Union.*

European Union (2014c). Solvency II Directive Delegated Regulation, i.e. Commission Delegated Regulation (EU) 2015/35. *European Union.*

European Union (2014d). Proposal for a Directive of the European Parliament and of the Council on the Activities and Supervision of Institutions for Occupational Retirement Provision (draft 17 September 2014). *European Union.*

FCA (2014b). Markets in Financial Instruments Directive II Page. *UK Financial Conduct Authority*, viewed 18 September 2014.

Ferguson, N. (2009). *The Ascent of Money*. Penguin Books, London, UK.

Financial Times (2016a). Negative Rates 'poison' German Pension Funds. *Financial Times*, 16 May 2016.

Financial Times (2016b). Fears Mount for Pensions as Gilt Yields Touch Negative Territory. *Financial Times*, 11 August 2016.

Financial Times (2016c). US Companies' Cash Pile hits $1.7 tn. *Financial Times*, 20 May 2016.

Fisher, M. (2002). Special Repo Rates: An Introduction. *Federal Reserve Bank of Atlanta Economic Review*, Second Quarter 2002.

French, A., Vital, M. and Minot, D. (2015). Insurance and Financial Stability. *Bank of England Quarterly Bulleting* 2015 *Q3.*

FSB (2013). Strengthening Oversight and Regulation of Shadow Banking: An Overview of Policy Recommendations. *Financial Stability Board.*

FSB (2014b). Regulatory Framework for Haircuts on Non-Centrally Cleared Securities Financing Transactions. *Financial Stability Board.*

FSB (2016). Consultative Document: Proposed Policy Recommendations to Address Structural Vulnerabilities from Asset Management Activities. *Financial Stability Board.*

Galbraith, J.K. (1955). *The Great Crash 1929*. Penguin Books, London, UK.

Gibson, R.C. (2013). *Asset Allocation: Balancing Financial Risk*. McGraw Hill, New York, USA.

Giraldi, C., Susinno, G., Berti, G., Brunello, J., Buttarazzi, S., Cenciarelli, G., Daroda, C. and Stamegna, G. (2000). Insurance Optional. *Risk Magazine*, April 2000.

Gray, A. (2016). US Life Insurers Shaken by Rock-Bottom Rates. *Financial Times.*

Green European Foundation (2014). The Price of Doing Too Little Too Late: The Impact of the Carbon Bubble on the EU Financial System. *Green European Foundation.*

Haldane, A.G. (2014). Halfway up the Stairs. *Bank of England.*

Hanouna, P., Novak, J., Riley, T. and Stahel, C. (2015). Liquidity and Flows of US Mutual Funds. *SEC.*

Harrison, D. and Blake, D. (2015). The Greatest Good for the Greatest Number. *The Pensions Institute, Cass Business School.*

House of Commons (2015). Greek Debt Crisis: Background and Developments in 2015. *UK House of Commons Library.*

IAIS (2015). Higher Loss Absorbency capacity for Global Systemically Important Insurers: Public Consultation Document. *International Association of Insurance Supervisors.*

IAIS (2015b). Global Systemically Important Insurers: Proposed Updated Assessment Methodology: Public Consultation Document. *International Association of Insurance Supervisors.*

IMF (2009). Guidance to Assess the Systemic Importance of Financial Institutions, Markets and Instruments: Initial Considerations: Report to the G-20 Finance Ministers and Central Bank Governors. *International Monetary Fund, Bank for International Settlements and Financial Stability Board.*

IMF (2015a). Global Financial Stability Report, April 2015, Chapter 3: The Asset Management Industry and Financial Stability. *International Monetary Fund.*

Impavido, G., Al-Darwish, A., Hafeman, M., Kemp, M. and O'Malley, P. (2011). Possible Unintended Consequences of Basel III and Solvency II. *IMF Working Paper No. 11/187.*

Insurance Europe (2014). Why Insurers Differ from Banks. *Insurance Europe.*

Jenkins, P. (2016). Banking can Learn from Insurance on Systemic Risk. *Financial Times.*

Johnson, S. (2014). Money Market Investors Consider Private Funds. *Financial Times*, 8 December 2014.

Kemp, M.H.D. (2009a). *Market consistency: Model Calibration in Imperfect Markets*. John Wiley & Sons, Chichester, UK.

Kemp, M.H.D. (2011a). Sponsor covenants in risk-based capital. Life & Pensions Risk.

Lowenstein, R. (2001). *When Genius Failed: The Rise and Fall of Long-Term Capital Management*. Fourth Estate, London, UK.

NOLGHA (2011). NOLGHA Testimony before the House Financial Services Committee on "Insurance Oversight and Regulatory Proposals". *National Organization of Life and Health Insurance Guaranty Associations.*

PPF (2014). The Purple Book: DB Pensions Universe Risk Profile 2014. *UK Pension Protection Fund and UK Pensions Regulator.*

PRA (2015). The Impact of Climate Change on the Uk Insurance Sector. *The Bank of England, Prudential Regulation Authority.*

Reinhart, C.M. and Rogoff, K.S. (2009). *This Time is Different: Eight Centuries of Financial Folly*. Princeton University Press, Princeton, New Jersey.

S&P (2013a). What May Cause Insurance Companies To Fail – And How This Influences Our Criteria. *Standard & Poor's.*

Schoenmaker, D., van Tilburg, R. and Wijffels, H. (2015). What Role for Financial Supervisors in Addressing Systemic Environmental Risks? *DCF Policy Paper Series, No. 50.*

SEC (2016). Investment Company Liquidity Risk Management Programs. *US Securities and Exchange Commission.*

Singh, M. and Aitken, J. (2010). The (sizeable) Role of Rehypothecation in the Shadow Banking System. *International Monetary Fund.*

Timmer, Y. (2016). Cyclical Investment Behavior across Financial Institutions. *Bundesbank Discussion Paper 08/2016.*

Tirole, J. (2015). Financial Regulation: A Focus on the Non-Bank Sector. *Banque de France-ACPR Conference: Financial Regulation – Stability versus uniformity: A focus on non-bank actors.*

UNEP (2015). Aligning the Financial System with Sustainable Development: Pathways to Scale. *United Nations Environment Programme.*

5

Measuring Systemic Risk

Once a risk has been identified, a natural next step is to try to measure it. Warning bells typically trigger in the minds of risk managers if they come across a material measurable risk that is not actually being measured. Usually this is a sign that the risk is not being given as much attention as it warrants.

The relative lack of attention given to systemic risk prior to the 2007–09 Credit Crisis is a case in point. Most practitioners and policymakers were seduced by arguments that a new paradigm was emerging. Most failed to appreciate the build-up of liquidity risk that was then taking place. They are keen to ensure that they do not make the same mistake again.

This has led to a considerable amount of research on how best to measure systemic risk. We describe below some of the techniques that have been proposed by academics for this purpose. We also describe some of the techniques that policymakers including central bankers seem to focus on in their published financial stability reports.

Once again, we uncover a dichotomy. Academic research tends to focus on (direct) interconnectedness but actual output from macroprudential bodies tends to focus on uncovering previously underappreciated vulnerabilities.

M.H.D. Kemp, *Systemic Risk*, DOI 10.1057/978-1-137-56587-7_5

5.1 Conceptual Components

5.1.1 Risk Measures

When risk managers (and academics) think of risk measures they tend to think of metrics such as *value-at-risk* (VaR), *tail value-at-risk* (TVaR), *expected shortfall* (ES), *ex-ante tracking error* (TE), *drawdown* etc. These are statistical measures that ultimately derive from the statistical distribution that is believed to characterise the likelihood of different future outcomes (or in the case of drawdown has characterised past outcomes). A summary of these risk measures (including definitions) is given in Box 5.1.

In principle, these sorts of statistical risk measures can be applied to any organisation or indeed to any grouping of organisations. The entire financial system is a special case of a particularly large group, i.e. containing every financial organisation, so these statistical risk measures can also be applied to it. The natural focus is then to characterise the likelihood of a systemic risk loss of a given amount. Some effort may be needed to identify what we mean by 'loss', e.g. is it the quantum of capital the state might need to put into the financial system to stabilise it in particularly adverse scenarios, or the total loss to the economy including second order effects etc.

Risk managers also tend to be interested in decomposing the total risk (measured using one of the above risk measures) into its constituent parts. Fortunately for most common risk measures it is conceptually simple to identify a decomposition that adds up, see Box 5.2. Papers that provide insights into how this can more specifically be done for systemic risk include Staum (2011). Decompositions such as these in principle allow us to determine the proportion of the total risk that derive from a specific organisation. They are therefore potentially relevant when assessing whether a specific firm should be deemed to be a G-SIFI. We discuss the characterisation of firms as G-SIFIs in Section 6.5.

Box 5.1: Statistical risk measures

Several (quantitative) measures of risk are used in the financial industry and increasingly in other sectors of the economy. Perhaps the best known is *value-at-risk* (VaR). For a portfolio (of risks, investments, ...) it is the loss which will be exceeded in some fraction, α, of outcomes if the portfolio is held for a given length of time, i.e. for a given time horizon, say *t*.

Suppose a portfolio consists of monetary amounts $\mathbf{a} = (a_1, \ldots, a_n)^T$ invested in n exposures. Here T is the (vector) transpose sign. Let x_i be the loss (i.e. negative payoff) on the i'th exposure, and $\mathbf{x} = (x_1, \ldots, x_n)^T$. Let $L = \mathbf{a}.\mathbf{x} = \sum_{i=1}^{n} a_i x_i$ be the total portfolio loss. For a portfolio with total losses over a holding period t equal to a random variable L, the value-at-risk with confidence level α $(0<\alpha<1)$, denoted VaR_α, is defined mathematically as:

$$VaR_\alpha = \inf\{z : Pr(L \geq z) \leq \alpha\}$$

In this formula 'inf' means the largest value of z such that the probability of L exceeding z is less than or equal to α.

For a continuous distribution VaR_α is implicitly defined by:

$$P\mathrm{r}(L \geq VaR_\alpha) = \alpha$$

Or, if the probability density function (pdf) of payoff X is $p(x)$ then VaR_α is defined implicitly using integration (noting that losses are negative payoffs) as the value k such that:

$$VaR_\alpha(X) = k \ \ where \int_{-\infty}^{-k} p(x)dx = 1 - \alpha$$

Points to note include:

(a) VaR_α is mathematically equivalent to the $(1-\alpha)$-quantile of the payoff distribution, or in mathematical notation $VaR_\alpha(X) = -F^{-1}(1-\alpha)$ where $F^{-1}(x)$ is the quantile function, also called the inverse cumulative distribution function or just the inverse function of the distribution with density $p(x)$. This indicates that when we are estimating and using VaR_α we may draw on an extensive body of statistical knowledge relating to quantile estimation.
(b) The above sets out the usual definitions of VaR_α but sometimes the sign is flipped and/or α is replaced by $1-\alpha$, since usually the VaR refers to the *downside* tail irrespective of how α is expressed.
(c) VaR_α has a natural connection with the amount of capital needed by a firm. Capital is (usually) defined as the excess of assets over liabilities. Nowadays the tendency is to value assets and liabilities in such calculations by reference to economically relevant market values rather than, say, book or purchase costs. If a firm holds (market-value) capital equal to VaR_α (calculated for a holding period of t) then it should experience losses exceeding its capital at time t with probability α (if it does not alter its portfolio or its asset or liability bases in the meantime).
(d) Some axioms need to apply for the mathematics underlying these computations to work, namely uniqueness, additivity and scalability, i.e. here that L is

well defined (in the above we are referring to the value that we place on the loss, typically its monetary value), and that if we have two losses x_1 and x_2 then $k(x_1 + x_2) = kx_1 + kx_2$

(e) Mathematical risk measurement theory almost always concentrates on the *market value* of the exposures or some reasonable economic proxy. This is partly because such values, if suitably defined, should adhere to the axioms in (d) if we adopt the principle of no arbitrage.

(f) We may conceptually split VaR_α into two parts, the expected loss, $EL = E(L)$, where $E(X)$ is the expected value of X, and the unexpected loss, $UL = VaR_\alpha - EL$. Some commentators argue that, say, banks should only hold capital to cover the UL since the expected losses on, say, a bank's loan portfolio should be offset by anticipated profit margins included in loan pricing. The potential flaw in this logic is that firms do not necessarily estimate EL correctly (or necessarily price the loan consistently in relation to the EL even if they have estimated EL accurately, e.g. their pricing may be driven by market forces). Moreover, the EL on existing loans may change through time (as economic conditions change) but loan rates for these loans may not move in tandem. Some capital is potentially required to protect against these risks.

(g) VaR as defined above focuses on monetary losses, but in many contexts (especially in an asset management context) the focus is on loss relative to a benchmark outcome. If this is the definition of loss then the resulting VaR may be referred to as a 'relative' VaR. If we are measuring returns relative to those generated by a benchmark, $\mathbf{b} = (b_1, \ldots, b_n)^T$ then we can define the *relative* VaR_α using vector algebra by reference to losses L given by $L = (\mathbf{a} - \mathbf{b}).\mathbf{x} = \sum_{i=1}^{n} (a_i - b_i)x_i$ where the losses per unit exposure of each underlying instrument are $\mathbf{x} = (x_1, \ldots, x_n)^T$.

A commonly proposed alternative to VaR is *tail value-at-risk* (TVaR). For a given confidence level α and time horizon t it is defined (if it exists) as:

$$TVaR_\alpha = E(L/L) \geq VaR_\alpha$$

Or, if the pdf of payoff X is $p(x)$ and $p(x)$ is continuous then $TVaR_\alpha$ is:

$$TVaR_\alpha(X) = E(-X|X \leq -VaR_\alpha) = -\frac{1}{1-\alpha} \int_{-\infty}^{-VaR_\alpha} xp(x)dx$$

TVaR is also sometimes called *conditional value-at-risk* (CVaR), because it involves a conditional probability, or *conditional tail expectation* (CTE). Occasionally TVaR (less commonly CVaR) is ascribed the same meaning as expected shortfall, see below, in which case the $1/(1-\alpha)$ factor is ignored, or is defined relative to some specific limit $-k$ that in effect defines the α to be used in the above formula.

Some probability distributions that are particularly prone to extreme events (i.e. particularly fat-tailed) do not have a first moment or mean in which case their TVaR is not defined (i.e. infinite) whereas all statistical distributions have finite VaRs. TVaR can also be somewhat harder to estimate robustly than VaR even for less pathological distributions.

Whilst VaR has been the industry standard risk measure for some time now (at least in the banking industry, see below regarding the asset management industry), there seems to be some regulatory drive towards greater use of Tail VaR (or expected shortfall) in the future. Reasons why TVaR may be preferred for this purpose include, see e.g. Kemp (2009b):

(a) VaR provides no guidance on how severe losses might be *beyond* the VaR cut-off point. This is potentially particularly important for some stakeholders (such as regulators, supervisors, customers and governments). If the VaR cut-off point is set at a level comparable with the point at which the firm defaults then losses up to the VaR cut-off point will (we might argue) be borne by shareholders. It is only when losses start to exceed this cut-off that costs fall to these wider stakeholders. So VaR in this sense can be viewed as overly shareholder-focused and insufficiently sensitive (as far as non-shareholder stakeholders are concerned) to magnitude of loss beyond the VaR cut-off.
(b) VaR does not in general exhibit desirable features we might expect a risk measure to exhibit in relation to diversification (it does not in general satisfy *sub-additivity*). For example, suppose there are two portfolios. One is (only) exposed to one risk that has a 0.3% chance of occurring and if it does then it will lose \$1m. The other is exposed to five independent risks each of which has a 0.3% chance of occurring and each involves a loss of \$0.2m. Then we would 'expect' a risk measure to show the second portfolio to be less risky than the first one (because it is better diversified). However, the 99.5% VaR of the second portfolio is \$0.2m which is *more than* the VaR of the first portfolio (which is 0 because its risk of loss has a likelihood of occurrence less than 0.5%). TVaR does behave as we might intuitively expect in the presence of diversification (as it, like ex-ante tracking error, see below) does satisfy sub-additivity.

In latest banking regulatory developments, TVaR is typically not referenced as such. Instead the focus is on its very close analogue, *expected shortfall* (ES), which is merely a constant multiple or fraction of TVaR. The expected shortfall, $ES(Q)$, given some trigger level, Q (usually $Q = -VaR_\alpha$), is normally defined as:

$$ES(Q) = -\int_{-\infty}^{Q} xp(x)dx$$

Although VaR and variants such as TVaR are probably the most commonly used risk measures in the banking world this is less true in the asset management world. Here, a particularly common risk measure is (ex-ante) *tracking error* (TE). If X is a random variable (e.g. a portfolio return) with (assumed forward looking) pdf $p(x)$ then its ex-ante tracking error (if it exists) is σ where $\sigma^2 = var(X)$, i.e. the variance of the forward-looking return.

Implicit in a focus on (ex-ante) tracking error is the view that we should not when monitoring the risk characteristics of an actively managed portfolio take credit for any assumed (expected) outperformance the manager might deliver.

Tracking error has a nice intuitive geometrical analogy. This arises because the formula for the (ex-ante) tracking error of the sum of two sets of exposures, i.e.:

$$\sigma_{\mathbf{a+b}}^2 = \sigma_{\mathbf{a}}^2 + 2\sigma_{\mathbf{a}}\sigma_{\mathbf{b}}corr(\mathbf{a},\mathbf{b}) + \sigma_{\mathbf{b}}^2$$

has a natural analogy with the relationship between the lengths of sides of a triangle two of whose sides are formed by vectors $\mathbf{a}$ and $\mathbf{b}$ if $corr(\mathbf{a},\mathbf{b})$ is associated with $\cos\theta$ where θ is the angle between these two vectors.

If losses are normally distributed then VaR, TVaR, ES and TE all end up being scaled multiples of each other (with perhaps additional shift terms). Once a suitable scale /shift adjustment is applied they should all in effect lead to the same risk management decisions being taken. Such decisions ultimately always involve selection between different alternatives.

Drawdown is a measure that is perhaps most commonly seen in the hedge fund industry. It measures the magnitude of losses a portfolio has suffered in the past over some specified period. Hedge fund managers might quote their worst historic drawdown since their trading strategy commenced or e.g. quote their cumulative drawdown peak to trough as the maximum fall their fund unit price has suffered at any time in the past.

Acharya et al. (2010) introduce the concept of a bank's *systemic expected shortfall* (SES). The SES of a given institution is defined as the expected amount by which a particular bank's equity, w_i (for the i'th bank in the financial system), drops below its target capitalisation level (in their paper defined as a fraction z of its assets a_i, where z is assumed to be defined externally), in the case of a systemic crisis, i.e. when aggregate banking capital $W = \sum w_i$ is less than z times aggregate bank assets $A = \sum a_i$, i.e.:

$$SES_i = E(za_i - w_i | W < zA)$$

Here $E(X|Y)$ is the expected value of X conditional on Y happening.

We have noted previously the importance accorded to interconnectivity by many in the macroprudential community. One way of quantifying the impact that interconnectivity has on risk is via the concept of CoVaR, see Adrian and Brunnermeier (2011). If we have two institutions i and j (or combinations of institutions, including the whole system) then we define $CoVaR_q^{j|i}$ to be the VaR of institution j conditional on some event $C(X_i)$ of institution i (they concentrate on the event being that the institution is in distress) i.e. so that (if we define signs etc. appropriately):

$$Pr\left(X_j \leq CoVaR_q^{j|C(X_i)} | C(X_i)\right) = q$$

They in practice focus mainly on the case where j corresponds to the whole system. They also note that the same sort of approach can be used with other risk measures mentioned above, such as expected shortfall, leading to 'co-expected shortfall' or Co-ES, etc.

Box 5.2: Contributions to risk

For many purposes, it is desirable to identify how individual risks contribute to an overall risk measure. For example, we might ask how we can best apportion a total portfolio value-at-risk into contributions from individual positions. The most common approach is to use *marginal value-at-risk* (MVaR).

Suppose we have the same total loss as per Box 5.1, i.e. $L = \mathbf{a}.\mathbf{x} = \sum_{i=1}^{n} a_i x_i$ where the amount of the i'th exposure is a_i and the loss arising from a unit amount of the i'th exposure is x_i. Then the marginal value-at-risk (with confidence level α and time horizon t) for the i'th exposure is denoted $MVaR_\alpha^{(i)}$ and is usually defined as (using partial differentials):

$$MVaR_\alpha^{(i)} = \frac{\partial}{\partial a_i}\left(VaR_\alpha\left(\sum_i a_i x_i\right)\right)$$

As risks arising from individual positions interact there is no universally agreed way of subdividing the overall risk into contributions from individual positions. However, a commonly used way is to define the *contribution to value-at-risk*, c_i, of the i'th position, as:

$$c_i = a_i MVaR_\alpha^{(i)}(\mathbf{a})$$

The c_i then sum to the overall VaR. This summation behaviour is a special case of a more general result called Euler's capital allocation principle that applies to any risk measure that is homogeneous (of order 1). A function $f(u_1, \ldots, u_n)$ is said to be homogeneous of order g (constant) if it satisfies $f(ku_1, \ldots, ku_n) = k^g f(u_1, \ldots, u_n)$.

VaR, TVaR, ES and TE are all (first order) homogeneous. The same sort of approach can therefore be used to decompose each into marginal VaR (MVaR), marginal TVaR (MTVaR), marginal ES (MES) and marginal TE (MTE) and hence into contributions to risk from individual positions that add up to the total risk of the portfolio (or firm, sector, financial system . . . etc., depending on aggregation level).

Sometimes attention is focused on an alternative measure called incremental VaR that does not (in general) add to the total portfolio VaR. It is the change in the VaR if the whole of a given position is removed from the portfolio, i.e. $IVaR_\alpha^{(i)}$ where:

$$IVaR_\alpha^{(i)} = VaR_\alpha\left(\sum_i a_i x_i\right) - VaR_\alpha\left(\sum_{j, j\neq i} a_j x_j\right)$$

IVaR thus involves calculating (i) the total VaR, (ii) what the VaR would be if we knocked out an entire position from the portfolio and (iii) the difference between (i) and (ii). This type of methodology can be simpler to estimate or calculate than a MVaR type of computation. It is in effect the sort of computation used by Adrian and Brunnermeier (2011) when trying to estimate the contribution of an individual institution to CoVaR, their proposed measure of systemic risk, see Box 5.1. Using terminology as in Box 5.1 they denote institution i's contribution to the systemic risk of institution j (or of the system as a whole) by, e.g.:

$$\Delta CoVaR_q^{system|i} = CoVaR_q^{system|X_i = VaR_q^i} - CoVaR_q^{system|X_i = median_i}$$

5.1.2 Stress Testing

An alternative risk management technique that has been gaining ground of late relative to statistical risk modelling is the concept of stress testing or reverse stress testing. Stress tests are typically less statistical in nature (i.e. they focus less on the likelihood of a specific outcome) and more orientated towards identifying configurations of events (and hence magnitudes) that might most plausibly put the organisation at a disadvantage. Further comments on stress tests are given in Box 5.3.

Stress testing is also a means by which regulators can promote market transparency and discipline. In some parts of the financial sector (such as banking) regulators regularly require firms to carry out stress tests using specifically mandated stresses chosen by the regulator, and firms may then be required to publish summary results of applying these stresses to themselves.

This type of exercise is therefore a way of forcing firms to quantify their exposures and then to communicate them to the market. If a firm appears to be put at a disadvantage relative to its peers in a specific stress scenario, and if the market is worried about that scenario (or some other scenario that is illuminated by the results of the scenario being tested), then the market value of the firm can react appreciably to publication of these results, if the results are not as expected by the market based on earlier disclosures.

Statistical risk measures tend not to include notions of causality. This is because with them we are usually agnostic as to what causes the statistical distribution of outcome, we are just interested in the probabilities to assign to different outcomes. Some straightforward stress tests, such testing the impact of the equity market declining by 40%, are also largely agnostic to causation. However, it is possible to introduce causation within stress tests, see Box 5.3.

Box 5.3: Stress testing and reverse stress testing

Stress testing involves analysing the impact of a (generally adverse) shock to a set of well-defined market prices or other economic factors, generally within the range of deemed plausible outcomes. To the extent that stress testing and scenario testing differ, stress testing presumes that the business is being stressed (adversely) whereas scenario testing does not necessarily include this presumption. Stress testing typically places less emphasis on the exact likelihood of a specific scenario occurring and more emphasis on the magnitude of impact. It can therefore be thought of as a way of mitigating some of the model risk exhibited by statistical risk models (because they do need to identify likelihood of occurrence). A corollary is that they are not necessarily so useful for reaching economic decisions, which necessarily do need to take some heed of the likelihood of different events occurring.

There are three main interpretations of the term 'stress testing':

(a) It may be equated with an analysis of the impact on a portfolio (or a firm, sector or even the entire financial system) contingent on movements in specific drivers, the sizes of these movements being selected by the participants in a manner that is considered by them to be appropriately adverse in the context of a plausible distribution of outcomes. These are the sorts of stress tests that are most commonly used within own risk assessments such as ICAAPs and ORSAs.
(b) It may be equated with specific industry-wide stress scenarios mandated by a regulator, typically either for industry-wide stress tests (see below) or for use in standard formula elements of regulatory capital computations. These sorts of stress tests are akin to (a) but with the stresses involved set by the regulator.
(c) It may be equated with a greater focus on the sorts of configurations of market events that might lead to large losses.

The most common form of (c) is reverse stress testing. This is now in effect mandatory for many financial organisations in many jurisdictions. It was originally popularised by CRMPG-III (2008) in August 2008, shortly before Lehman Brothers defaulted (tellingly in a document titled '*Containing Systemic Risk: The Road to Reform*', suggesting that the authors thought that it had specific relevance to systemic risk). It involves:

(1) starting out with a specified outcome (typically that the firm's business model becomes broken);
(2) working out some potential scenarios which might be expected to give rise to this outcome (with the scenarios being plausible conditional on the outcome materialising); and
(3) working out what mitigating steps can be taken to reduce the likelihood of these scenarios occurring and/or their impact if they did materialise.

When it is necessary to identify the level of plausibility of a specific stress test then the most common way of doing so is to use Mahalanobis distances, see e.g. Kemp (2011). The Mahalanobis distance between two multivariate vectors

$x = (x_1, \ldots, x_n)^T$ and $y = (y_1, \ldots, y_n)^T$ coming from the same distribution (with covariance matrix V) is calculated as $d(x,y) = \sqrt{(x-y)^T V^{-1}(x-y)}$ and can be thought of as a measure of their dissimilarity.

Since the 2007–09 Credit Crisis regulators have required banks to carry out specific stress tests laid out by the regulator. Originally (in the USA, for banking) this was called the Supervisory Capital Assessment Program (SCAP), see Section 5.2.5 but it is now included in the Comprehensive Capital Analysis and Review (CCAR) that firms above a certain size are required annually to submit to the regulator. Dodd-Frank requires the CCAR to include at least three different supervisory scenarios provided by the Fed, a baseline, an adverse and a severely adverse scenario.

Industry-wide stress tests have also been mandated for other parts of the financial sector, e.g. for the EU insurance industry. Often contribution to these stress tests has been notionally voluntary, but larger firms may be expected to contribute to them. Some have been carried out to help the regulator understand the broader position of the industry to some specific regulatory feature (e.g. recent EU insurance industry stress tests have included quantification of the impact of some of the features in Solvency II referred to in Box 4.5 such as the impact of transitional provisions).

Requirements to carry out suitable stress and scenario testing also appear in guidance regulators have published in relation to firms' regulatory requirements. For example, (proposed) guidance on measurement of interest rate risk in the banking book (IRRBB), see BCBS (2016e), includes '*Principle 4: Measurement of IRRBB should be based on outcomes of both economic value and earnings-based measures, arising from a wide and appropriate range of interest rate shock and stress scenarios*'.

Some commentators believe it is helpful to develop stress tests in ways that include causative links, see e.g. Rebonato (2010). Techniques used for such purposes include Bayesian networks and other ways of mapping out causation chains. These can make the stress testing more credible. This is often important in ensuring that the results receive adequate attention from senior management and other stakeholders.

5.1.3 Macroprudential Overlays

Policymakers are interested in statistical risk measures and stress testing, but they are perhaps more interested in causal linkages than a typical risk manager outside the regulatory community. Policymakers have a wide range of possible tools to mitigate systemic risk (see Chapter 6). They are also usually conscious of trade-offs that exist with different possible tools. They are interested in which tools impact systemic risk the most but have the least side-effects in other contexts. They also need to justify their actions to sceptical industry participants, politicians and others. To justify introduction of some specific policy measure it helps if you can establish causal linkages for why any policy action is needed and why the proposed course of action is the one that is most appropriate to address the specific risk in question.

Statistical risk measures are potentially more relevant for analysing the contributions of individual organisations to systemic risk and hence, for example, the identification of G-SIFIs. However, firms change nature through time, whilst their inclusion in (or removal from) the list of G-SIFIs involves a relatively long process.

The usual way risk managers outside the macroprudential community characterise a firm that can change through time is by a set of exposures to a series of factors (plus an idiosyncratic term), see Box 5.4. The theoretical basis underlying the metrics used to identify SIFIs is a special case of factor-based risk modelling.

Box 5.4: Factor-based risk modelling and its relevance to systemic risk

Most investment portfolios expressing market risk contain many different instruments each of which is expected to behave somewhat differently. A major aspect of *portfolio* risk models is to provide some simple but not overly simplistic way of aggregating the impact of these individual exposures. The most common way in which this is done is via factor models.

There are three main ways of estimating such models using past time series data:

(a) A *fundamental risk model* ascribes certain fundamental factors (such as price to book) to individual securities. These factor exposures are exogenously derived, e.g. by reference to a company's annual report and accounts. The factor exposures for a whole portfolio (and for a benchmark, and hence for a portfolio's active positions versus a benchmark) are the weighted averages of the individual position exposures. Different factors are assumed to behave in the future in a manner described by some joint probability distribution. The overall portfolio risk (versus its benchmark) can then be derived from its active (i.e. net) factor exposures, this joint probability distribution and any additional variability in future returns deemed to arise from security specific idiosyncratic behaviours.

(b) An *econometric risk model* is like a fundamental model except that the factor exposures are individual security-specific sensitivities to certain pre-chosen exogenous economic variables, e.g. interest rates, currency exchange rates or oil prices. The sensitivities are typically found by regressing the returns from the security in question against movements in the relevant economic variables, typically using statistical multivariate regression techniques.

(c) A *statistical risk model* eliminates the need to define any exogenous factors, whether fundamental or econometric. Instead we identify a set of otherwise arbitrary time series that in aggregate explain well the past return histories of a high proportion of the relevant security universe. We ascribe elements of this set the status of 'factors'. Simultaneously we also derive the exposures of each security to each factor. A common approach is to use *principal components analysis*. This seeks to identify statistical factors that in some suitable sense explain the maximum amount of observed variability of returns in the universe of instruments under consideration.

If the risk measure of interest is ex ante tracking error (i.e. the standard deviation of future returns /losses) then it is often possible to express the output of a factor based risk model very succinctly using matrix algebra. If the instrument is assumed to react linearly to movements in each factor then the return $r_{j,t}$ on the j'th instrument in period t will have the following form where $\beta_{j,k}$ is the exposure ('beta') of the instrument to the k'th factor, $x_{k,t}$, and $\varepsilon_{j,t}$ are residual (idiosyncratic) components to $r_{j,t}$ (which are assumed to be independent of the $x_{k,t}$).

$$r_{j,t} = \alpha_j + \sum_k \beta_{j,k} x_{k,t} + \varepsilon_{j,t}$$

If the portfolio is described by a vector of (active, i.e. net) weights $a = (a_1, \ldots, a_n)^T$ then the portfolio has an expected return of $a.\alpha$ (where $\alpha = (\alpha_1, \ldots, \alpha_n)^T$and an ex-ante tracking error of σ where:

$$\sigma^2 = a^T V a = a^T \left(\beta^T \bar{V} \beta\right) a + a^T Y a$$

where V is the correlation matrix between the individual instrument return series, $\bar{V}$ is the correlation matrix between factors and Y is a very sparse (almost diagonal) matrix identifying the idiosyncratic terms.

A major advantage of this type of risk model is that it can be much more parsimonious (i.e. involve far fewer parameters that need estimating) than a model that does not have a factor structure. For example, if there are n securities and m factors and $n \gg m$ (as is usually the case) then V has $n(n+1)/2$ terms, which is typically much larger than the number of terms in $\bar{V}$ (i.e. $m(m+1)/2$ plus the number of terms in Y (possibly a little higher than n, as a few securities may be deemed to have correlated idiosyncratic terms, e.g. because they represent dual listings of the same underlying security).

The main ways in which factor models appear in systemic risk measures are:

(a) *Directly*: the whole system may be modelled by assuming it depends on a small number of easier to identify factors
(b) *Indirectly*: we may proxy the systemic risk characteristics of a specific firm by a set of factors.

A decision about whether a firm should be deemed a G-SIFI is in practice reached by ascribing it exposures to a set of factors (such as leverage, size, . . .), see e.g.

Table 4.5 and Section 6.5. It is therefore a form of indirect factor modelling. It only has theoretical validity if the factors involved discriminate between firms in terms of their systemic risk characteristics and if the weights given to different factors align with the actual characteristics of the firms in question.

5.1.4 Estimating Risk Measures

As with any other risk measure, it is one thing to define a risk measure in mathematical terms but quite another thing to identify how to estimate it robustly. This requires the development of a suitable risk model that can be used to estimate the risk measure of interest. There are many ways of developing risk models that can be used to estimate risk measures. Some common (factor based) approaches are described in Box 5.4. For credit risk modelling it is more common to use models that take as inputs credit ratings and then simulate how these might change through time using a suitable transition matrix, akin to the approach used in Box 3.9, or to use insights gains from the Merton theory of the firm and option pricing theory. The latter ultimately derives from the insight that both equity and debt are parts of the same balance sheet, and hence uses an approach conceptually aligned with material in Box 3.6.

5.2 Risk Analytics Proposed by Academics

5.2.1 Introduction

In this Section, we explore risk measures that have been proposed by academics either when seeking to identify the systemic riskiness of an individual organisation or when seeking to identify the riskiness of the whole financial system.

Bisias et al. (2012) was perhaps the first paper to seek to classify the techniques that commentators have proposed as ways of measuring or analysing systemic risk. That this paper was written as recently as 2012 highlights the relative novelty of this discipline. It was the first working paper published by the Office of Financial Research (OFR), a US-based body set up under Dodd-Frank. Its authors identified 31 different types of measure or technique that had already by then been proposed to aid in the assessment of systemic risk. They noted:

> Thanks to the overwhelming academic and regulatory response to the Financial Crisis of 2007–2009, we face an embarrassment of riches with respect to

> systemic risk analytics. The size and complexity of the financial system imply a diversity of legal and institutional constraints, market practices, participant characteristics, and exogenous factors driving the system at any given time. Accordingly, there is a corresponding diversity of models and measures that emphasize different aspects of systemic risk.

Summary descriptions of the measures and techniques they refer to in their paper are set out in Table 5.1 and Table 5.2. Table 5.1 covers general, i.e. non-sector specific, techniques. These can be system-wide or firm-specific. Table 5.2 covers sector-specific techniques. Some of the techniques they mention, such as principal components analysis (see Box 5.4), contingent claims analysis (i.e. option pricing techniques) and Mahalanobis distances (see Box 5.3), are generic statistical or financial tools whose application is not limited to systemic (or other) risk measurement. In many cases the 'sector-specific' techniques they refer to (such as serial correlation) can also be applied to other parts of the financial system without much modification.

In practice, even the impressive range of techniques described by Bisias et al. is now only a reasonably modest fraction of the full range of techniques that have been co-opted to study systemic risk. For example, credit risk modelling techniques such as those illustrated in Box 3.9 can be used to model the specific impact that some change might have to

Table 5.1 Non-sector specific techniques proposed by academics to measure systemic risk

	System-wide	Firm-specific
A. Macroeconomic measures	Asset-price boom/bust cycles Property price, equity-price and credit-cap indicators	
B. Granular foundations and network measures		Default intensity model
C. Forward-looking risk measures	Principal components analysis	Contingent claims analysis Mahalanobis distance
D. Stress-test measures	GDP stress tests	
E. Cross-sectional measures		CoVaR Co-Risk Marginal and systemic expected shortfall
F. Measures of illiquidity and Insolvency	Noise as information for illiquidity	Risk topography Leverage cycle

Source: Nematrian. Adapted from Bisias et al. (2012).

Table 5.2 Sector-specific techniques proposed by academics to measure systemic risk

	Securities/Commodities/ Asset management	Banking and Housing	Insurance and Pensions
A. Macroeconomic measures			
B. Granular foundations and network measures		Network analysis and systemic financial linkages Simulating a credit scenario Simulating a credit-and-funding-shock scenario Bank funding risk and shock transmission	Granger causality networks Mark-to-market accounting and liquidity pricing
C. Forward-looking risk measures		The Option iPoD Multivariate density estimators Simulating the housing sector Consumer credit	
D. Stress-test measures		SCAP 10-by-10-by-10 approach	
E. Cross-sectional measures		Distressed insurance premium	
F. measures of illiquidity and insolvency	Crowded trades in currency funds Equity market illiquidity Serial correlation and illiquidity in hedge fund returns Broader hedge fund-based systemic risk measures		

Source: Nematrian. Adapted from Bisias et al. (2012)

parts of the financial system. At a sufficiently granular level, nearly any quantitative risk technique used elsewhere in finance is in principle relevant to the analysis of systemic risk. The range of techniques potentially usable is therefore very broad, see e.g. McNeil et al. (2005) and Sweeting (2011).

In the next few Sections we describe some of these approaches in more detail.

5.2.2 Systemic Expected Shortfall

Systemic expected shortfall (SES) is explained further in Box 5.1. SES is defined as the amount by which a specific bank's equity drops below its target capitalisation level in the case of a systemic crisis.

SES was first developed by Acharya et al. (2010). Their approach was motivated by the insight that deposit insurance is a potential drain on the state's finances. To get banks to internalise the risks that give rise to such costs, banks should implicitly be charged a fee for access to deposit insurance. The authors show that ideally the fee should be computed as the sum of two components (if it is to represent a 'fair' premium for the insurance being provided):

(a) an institution-specific component related to the expected cost of providing such an underpin in the absence of systemic risk; plus
(b) a systemic risk component based on that firm's percentage contribution to expected systemic-wide undercapitalisation arising in a systemic risk event.

In their paper, they demonstrate, if a variety of assumptions apply, that the SES of a bank has three components:

(1) Excess ex ante leverage (i.e. the amount by which the bank is already undercapitalised);
(2) Marginal expected shortfall (MES), measured by reference to the bank's contribution to the overall expected shortfall of the whole financial sector (during 'normal' times), scaled up by a factor to account for worse performance of the system during a crisis; and
(3) An adjustment term consisting of a part which depends on the excess costs of financial distress (since these will typically be underrepresented in any sample used to estimate MES) plus a part representing expected profits the bank would otherwise generate over the holding period under consideration.

Acharya et al. (2010) estimated the MES of a range of banks and compared it with several other approaches to measuring a bank's contribution to systemic risk including:

(a) *Results of the Fed's February 2009 Supervisory Capital Assessment Program (SCAP)*, see Section 5.2.5.

The SCAP involved the Fed asking each of the 19 largest US banks to identify how much of an additional capital buffer, if any, each bank would need to make sure that it had sufficient capital if the economy got even worse. Acharya et al. (2010) concluded that their MES estimates were relatively highly correlated with SCAP results, after allowing for SCAP results that revealed no additional required capital buffer.

(b) *Actual experience from July 2007 to December 2008 based on MES and leverage data estimated using other data from the year prior to the crisis (June 2006 to June 2007).*

In risk management circles this would be called an *out of sample backtest*. It is a 'backtest' because it is trying to work out how a risk modelling approach would have performed in the past had it been in use then. It is 'out of sample' because it estimates a model only using data that would have been available at the time the model is assumed to have been used, rather than data that only became known subsequently. It is generally considered preferable to use out-of-sample backtesting rather than in-sample backtesting (which uses all available historical data whether it would have been available to modellers at the time) because lookback bias should be less with out-of-sample backtesting.[1]

They contrasted the explanatory power of their estimated MES with other individual measures of a firm's riskiness, such as the firm's estimated beta to the market (as per the Capital Asset Pricing Model), its volatility and an estimate of the individual firm's (non-systemic) expected shortfall. Most of these measures offered reasonably high explanatory power when viewed across the universe of US financial firms that they considered. The measures generally seemed to imply that insurance firms were typically less systemically risky than other financial firms. Essentially all these measures suggested that securities dealers and brokers were riskier than other classes of firm. However, Acharaya et al. concluded that MES was generally *better* at estimating individual performance (i.e. provided a better 'cross-sectional' regression fit) than other measures, if it was estimated using data that was reasonably up-to-date, i.e. not particularly lagged.

---

[1] Lookback bias is the tendency to select models that (had they been used in the past) would appear to have predicted results immediately thereafter artificially well, by including knowledge about what then happened.

However, there were some cases where Acharya et al. (2010) thought that their MES estimates seemed to be producing misleading results. For example, their MES estimates for exchanges were relatively high, even though such organisations do not have the same leverage characteristics as investment banks. Likewise, their MES estimates for insurers like AIG and Berkshire Hathaway were relatively low, but the two had substantially different leverage positions and substantially different experiences through the 2007-09 Credit Crisis. They therefore proposed a refined approach that combined MES and leverage as a predictor of likely contribution to systemic risk.

(c) *Credit default swap (CDS) data*.

In risk management terms this involves using a model that includes in its estimation process more contemporaneous *market-implied* data than is typical for risk models (most of which ultimately estimate model characteristics using merely historic data series).

Acharya et al. (2010) sought to identify if incorporation of information from CDS markets could enhance the predictive capacity of their MES estimates. The CDS premium resembles the spread between risky and riskless floating rate debt, so can help to inform for a given name which periods that name appears to be under most stress. The methodology Acharya et al. (2010) used to estimate MES focuses on such periods. Within specific sectors of the market (insurance companies, depositary institutions, broker-dealers and others) nearly all firms with the highest estimated CDS MESs were ones that suffered major stress or effectively failed during the Crisis. This is probably to be expected, since CDS premiums provide a market view on the creditworthiness of a firm. If we believe that market participants have insights into the robustness of other market participants then we should expect CDS premiums to be correlated with their actual robustness in times of stress (particularly if there is a sufficient level of market transparency and disclosure by market participants to others, e.g. a sufficiently robust Pillar 3 disclosure framework).

5.2.3 CoVaR

CoVaR was proposed by Adrian and Brunnermeier (2011) and was motivated by research they carried out on externalities such as those that amplify liquidity spirals. It aims to measure the extent to which distress at a given institution contributes to the riskiness of the whole financial system. It can be thought of as somewhat akin to SES but focusing on a value-at-risk measure, rather than on an expected shortfall risk measure. More precisely it is the VaR of the financial system conditional on a specified institution being in distress.

Adrian and Brunnermeier (2011) define an institution's 'contribution to systemic risk', i.e. $\Delta CoVaR$ as the difference between the CoVaR conditional on the institution being in distress and the CoVaR in the 'normal' (i.e. unstressed) state of the institution, see Box 5.2. We note immediately that an implicit assumption they adopt is that in normal times institutions should not be in distress, an assumption which might not be true, see Section 3.2.5.

They derived estimates of $\Delta CoVaR$ for the universe of publicly traded financial institutions using quantile regressions (which they thought should be appealing because of their simplicity and efficient use of data). They also quantified the extent to which characteristics such as leverage, size, and maturity mismatch seemed to drive these estimates and hence predict contribution to systemic risk. Although most of their focus was on CoVaR of the system conditional on institution i being in distress, sometimes they focused on what they call 'exposure CoVaR', which is the opposite way round, i.e. the CoVaR of institution i conditional on the system being in distress. This is conceptually akin to systemic risk related stress tests such as SCAP performed by individual institutions, see Section 5.2.5.

Features of CoVaR include:

(a) It does not distinguish between contributions that are causal (caused by institution i being in distress) or simply driven by a common factor.
(b) It is relatively sensitive to /dependent on the tail distribution (because CoVaR is more extreme than unconditional VaR as it characterises what might happen conditional on a 'bad' event having already happened).
(c) It illustrates the endogeneity of systemic risk, i.e. each institution's CoVaR depends on other institutions' risk taking. This reflects its implicit focus on interconnectivities between different institutions as the principal source of systemic risk.

5.2.4 10 by 10 by 10

The aim of this approach, set out in Duffie (2011), is to provide a relatively straightforward way of monitoring interconnectedness in the financial system. It does this by focusing on a small number of the most important elements of interconnectivity, e.g. the 10 most systemically important firms, who are each asked to identify their 10 largest counterparties by exposure magnitude and to quantify for each of these counterparties the results of 10 stress scenarios deemed most important at the time.

This information should be much easier to derive than a complete map of interconnectedness of the financial system. In effect, the assumption being made is that we can derive a suitably complete map of the system by extrapolating from interconnectedness features of just the 10 largest nodes in the system. Particularly large firms are likely to be more interconnected than others, so the way in which interconnectedness alters as we go down the list of major firms should provide some guide as to how the system interconnects for smaller organisations, perhaps using some form of power law extrapolation. If 10 by 10 by 10 is insufficient for this purpose then the approach could be modified to an equivalent $n_1 \times n_2 \times n_3$ as long as each of the n's is kept manageably small. The challenge therefore is to choose the right level of granularity to provide the maximum insight at minimum cost to the whole industry. We also need to minimise the risk that there is a small but highly contagious pocket that is below the granularity threshold used to create the map but above the importance threshold we would like the map to capture.

Some of the (industry-wide) stress tests that regulators have introduced, see Section 5.2.5, have some similarities with this approach. For example, some of these stress tests have asked industry participants to quantify their exposure to counterparties such as specific peripheral Eurozone countries (and/or their banking sectors) under specific scenarios.

However, a significant problem with this type of approach is that it concentrates on *direct* interconnectedness, i.e. links in the system represented by explicit contractual relationship. Its usefulness therefore depends on the extent to which systemic risk follows the domino model versus the tsunami model as explained further in Chapter 2.

5.2.5 Supervisory Capital Assessment Program (SCAP)

As noted in Box 5.3, one response to the 2007–09 Credit Crisis has been to impose requirements on firms to carry out specific stress tests and for summaries of these results to be made public. One reason for doing so is that it provides information to other market participants that can add to market confidence and hence financial stability. Publication of these results may not necessarily be deemed to form part of the Pillar 3 of the regulatory framework formally applying to the relevant sector, but achieves much the same ends, i.e. better market transparency and hence market discipline.

The original (US banking) *supervisory capital assessment program* (SCAP) involved the Fed asking each of the 19 largest US banks to identify how much of an additional capital buffer, if any, each bank would need to make sure that it had sufficient capital if the economy got even worse. The results were made public in May 2009. Ten banks were then required to raise $74.6bn in capital.

Subsequent US bank stress tests, now under the auspices of Dodd-Frank, are explained further in Box 5.3. The results appear to be sufficiently detailed to allow other parties, such as IMF to analyse in some detail the robustness of the US banking system, see e.g. IMF (2015b).

5.3 The Cloning Property

One desirable property Adrian and Brunnermeier (2011) think a systemic risk measure should possess is the cloning property. This is the property that if we split one large (individually systemic) institution into n smaller clones, then the measure of the systemic risk of the large institution should be the same as the sum of the measures for the n clones. Or as they note '*Put differently, conditioning on the distress of a large systemic institution [should be] the same as conditioning on one of the n clones.*' This property is akin to the property of homogeneity that is needed for statistical risk measures to be able to be decomposed into contributory parts as per Box 5.2.

There are some interesting interactions between the cloning property, TBTF and factor based risk modelling as applied to systemic risk. If the cloning property applies fully then TBTF ought to lose much of its relevance, as a TBTF firm should be ascribed the same total risk exposure as the combination of several smaller otherwise identical clones none of which should be large enough to be deemed to be TBTF.

This conundrum can be solved by assuming each of the clones somehow instantaneously acquires different idiosyncratic risk characteristics as soon as cloning takes place. In the language of factor-based risk modelling, firms are assumed to have factor exposures (which they can share with other firms, but potentially in different proportions) and idiosyncratic exposures (which are unique to themselves).

One such idiosyncratic factor that does depend on size is the cost of liquidating assets or liabilities that might be incurred by such an organisation if it were in distress. This is because transaction costs including market impact are in general not independent of the size of the portfolio being bought or sold, see Box 8.4.

5.4 Risk Analytics Used by Policymakers

5.4.1 Introduction

The picture looks different when we consider what policymakers actually analyse, or at least publish. There is a greater focus on the macro-economic picture. The analysis also tends to focus on characteristics of the whole financial system, rather than on individual firms within it. If the macroprudential body has a remit that focuses on a specific geography then the analysis will primarily focus on the financial system within that geography, although there may be reference to wider financial stability concerns (if it is believed that they might spill over to the geography in question).

The sorts of analytics that such policymakers might concentrate on if, say, they are exploring the impact of a low interest environment on the financial system (and hence on something that might encompass the entire financial system) are set out in Table 5.3. We select such a scenario because it might be expected to have an influence on nearly all parts of the financial system, and hence we might also expect the metrics considered to span the whole system.

Regular 'general purpose' financial stability reports would usually have a greater bias towards banking, reflecting the belief amongst most in the macroprudential community that banking is the sector that usually presents the greatest potential financial stability risks (based on past experience). Financial stability reports produced by regulators with specific sectoral responsibilities (e.g. in the EU, EIOPA for insurance and pensions, or ESMA for securities and financial markets), naturally tend to spend more time focusing on risks that are considered to have greater relevance to their particular sector. Most such regulators now seem to have a financial stability directorate or the equivalent, reflecting the increased importance placed on macroprudential matters across the whole of the regulatory community.

Table 5.3 Systemic risk analytics relating to low interest rates

Banks	**Banks**
(a) Net interest income /assets	(q) Life insurance as % of household financial assets
(b) Return on assets	(r) Life insurance as % of GDP
(c) Cost-to-income	(s) Insurers' duration gap (between assets and liabilities)
(d) Coverage ratio	(t) Approx. share of guaranteed life insurance
(e) Loan to deposit ratios	**Pension funds**
(f) Percentage of impaired loans	(u) Pension funds as % of household assets
(g) Forbearance ratio for total loans	(v) DB vs DC split
(h) Tier 1 capital ratio	(w) DB average liability coverage ratios
(i) Interbank market dependence	(x) (Recent) return on assets
(j) Credit-to-GDP gap	**Overall indebtedness metrics**
(k) Banks assets as % of GDP	(y) Sovereign debt to GDP ratio
(l) Share of domestic credit institutions	(z) Household debt to GDP ratio
(m) Share of the top 5 credit institutions	(aa) Non-financial corporations' debt to GDP ratio
Financial markets	
(n) Recent stock price growth	
(o) Share of market-based financing in the economy	
(p) Share of new real estate loans to households on floating rates	

5.4.2 Structure of Typical Financial Stability Reports

A typical financial stability report issued by a regulatory body that has systemic risk responsibilities often takes the following structure:

(a) A summary of the overall perspective of the body including its views on the most important risks to financial stability at that time. These could be specific economic or political developments, or could be perceived market bubbles e.g. in housing or credit markets. They rarely mention individual companies (perhaps for confidentiality reasons, or perhaps because such matters are primarily the remit of microprudential supervisors).

(b) Analytics believed to be relevant to assessing overall financial stability risks (and a description of what they are perceived to be saying at the time of publication).

(c) One or more feature articles that explore in more depth some specific aspect of financial stability. This may be partly to demonstrate to third parties that the body concerned is giving or has given due consideration to issues within its remit even if the analysis suggests that the risks in question are not large.

The sorts of analytics that might be covered under (b) include:

(1) Ex-ante measures that aim to provide early warning indicators
(2) Ex-ante measures that involve 'counterfactual' simulations and stress tests (by 'counterfactual' we mean expressing what has not happened but which could, would or might under differing conditions);
(3) Contemporaneous measures seeking to capture current fragilities and vulnerabilities;
(4) Contemporaneous measures that involve crisis monitoring (if a crisis has already struck);
(5) Ex-post measures designed to carry out forensic analyses of past crises; and
(6) Ex-post measures that aim to assess how orderly a resolution has been, should one have been needed.

A possible downside of a publication strategy such as in (c) is that many macroprudential analysts seem unconsciously (like most researchers) to prefer positive conclusions rather than negative ones, i.e. here concluding that there are potential systemic risks present in whatever they have just analysed. This can lead to the impression that systemic risks are everywhere. Very few publications I have read in this space seem to conclude that there is no need to worry about systemic risk.

Of course, in one sense risks *are* everywhere. However, over-focus on leaving no stone unturned in publicising these risks may bias the readers of such reports (including those within the macroprudential bodies concerned) towards stamping on risks however small they might be. They may then inadvertently forget that to maximise the benefit to society a financial system can bring may require acceptance of some risks within suitable limits, see Section 2.7.

To get a better handle on risks versus rewards, macroprudential bodies may try to create composite analytical measures that highlight the current overall systemic stress level of the financial system. For example, ECB (2010) proposed a measure which the ECB called a *composite indicator of systemic stress* (CISS). Further details are given in Hollo et al. (2012). This measure covered money, bond, equity and foreign exchange markets, as well as financial intermediaries. For each of these five components, stress was measured through several sub-measures such as volatilities, cumulative price declines, risk spreads or recourse to central bank emergency facilities. Input values for these sub-measures were normalised by replacing the actual observations with their quantiles (e.g. if at a specific point in time the input variable had reached its 90th highest observation in a sample of 100 observations then the original observation would be transformed to a value of 0.9), and then weighted in a suitable manner. It reached close to its maximum possible value of 1 at the height of the 2007–09 Credit Crisis. It was generally much lower before 2007, although it did show increases around the 9/11 New York terrorist incident (11 September 2001), the Enron bankruptcy and the Iraq War.

5.4.3 Visualisation of Systemic Risk Exposures

It is said that a picture paints a thousand worlds. As a sweeping generalisation, reports and other material produced by macroprudential authorities and policymakers tend to make more use of visualisation techniques such as charts than is typical for academic papers. Perhaps this is because such authorities have greater practical access to data presentable in this fashion. Also, if the aim is to present a picture of the current characteristics of the financial system then arguably a suitable visual presentation is likely to be a particularly effective way of doing so. For example, if our aim is to map out the extent of direct interconnectedness within a system then network spider-charts can help to highlight concentrations to specific players, see e.g. Fig. 5.1.

Surveys of perceived threats to financial stability in different locations can be heat-mapped (i.e. colour-coded perhaps with red showing perceived higher risks in a specific location and green showing perceived lower risks), if the aim is simultaneously to show visually the most important risks and how these vary by location. Elevated levels of specific indicators can also be effectively highlighted by graphical presentations of indicator levels through time.

5.5 Data and IT System Requirements

5.5.1 Introduction

As Bisias et al. (2012) note, most of the techniques that have been proposed to date impose additional data burdens on regulated industries. It is hard to argue against provision of some additional data to regulators for such purposes. If governments and regulators are implicitly on the hook for financial stability then they have a vested interest in receiving data that might help them

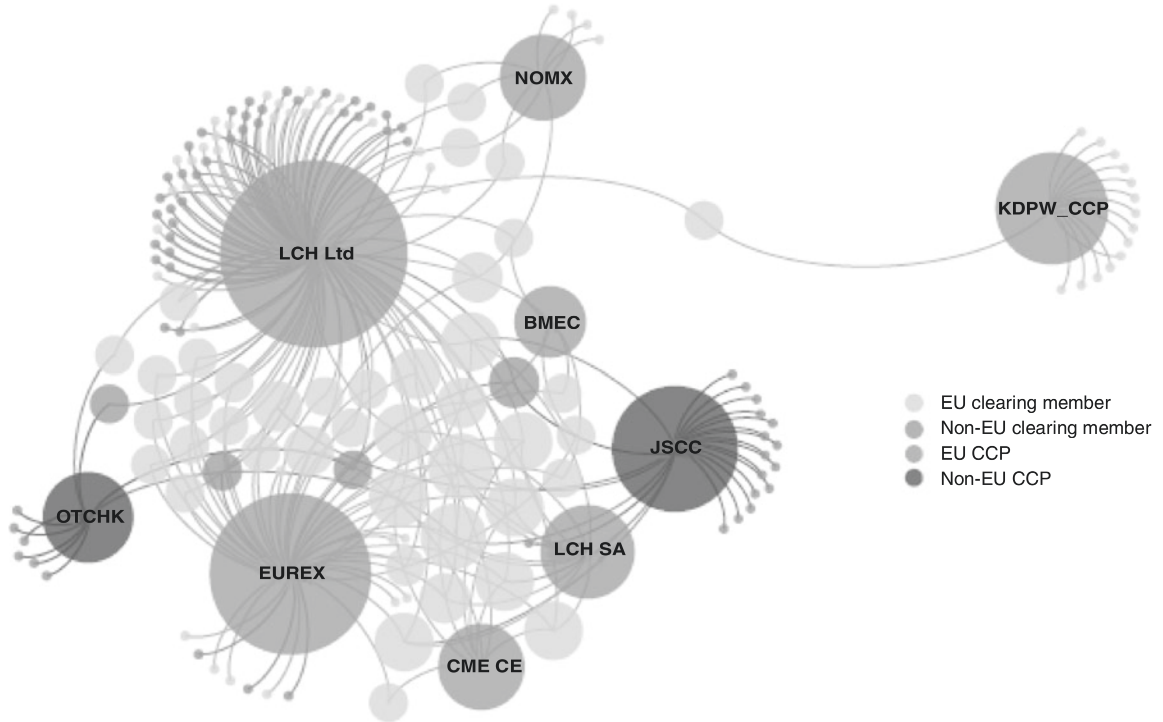


Fig. 5.1 EU Central Counterparty (CCP) network structure

Source: ESRB and Nematrian, reproduced by kind permission of ESRB. Shows clearing links between CCPs and ESMA category 1 Clearing Members as at late 2016, based on ESMA Category 1 reporting data as published on relevant CCP websites. Size of node represents number of links to other nodes in the network chart, not in this instance volumes traded understand (and price) such exposures better. If firms were insuring each other then they would probably want as much data as possible.

But this doesn't make the actual provision of such data any easier or less costly to supply. Particularly vexing is if you don't think that the problem being illuminated by the data is relevant to you. This explains much of the lobbying that some parts of the financial services industry have carried out to avoid being deemed within the scope of systemic risk.

5.5.2 Will the Extra Data Prove Useful?

We might also ask whether it is likely that society will obtain value-for-money from the provision of this extra data. Ultimately, the extra costs of additional data provision will fall on customers. Moreover, Bisias et al. (2012) note the critique given in Lucas (1976) that '*any change in policy will systematically alter the structure of econometric models*', thus invalidating policy advice that might be derived from conclusions drawn from such models. If the data that is being collected is being *primarily* used for econometric modelling purposes then this might significantly weaken the merits of collecting copious additional data.

Ideas in a similar vein to Lucas's (and also published about the same time as his ideas) include Campbell's law ('*the more any quantitative social indicator (or even some qualitative indicator) is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor*') and Goodhart's law ('*as soon as the government attempts to regulate any particular set of financial assets, these become unreliable as indicators of economic trends*').

To try to head off such potential criticisms, regulators tend to emphasise collation of data that they think firms ought to have easily available internally anyway. This includes data on assets held, which regulators think firms need to have readily to hand to understand their own business. Part of the drive for more data can therefore be viewed as regulators believing that firms previously had inadequate risk management disciplines in place.

Transaction data is also seen as potentially reasonable data to request, since businesses clearly need to record trades to manage and settle them. What this view doesn't directly reflect is the effort required to link systems together, which is often much higher than outsiders might expect.

Conversely, more general collation and analysis by firms of data that they use in their day to day business activities may offer them some competitive advantage. Clever analysis and use of data is core to the business models of some of the tech titans, who have created extraordinary value for their shareholders in the process. Is there a role for governments and regulators to foster a more data-centric style of business? Regulatory pressures certainly influence business activities. Some press reports suggest that capital rules applicable to their parents appear to have incentivised bank- and insurance company-owned asset managers to enhance their operational risk analyses more than is the case for independent asset managers (who have typically been subject to less sophisticated regulatory regimes).

If we believe that good risk management disciplines as promoted by regulators adds value to firms (rather than being treated by the firms as just another cost of doing business in a regulated environment) then these sorts of enhancements have the potential to add value to firms willing to undertake them.

But if effective data management (that would make it easier for firms to supply systemic risk data to regulators) is so inherently valuable, why aren't firms doing more of it anyway? The suspicion is that much of the information businesses are being asked to provide is tangential to firms' actual business needs.

5.5.3 Information Technology (IT) Challenges

From the perspective of regulators, analysing and responding to systemic risk involves enormous logistical challenges. Lots of information provided by lots of different organisations needs to be collated into a single overarching picture that illuminates the relevant risks, interconnectivities and vulnerabilities.

Regulators can in principle require firms to provide information in a single standardised machine-readable format. The one that seems to be preferred currently is *eXtensible Business Reporting Language* (XBRL). However, even after agreeing on a suitable machine readable format there is still the need to identify:

(a) What data should be provided, at what frequency, by whom and to whom;
(b) What is meant by a particular piece of data. This generally requires a 'taxonomy', i.e. some sort of classification of the data, and an 'ontology', i.e. a formal logic that applies to the taxonomy (e.g. that one sort of data has some specific relationship to another sort of data); and

(c) What to do because the financial system, like the rest of business life is not static. Methodologies that are appropriate now may not stay appropriate in the future. Financial markets have been very innovative over the last few decades. Some of these innovations have been fingered as contributing to past systemic risk episodes. Effective tools and techniques for analysing and responding to systemic risk information are also likely to evolve through time.

We can illustrate some of these challenges by considering the amount of effort needed merely to handle say interest rate swaps.

First, our data taxonomies and ontologies would need to include the notion of an 'interest rate swap' and the notion of a 'cash flow leg' (since a swap involves parties exchanging different legs). We also need to capture information about the cash flow legs involved, including frequencies of cash flow payments, currencies and reference rates used (e.g. LIBOR versus OIS). To capture (direct) interconnectivities we also need to capture counterparty information, the structure of the collateralisation framework applicable to any given swap and the collateral posted at a specific point in time. One important innovation in derivative pricing since c. 2008 has been the closer attention paid to collateralisation arrangements. Optionality in what collateral can be posted alters the economics of the overall contract although not the underlying cash flows (and this is now recognised better in how swaps are priced).

Even something as conceptually simple as who the counterparty might be offers many practical challenges. Different firms may use different codes or names for the same organisation. Many firms are groups that consist of a range of different legal entities. Distinctions between them can be quite important in terms of how systemic risk might be transmitted. Strict 'ownership' may also not be the sole feature we are interested in from a systemic risk angle. This was highlighted during the recent financial crisis by the impact on banks of notionally off-balance sheet SIVs that they had set up prior to the crisis and into which they placed some of their risks. At the height of the crisis these SIVs proved less unrelated to the originating bank than many banks had hoped would be the case.

Ensuring clear identification of counterparties is seen by regulators as of key importance to building up a meaningful systemic risk picture. Regulators are therefore mandating the introduction of unique industry-wide *legal entity identifiers* (LEI). But even this conceptually simple step offers a rich vein of complexity. To implement it effectively there needs to be agreed ways of creating such LEIs and assigning them to individual companies. Processes are needed to handle situations where organisations subdivide or merge. This initiative is being coordinated by the LEI Regulatory Oversight Committee. Over time we may also find further subtleties that need handling, e.g. ring-fenced funds within the same legal entity and therefore presumably carrying the same LEI but exhibiting different economic characteristics in a systemic risk situation.

5.5.4 'Unstructured' IT Approaches

Building up a clear picture of the structure of systemic risk to the level of robustness envisaged in the previous paragraph is likely to prove very expensive, if not in some cases impossible. Moreover, the financial system does not stay still. This raises the question of whether society will get value-for-money from the effort incurred and how we might arrange incentives within the system so that the desired outcome actually comes to pass. Fouque and Langsam (2013) devote the first few chapters of their *Handbook of Systemic Risk* to the IT and other organisational challenges of carrying out systemic risk analysis. This suggests that enhanced data provision may be one of the biggest practical impacts that the current trend towards greater focus on systemic risk has on most businesses within the financial system.

Most risk managers within most financial organisations can relate to the very substantial amount of effort needed to create risk analytics that are sufficiently reliable to form the basis for robust decision making. Why should systemic risk analytics be any different, except to the extent that the picture needed is even more wide-ranging and therefore presumably even more challenging to paint effectively?

Perhaps, though, we are asking the wrong sorts of questions. Is it possible to make do with much less data and end up with a not much less complete answer? This is the underlying logic behind Duffie's 10 by 10 by 10 approach, see Section 5.2.4. It seeks to provide a broad picture of the nature of interconnections within the system by sampling just a small number of the (largest) interconnections.

Just possibly, the IT picture we have painted above is also flawed, but from a purely IT perspective. Focusing on data taxonomies, ontologies, flow processes and the like sounds very much like a classical 'structured database' way of doing IT. It assumes all information used should have precise meanings and data flows should be capable of being precisely tracked if needed. This arguably doesn't reflect the current direction of travel of many IT technologies especially ones centred on the largest IT network of them all, the World Wide Web.

One of the features of the Web is that there is so much data now within it, flowing around it or being added to it that trying to capture a rigidly structured picture of this data is essentially impossible. Instead, the best we can hope for is to develop techniques that take the 'messy' and rather unstructured data that is held in the Internet and draw necessarily imprecise conclusions from it.

Take for example modern Internet search engines. They often involve very large back end databases that create some structure on which the output of the search engine relies. Overlaid on these databases are relatively heuristic techniques that aim to return information that is as relevant as possible. There is no exact definition of 'relevance' that applies in all circumstances. It differs between users. It also differs for any given user depending on what he or she wants at any specific point in time. So, the search engine can only guess at what might be most relevant. Its guesses are coloured by what seems to have been found helpful by previous users and/or is expected on intrinsic grounds by the search engine provider to be helpful to the user. Sometimes search engines hone in on information the user finds helpful very quickly. At other times, they can be frustratingly ineffective at doing so.

These insights suggest that there is a trade-off between precise manipulation of 'pure' data and imprecise manipulation of 'messy' data. The former might be more desirable from an accuracy perspective, but the latter may cost much less to implement. How much of the former versus the latter should analysis of systemic risk aim to encapsulate?

My hunch is that it is a bit of both, but probably less of a focus on the precise manipulation of 'pure' data than many regulators, academics and third party consultants are currently pushing for. Again, this is related to whether systemic risk is primarily about analysing (direct) interconnectedness in detail or more about searching out underappreciated vulnerabilities. If it is the former then extra effort to map out these interconnectivities is crucial to mitigating systemic risk. But if it is the latter then this extra effort may not be quite so helpful. More important may be to develop the capability to follow lines of enquiry that can search for currently hidden vulnerabilities as efficiently as possible.

5.6 Key Takeaways

This Chapter has explored ways in which systemic risk can be measured, and the consequences that trying to measure it has on IT systems. Key points noted include:

(a) Practical ways of quantitatively measuring an individual firm's contribution to systemic risk commonly build on methods already used by risk managers to measure other sorts of risk. Statistical techniques such as value-at-risk, tail value-at-risk, expected shortfall (and corresponding contributions to these risk measures) are all potentially applicable to measurement of systemic risk. Likewise, many of the tools and techniques used elsewhere within the risk management industry are relevant, such as factor-based risk modelling.

(b) As in other areas of risk measurement, we also see a strong focus on stress testing and reverse stress testing. These techniques are perhaps most applicable to extreme risks and for which it is difficult to work out the likelihood of occurrence, in line with most types of systemic risk.

(c) Some quantitative techniques for measuring systemic risk are more specifically focused on the network nature of the financial system. These include network maps like those used to analyse other types of network, as well as practical simplifications such as the 10 by 10 by 10 approach that aim to characterise such maps at much reduced overall effort but hopefully not much reduced analytical insight. However, many systemic risk events have been characterised by significant indirect interconnectivities. These types of risk are unlikely to be captured well by such maps however granular they are.

(d) An important theoretical property exhibited by most ways of quantitatively measuring a firm's contribution to systemic risk is the cloning property. Some mental gymnastics is required to align the cloning property with the domino theory of systemic risk, i.e. the thesis that the primary cause /propagator of systemic risk is the propensity of failure of one (large) firm to lead to failure of others via a chain-like reaction.

(e) A dichotomy soon comes to light when we explore policymaker publications in this area. Most of the analytics quoted in such documents are more macro-economic and system-wide (or at least sector-wide) in nature. This may reflect confidentiality concerns but more probably reflects the greater weight given by policymakers to uncovering previously hidden vulnerabilities and explaining the current state of the whole system. If the aim ultimately is to facilitate efficient searching out of these vulnerabilities then greater use of some of the less structured IT techniques nowadays used by internet search engine providers may be desirable.

(f) All likely evolutions of systemic risk trends appear to include capturing and processing more data. This has important implications for resourcing within firms caught within the systemic risk net, explaining why some are particularly keen to extricate themselves from it.

References

Acharya, V.V., Pedersen, L.H., Philippon, T. and Richardson, M. (2010). Measuring Systemic Risk. *SSRN.*

Adrian, T. and Brunnermeier, M.K. (2011). CoVaR. *Federal Reserve Bank of New York Staff Report No. 348.*

Bisias, D., Flood, M., Lo A.W. and Valavanis, S. (2012). A Survey of Systemic Risk Analytics. Office of Financial Research, Working Paper #0001.

BCBS (2016e). Standards: Interest Rate Risk in the Banking Book. *Basel Committee on Banking Supervision.*

CRMPG-III (2008). Containing Systemic Risk: The Road to Reform. *Report of the Counterparty Risk Management Policy Group III*, 6 August 2008.

Duffie, D. (2011). Systemic Risk Exposures: A 10by10by10 Approach. www.darrellduffie.com

ECB (2010). Analytical Models and Tools for the Identification and Assessment of Systemic Risks. *ECB*, Financial Stability Review, June 2010, p. 138.

Fouque, J-P. and Langsam, J.A. (2013). *Handbook of Systemic Risk*. Cambridge University Press, Cambridge, UK.

Hollo, D., Kremer, M. and Lo Duca, M. (2012). CISS – A Composite Indicator of Systemic Stress in the Financial System. *SSRN.*

IMF (2015b). United States Financial Sector Assessment Program: Stress Testing – Technical Note. *International Monetary Fund.*

Kemp, M.H.D. (2009b). The VaR vs Tail VaR Mindset: Presentation to the Institute and Faculty of Actuaries Open Forum on Capital Adequacy – Is VaR the best approach. *Institute and Faculty of Actuaries.*

Kemp, M.H.D. (2011). *Extreme Events. Robust Portfolio Construction in the Presence of Fat Tails*. John Wiley & Sons, Chichester, UK.

Lucas, R.E. (1976). Econometric Policy: A Critique. *ScienceDirect.*

McNeil, A.J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton University Press, Princeton, New Jersey.

Rebonato, R. (2010). *Coherent Stress Testing*. John Wiley & Sons, Chichester, UK.

Staum, J. (2011). Systemic Risk Components and Deposit Insurance Premia. *SSRN*.

Sweeting, P. (2011). *Financial Enterprise Risk Management*. Cambridge University Press, Cambridge, UK.

6

Designing and Implementing Macroprudential Policy

In this Chapter, we explore tools that policymakers and regulatory bodies can deploy to mitigate systemic risk vulnerabilities, if they have reached a level where policy action seems to be warranted.

In most walks of life responding quickly but in a proportionate manner to changing circumstances is helpful. For example, when approaching a bend in a road, the usual response of a car driver is to turn the steering wheel smoothly as the bend progresses, rather than suddenly jerking the steering wheel half-way through the bend. Reacting too suddenly runs the risk of flipping the car over with catastrophic consequences.

This analogy suggests that if central authorities telegraph early their planned actions (or even the suite of possible actions they might follow in a variety of yet to happen circumstances) then market participants can plan more effectively. This 'forward guidance' should itself aid financial stability, because it results in more predictable outcomes.

Some central authorities such as central banks are familiar with forward guidance. It is a core way in which they guide market expectations regarding monetary policy. Regulatory bodies also commonly propose approaches that they might adopt, via consultation papers and the like. This too can be thought of as a form of forward guidance. It allows the regulator to get feedback on tentative ideas. It flags up ideas that might be particularly contentious, helping the regulator to work out what ideas to drop or to focus more robust analysis on. It also gives the industry more time to respond to change, since it becomes easier to see the direction of travel.

Sometimes, policy is deliberately introduced abruptly, but this is usually to address clearly inappropriate behaviours or urgent situations once attention

M.H.D. Kemp, *Systemic Risk*, DOI 10.1057/978-1-137-56587-7_6

has been drawn to them. Rapid implementation may also be called for if policymakers think that the industry will arbitrage the proposals in some way were a more gradual approach to be adopted. Hopefully such instances will be relatively rare.

A caveat with forward guidance is that if everyone knows in advance what central authorities will do then they might be able to game the system, taking advantage of these predictabilities for their own ends. For example, if too many people rely too heavily on central authorities to bail them out in times of difficulty then this can itself lead to herding behaviour that pushes out the envelope potentially to beyond a stable equilibrium, given the complex adaptive nature of the financial system, see Box 2.4.

The challenge for central authorities tasked with macroprudential responsibilities is to be transparent over how the financial system is being (and will be) 'managed' (to the extent that management is practical), whilst avoiding implicitly encouraging behaviours that might hinder financial stability.

6.1 The History of Macroprudential Policy Making

If the term 'macroprudential policy' is deemed to relate to responses to systemic risks that have become evident since the 2007–09 Credit Crisis then by definition macroprudential policy will have had only a short history.

For example, Hellwig (2014), in a paper on Systemic Risk and Macroprudential policy presented to the DNB High Level Seminar on '*Making Macro-Prudential Policy Work in Practice*', noted that the EU created the ESRB with a specific mission to '*conduct…macroprudential oversight at the level of Union*', but provided no clear definition of what 'macroprudential policy' actually means. He noted that the term first seems to have been coined by Andrew Crockett in 2000 when warning that trying to ensure the safety and soundness of individual institutions might deliver too little because authorities might then fail to monitor risk and take remedial action linked to their collective behaviour.

In Hellwig's view, prior to the 2007–09 Credit Crisis few specifically macroprudential policies seem to have been adopted by any country. One example that some refer to is Spain's introduction of a countercyclical capital provision for its banks but Hellwig notes that it appears to have been motivated primarily by monetary policy concerns linked to interest rate falls arising from Spain's entry to the European Monetary Union /Eurozone rather than macroprudential concerns. It also didn't stop some parts of Spain's financial system from struggling later in the 2007–09 Credit Crisis and subsequent Eurozone sovereign debt crisis, see Box 4.14 (but is believed by many commentators to have reduced the severity of their impact on Spain).

Presumably, by ‘prior to the Crisis’ he means in, say, the decade preceding the Crisis. Introduction of e.g. deposit insurance in the 1930s (FDIC etc.) can be viewed as a macroprudential policy, or at least one motivated by financial stability, even if not called that at the time.

Lack of prior history is, of course, not a good reason for not responding when the need arises, as Ekholm (2014) notes. She agrees with Hellwig that there are many unresolved issues with macroprudential policy. Ones that she highlights include:

(a) What kind of systemic risk is it supposed to mitigate?
(b) Is the issue financial stability or macroeconomic stability?
(c) How do we assess systemic risk exposure?
(d) How can we establish the degree of commitment needed to ensure that policies are implemented in a timely fashion to prevent the build-up of excessive risks?
(e) What is the relation between macroprudential and microprudential supervision?
(f) How should we think about the trade-offs involved in regulatory forbearance?

But despite uncertainties in how to answer these questions, she stresses the need for authorities to make policy in this area, even if all the relevant effects of such policies are yet to be fully understood. As we have noted earlier, the financial system is a complex adaptive system so intrinsically difficult to model. If we wait until we have a full understanding of how it will react we will wait for ever!

6.2 Longer-Term Implications of Increased Focus on Macroprudential Policy

Central bankers also recognise that an increased focus on macroprudential policy has some major longer term ramifications. For example, Haldane (2014) notes that:

> Macro-prudential policy is gaining ground every bit as quickly as central bank independence did in the 1990s. It has quite radical implications. Pre-crisis

> credit cycles were allowed to operate largely unconstrained. Macro-prudential policy overturns that orthodoxy, with policy instead leaning against the credit cycle to moderate its fluctuations, both during the upswing and the downswing. It, too, is a big step forward.

He thinks that a likely consequence of the 2007–09 Credit Crisis, and the resulting regulatory response that has seen a clampdown on bank capital and liquidity rules, is that the financial system '*will reinvent itself*' with financial activity and risks migrating '*outside of the banking system*'. He is hopeful that the financial system and economy may become less prone to the low-frequency, high-cost banking crises seen in the past. However, he thinks that the financial system could '*exhibit a new strain of systemic risk – a greater number of higher-frequency, higher-amplitude cyclical fluctuations in asset prices and financial activity, now originating on the balance sheets of mutual funds, insurance companies and pension funds*' which could in turn be transmitted to, and mirrored, in greater cyclical instabilities in the wider economy.

Largely absent from such discussions to date have been the systemic risks, if any, posed by pension funds, although they are referred to in passing by Haldane (2014), see above. Most of those in the pension fund industry would argue that pension funds are even less likely to contribute to systemic risk than insurers. However, insurers didn't win this argument with the regulators. It is therefore not clear whether pension funds will do so either. The very largest global (DB) pension funds are mainly sovereign funds, see e.g. Towers Watson (2013). It is tricky to see how in practice sovereign funds might be brought within the scope of macroprudential supervisory principles without creating sizeable political issues. However, there are a handful of private sector funds and somewhat more local government funds that might be large enough to fall within the scope of some of the tests proposed by FSB (2014a).

6.3 Differentiating between Macroprudential, Microprudential and Monetary Policy

A complication we immediately face is to define exactly what counts as a macroprudential policy. Modern thought generally seeks to delineate such policies from monetary policies or microprudential policies. It is perhaps easiest from a conceptual perspective to differentiate between macroprudential policy and microprudential policy although even this is potentially unsound, see Section 2.9.

Microprudential policy is typically defined to be the framework a jurisdiction establishes for formulating and implementing ways of regulating individual (financial) firms. The general thesis is that there is an information asymmetry between a (financial) firm and its customers. Microprudential regulation seeks to address this asymmetry by e.g. requiring individual firms to adhere to minimum capital requirements and minimum standards of behaviour.

Of course, *all* microprudential policy in some sense has a macroprudential angle. The reason we worry about the existence of this information asymmetry is because we believe that if left unchecked it will result in loss of confidence in financial firms, i.e. in some loss of overall financial stability.

In contrast, macroprudential policy can in this context be viewed as the framework we apply to firms in aggregate. It seeks to respond to system-wide factors that might affect large numbers of firms at the same time.

It is somewhat more difficult to disentangle, at least from a banking perspective, macroprudential policy and monetary policy. At its most basic, monetary policy involves actions that determine the size and rate of growth of a country's money supply. In practice the boundaries are blurred, see Box 6.1.

Perhaps responding to all this blurring, a common approach nowadays to defining regulatory responsibilities involves putting all these responsibilities under a single roof. Without careful management, this does not perhaps foster the sort of transparency and forward guidance regarding macroprudential policy that might be most desirable, but it does make it more likely that all three types of policy will be implemented in a reasonably coherent fashion.

Box 6.1: Monetary policy

Monetary policy may be defined as the actions of a central bank, currency board or other regulatory committee that determine the size and rate of growth of a country's money supply, which in turn affects the interest rates prevailing in that economy. Monetary policy is maintained through actions such as modifying interest rates, buying or selling government bonds, and changing the amount of money banks are required to keep in bank reserves.

Monetary policy can be expansionary or contractionary. Expansionary monetary policy typically increases the money supply with the aim of lowering unemployment, boosting private-sector borrowing and consumer spending and hence stimulating economic growth. Contractionary monetary policy typically slows the rate of growth of the money supply or decreases it to control inflation, accepting that this can slow economic growth, increase unemployment and depress borrowing and spending by consumers and businesses.

Some of the 'conventional' tools that central banks use to shape monetary policy involve direct interaction with the money markets, e.g. a central bank's open market operations involving direct purchase or sale of short-term government bonds. There is some potential blurring with macroprudential policy, as monetary policy actions are usually attempting to modify the economic cycle, and hence to provide greater economic stability, which overlaps to some extent with financial stability. However, as these two cycles probably operate to different lengths (see Box 2.5), the blurring is not very large.

In recent years, use of 'unconventional' monetary policy has become more common. This includes quantitative easing, in which central banks buy various types of financial assets from banks (and others), e.g. longer-dated government or private-sector debt. Quantitative easing (when applied to government debt) raises the price of securities and increases total monetary supply. It also in principle adds to what in the derivatives world might be called 'gap' risk, since there is a finite limit to the amount of quantitative easing any central bank can carry out. It cannot buy in more than the total amount of debt its sovereign has already issued. There would be a sudden discontinuity if it reached this limit. 'Credit easing' is another 'unconventional' monetary policy that involves the purchase of private-sector debt from banks (or allowing banks to borrow liquid government debt from the central bank in exchange for depositing less liquid private-sector debt with the central bank).

Unconventional monetary policy would not normally be viewed as an example of macroprudential policy as such. However, quantitative easing has driven down long-dated yields. This strains the balance sheets of some types of organisation, including some types of life insurer and some types of pension fund, leading to potentially unintended macroprudential consequences. Credit easing can also be viewed as a form of regulatory forbearance for the affected credit institutions, and if implemented broadly (as it has been in some jurisdictions recently) can therefore also be viewed as a type of macroprudential policy.

Another policy that would classically have been described as a 'monetary' policy involves imposing (temporarily) higher (or lower) reserve requirements on banks, e.g. requiring that they temporarily hold more (less) of their assets in government debt and/or that they temporarily hold higher (lower) amounts of capital overall. These changes alter their ability to advance credit. If implemented in the name of financial stability then such a policy response would normally be deemed an example of macroprudential policy, indicating that in some spheres there is a very strong overlap between 'macroprudential policy' and 'monetary policy'.

6.4 Banking Sector Macroprudential Policies

6.4.1 Introduction

The most desirable macroprudential policies are ones that directly affect sources of systemic risks without having many undesirable side effects. Macroprudential tools can be applied at the level of:

(a) The financial system, e.g. imposing a system-wide countercyclical capital buffer;

(b) The institution, e.g. declaring that an institution is systemically important, see Section 6.5, or imposing a systemic risk buffer on it;
(c) The individual contract, e.g. imposing loan-to-value (LTV) limits.

Many tools tend to be linked to the credit cycle, since excesses in credit provision often seem to have had a hand in stoking systemic risk fires.

6.4.2 Tools Used to Date

The countercyclical capital buffer (CCB) is a key macroprudential instrument agreed under the Basel III framework. It is designed to counter procyclicality in the financial system. The intention is that by strengthening the capital base during periods of excessive credit growth, the banking system can absorb losses during the downswing of the financial cycle without constraining the flow of credit to the economy. Ideally, during a financial cycle upswing, the CCB is built up, dampening excessive credit growth. Ideally, it is then released during periods of stress or when systemic risks abate, to limit undesirable procyclical behaviours.

ESRB (2014a) suggests that for the whole EU the most appropriate way of setting countercyclical capital buffers is via a credit-to-GDP gap measure (i.e. the deviation of the ratio of credit to GDP from its long-term trend). However, it also notes that this measure does not necessarily perform well in all cases. Other measures that the paper suggests could potentially add to or substitute for it when it seems to be working less well in isolation include:

– residential property price-to-income ratio
– residential and commercial property price gaps (i.e. deviations from long-term trends)
– debt service-to-income ratio for households
– real bank and household credit growth
– deviation of the (deflated) broad monetary aggregate M3 from its trend.

In the EU, the ESRB has a specific mandate to consider systemic risk but its powers are primarily persuasive rather than specific, i.e. it can propose actions but ultimately decision to implement selected actions lies with competent authorities at (usually) the member state level.

To help with its mandate, the ESRB maintains (and publishes) lists of macroprudential measures that have been implemented at the member state level (to the extent that these are communicated to it). These indicate that in 2015 over one hundred individual macroprudential measures were implemented across the EU (more if you include states that access the EU single market but are not actually members of the EU). However, only c. 60 were deemed economically substantial. Virtually all the macroprudential measures it observed at the time related to the banking sector and most derived from EU-wide powers as set out in CRD /CRR rather than national rules.

A high proportion of implemented policies related to the credit cycle. Essentially all at the time had been either neutral (typically procedural, involving formal adoption of powers that may bite further down the road) or tightening. At the time, credit cycles appeared to be getting more unbalanced but economic growth was remaining muted.

Many of these measures have been in the residential real estate area, e.g. introduction of LTV limits. Most of these types of measures are only practically applicable to new loans. However, some were linked to treatment of existing foreign currency loans (which were prevalent in some member states, particularly loans denominated in Swiss Francs). Most measures were capital based, highlighting the fundamental links between microprudential and macroprudential policy.

Most of the remaining policies related to categorisation of banks as global systemically important institutions (G-SIIs) or other systemically important institutions (O-SIIs). These measures are specifically designed to address the 'systemic-ness' of a given institution. However, in some instances it appears that O-SII classification had been used more as a means of applying a capital add-on to an individual firm than for particularly obvious systemic risk reasons. The suspicion is that this tool provides a more flexible way of introducing an add-on than other less explicitly macroprudential powers. At issue is that it is possible to argue that practically any reasonable sized firm has some systemically important characteristics, provided the part of the system to which it relates is not too large.

6.4.3 Identifying a Macroprudential Policy Stance

As we have noted above, it can be desirable to telegraph in advance what powers a central authority is likely to use and when. We might also expect it to be desirable for macroprudential bodies to communicate the overall 'stance' of macroprudential policy, i.e. whether it is likely to be tightened or loosened in the near term. Sometimes merely saying that you are intending to do something can be as effective as actually implementing an action.

There are some (indeed probably many) macroprudential topics where this is likely to be impractical. Where systemic risk arises from hidden vulnerabilities then only when these vulnerabilities become clearer does it become possible to identify what policy stance should be adopted in relation to them.

Where a macroprudential stance is more likely to be practically identifiable is in the credit space. It is no coincidence that this area aligns with where most policies have been implemented in practice, see Section 6.4.2, and where it is more practical to measure the risks involved, see Chapter 5. Policymakers can more practically form the view that bank lending in a specific area is becoming excessive and may therefore wish to place brakes on it. At the other extreme, policymakers might feel (for macroeconomic reasons) that lending is too limited, potentially damaging the wider economy, and might want to take actions to expand lending.

Stated stances might also, for example, align with any intermediate objectives set by the macroprudential body (and any measures it thinks align to these objectives) to help it better achieve any financial stability mandate it has been given. For example, ESRB (2013) lists five intermediate objectives the ESRB has defined in this context (for each of which it has identified some representative macroeconomic measures that relate to that objective). These intermediate objectives are:

1. Mitigate and prevent excessive credit growth and leverage
2. Mitigate and prevent excessive maturity mismatch and market illiquidity
3. Limit direct and indirect exposure concentration
4. Limit the systemic impact of misaligned incentives with a view to reducing moral hazard
5. Strengthen the resilience of financial infrastructures

6.4.4 Challenges

The biggest economic challenge with macroprudential policy setting involves the inherent difficulty of predicting build-ups of systemic risk pressures. In principle, this can be addressed by measuring systemic risk exposures and forecasting how they might develop.

There can also be some implementation challenges, particularly for measures that target systemic risks relating to residential or commercial real estate:

(a) Loans may be primarily sourced through banks. Controlling bank lending will therefore affect real estate markets, but may take some time to make much difference (as most measures can only practically be introduced for new lending).
(b) Non-banks may also provide loans. This lessens the effectiveness of using controls on banks as policy levers. Push banks too hard to limit lending and substitution effects may occur, increasing the proportion of loans being provided by non-banks.

Sometimes broader political developments also change these dynamics. For example, in the EU at the time of writing there is a push in favour of developing capital markets and other non-bank sources of lending. This means that macroprudential policies targeting lending and applied through banks may become less powerful through time, if the fraction of lending that comes from banks declines.

6.5 Identifying Systemically Important Firms

6.5.1 Introduction

We have already noted that politicians and regulators appear to view with increased suspicion the idea that different components of the financial sector are necessarily disjoint when it comes to potential to create, amplify or transmit systemic risk. They worry that any type of firm can contribute to systemic risk.

This is perhaps most evident when it comes to firms that are formally classified as systemically important.

The most systemically important such firms, at a global level, are called global systemically important financial institutions (G-SIFIs). Around 10 insurers have already been classified as global systemically important insurers (G-SIIs) and around 30 banks as global systemically important banks (G-SIBs). G-SIIs and G-SIBs, collectively global systemically important financial institutions (G-SIFIs), are considered potentially 'too big to fail' (TBTF) based on size, interconnectedness, complexity, lack of substitutability, global scope (for banks) as well as volume of non-traditional and non-insurance activities (NTNI) activities (for insurers).

The global insurance industry has generally sought to refute the argument that insurers can create systemic risk, see e.g. Geneva Association (2010) and Insurance Europe (2014). More nuanced are commentators such as Cummins (2013) and Cummins and Weiss (2014) who conclude that although the core activities of life insurers do not pose systemic risks, (US) life insurers are vulnerable to intra-sector crises and both life and non-life (i.e. property & casualty) insurers are vulnerable to reinsurance crises.

Regulators also seem to be worried that non-core (i.e. NTNI) activities of insurers might create or at least transmit systemic risk. They are worried that insurers (and other non-banks) might increasingly carry out activities that are banking-like in nature. For example, the Bank of England in its June 2014 Financial Stability Report, Bank of England (2014b), noted that:

> In the CRE [Commercial Real Estate] sector, data from the De Montfort survey suggested that non-banks originated nearly a quarter of all loans during 2013 H2. Some non-bank lenders are also important providers of household credit. For example, finance companies provided finance for around 75% of new car purchases in 2013 … Lending by insurance companies and pension funds grew further during 2013. Loans to UK businesses from these companies rose to around £35 billion at end-2013 (Chart 1.24). That was equivalent to 8% of outstanding loans to UK businesses, compared with 4% in 2009

Groups that are primarily insurance focused can often have banking subsidiaries and vice-versa. Taking this one step further we can conceive of business models that deliberately seek to offer insurance and banking products alongside each other. This is the concept of *bankassurance*. It was much talked around a decade ago. It is prevalent in some jurisdictions but globally it has somewhat fallen out of fashion as a business model. But who is to say that it won't come back into fashion sometime in the future?

Like leading global insurers did before them, leading global asset managers have been pushing back on the notion that asset managers can pose, create or amplify systemic risk, partly in response to papers such as FSB (2014a) setting out possible assessment methodologies for identifying non-bank non-insurer (NBNI) G-SIFIs. The FSB based its proposals in that paper on the following principles:

> (i) The overarching objective in developing the methodologies is to identify NBNI financial entities whose distress or disorderly failure, because of their size, complexity and systemic interconnectedness, would cause significant disruption to the global financial system and economic activity across jurisdictions.
>
> (ii) The general framework for the methodologies should be broadly consistent with methodologies for identifying G-SIBs and G-SIIs, i.e. an indicator-based measurement approach where multiple indicators are selected to reflect the different aspects of what generates negative externalities and makes the distress or

disorderly failure of a financial entity critical for the stability of the financial system (i.e. 'impact factors' such as size, interconnectedness, and complexity).

The FSB's consultation paper deliberately aimed to be comprehensive, covering proposed methodologies for (i) finance companies, (ii) market intermediaries (securities broker-dealers) and (iii) investment funds (including hedge funds). It also included a 'backstop' methodology applying to all other NBNI financial entities (or entity types) to be used to identify any potential NBNI G-SIFIs not otherwise captured under (i) to (iii). The paper did not propose any specific entities for designation. Neither did it propose any specific policy measures that would apply to NBNI G-SIFIs. Excluded from the FSB's consultation paper were financial market infrastructures (FMIs). This is because under the CPSS-IOSCO *Principles for Financial market Infrastructures*, see CPSS-IOSCO (2012), there is a presumption that all FMIs, as defined in the principles, are systemically important or critical, at least in the jurisdiction in which they are located.

6.5.2 Impact to a Sector of Having Some G-Sifis within It

The global asset management industry (amongst others) is right to be interested in the whether some non-bank non-insurer entities will be classified as G-SIFIs. The longer-term implications for the insurance industry of some insurers being classified as G-SIIs are only now becoming apparent. Specifically, the existence of some G-SIIs is leading to the development by the IAIS of an international insurance capital standard (ICS).

One might have been forgiven in an EU context three or four years before this book was written for ignoring potential longer-term developments in global insurance capital requirements. Even an EU-wide insurance capital standard in the form of Solvency II then seemed in the balance. But Solvency II is now live, allowing greater focus on what might come next.

Once the view is reached that some insurers are G-SIFIs (and should be subject to higher capital requirements) then the following, difficult to fault, chain of argument kicks in, see Fig. 6.1. It implies (potentially major) changes in global insurance capital requirements:

(a) If some entities in a specific financial services sector are deemed globally systemically important then we might expect their regulatory capital framework to require them to hold more capital than less systemically

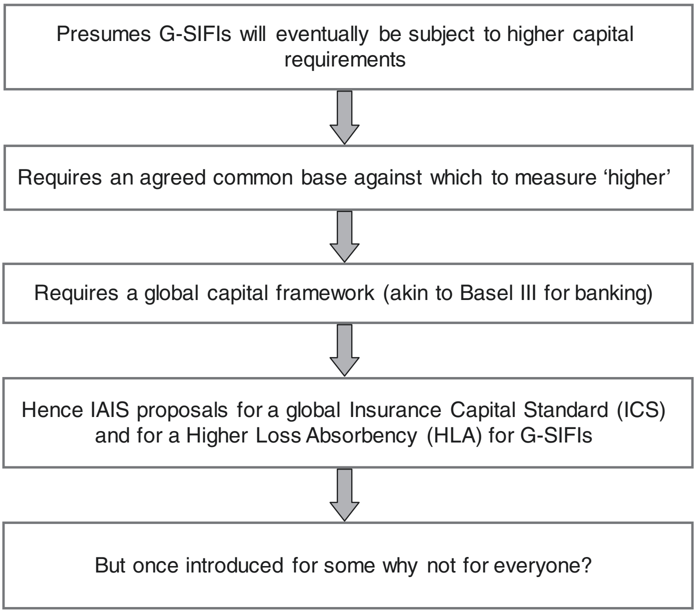


Fig. 6.1 Wider impact of some insurers being deemed globally systemically important

Source: Nematrian

important entities. Not all systemic risks might be mitigated by extra capital, but some may be.

(b) To be able to demonstrate that a G-SIFI holds extra capital, we need to be able to identify how much capital it should hold if it wasn't a G-SIFI and separately identify how much extra it should hold because it is a G-SIFI. The former requires the identification of a suitable (globally consistent) baseline onto which a capital add-on for G-SIFIs can be applied.

(c) To identify such a baseline, you need some form of comparable capital standard that applies across jurisdictions for the sector in question. You therefore need a global standard, rather than e.g. one applying in the EU (Solvency II) and a different one applying in the USA.

This is the backdrop to the proposed new Insurance Capital Standard (ICS) that the IAIS has started to explore. The IAIS is committed to developing the ICS over the next c. 4 years. The ICS is targeted to apply to the 50 or so insurers that are deemed to be internationally active as well as to the subset of them that are G-SIIs. However, many commentators expect that the standards will in due course percolate more widely across the insurance sector, just as the Basel Accords have done in the banking sector. The ICS is expected to be more risk sensitive than the (interim) Basic Capital Requirement (BCR) that IAIS has been developing, see IAIS (2013b) and IAIS (2014). Both the ICS and the BCR lie within the wider backdrop of COMFRAME, the common supervisory framework the IAIS has been developing which is currently undergoing field testing.

Another likely consequence of an added focus on systemic risk is greater longer-term harmonisation of capital requirements across sectors within the financial services industry.

The direction of travel is apparent from the principles being adopted to identify what might constitute a systemically important NBNI institution as set out in FSB (2014a). These principles have been deliberately chosen to be broadly consistent with the corresponding principles for banks, see BCBS (2013), and insurers, see IAIS (2013a).

6.6 Entity-Based versus Activity-Based Regulation

A point that is not so often recognised in debates about capital requirements is that the theoretically correct approach to cater for systemic risk depends partly on the extent to which systemic risk follows a domino or a tsunami model. As we noted in Section 3.1.5, if systemic risk follows primarily a domino model then, all other things being equal, it makes sense to impose relatively higher capital requirements on larger firms. Larger firms are then more likely to contribute to domino chains of failures. Conversely, the case for higher capital requirements becomes weaker if systemic risk follows primarily a tsunami model. The same overall loss can arise from lots of smaller clones, using terminology from Section 5.3.

At the time of writing, there are some insurance specialists who are actively pushing the notion that 'activity-based' regulation is more appropriate for addressing systemic risk associated with the insurance sector than 'entity-based' regulation. Activity-based regulation focuses on regulation of the activities that firms undertake, e.g. imposing controls on firms' securities lending and derivatives activities irrespective of the nature of the firm itself (since these types of activities may potentially introduce systemic vulnerabilities, see Sections 4.7 and 4.8). In contrast, entity-based regulation focuses on regulation of the firm, and is more aligned to how most microprudential regulation is currently structured.

There are some attractions (particularly to those in the insurance industry seeking to escape the systemic risk regulatory net) from running with this type of argument, e.g.:

(a) Activity-based regulation is agnostic about firm type, so may reduce regulatory arbitrage and may foster more level competition
(b) Activity-based regulation appears to fit with the probable direction of travel of e.g. Fed policy with NBNI entities such as asset managers and investment funds. As explained in Sections 4.4 and 4.5, most macroprudential regulatory proposals to date for asset managers and investment funds have been activity-based (partly because it is not obvious how an entity-based framework might work with firms that typically act as agents on behalf of others and so do not typically have large own capital bases).

If systemic risk follows a tsunami model then this would add further weight to arguments favouring activity-based over entity-based macroprudential regulation, but only up to a point. Issues not fully addressed solely by activity-based regulation include:

(1) Asset managers generally behave as agents for others in financial transactions. Do the reasons that seem to favour use of activity-based regulation for them primarily derive from this characteristic of their business model? As explained in Section 4.2, some insurance business models also have an agent-based business model (e.g. unit-linked life insurance). Maybe for them, the argument in favour of activity-based macroprudential regulation is strongest. However, other insurance business models don't share this characteristic (e.g. non-life, non-participating business or life business involving guarantees, if the guarantees may prove onerous). Maybe the case in favour of entity-based macroprudential becomes stronger if a firm typically enters into financial transactions as a principal rather than as an agent.
(2) We may expect macroprudential regulators to be interested in the cost of sorting out a firm if it runs into distress, since this cost might end up being borne by the public purse. If the firm has positions in its own right (i.e. is acting as a principal rather than as an agent) then the costs of unwinding these positions becomes relevant. As explained in Box 8.4, these costs (in particular, market impact) would typically increase (as a fraction of total position value) as the portfolio of positions gets larger.

Applying the conceptual model of regulatory capital set out in Box 3.6, we find that size *is* a hindrance, as long as there is some risk of a failing firm becoming a forced seller, irrespective of the nature of any systemic risk event that might have triggered this failure.

6.7 Central Clearing

6.7.1 Introduction

A good example of an activity-based macroprudential policy involves *central clearing*.

The proposal to require central clearing of standardised derivatives was one early idea for change to arise out of the experience of the 2007–09 Credit Crisis. One key problem central banks and regulators faced was working out who had exposure to whom and how these exposures might affect the firms they might need to bail out. Central authorities felt that they were often operating in the dark. Even when they did have a better idea of the exposures involved they often felt that they were not practically or legally capable of resolving failing firms in a manner that avoided undue drain on the public purse. A perceived major issue here was the huge size in nominal terms of over-the-counter (OTC) derivatives and the opacity to outsiders such as regulators of the different parties to these contracts. This applied even to ones that were relatively straightforward in nature, such as traditional interest rate swaps and simpler credit default swaps.

G20 governments committed in Pittsburgh in September 2009 to require central clearing of standardised derivatives. Specifically, the G20 governments agreed that:

> All standardised OTC derivative contracts should be traded on exchanges or electronic trading platforms, where appropriate, and cleared through central counterparties by end 2012 at the latest. OTC derivative contracts should be reported to trade repositories. Noncentrally cleared contracts should be subject to higher capital requirements. We ask the FSB [Financial Stability Board] and its relevant members to assess regularly implementation and whether it is sufficient to improve transparency in the derivatives markets, mitigate systemic risk, and protect against market abuse.

Those in favour of such approaches argue that they should aid transparency and limit systemic risks posed by large interconnected counterparties (e.g. Lehmans and AIG). This could increase the size above which an entity becomes 'too big to fail' and hence reduce the likelihood and/or quantum of government bailout that might thus occur. Those arguing against such approaches point to the increased risk arising with the central counterparty itself (would we merely be putting all our eggs into one basket?). They may also suggest that there may be good reasons (e.g. uncompetitive pricing by central counterparties) why the earlier, more diffuse, market structure had developed.

6.7.2 Lobbying against Such Changes

Some large non-financial organisations active in derivatives markets (who can be thought of as examples of end customers of the financial system) have lobbied against such proposals because they are worried that they might be forced to post cash margin daily. They think that this would hinder their mainstream business activities. Margin is likely to need to be posted in the form of cash because this is how the CCPs seem likely to operate. Some pension funds have also lobbied against such proposals, because they believe that the opportunity cost of them holding extra cash assets (or being able to access such cash assets from third parties) to be able to post cash collateral will also hinder their activities.

One result of this lobbying is that different regions are implementing central clearing at different speeds. This has created challenges for business models predicated on rapid introduction of central clearing.

In a world in which derivatives are traded bilaterally and if there are n market participants then in principle there may be $n(n-1)/2$ possible pairs of participants who might enter into separate derivative transactions with each other. Moreover, although it is common to net the exposures on different OTC contracts between the same two counterparties (to reduce the counterparty risk each can have to the other) such netting is not universal. When it is applied, it may only relate to specific types of trade. Any such contracts are unique to the counterparties in question, so are not explicitly fungible with contracts entered into with any other counterparties. Moreover, such contracts do not have to follow standardised formats. Capturing a complete picture of the interconnectivities between different market participants created by such contracts can therefore be very difficult. Somewhat mitigating this complexity is that economic factors favour standardisation and centralisation of most such transactions. Even before the crisis, the bulk of such derivative contracts involved deals that had just a handful of leading global investment and/or commercial banks as a counterparty. Terms and collateralisation protocols were generally consistent with master agreements and credit support annexes promulgated by the International Swap Dealers Association (ISDA).

The picture changes in a world in which derivatives are traded on exchanges or electronic trading platforms and centrally cleared. Contracts that would previously have been bilateral are novated very quickly after execution into contracts between each market participant and the clearing house or CCP (or possibly with a market participant's central clearer who then fronts the contract with the CCP). Such an arrangement reduces very substantially the number of pairings it is possible for contracts to exhibit. It requires the contracts involved to be relatively standardised. It should make the contracts more fungible, which ought perhaps to increase liquidity. It should also make it much easier to build up a map of interconnectivities when needed. All this information should be available with minimal delay and in a standardised form by referring to the CCP's position records.

6.7.3 The Risk Consequences of Central Clearing

The actual impact that imposition of central clearing will have on the systemic risk characteristics of derivatives markets is actively debated. At issue is that it is not clear that central clearing explicitly reduces systemic risk per se. Instead it may merely redistribute this risk, including potentially centralising risk into a small number of specialist regulated counterparties, i.e. the CCPs.

For example, Pirrong (2014) notes that arguments put forward for central clearing include:

(a) By allowing more extensive netting, CCPs reduce risk exposures in the financial system.
(b) CCPs will implement rigorous collateralisation (margining) of derivatives transactions. This will reduce both counterparty risk in the system and the potential for the insolvency (or illiquidity) of one major derivatives trader to cause the insolvency (or illiquidity) of other major financial institutions.

(c) Clearing will reduce the interconnectedness of the financial system, thereby reducing the potential for contagion.

However, he argues that none of these views about how clearing reduces systemic risk stands up to scrutiny when we analyse the effects of clearing from a truly systemic perspective. He argues that these views typically evaluate clearing and derivatives markets in isolation from the rest of the financial system and do not consider how the financial system will change in response to introduction of central clearing.

For example, netting through CCPs is typically considered by policy-makers to be systemically stabilising. It reduces the derivatives exposures of SIFIs such as the major global investment and commercial banks. They were previously counterparties to a high proportion of earlier bilateral trades. However, Pirrong notes that increased netting may merely redistribute risk exposures to non-financial firms away from these leading derivatives counter-parties and towards other creditor types. Some of these creditors (e.g. MMFs) may themselves be systemically important. Within the financial sector, exposures to SIFIs may reduce but exposures to CCPs may increase. Moreover, CCPs may themselves have default funds partly supported by SIFIs. This introduces 'wrong way' risk. The SIFIs may be most likely to be called upon to support the CCPs during periods of severe financial turbulence when they may be most vulnerable.

Increasing collateralisation is also typically considered to be systemically stabilising because it typically reduces the amount of leverage (and hence counterparty credit risk) in derivatives transactions. However, the way it reduces counterparty credit risk in effect elevates the priority of derivatives claims on a firm in a distressed situation. Distressed firms will need to post more collateral (and typically more quickly) as any distressed situation unfolds. So again, arguably at a system-wide level it primarily results in a redistribution of risk rather than risk reduction per se. Of course, maybe ultimately redistribution is what governments want, or more specifically redistribution away from the public purse towards other market participants.

At its heart, mandating central clearing is perhaps less about direct industry-level risk reduction per se and more about transparency and ease of resolvability, as far as governments and regulators are concerned. There is little reason to believe that the industry would have embraced it any time soon, without a substantial amount of prodding. It offers too few attractions to too many individual industry participants. Some improved transparency might be feasible without increased central clearing. However, governments and regulators do not seem to believe that other approaches could feasibly have delivered the increased transparency they were seeking within a meaningful timescale. Governments seem to be finding it equally difficult to get firms to make themselves easier to resolve of their own accord, see Section 3.2.5. Again, this is probably because it is not obviously in the interest of an individual firm's shareholders, even if it may lead to a better overall outcome for society.

Central clearing therefore appears to be a change that the industry is only adopting under compulsion. It is creating business opportunities for some at the same time as disrupting existing business models for others. Winners and losers at a firm level depend in part on how regulatory change is implemented and how quickly, see e.g. Sourbes (2014). As Knight (1921) noted, business ventures are subject to inherent uncertainties. Businesses in the financial services industry are not immune from such drivers. Some of this uncertainty derives from uncertainties in how regulatory frameworks may develop.

6.8 Key Takeaways

This Chapter has explored policies that bodies with financial stability responsibilities have either implemented or have actively suggested that they might implement in the near term. Key points noted include:

(a) It is often unclear exactly what impact a given macroprudential policy may have and whether it may suffer from unintended consequences. This complicates decision-making. But not doing anything is not a realistic option either. There is a desire to develop appropriate ways of identifying and communicating a macroprudential stance, akin to those already used by central banks for monetary policy. This is not always practical, particularly for macroprudential risks that involve hidden vulnerabilities. But it is more practical for some areas such as tackling credit bubbles (which might previously have been deemed within the remit of monetary policy before the concept of macroprudential policy became more widely entrenched).

(b) Most macroprudential policies already implemented in practice have focused on the banking sector, given its dominant role in the 2007–09 Credit Crisis. Tools here include raising capital requirements for bank lending, increasing risk weights applied to specific assets and putting limits on some types of lending.

(c) A common tool that has been applied to the EU banking sector is to require firms to hold additional systemic risk capital buffers or to put them on notice that they may be required to do so. However, some of this activity appears to reflect the relative ease of introducing such buffers versus other ways of applying microprudential overlays to capital requirements, highlighting the fluid boundaries that exist between microprudential policy and macroprudential policy.
(d) A particularly important tool for the firms affected is the classification of some firms as systemically important. Such a classification typically leads to enhanced capital requirements and/or greater supervisory attention. The mere fact that some firms in a sector can be classified in this manner creates tailwinds for change for other firms in the same sector. For the insurance sector these are likely to include likely greater harmonisation of capital requirements across the globe. Classification of systemically important firms is not limited to the banking and insurance sectors, although the precise implications for other firms and organisations classified as SIFIs is still a work in progress.
(e) For the broader spectrum of capital markets participants another important development since the 2007–09 Credit Crisis has been the introduction of mandatory central clearing for some types of derivative instrument. This has created business threats for some players and business opportunities for others. Opinions differ as to whether this change will lead to a safer financial system, but it is in due course likely to lead to a more transparent system, as far as policymakers are concerned.

References

Bank of England (2014b). Financial Stability Report, June 2014. *Bank of England.*

BCBS (2013). Global Systemically Important Banks: Updated Assessment Methodology and the Higher Loss Absorbency Requirement. *Basel Committee on Banking Supervision.*

CPSS-IOSCO (2012). Principles for Financial Market Infrastructure. *Bank for International Settlements Committee on Payment and Settlement Systems and Technical Committee of the International Organization of Securities Commissions.*

Cummins, J.D. (2013). Systemic Risk and the U.S. Insurance Sector. Presentation to The Brookings Institution Conference on regulating Non-Bank SIFIs.

Cummins, J.D. and Weiss, M.A. (2014). Systemic Risk and the U.S. Insurance Sector. *Journal of Risk and Insurance*, 81, 489–528.

Ekholm, K. (2014). What should be the Ambition Level of Macroprudential Policy? *Paper Presented to DNB High-Level Seminar 'Making Macroprudential Policy Work in Practice'.*

ESRB (2013). Recommendation of the ESRB of 4 April 2013 on Intermediate Objectives and Instruments of Macro-Prudential Policy (ESRB/2013/1). *European Systemic Risk Board.*

ESRB (2014a). Operationalising the Countercyclical Capital Buffer: Indicator Selection, Threshold Identification and Calibration Options. *European Systemic Risk Board, Occasional Paper 5.*

FSB (2014a). Consultative Document: Assessment Methodologies for Identifying Non-Bank Non-Insurer Global Systemically Important Financial Institutions: Proposed High-Level Framework and Specific Methodologies. *Financial Stability Board.*

Geneva Association (2010). Systemic Risk in Insurance: An Analysis of Insurance and Financial Stability. *Geneva Association.*

Haldane, A.G. (2014). Halfway up the stairs. *Bank of England.*

Hellwig, M. (2014). Systemic Risk and Macro-Prudential Policy. *Paper for Nederlandse Bank High-Level Seminar on 'Making Macroprudential policy Work in Practice'*, June 10, 2014.

IAIS (2013a). Global Systemically Important Insurers: Initial Assessment Methodology. *International Association of Insurance Supervisors.*

IAIS (2013b). Capital Requirements for Global Systemically Important Insurers (G-SIIs): Proposal. *International Association of Insurance Supervisors.*

IAIS (2014). Basic Capital Requirements for Global Systemically Important Insurers. *International Association of Insurance Supervisors.*

Insurance Europe (2014). Why Insurers Differ from Banks. *Insurance Europe*

Knight, F.H. (1921). *Risk, Uncertainty and Profit*. Cosimo Classics, New York, USA.

Pirrong, C. (2014). A Bill of Goods: Central Counterparties and Systemic Risk. *Journal of Financial Market Infrastructures* 2(4), 55–85.

Sourbes, C. (2014). EU clearing timeline could drive banks out of the business. *Risk*, September 2014.

Towers Watson (2013). The World's 300 Largest Pension Funds – Year End 2012. *Towers Watson.*

7

Network Effects and Societal Shifts

No discussion on systemic risk would be complete without discussion of general trends in systems and interconnectivity across society. These loosely go under the name of network effects. Perhaps the most obvious of these are ones promulgated by modern IT tools such as the internet and social media.

There is arguably an entire branch of risk management, namely 'cyber risk' linked to this topic. Typically, 'cyber-risk' is most commonly used in a narrow sense to refer to cases where the systemic risk event is intentionally created by players wishing to profit from its consequences. Here 'profit' is used in the widest sense. It may involve merely a political desire to see the system run into difficulties or it may involve a financial criminal profit motive.

The largest 'network' of all from a human perspective is the whole of human civilisation. No discussion of network effects would therefore be complete without consideration of wider societal trends. This is too broad a topic to address fully in a book just on (financial) systemic risk. We therefore focus our attention on societal trends that seem to relate to systemic risk and/or to the financial system, paying most attention to evolving notions of 'fairness' (since these underpin the social contracts that form the bedrock on which any financial system is built).

7.1 Cyber Risk

The relative importance of cyber risk to the risk management community can perhaps be gauged by identifying the proportion of entries on this topic in cross-practice risk management knowledge databases. For example, the

M.H.D. Kemp, *Systemic Risk*, DOI 10.1057/978-1-137-56587-7_7

RIMS Risk Knowledge database (see www.rims.org/RiskKnowledge/RiskKnowledgeMain.aspx) contained 570 entries (articles, white papers, webinars etc.) and as at 2 September 2014 of which roughly 7% were focused on cyber risk at that time. This is not an insignificant proportion bearing in mind that most of the categories with more entries were more general in scope (e.g. 'Risk Management (General)', 'Global' and 'General Management').

The rationale for this interest in cyber risk is summarised by e.g. Rudolph (2012):

> Extraordinary online business benefits have revolutionized business and, as digital interconnectedness continues growing daily around the globe, so too do the implications of its power. Managing assets and financial risk in business today relies heavily on the speed and ubiquity of computer connections and networks globally … But, for the nation's risk managers, it is clear that cyber-risk has become the revolution's menacing dark side. Increasingly, headlines spotlight massive credit card privacy breaches, allegations of sovereign espionage, and 'hacktivists' penetrating the firewalls at the Department of Justice and other federal agencies, sending shudders through risk officers charged with protecting corporate assets, regardless of whether those assets are intellectual property, financial transactions, customer data, supply chains or infrastructure.

How important is cyber risk to financial stability? Reasons for believing that it is important include:

(a) IT is an increasingly important and complex component of much economic activity, including activity within the financial services sector. High street banks are increasingly relying on telephone, mobile and internet banking and closing physical high street branches. They also appear to be exploring ways of changing themselves into more explicitly IT orientated businesses, see e.g. Financial Times (2014).
(b) Some sectors of the financial industry are very heavily reliant on IT. Indeed, some commentators argue that some leading investment banks and financial infrastructure players might as well be IT companies with a financial services spin given the relative size and importance of their IT activities.
(c) Business activities that create value by leveraging network effects are inherently sensitive to downside if these networks are disrupted or trust in them is compromised. Core components of the financial sector, e.g. stockmarkets, are useful to their participants precisely because they

leverage network effects. The financial system should therefore be peculiarly sensitive to such disruptions.

(d) We have become so used to continuous access to modern technology that even short outages or failures can have severe reputational consequences.
(e) Financial transactions involving IT are often reliant on similar security protocols to those used by other parts of the internet, e.g. Secure Sockets Layer (SSL) and Transport Layer Security (TLS) protocols and their cryptographic underpins. There could be hidden vulnerabilities present in these protocols which might create systemic risks. In Section 7.4.2 we refer to the possible impact quantum computers might have on the robustness of these protocols.

Conversely, others might argue that there is a danger of overreaction because:

(a) For all the claimed importance of IT to the financial services sector, ultimately its core business activities are not explicitly IT in nature.
(b) When things 'go wrong' authorities and/or courts ultimately have the power to cancel or unravel inappropriate trades. Ultimately financial service activity involves changing ownership apportionment of other more tangible contributors to economic cash flows. This ownership ultimately depends not on IT per se but on legal jurisprudence, legislative decisions and future economic developments. In other words, there are lots of other sources of financial risk. Some of these, like wars, can be expected to propagate through the financial system whether the system makes any use of IT.
(c) The same tendency to reinvent financial sector business models into ones with a more explicit IT focus was evident in the dot com boom. It mostly unravelled in the subsequent dot com bust.
(d) This is not a 'new' threat as such. The financial community already expends a significant effort to mitigate its potential impact.
(e) Behavioural finance argues that we all exhibit behavioural biases such as the 'framing' bias. We are all heavily influenced by what everyone else views as important. Within modern culture (e.g. films, books, TV shows etc.) there is a strong dystopian strand, e.g. action movies where the world is saved from disaster. 'Disaster' in such movies is increasingly likely to include an IT element. Perhaps we are merely projecting these fears into our working environment. Over-focus on 'cyber security' may be just as ineffective at adding value to society as over-focus on the Year 2000 Bug was during the dot com boom and bust.

On balance, it does seem likely that there are some important cyber security issues for the financial community to address. Regulators such as the UK's Prudential Regulatory Authority certainly seem to think so, particularly for those parts of the financial sector that are 'systemically important'. This is here typically equated with importance in relation to the operational network that underpins the financial system as we currently know it. For example, Gracie (2014) indicated that:

> But cyber presents new challenges. It is not a game against nature. Unlike other causes of operational disruption like fires and floods, we know there are agents out there – criminals, terrorist organisations or state sponsored actors – that have the will, if not necessarily the means, to attack the system. Motivations vary. More often than not they are economic – to defraud banks or their customers or to extract information. But we have seen cases where the motivation is to damage the system, either to destroy data or cause non-availability of systems or both.

Gracie argued that the financial community is likely to need to go beyond existing cyber security standards more generally applicable to the business sector, e.g. the '*Ten Steps to Cyber Security*' promoted by GCHQ (2012). In his speech, he introduced a new framework, CBEST, which focused on IT vulnerabilities within the financial services sector, see Bank of England (2014c). Firms or FMIs that have been identified as being core to the financial system will be expected to follow appropriate processes as laid out in CBEST to test their cyber security.

7.2 Entrepreneurialism Versus Conservatism

Many commentators perceive much of the overall value added from society from IT as having come from relatively entrepreneurial and experimental approaches to business. Whether IT developments present 'unacceptable' financial stability risk can also be framed as partly a discussion about where along the spectrum between gung-ho entrepreneurialism and stifling conservatism we want the financial services industry (or at least our bit of it) to be positioned.

Established business models have since the earliest of times been susceptible to disruption from new entrants. More recently these disruptions have increasingly included a strong IT element. In recent years, industries such as the music and book industries have been reshaped by IT companies such as

Apple and Amazon. Perhaps the business risk elements of firms' risk profiles are particularly sensitive to IT trends even if other elements of their overall risk profile (such as market and credit risk) are less obviously impacted. Some commentators argue that the financial services industry is heavily regulated, which introduces high barriers to entry (at least for firms proposing radically different business models). But other industries have been disrupted even though they seemed at the time to face (other) high barriers to entry.

Probably regulators and governments want the financial sector to be *both* entrepreneurial (provided the entrepreneurship is customer focused) and strongly focused on mitigating cyber risk and systemic risk. If this is achievable then it would offer the best of both worlds, providing the maximum benefit to society and sustaining the maximum trust in money as a medium of exchange.

7.3 Interconnectivity and Knowledge Sharing

Cyber security is not the only way in which networking and associated interconnectivity is likely to influence the financial sector. In the above discussion, we have focused primarily on the infrastructure on which IT software (and hardware) operates. Arguably even more important may be networking effects linked to what IT software (and hardware) is designed to accomplish.

Adopting a long-term perspective, we might view human history since the Stone Age as involving incremental accrual and dissemination of technological knowledge and expertise. At least that is the optimistic perspective. The pessimistic alternative is that ecological exhaustion, natural disaster, plague, major war, an artificial intelligence singularity or other catastrophe is waiting around the corner to trip up our ultra-highly specialised and interdependent society, taking us back to the Stone Age if not worse. Either way, within this broader context, modern information technology is just the latest tool (albeit a particularly effective one) that we have developed as a species to be 'hyper-social' and to share ideas and technology with each other. Following this line of thought, a propensity towards networking, interconnectivity and knowledge sharing can be argued to be in the human DNA. It can thus be expected to have a pervasive influence on how we act and think, in risk management as in other areas of life.

Many aspects of regulation can be viewed through this lens. For example, at a high level, regulators and politicians favour adoption of common regulatory structures, such as the three Pillar framework underpinning both Basel III and Solvency II, see Fig. 3.2. Kemp (2005) noted that this trend is amplified if the relevant regulators are 'unitary', i.e. regulate the whole (or large parts of) the financial industry. Adoption of common regulatory approaches is also facilitated by sharing of ideas and contacts within and across organisations. These networking activities can help build consensus on how regulation 'ought' to be structured and implemented. Introduction of modern financial regulatory frameworks require huge commitments of resources from both regulators and industry. Successful implementation requires broad agreement across multiple constituencies.

Why, also, do these regulatory frameworks include a third pillar that focuses on market transparency? The accepted view is that sharing of (some) knowledge across markets about the financial state of individual market participants is inherently desirable. It promotes trust between market participants. It is only by adoption of information sharing protocols, i.e. a 'network', that such dissemination can occur.

Equally relevant are impacts more directly associated with advances in IT. The computational aspects of risk measurement have changed dramatically over the last 20–30 years. The calculations involved have become much more sophisticated and detailed, as the computing power available to firms to apply to such tasks has expanded. The more general growth in computing power (of which the financial services industry is only one of many beneficiaries) has been facilitated by sharing of software approaches and hardware manufacturing techniques. Modern economic and academic activity has facilitated these developments. All these contributors have in turn been helped by the accumulation of human, physical and intellectual capital across the whole of society. The financial services industry has arguably played a major part in fostering the accumulation of this human, economic and intellectual capital.

Extending this line of logic, we might view the whole economy as one particularly large network involving a particularly large number of participants. The economic growth a high proportion of us have benefited from over the last few decades might then be viewed as a particularly compelling example of a network effect. Such a view is, of course, core to the concept that the economy can be disrupted by 'systemic' risks.

7.4 Can Advances in IT 'Solve' Systemic Risk?

7.4.1 Introduction

We have become used to steady advances in central processor unit (CPU) power and memory resources as epitomised by Moore's Law. These have led to extraordinary changes in our day-to-day lives. Problems that have taxed experts in the past, such as playing chess, developing autonomous cars and developing robotic surgery techniques, have proved easier to solve using computing than many might have forecast even a few years ago. Can we 'solve' systemic risk in a similar fashion, by throwing more (and cleverer) computer power at it?

I am sceptical that this will prove to be the case, at least in the near term.

Probably, potential enhancements to IT processing power will continue for a while yet. Researchers point out that advances implicit in Moore's Law continuing will inevitably eventually dry up. Scientists and engineers have in the past managed to circumvent what were perceived to be major challenges to Moore's Law, although for how long this will continue is debatable. Some commentators point to potential further exponential leaps in computing power, commonly highlighting the possible development of quantum computers. Other commentators such as Markov (2014) note that for many real-life computational tasks quantum computers offer relatively little theoretical scope for speed enhancement. Perhaps the biggest impact of quantum computers if they can be commercialised is that they appear likely to offer potentially significant speed enhancements over traditional computers in the factorisation of very large integers. The current difficulty of this mathematical problem underpins many existing cyber security protocols, taking us back to the discussion on cyber risks earlier in this Chapter.

Advances in CPU power and memory have already had a major influence on risk management toolsets and activities (and arguably therefore also on underlying regulatory frameworks) and this will probably continue. For example, the Standard Formula SCR under Solvency II involves application of multiple stress tests to a firm's balance sheet. The effort involved for complex firms is considerable, relative to what would have been considered practical even just say 20 years ago. Creating internal models with the level of credibility now required to get supervisory approval would have been difficult or impossible then. The same comments apply to the banking industry. The amount of information that firms are being required to publish (and the extent to which this information will need to be made IT-readable) is in the process of increasing dramatically.

And yet, some might question whether all these past IT advances have truly enhanced our ability to manage the risks of big financial institutions and whether imaginable future IT advances might also struggle to improve our understanding much further.

7.4.2 The Intrinsic Difficulty of Risk Management

This issue is linked to how intrinsically easy or difficult it is to answer questions of the sort underlying risk management. Ultimately, most risk management involves taking actions that depend on effective extrapolation of past behaviour into the future.

Extrapolation is an inherently challenging mathematical problem, as noted by Press et al. (2007). This is because we don't know for sure whether the data we base our extrapolation on will be representative of the future. Economists might reach much the same conclusions by referring to the seminal work of Frank Knight, see Knight (1921). He noted that most business activities are inherently uncertain, rather than merely being 'risky' in a statistically measurable sense. Hence we use the term 'Knightian uncertainty'. Indeed, in his view one of the core attributes of an entrepreneur is a willingness to take on such uncertainties. Something that is fully Knightian uncertain is inherently not mathematically measurable. No amount of computing power can be expected to fully answer questions we might have about such uncertainties.

Take, for example, credit risk modelling. Three common ways in which portfolio credit risk is modelled involve ratings-based models, equity-based models or mixture models, see e.g. Nematrian (2014). However, all these approaches require assumptions about correlations between different issuers if they are to allow for diversification effects. Often these correlations are in practice derived from correlations between the stock returns on the equities of the different issuers (to the extent that these are available). It is well known that such correlations are not very stable through time. Portfolio credit risk modelling, however it is done, faces the inherent difficulty of estimating what these correlations will be in the future (rather than merely what they have been in the past).

7.4.3 The Speed at Which Risk Management Needs are Becoming More Complex

The assumption that exponential increases in computing power will in due course make all relevant risk management problems amenable to analysis itself depends on another assumption. This is the assumption that the complexity of the problems to be solved is not itself changing. This too is doubtful.

A common way of deriving risk management sensitivities is to ‘bump and revalue’ the balance sheet by applying small shocks consecutively to each input value driving the end valuation. As instruments have become more complicated so to have these computations.

Application of clever mathematics can sometimes offer substantial computing benefits. For example, suppose we express balance sheet values primarily in an algebraic rather than a numerical manner (e.g. with the value of derivative positions expressed in terms of mathematical functions which only at the end of the process are then converted into numerical values). Then a theoretically much quicker way of calculating the sensitivities (and of calibrating the valuation to market prices) can be to derive the sensitivities algebraically. This is the basis of adjoint algorithmic differentiation, a branch of computational finance, see e.g. Homescu (2011). Such refinements are not necessarily easy to incorporate within existing risk management systems, reminding us that a firm’s risk management activities are subject to the same sorts of trade-offs between seeking returns on past investment and making new investments for the future as any other business activity the firm faces.

One of these clever techniques is the development of so-called ‘proxy’ models, see e.g. Cocke et al. (2014). These model the behaviour of other more complicated models to make it easier to apply risk management disciplines in near real time. We only need such models because the underlying models which they proxy take so long to run. The use of proxy models seems to have increased of late, despite advances in computing power that presumably can be thrown at the underlying models. This is great for consultants and others developing such models, but what does it tell us about the direction of travel of the broader risk management industry (of which systemic risk management is ultimately a part)? Maybe growth in complexity of modelling requirements (including requirements being imposed by changing regulatory frameworks) is outstripping available improvements in CPU power and memory resources.

7.4.4 Addressing privacy

A further question we should ask is whether societal choices will frustrate the ability of IT advances to address systemic risk issues.

As we have noted previously, some computer scientists are quite positive about the ability of firms to harness growing computing power to further business goals. This underlies the current enthusiasm for 'Big Data'. Firms such as Google and Amazon constantly monitor our electronic footprint. They collect so much of it that they can in effect analyse what 'everyone' does, rather than having to extrapolate from the behaviour of small and possibly unrepresentative samples of customers.

There is little doubt that Big Data will be an important strand in how some firms' business models develop, perhaps tempered by constraints imposed by regulators. However, it does not necessarily have so much to offer for some risk management purposes. Extrapolation remains an intrinsically challenging endeavour, however big the dataset.

Many advances in IT computer hardware and software come with a potential downside, in the form of reduced privacy, see e.g. Mayer-Schönberger and Cukier (2013). The extent to which people view privacy as important can vary by society. However, there is little doubt that public concerns in this area have been heightened by discovering that agencies such the U.S. National Security Agency have been undertaking mass electronic spy programs.

Take, for example, the use of *telematics* in the insurance industry, e.g. the use of data collected in real time as a car is being driven by the insured driver. The thesis is that this information can help the insurer identify whether a driver is a higher or a lower risk. The insurer should then be able to adjust premiums accordingly. If telematics turns out to be sufficiently good at differentiating between customers then it is likely to become widely adopted (provided societal norms do not consider it 'unfair', see later). But in life insurance the equivalent of telematics is applying own-genome testing to set life and health insurance premium rates. Is it 'fair' for insurers to use such approaches? Not everyone thinks it is, because it might result in some individuals being effectively unable to benefit from (protection-based) life insurance.

Concerns over privacy can perhaps explain the apparently relatively slow take-up of 'cloud' computing by financial services firms. Cloud computing could be another way firms and regulators could leverage IT advances to facilitate provision and analysis of systemic risk related data. Cloud computing involves execution of computer software steps largely 'in the cloud', i.e. on remote servers usually owned by or rented out from third parties. Financial services firms wanting to make extensive use of such techniques will typically need to transfer (potentially sensitive) data on individual customers to the cloud infrastructure. They may be more sensitive than other potential cloud users about privacy issues because of regulatory requirements imposed on them regarding the use of such data.

However, the apparently relatively slow take-up of cloud computing by financial services firms may just be an example of careful presentation of material to third parties. Some commentators believe that many financial sector firms are already making quite considerable use of cloud computing (e.g. using e-mail systems hosted in the cloud or other less contentious cloud based systems made available by larger and perceived to be more reliable IT vendors).

Finally, we might note that nowadays we have a lot more information to hand, but it is still hard to create competitive advantage from collating it. Everyone else is also seeking to enhance their management and risk information. Following this line of reasoning, we might expect some upper limit to apply to the amount of effort regulated firms can reasonably apply to risk management (including any associated with systemic risk). There still needs to be a core business generating revenues to be able to afford to invest in such activities!

7.4.5 Blockchain and Other Related Technologies

Like any other industry, the IT industry can sometimes experience hype. At the time of writing it seems to have become particularly fashionable to promote the potential usefulness of blockchain technology.

A blockchain involves a mutual distributed ledger. This is a record of all historic transactions which everyone using the blockchain can access (and which everyone actively involved in updating the blockchain keeps a copy). It potentially offers several advantages over other ways of recording and implementing (financial) transactions, including:

(a) *Potential to transfer value without a trusted intermediary*. Usually if you wish to participate in a significantly sized financial transaction you need to use a trusted third party (e.g. a bank). Blockchain avoids the need to do this because participants can simply reassign ownership of value directly on the ledger itself.
(b) *Potential for 'smart contracts'*. A smart contract is a piece of self-executing computer code that runs on the blockchain. If the process of updating of the ledger is designed in a suitable manner then we can arrange for it automatically to execute additional transactions as specified in earlier entries in the ledger, allowing e.g. automatic settlement of trades and other financial transactions.
(c) *Immutability*. As each new block of data is added to the blockchain ledger, an algorithm known as a cryptographic hash is run, producing a unique value based on the content of that block of data. It forms the start

of the next block making it in practice impossible unilaterally to amend earlier blocks, i.e. earlier entries in the ledger.

(d) *Cyber-resistance*. Every participant holds a complete copy of the synchronised ledger, considerably reducing the risk of cyber-attacks compromising the ledger itself.

(e) *Cryptographic security*. If needed, it is possible to control who has practical access to the meaning of different parts of the ledger, using combinations of public and private key encryption. This means that blockchains can be structured so that participants only have practical access to data relevant to themselves. This has some appeal to market makers who are sensitive to the so-called winner's curse, see Box 8.4. It may have less appeal to law enforcement officers who may lose the ability to identify who owns what and whether ownership rights have been gained illegally.

(f) *Scope for operational efficiency gains*. Organisations often spend a significant amount of time and effort carrying out reconciliation processes, checking that data in different ledgers within their own organisation or across organisations are consistent with each other. If there is just one (distributed) ledger then reconciliation can be simplified.

The best-known example of blockchain at the time of writing was the Bitcoin virtual currency. As at end September 2016 the total market value of Bitcoins held in its blockchain was approximately US$10 bn (source blockchain.info). The total size of all its blockchain headers and transactions (not including database indexes) was approximately 84 Gbytes.

Whilst blockchain technology does appear to offer some advantages to financial organisations, some of the hype surrounding it appears to me to be overstated. Specifically, it is worth asking how theoretically sound it is from an economic perspective to believe that financial transactions can be generally arranged without the need for *any* sort of trusted intermediary.

Marshall (1924) notes that 'money is not desired mainly for its own sake, but because its possession gives a ready command of general purchasing power for its own sake'. This suggests that Bitcoin and other outputs of blockchain technology ought to be capable of functioning as money, provided they offer a ready command of general purchasing power.

But there is a problem. Suppose someone has money in a bank account (or even in notes and coin). It only provides a ready command of general purchasing power if it is accepted by others as doing so. But what happens if society decides that the person shouldn't have access to this money, e.g. if the courts decide that the assets should be seized to meet previously unpaid tax liabilities, or that the assets were not rightfully owned by the person in the first place? The 'money' is (forcibly) redistributed even though all entries in the ledger corresponding to the bank account had up to then showed that it was owned by the person in question.

At its heart, society is organised with notions of fairness and propriety which imply that at times it will reject the outcome of automatically derived ledgers of who owns what. Some participants in Bitcoin appear to want to use this technology precisely because it offers scope to circumvent these notions of propriety and to profit from shady or illegal activities. Any type of money that has broad appeal (rather than being a one-on-one type of transaction) requires some form of broader trust. Almost certainly this broader trust can only be maintained if it includes scope for public authorities to override the supposed immutability that is the foundation of a distributed ledger system. Maybe this can be achieved using smart contracts giving overriding powers to these authorities, but introducing such features may then make the blockchain more susceptible to cyber-attack.

Another way of thinking about blockchain technology is to view it as in some sense providing the digital equivalent of a *bearer security*. A bearer security is an instrument carrying some value whose practical ownership resides with the physical holder of the security rather than anywhere else. If a bearer security is stolen then the thief can present the security and receive value in return. Notes and coins typically have these characteristics. Likewise, 'physical' ownership of a Bitcoin presumably resides with whoever holds the digital keys to the Bitcoin wallet in which it is held (of which there were apparently around 8–9 million as at end September 2016). To be more precise, it presumably resides with whoever seeks to withdraw units from the wallet first, if more than one person has possession of the digital keys to the same wallet.

The opposite of a bearer security is a *registered security*. Here, legal ownership in theory resides within a register (i.e. a ledger) controlled by someone (e.g. the company itself, if the 'security' is a share in that company, the mutual fund manager if the 'security' is a unit in a mutual fund, etc.). Holders of such securities are placing their trust in the organisation coordinating this register (and by implication the legal framework that maintains and constrains the actions of this organisation but at times can also override previously recorded entries in this ledger). Over time, securities have tended to shift from bearer to registered form (or even if legally have remained in bearer form have been dematerialised into registers held by custodians who then hold the actual physical securities in vaults), if only because investors no longer then need to worry (so much) about losing the bearer bonds or having them stolen. A major expansion in the use of blockchain technology would appear to require this longstanding societal trend to go into reverse.

Conversely, some financial instruments do currently exist in bearer form, e.g. cash in the form of notes and coins. Moreover, maybe many of the benefits capable of being captured using blockchain technology do not need *every* participant to keep separate copies of the distributed ledger but just to have the scope to access a trusted copy if they so wish. In recent years, many IT innovations have involved the implementation of new ideas by firms and individuals with only relatively hazy notions of how the new idea might eventually get used or commercialised. Take-up of most of these innovations is low or nil, but a few are extremely successful. If you happen to bet right then the rewards can be great, as entrepreneurs in the FinTech community, see Box 2.3 (including ones betting on blockchain) are hoping will be the case for the innovations they are backing.

7.4.6 Open Source and 'Community' Based Software

One feature that blockchain shares with many other IT innovations is the emphasis it places on open source (or 'community') software, i.e. software that is developed and maintained in whole or in part by voluntary contribution of effort from a community of individuals and firms. Murphy (2017) describes how at the time of writing DTCC was planning to transfer one of its trade databases onto a blockchain distributed ledger. Her article also notes that the plan is to submit the distributed ledger infrastructure and smart contract applications involved to an open source platform when the distributed ledger goes live.

Many elements of the worldwide IT infrastructure now depend on open source software. The underlying code associated with the software is then usually freely accessible by essentially anyone (with sufficient technical expertise), and usually with essentially anyone (with this level of expertise) in a (largely) self-selected community of interested parties (some individuals but often some corporates) able to propose enhancements. The software itself forms the information stored in a distributed updatable ledger, so to speak.

Related to (but not identical to) open source software are open standards, which define how software (whether open source or proprietary) should operate. Documents designed to be viewed in browsers over the internet adhere to such standards (via the appropriate standards that define HTML and JavaScript). If enough important players adopt a selected standard or give it their seal of approval then network effects create a strong tailwind in favour of the standard becoming essentially universal within some specific playing field. Standards are of course not limited to software. Legal frameworks often involve features chosen largely arbitrarily (such as whether in a specific country cars should drive on the left or the right side of a road) because universal agreement on a single choice is of benefit to essentially everyone. The most potent standards are the ones that confer the great practical benefit for the least effort amongst the community involved. Recent experience of some software companies has shown how extraordinary value can accrue to firms that successfully define and implement standardised ways of doing things if the firm can also somehow capture some of the value accruing from the adoption of the standard.

From the perspective of financial stability, several points are worth noting:

(a) The usual thesis within the (open-source) software community is that open source software is likely to be more robust than proprietary software, because it is subject to greater external checks and/or is easier to correct as time progresses. Conversely, proprietary companies may argue that proprietary software is likely to be more robust, e.g. because in practice it may be subject to more rigorous checking and greater incentives for being right when first launched. Ultimately a leap of faith is required either way. Whatever the pros and cons of each approach, any software potentially contains flaws and hence vulnerabilities.

(b) Most software of any size is now a collaborative venture, perhaps just within a single firm, but often involving broader networks of collaborators. If the software is important enough (e.g. the internet itself) then this creates the potential for systemic risk even at the scale we have been concentrating on within the remainder of this book. Open source software has the added challenge that practical understanding of the software involved may reside mainly in the heads of a few key individuals. For whatever reason, they may lose enthusiasm for continuing their involvement. Of course, key-man dependency can also exist within individual firms. However, financial regulators probably have more ways of incentivising individual firms (via regulation) to address these risks than is the case with otherwise hard to pin down volunteer communities of software developers.

(c) Relevant software standards can sometimes create environments that foster certain types of flaws. The predecessor of the internet was originally designed to be robust against large parts of the network being destroyed (by nuclear war). It appears to be very robust against such attacks. However, at the time the relevant underlying standards were being developed, less focus was paid to some other aspects of network robustness, including verification protocols that facilitated establishing the true sender of any given message. A consequence is that the internet may be more exposed than some other network types to the possibility of message spoofing and related forms of identity hijacking.

7.5 Interpreting the Concept of 'Fairness'

We've just discussed how Blockchain highlights (for the digital age) the implicit tension that exists between the desire for monetary (and other financial) transactions to incorporate certainty of outcomes whilst also ultimately respecting notions of fairness and propriety that society considers 'ought' to apply to any such transactions.

This means that any broader view of financial stability needs to understand how these notions of fairness and propriety might evolve.

Identifying what we mean by 'fairness' is tricky. For example, the notion is not exactly the same as 'equality', as past debates on unisex annuity rates show. The EU Gender Directive now bars EU insurers from setting annuity rates that differ between men and women. The overall effect has been to increase the price of an annuity for men and to reduce it for women, because women on average live longer than men. The EU Gender Directive explicitly provides 'equality' between men and woman in this respect, but whether this is 'fair' is more debatable. It is 'fair' in one respect, i.e. here achieving equality between sexes. However, it is 'unfair' in another respect. It deliberately prohibits the use of a risk factor (i.e. gender), even though this risk factor is generally considered to have a scientific basis as a means of differentiating between risks. This prohibition results in financial detriment for some members of society relative to what would otherwise have prevailed.

Conceptually the same sorts of issues arise in lots of other ways relevant to the financial services industry. Sometimes they are strongly linked to privacy issues. If I am exposed to some health condition and this information becomes freely available to health insurers then might this stop me being able to get insurance cover? What information should a bank be allowed to collect when deciding on whether to make a loan to me?

Teasing out what constitutes 'fairness' is particularly important for the financial services industry as there are often regulatory requirements to adopt behaviours that involve *treating customers fairly* (TCF). When providing a financial product or service to a customer it is clearly possible to adhere to

TCF whilst still making a reasonable profit. However, the more 'excessive' the profit can be construed to be, the more debatable the product or service becomes in terms of TCF. At what point is the boundary reached? And how might the definition of 'reasonable' or 'fair' change through time?

Balancing the interests of different parties is arguably particularly relevant to disclosure of information. The general regulatory view is that there is an information asymmetry between the firm and its customers, between the firm and third parties and between the firm and its regulators which should be rectified by requiring the firm to make available information it might not otherwise have provided. For example, firms need to provide extensive Pillar 3 disclosures, both quantitative and qualitative.

But provision of too much information might offer competitive advantage to others. Would this be in line with 'fairness' (to the firm itself)? For example, originally it was proposed under Solvency II that insurers should from time to time make publicly available line by line information on their individual asset holdings, including market values. But suppose the holding is not very liquid. At what level of illiquidity might such market value data become a potential millstone if the firm wanted to sell the asset? Some of the challenges that LTCM faced when it ran into difficulties (see Box 4.9) were due to detailed knowledge of its (illiquid) positions becoming too widely known.

Provision of information can also be costly, and the industry is always quick to argue that such costs are ultimately borne by customers. What is the correct balance between these competing arguments? How might it change through time? How should proportionality be (fairly) interpreted in such a context?

Added focus on 'fairness' also tends to increase focus on 'fair' values, also called 'market consistent' values in the insurance world. As noted in Section 3.4.2, these emphasise information extracted from market prices. Whilst some might claim that referring to such values as 'fair' can be misleading, the terminology does, as Kemp (2009) notes, still encapsulate an important truth. If the aim of a valuation is to provide equity between different parties (which in many cases it is, implicitly or explicitly) then those carrying out the valuation ignore such values at their peril. For example, in court sanctioned work on insurance company restructurings there may be an explicit need for the outcome to be equitable between different policyholder interests. Suppose we were to use demonstrably off-market values in such work. Then one or other party interested in the valuation might object that they were being short-changed relative to the other party. They might argue that we had inappropriately favoured the other party by ascribing a subjective (and by implication potentially 'wrong') value to the asset or liability.

Even clearer is the situation where we are trying to value units in a collective investment scheme with many different investors. It would be very uncommon to use other than market values (if they are available) or close proxies to identify transaction prices at which one investor can sell to another. Any other approach would be viewed as favouring one client over another.

Elsewhere in the financial arena we have also seen a greater desire for prescription in advance on how 'fairness' should be interpreted. For example, insurers in the UK offering with-profits (i.e. participating) policies are now required to set out in advance how they expect to run the funds backing these contracts. Those parts of the financial sector most involved in (long term) savings arrangements are probably the most likely to be caught up in this specific trend. Many of the conduct of business rules applied to asset managers can be viewed from this perspective. Banks have also come under greater pressure to be seen to treat different parts of society 'fairly' in relation to their lending activities.

7.6 Key Takeaways

This chapter has explored a range of topics linked to network effects applicable both to the financial system in isolation and to the broader economy and society of which it is a part. Key points noted include:

(a) Many macroprudential bodies interpret their financial stability mandates broadly and consider cyber risk and related aspects of how a financial system operates in practice as within their remit. This implies that the regulatory focus given to cyber risk and IT network resilience is likely to grow.
(b) At the same time, as was noted in the previous Chapter, policymakers also want firms to provide more data and in a more transparent way, adding to IT challenges that firms face.
(c) Some believe that technological advances can help. At the time of writing, some commentators are hoping that blockchain technology will assist with some of the internal business challenges some firms face. However, in some other areas, regulatory developments appear if anything to be demanding progress more quickly than is easy for IT systems to keep up with.

(d) An important issue firms face in the IT space is the balance to draw between privacy and the economies of scale that might come from sharing IT resources and data (e.g. via increased use of cloud computing). More generally, financial firms need to identify how to ride current technological developments successfully for the benefit of customers and shareholders. These mirror similar dilemmas faced by firms in other industries.

(e) More broadly, organisations within or touching the financial system need to contend with the continuing evolution of the concept of 'fairness' across society. Most aspects of finance ultimately depend on societal norms that extend beyond the financial system. The correct meaning to ascribe to financial promises is ultimately decided not by financiers themselves but by broader society. Firms that ignore this insight do so at their peril, as illustrated by the sizes of some fines and compensation payments financial sector firms have suffered by not respecting these norms.

References

Bank of England (2014). An Introduction to CBEST. *Bank of England.*

Cocke, M., Hannibal, C., Hursey, C., Jakria, P., MacIntyre, I. and Modiset, M. (2014). Heavy Models, Light Models and Proxy Models – A working paper. *Institute and Faculty of Actuaries Sessional Meeting*, 24 February.

Financial Times (2014). Banks Want to Keep Your Digital ID in Their Vaults. *Financial Times*, 2 September.

GCHQ (2012). *10 Steps to Cyber Security*. CESG, the Information Security Arm of GCHQ, Cheltenham, UK.

Gracie, A. (2014). Managing Cyber Risk – The Global Banking Perspective. *Bank of England.*

Homescu, C. (2011). Adjoints and Automatic (Algorithmic) Differentiation in Computational Finance. *SSRN.*

Kemp, M.H.D. (2005). Risk Management in a Fair Valuation World. *British Actuarial Journal*, 11, No 4, pp. 595–712.

Kemp, M.H.D. (2009). *Market Consistency: Model Calibration in Imperfect Markets*. John Wiley & Sons, Chichester, UK.

Knight, F. H. (1921). *Risk, Uncertainty and Profit*. Cosimo Classics, New York, USA.

Markov, I.L. (2014). Limits on Fundamental Limits to Computation. *Nature* 512, pp. 147–154.

Marshall, A. (1924). *Money, Credit and Commerce*. Macmillan, London, UK.
Mayer-Schönberger, V. and Cukier, K. (2013). *Big Data*. John Murray (Publishers), London, UK.
Murphy, H. (2017). Database Move Gives Blockchain Its First Big Test Case. *Financial Times*, 10 January.
Nematrian (2014). Foundation ERM Session 5: Credit Risk. *Nematrian.*
Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (2007). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, UK.
Rudolph, M. (2012). Preparing to Deal with Digital's Looming Dark Side, *Risk Management Magazine*, October.

8

Responding to Systemic Risk

In this final Chapter, we summarise the main trends that appear to be driving systemic risk developments and we explore how individuals and organisations in the financial sector might best respond to these trends. In Section 8.1 we categorise broad regulatory trends with some systemic risk relevance into three core strands differentiated by the likely timescale over which they will play out. We then explore what firms can do to respond to these trends in the following areas:

(a) *Managing interaction with regulators and supervisors*. In Section 8.2 we concentrate on what firms can do to influence the likely direction of travel of financial regulation and how they might best interact with regulators and supervisors in response to these trends.
(b) *Managing data*. One clear outcome of systemic risk deliberations is an increased focus on data to understand the bigger picture. This is imposing additional data requirements on firms. In Section 8.3 we illustrate some of these tensions and comment on how firms can best respond to them.
(c) *Risk measurement and risk management*. Many readers of this book are likely be involved in risk measurement and management. In Section 8.4 and Section 8.5 we explore some likely impacts of trends in systemic risk in these areas.
(d) *Internal risk management team structures*. Some organisations (particularly regulators but also some FMIs) have designated specific individuals or teams whose focus is on systemic risk, see Section 8.6.

M.H.D. Kemp, *Systemic Risk*, DOI 10.1057/978-1-137-56587-7_8

(e) *Market structure*. Systemic risk events, once they materialise, can be expected to influence the way markets operate, for all investors and not just ones directly at risk from systemic domino effects or tsunamis. In Section 8.7 we discuss changes in market microstructure that such events can bring.

8.1 Broad Regulatory Trends

We may group current trends in financial service regulation into three broad groups:

(a) *Increased focus on systemic risk following the recent financial crisis*. This is already overturning some existing business models and creating opportunities for others. It seems likely that the outcomes of debates about what types of entity potentially create, amplify or transmit systemic risk will have major implications for affected parts of the financial community for at least the next c. 3–5 years. For example, large global insurers are facing the introduction of a global Insurance Capital Standard (ICS) specifically as an outcome of this focus, which may percolate down to the rest of the industry in due course, if Basel Capital Accords are any guide. In the longer term, memories of the 2007–09 Credit Crisis will no doubt fade and this strand of regulatory development is likely to abate, unless a new systemic crisis hits in the meantime.

(b) *Increased scepticism amongst regulators and governments that different parts of the financial services industry are inherently different*. This can be expected to lead to increased harmonisation and cross-fertilisation of risk and regulatory techniques and practices across the industry. These trends are amplified by those in (a) but probably will continue even after those in (a) have died down.

(c) *Continuing societal change driven by IT and other technological developments and by how societies interpret 'fairness'*. Even in the absence of financial crises we can expect regulatory frameworks to change as societies change and technology develops. Information and business asymmetries are inherent in many financial services activities. So is the desire to modify regulatory frameworks to try to limit the asymmetries that are most in the public eye at any given point in time. Some recent changes to regulatory frameworks have sought to prohibit or limit incentive structures within the industry that regulators have deemed inappropriate. In due course, new types of inappropriate incentive structures will no doubt materialise.

One consequence of these broad trends that is likely to be of specific interest to readers of this book is its implications for the sorts of skill sets that might be most appealing to employers in this industry. The trend mentioned in (b) suggests that there will be a blurring of applicable staff skillsets across different parts of the financial sector. The permeation of ideas and approaches across the industry has made it easier for individuals to move between different types of firm. This arguably facilitates other types of harmonisation across the financial services industry.

The increasing tendency to view firms (and other entities) across the financial services sectors as forming a single overarching 'industry' has other self-reinforcing aspects. For example:

(1) Firms' business models (and/or owners) may change through time. This increases demand for staff who understand business models adopted in other parts of the industry.
(2) The greater the extent to which firms in different sectors are perceived to be inherently similar, the greater the incentive and rationale for adopting similar regulatory frameworks for them. Also greater is the incentive for adopting 'unitary' regulators or supervisors whose remit spans different industry sectors. Adoption of unitary regulation in turn promotes similar behaviour patterns across the affected sectors (hopefully desirable behaviours!). It also encourages journalists, politicians and other commentators to view such entities as similar.
(3) Academics and other thought leaders can be increasingly expected to seek common strands between sectors. The entire financial services industry in some sense derives from the invention of money and the uses societies have made of this invention. It is therefore highly likely that when we seek such common strands there will be some to be found.
(4) It increases the tendency of disciplines such as risk management to disseminate techniques and ideas across the relevant sectors. For example, if an approach to market or credit risk is perceived to be useful in one sector then it is likely to be perceived to be useful in other sectors. Consulting and software firms supporting such activities have a natural incentive to market their services as widely as possible.

8.2 Managing the Interaction with Regulators and Supervisors

A common method for a business sector to respond to a regulatory trend it considers to be unhelpful is to lobby against it and /or commission research and other material highlighting its flaws. Financial firms have been doing this for many years and will no doubt be doing so for many years to come.

However, it should be recognised that such lobbying hasn't been particularly effective at the top level for the financial industry since the 2007–09 Credit Crisis. Many politicians still seem to view the banking industry with scepticism at best and pariah-like status at worst. The insurance industry has been unable to stop the regulatory community pushing ahead with plans for global capital standards and other responses to systemic risk concerns. The pensions industry may have been more adroit at sidestepping systemic risk trends, but is facing emerging challenges that may leave it susceptible to its own sorts of systemic risk events. The asset management industry has also arguably ended up relatively unaffected, but this is probably mainly down to the difficulty of identifying what actions might usefully be taken to address systemic risk concerns (given that the assets involved are not owned by the investment managers themselves but by their clients).

Ultimately, the financial services industry has a synergistic relationship with central banks and other financial authorities present within different jurisdictions. The ultimate usefulness and value of this industry derives in part from central edicts imposed by governments. It cannot therefore expect to avoid being caught up in regulation designed to address specific issues if these issues are perceived important enough by the state.

So probably at least one of the strategies for success is to accept that systemic risk concerns on the part of regulators and politicians are here to stay (at least for quite a while) and to get on with managing the implications.

Larger firms can expect this to involve more interaction with supervisors than smaller ones, but all are likely to benefit from trying to make the interaction as productive as possible, see Box 8.1.

Box 8.1: Treating the regulator /supervisor as a key stakeholder

An important aspect of effective risk management for any financial organisation is stakeholder management. Nowadays an important stakeholder for any such organisation is its regulator or supervisor. Set out below are some general pointers on what this specific stakeholder management exercise might involve, based in part on material from IAA (2009):

(a) Prudential supervision is nowadays accepted worldwide as an integral component of the regulation of financial institutions. So, if an organisation wants credibility with its regulator it needs to take prudential supervision seriously.
(b) Usually, prudential supervision starts from the premise that primary responsibility for financial soundness and prudent risk management within a supervised institution rests with the Board and senior management of the institution. In this context, the Board is generally considered to be the body that ultimately carries responsibility for the organisation (including the hiring and firing of its most senior management when necessary) whilst senior management is the group of individuals who day-by-day manage the firm in accordance with parameters set by the Board. As approaches to governance can vary by jurisdiction, it can sometimes be unclear exactly what body most aligns with the 'Board' in such a framework. So, clear demonstration of who has what roles and responsibilities (and that none are falling into cracks within the organisation's governance structure) is also important to gain credibility with the regulator.
(c) Prudential supervision typically includes elements relating to:

   a. Financial oversight
   b. Licensing of what the organisation can do
   c. Review of how it operates
   d. Procedures and processes for monitoring compliance with licence conditions etc. and ongoing operational requirements
   e. Where necessary, undertaking enforcement action either to force a non-compliant firm into compliance or to remove it from the industry.

   So, prudential supervision is multi-faceted. Organisations need to avoid giving the impression they are only focusing on a sub-set of the regulators' overall requirements.
(d) Supervisors typically adopt a risk-based approach to supervision (to maximise the cost-benefit trade-off from their activities). So, institutions perceived to face greater risks can expect to receive closer supervisory attention.
(e) Supervisors can only categorise firms by risk if they can form their own views of risks that the supervised institution presents (and of the effectiveness of its management of these risks). To do their job effectively, supervisors ultimately require openness and transparency from the firms they supervise and will take a dim view if these behaviours do not appear to be present.
(f) Supervisors are exposed to the full spectrum of worst to best practices. So, engaging effectively with supervisors may help firms to improve their own risk management.
(g) If a supervisor does not have an adequate level of comfort about the strategic and higher level aspects of a firm's risk management framework then it is likely to adopt a more intensive supervisory approach than would otherwise be the case. So, firms should aim for ongoing and transparent dialogue with supervisors about strategy and framework.
(h) There are lots of different ways in which larger firms interact with their supervisors. These include operational interactions (such as submitting

standardised, periodic returns and statistics and responding to routine queries relating to standard operations) as well as more one-off interactions such as consulting with supervisors in relation to strategic initiatives such as acquisitions or other corporate transactions. Most large financial firms/groups therefore develop accountability mechanisms and protocols to ensure the right people are engaging supervisors appropriately.

(i) Firms often engage with supervisors in relation to policy development. Supervisors often look for constructive feedback on their proposals and look to firms to test the robustness and proportionality of new proposals (although firms may use industry bodies to coordinate submissions on proposed new policy).

(j) Supervisory visits provide the supervisor with an opportunity to delve deeper into specific aspects of a firm's operations and risk management processes. Often these will run more smoothly if firms work with supervisors to coordinate site visits, e.g. agenda development, document submission and overall visit logistics, strengthening the relationship at an operational level.

(k) Requirements and recommendations arising from supervisory visits should be welcomed, and taken seriously. Supervisors may view firms that unreasonably challenge supervisory requests and requirements as having adverse cultural issues. This may lead to more intensive supervision.

(l) A key test of an effective supervisory relationship is how a firm handles the management and reporting of regulatory breaches, even though many or even most such breaches may be inadvertent human and/or process errors rather than blatant disregard of rules. The identification, management and reporting of breaches is an opportunity for process improvement. No one expects zero breaches. Ironically, an absence of breach reporting to supervisors for an extended period could be viewed as an indicator of ineffective risk management and/or culture.

A complicating factor when it comes to systemic risk is that countries usually give specific bodies responsibility for systemic risk and macroprudential policy. These bodies may not be the same as the firm's main supervisor (and even where they are the same, the individuals involved may differ). The flip side is that work done by systemic risk bodies tends to be higher-level and non-firm specific, so primary effort in managing the regulatory relationship presumably needs to be with individuals the firm interacts with, accepting that these individuals in turn are operating within frameworks set by others.

Particularly high up the agenda for some supervisors is their ability to respond in a timely and productive fashion if a firm gets into difficulties. This explains the focus on recovery and resolution planning that is in the vogue at present, at least among firms in sectors that have had material numbers of business failures in recent years. Put bluntly, every firm will come to an end one day, just as every individual will die. Supervisors don't need to have come across many cases which involved a lot of hassle to be extremely keen for firms to carry out some planning for such an eventuality.

What about unregulated firms, such as shadow banks? They don't have supervisors and hence a supervisory relationship to manage. But they do have other stakeholders, e.g. shareholders, bondholders, managing agents and institutional clients many of whom are themselves regulated entities. Moreover, one of the outcomes of greater concern over systemic risk is a greater focus on the regulatory perimeter and potential regulatory arbitrage between regulated and unregulated bodies. So, such firms are still likely to find it helpful to understand and position themselves in ways that look tolerably regulator friendly, even if they are not themselves formally regulated. This mitigates the risk that they might find themselves moved into the regulatory net at an unexpected and inconvenient time for themselves.

8.3 Data Management Activities

One issue that seems unlikely to go away any time soon is the issue of 'too big to fail'. Central bankers and regulators have been at pains to claim that TBTF has been tackled by responses to the 2007–09 Credit Crisis and it does seem likely that some significant progress has been made. However, to believe that the issue is completely eradicated seems optimistic. Indeed, TBTF was thought to have been tamed prior to Continental Illinois's failure as well as prior to some other failures since then, suggesting that a full solution is tricky or impossible to achieve. Almost certainly firms in the financial sector will continue to fail from time to time. If, for whatever reason, a failure is expected to have sufficiently adverse externalities then pressure will be placed on governments to bail out the firms concerned irrespective of the nature of any arrangements previously established that were supposed to ensure otherwise.

However, what does seem to be different this time is a desire to impose unprecedented levels of transparency and market disclosure on regulated firms (sometimes just to supervisors). Perhaps subconsciously, this is recognising that if the network of relationships between market participants is sufficiently clear then maybe governments will feel less need to step in when things go wrong, to contain uncertainties. There is an explosion of data provision happening elsewhere in society, so why shouldn't the financial services industry be subject to the same underlying drivers?

Examples of this are central clearing requirements, including those being introduced in the EU by the European Market Infrastructure Regulation (EMIR), see Box 8.2. We can also see this trend at work in the increasing data provision requirements being imposed on regulated firms, such as banks and insurers, see e.g. Box 8.3.

The message is clear. Firms need to beef up their ability to manipulate, process and make available to others data about their assets and liabilities and how these interact with other market participants. This should provide business opportunities for the entrepreneurial, since the same underlying drive for more data and better data manipulation capabilities is shared across the industry.

Box 8.2: Mandatory Central Clearing and EMIR

In 2009 the G20 pledged to undertake reforms aimed at increasing transparency and reducing counterparty risk in the OTC derivatives market. This is most commonly referred to worldwide as the introduction of mandatory central clearing, see Section 6.7. EMIR implements most of these pledges in the EU and covers OTC derivatives, central clearing (including CCPs) and trade repositories (TRs), see FCA (2014a). It applies both to financial counterparties (FCs), including banks, insurers, investment firms and fund managers and to non-financial counterparties (NFCs). NFCs cover any counterparty that is not classified as a financial counterparty, including entities not involved in financial services. Its main requirements are:

(a) *Reporting*: All counterparties with outstanding derivative contracts will need to report details of those contracts (and any new contracts they enter into) to an authorised *trade repository* (TR).
(b) *Clearing*: The European Securities and Markets Authority (ESMA) can impose mandatory clearing obligations for OTC derivative contracts of a specific type if an EMIR-authorised CCP exists for that type of contract.
(c) *Specific operational risk management requirements for non-cleared transactions*: All counterparties are required to comply with certain operational requirements (including timely confirmation, valuation, reconciliation, trade compression and dispute resolution).
(d) *Collateral*: Contracts not cleared through a CCP will also be subject to bilateral collateral requirements for FCs.

NFCs will only be subject to clearing and bilateral collateral requirements if their OTC derivatives positions are large enough and are not directly reducing commercial risks or related to treasury financing activity.

Box 8.3: Additional Data: COREP and FINREP (banks) and QRTs (insurers)

Large parts of the CRD and CRR relate to the introduction of Basel III capital requirements for EU banks. The CRD also covers capital requirements for certain MiFID investment firms not explicitly covered by Basel III, see e.g. FCA (2014b). It also introduces a ‘bonus cap’ and standardised EU regulatory reporting requirements.

The reporting frameworks introduced by the CRD are referred to as COREP and FINREP and specify the information affected firms must report to supervisors in areas such as own funds, large exposures and financial information. In some member states the sole reporting format for this data will involve XBRL. XBRL is also the typical reporting format that will apply to insurers under Solvency II.

Quantitative data provision that EU insurers need to provide to their supervisors has expanded considerably with the introduction of Solvency II. The forms concerned are called Quantitative Reporting Templates (QRTs). Particularly challenging for some insurers is a QRT that contains a look-through to the underlying holdings within collective investment schemes in which the insurer invests. Such look-throughs only generally need to be supplied to supervisors once a year, but the rationale for expecting firms to be able to supply such data is that firms ought to know what they have invested in, so ought to be getting this information regularly. Of course, it is one thing to receive such information in some shape or form from an individual investment manager, it is potentially something else to be able to manipulate this information effectively and to consolidate it with other similar information potentially coming from lots of different investment managers.

Add up these additional data requirements across the entire financial services industry and you end up with a significant data management challenge. Looked at from a different angle, you also end up with some associated business opportunities, if your field of expertise happens to include such data management or you are a FinTech firm.

8.4 Risk Modelling

Data provision and manipulation is not the only type of IT being affected by systemic risk trends. Another major consumer of IT budgets in the financial sector is risk modelling, both modelling used in day to day business and investment management and modelling used to derive firms' capital requirements (either formal Pillar 1 regulatory capital requirements or firms' own views of what underlying amounts of capital they need to face the risks to which they are exposed).

Systemic risk considerations are influencing not just high level principles but also basic risk modelling approaches. Its influence here includes:

(a) Likely greater emphasis on Expected Shortfall (ES) relative to Value-at-Risk (VaR);
(b) Greater emphasis on reverse stress testing
(c) Greater emphasis on mathematically simpler approaches to analysing risk such as stress testing relative to more complex statistical VaR-like or ES-like modelling; and

Despite some industry pushback, BCBS in its FRTB, see BCBS (2012) has continued to promote the adoption of risk measures that it sees as better able to capture ‘tail risk’ than VaR. Whilst BCBS recognises that moving to ES could entail certain operational challenges, it nonetheless believes that these are outweighed by the benefits of replacing VaR with a measure like ES that better captures tail risk.

The BCBS proposal can be directly linked to worries about systemic risk, see Box 5.1. We might therefore expect the use of ES (or some variant) eventually to percolate into capital adequacy computations for other parts of the financial services industry. This is likely to happen faster with sectors that regulators have decided can pose or transmit systemic risks or which contain systemically important financial institutions (SIFIs).

The association between systemic risk and tail risk also implicitly favours placing greater attention on stress testing. In the sorts of extreme circumstances that arise when systemic risk strikes, robust statistical risk modelling becomes more challenging. To mitigate against model risk, regulators have been increasingly promoting use of (less statistical) risk modelling approaches such as stress testing and placing not quite so much reliance on potentially less robust statistically based measures such as VaR, TVaR and ES.

Alongside the increased focus on stress testing there is an increased focus on reverse stress testing. As noted in Box 5.3, it was proposed in August 2008, i.e. shortly before Lehmans defaulted. It has since become widely required across most of the financial services industry.

There are several other examples of increased scepticism being placed on reliability of statistical risk modelling techniques in the light of systemic risk, including:

(a) Introduction as part of Basel III of the liquidity coverage ratio (LCR) and the net stable funding ratio (NSFR). These are measures designed to mitigate liquidity risk, a type of risk that is notoriously difficult to quantify accurately. Broadly speaking, the LCR requires banks to hold a minimum fraction of their assets in types deemed to be ‘liquid’ as defined by regulator mandated requirements. The aim is to mitigate liquidity risk in the event that there are unexpectedly high amounts of withdrawals by depositors. Broadly speaking, the NSFR requires banks to source a minimum level of their funding requirements from liabilities that are deemed to be ‘stable’. The aim is to mitigate liquidity risk due to flighty types of depositors and other providers of funds. Both are in effect relatively simple regulator mandated formulae that are on top of any other capital or liquidity risk assessments that banks are required to carry out. Banks are also now subject to ‘internal liquidity adequacy assessment process’, i.e. ILAAP, requirements, which are akin to internal capital adequacy assessment process, i.e. ICAAP, requirements but focusing on the firm's liquidity rather than its capital needs.

(b) Planned abandoning by BCBS of the ability of banks to use the 'advanced measurement approach' (AMA) for determining the amount of regulatory capital they should hold for operational risk modelling.

(c) Doubts about the applicability of internal models in the planned new global insurance capital standard (ICS). Initial versions of the ICS are focusing solely on standard formula styles of determining capital requirements, with a decision on whether to permit use of internal models being deferred to later in the process of rolling out the ICS. This apparently reflects doubts expressed by some regulators about the practicality of developing a framework for agreeing firms' internal models that can achieve sufficiently robust comparability across firms (and can sufficiently eliminate scope and incentives for firms to 'game' the resulting framework).

8.5 Risk Management and Governance

Another response to potential systemic risks in advance of them materialising is to ensure that as far as possible processes are in place that can react appropriately to such events as and when they do arise. Regulators generally believe that firms that had better risk management frameworks tended to navigate the 2007–09 Credit Crisis more effectively than firms with less developed ones. This has led them to place a strong emphasis on firms to improve their risk management disciplines, see e.g. HM Treasury (2009). A substantial number of firms have carried out governance reviews or other activities aimed at ensuring that effective risk management and governance disciplines are in place within the firm. Numbers of staff involved in risk management and compliance functions have risen.

An alternative way of interpreting regulatory views on risk management (and risk modelling) is to view it as an example of a more general societal trend favouring increased emphasis on risk management. There are many other examples of this including the greater focus given to risk management in Solvency II, increased emphasis on risk management and systems of governance for EU pension funds in the proposed IORP II Directive and, more generally, revisions to corporate governance codes in e.g. the UK to emphasise risk management, see FRC (2014). If politicians view enhanced risk management as intrinsically a 'good thing' then it is hard to see them wanting to ignore systemic risk in this picture.

8.6 Systemic Risk Officers

Another response to systemic risk that some organisations appear to be adopting is to identify individuals or teams who are specifically responsible for that organisation's analysis of and response to systemic risk. Many regulatory bodies now have financial stability departments. This reflects the importance now being accorded to systemic risk and financial stability by politicians and others ultimately responsible for the mandates given to these regulators.

Some other bodies are also specifically carving out roles that focus on systemic risk. For example, the foreword of DTCC (2015) is co-authored by its DTCC's Group Chief Risk Officer and its Chief Systemic Risk Officer and explains:

> risk managers can no longer view financial firms as stand-alone entities because, in reality, they are now a diverse set of interconnected components that distribute risk and are exposed to it, oftentimes in ways that are not transparent or expected. Furthermore, the openness and complexity of the financial ecosystem and the likelihood that breakdowns will occur mean that firms must do more than monitor and mitigate these risks – they also need to focus on building resiliency so they can detect potential systemic shocks before they strike or recover from them as quickly as possible … We have incorporated this thinking into the organization's risk management framework, beginning with creating a Systemic Risk Office and appointing a Chief Systemic Risk Officer to conduct industry outreach, map emerging risk trends and engage with regulators and clients to identify and report on internal and external sources of systemic risk.

To date, DTCC seems to be relatively unusual in specific identification of a chief systemic risk officer. It seems to be more common for such responsibilities to fall within the remit of broader risk teams.

8.7 Responding to Changes in Market Structure

It is difficult to plan for how to deal with a systemic risk event, as we don't know in detail what it might look like. However, there do appear to be some relatively common strands, so it is worth rehearsing how one might best respond if these strands reappear for any given systemic risk event.

Systemic risk events are often associated with a drying up of market liquidity. One short-term consequence for market participants involved in buying and selling financial assets of such episodes is that it increases the cost of market transactions and hence the focus they should give to minimising transaction costs when they are carrying out such transactions. A significant fraction of Fouque and Langsam's '*Handbook of Systemic Risk*', i.e. Fouque and Langsam (2013), explores this interaction, highlighting its current apparent importance to the academic community. The way in which markets respond to and process individual transactions is usually called 'market microstructure', to contrast it with e.g. the market 'macrostructure' arising from firms in aggregate issuing different sorts of instruments and the mix of instruments traded changing as economic and business conditions change.

A rich literature has developed in recent years on optimal trade execution, i.e. how best to execute such transactions in the presence of transaction costs. The underlying theory is that any given market participant will on average suffer some price impact when executing such trades (see Box 8.4 for a stylised description of why). The more challenging is the liquidity position the greater this price impact can be expected to become. A trade execution strategy is a formalised way of responding to such an insight. Conceptually it applies to any sort of market participant (even a private investor, whose 'strategy' may in practice merely involve asking someone else to decide how to execute his or her desired trade). Formalisation of such a strategy becomes particularly important for very active traders, especially high frequency algorithmic traders whose decisions on when to buy and sell and in what quantities may be largely decided by computer algorithms.

Identifying a precise measure of market impact for such purposes is non-trivial. Many recent contributions to this literature build on the generic model outlined in Almgren and Chriss (2001) who formulate a measure of 'instantaneous price impact' combining a measure of market depth and resilience into one stylised quantity. By market depth we mean size of trades possible to execute and by market resilience we mean how big an apparent impact on market price /sentiment a trade of a given size might have. Authors such as Fruth et al. (2011) explore how trade execution strategies might be affected by (deterministic) changes in liquidity levels.

If we control a sufficiently large pool of assets then our behaviours can influence the prices of assets. Market participants in such positions not only need to avoid unfairly manipulating market prices, especially not in a way that is detrimental to their clients. They also need to demonstrate that they have not done so.

This begs the question of what we mean by (unfair) 'market manipulation'. Usually the precise way to interpret 'manipulation' is strongly influenced by regulation, e.g. the MiFiD Directives in the EU. Behind such regulation is usually some notion of fairness. For example, modern financial regulation views it as unfair for asset managers to front run customer orders, e.g. by buying immediately in advance of a large customer buy order that the manager could be expected to think would lead to an increase the price of the asset. Such a strategy will typically exacerbate the market impact suffered by the customer whilst enriching the asset manager.

Conversely, it is not normally considered unfair for an asset manager to take a view on an asset going up or down in value (provided the view does not make use of inside or otherwise privileged information). This is the case even though someone (i.e. the other side of the transaction) presumably will be relatively speaking worse off if the asset manager makes a correct call. At least it is not normally considered unfair in modern capitalist economies that expect diversity of views across market participants and believe that the market mechanism is an effective longer-term way of allocating economic resources.

Systemic risk considerations interact with these issues in the following ways:

(a) In extreme circumstances, market microstructure can become dislocated or can become viewed by policymakers as part of the problem rather than part of the solution. For example, Gale and Yorulmazer (2011) refer to liquidity hoarding and explain how some types of market structure can encourage banks to hoard liquidity. They also conclude that introducing specific policies such as liquidity and lending requirements can reduce such hoarding. Part of the reason for such behaviour is that market participants may hold assets both from a precautionary and from a speculative motive. In certain circumstances, markets can enter an 'inefficient' state in which some entities may hoard liquidity even whilst others have unmet liquidity needs, i.e. it can become rational (at least from the perspective of individual market participants) to adopt strategies that in aggregate result in markets failing to function as expected. This is the market microstructure equivalent of propagation of systemic risk. The market may not have become unstable as such (so it can't specifically be said to have suffered a bout of financial instability), but it may have got locked into an 'unhelpful' state because of the original stress.

(b) In extreme circumstances, the view that capital markets are effective ways of allocating capital can also be challenged, especially in relation to the state's own finances. We saw this in the 2007–09 Credit Crisis and in the subsequent Eurozone sovereign debt crisis when governments took a dim view of investors speculating against the health of the governments themselves (and in some cases banned activities such as short-selling of government credit risk using CDS). More generally, if the public becomes sufficiently disenchanted with how well the financial system is working in its favour then the political process can result in all sorts of restrictions being placed on the financial sector. In even more extreme circumstances trust in social order more generally may break down and social revolution can result or those in power may 'gamble for resurrection', e.g. by initiating a war to divert public attention elsewhere, which might then go catastrophically wrong.

Debates about market microstructure can also trigger discussions about possible price manipulation in dark pools or equivalent elements of financial markets or other conduct related issues. A dark pool is a private forum for trading securities, carried on outside the remit of public exchanges (usually so that extra confidentiality applies to the trade). Some of the rules introduced by MiFID (see Box 4.13) aim to ensure that suitable regulatory oversight is being applied to such trading venues. There is an echo here with debates about shadow banking, see Section 4.6, since shadow banking is also an example of an activity that may be taking place under less regulatory scrutiny than would be the case in other contexts.

Systemic risk considerations can also in time influence what economic exposures are traded in financial markets. Regulators have for some time been trying to encourage the interest rate swap market to move away from London interbank offered rate (LIBOR) floating rates to overnight indexed swap (OIS) floating rates. This is in part because the OIS rates are viewed as a purer measure of risk-free rates and hence less entangled with the banking system than LIBOR rates. Most such ideas involve considerable structural challenges, which in this case include the very large back book of swaps already referenced to 3 /6 month LIBOR rates, although there is some evidence that a shift has taken place towards approaches desired by regulators, see e.g. Khwaja (2016).

The increased focus regulators and governments are placing on systemic risk is not always leading to results that governments may themselves be comfortable with. It can turn the spotlight onto the risks firms face because sovereigns can default. An example of the new normal in this regard might be the IAA's paper on stress testing and scenario analysis, i.e. IAA (2013). It includes three 'case studies', one of which, tellingly, covers sovereign default. As Reinhart and Rogoff (2009) note:

> Throughout history, rich and poor countries alike have been lending, borrowing, crashing – and recovering – their way through an extraordinary range of financial crises.

Box 8.4: Transaction costs and market impact

When securities are bought or sold in a market the parties involved in the transaction usually incur some costs. For example, markets may levy transaction fees, there may be some taxes payable and market participants may employ (for a fee) brokers who act as agents to identify parties willing to be the other side of the transaction.

In addition, transactions also typically have *market impact*. For example, suppose one market participant sells a block of shares at the then prevailing market price. Suppose hypothetically another market participant comes along instantaneously afterwards seeking to sell the same share. Then as part of the process of price discovery that is going on, the first seller will have caused market prices to drop by the time the second seller tries to sell.

Closely allied to market impact is the so-called *winner's curse*. A market-maker may offer to buy securities from one market participant in an auction. The winner is left with the holding, which all other things being equal may need to be marked down in value (reflecting the lower price at which the market maker will now be able to sell the securities to another market participant). This means that market-makers need to include some form of bid-offer spread (also called bid-ask spread) in how they price market transactions, the bid-offer spread potentially being larger for particularly large deals. In a sense, bid-offer spreads delay the immediate recognition of market impact until the market-maker unwinds its position via subsequent deals. If the market-maker has enough offsetting positions to find both a willing buyer and a willing seller at essentially the same time then it will pocket the bid-offer spread. Conversely, if liquidity declines then it can be left holding unwanted positions, i.e. part of the bid-offer spread is recompense for it carrying the risk that market impact is larger than expected.

Some exchanges have mechanisms which mean that purchase and sell orders are as far as possible automatically matched, with the price quoted by the exchange being the price at which deals are matched. They may therefore appear to have no bid-offer spread. However, they are still subject to market impact; it is just that this does not show up via bid-offer spreads but in how prices move after any trade is executed.

In recent years, market makers have become capital constrained and are less willing to put up risk capital for large trades. Increasingly, the provision of liquidity within markets has been taken up by others, including high frequency trading funds and others using algorithms that try to work out the best way to execute larger orders. From time to time these algorithms can become self-reinforcing, leading to so-called flash crashes where markets suddenly fall (or the opposite, suddenly rise) by unexpectedly large amounts over a very short window, with the effect usually largely reversed shortly afterwards. Some commentators view market-wide disturbances such as the October 1987 stockmarket crash, see Box 4.3, as akin to these types of flash crashes, just happening a little slower given the slower technology then prevalent.

8.8 Key Takeaways

In this Chapter, the last in the book, we consider ways that individuals and organisations can best respond to systemic risk and the trends that are giving it added importance at the current time. Key points noted include:

(a) Desired skillsets across different parts of the financial system are likely to become more blurred, at least across those parts of the financial system typically deemed to be exposed to systemic risk. Increased emphasis on systemic risk increases the degree to which firms and others seek out commonalities across organisations, and hence want staff able to interpret these commonalities. It remains important to engage effectively with supervisors.

(b) As was also highlighted in earlier Chapters, a key consequence for individuals working in the financial sector is the added impetus systemic risk gives to trends towards greater data provision and transparency. Successful firms in this space will need to be on top of these trends, which in practice will draw heavily on IT capabilities.

(c) Some of the changes underway in risk modelling required of firms, such as increased emphasis on use of expected shortfall rather than value-at-risk, can also be pinned on the extra focus now being given to systemic risk. Different risk measures have different appeals to different stakeholders, and increasingly it is regulators and supervisors who are calling these shots.

(d) Another strong focus of policymakers has been to enhance governance disciplines within regulated firms. This has increased the level of risk management capabilities that firms are expected to exhibit, as well as the focus needed to manage different stakeholders effectively. Some systemically important organisations have gone so far as to appoint systemic risk officers, mirroring some of the internal organisational changes introduced by policymakers to respond more effectively to their financial stability mandates.

(e) Changes being introduced to the financial system because of the greater focus on systemic risk also influence market microstructure. This has implications for market facing individuals and firms, who need to adapt to handle these changes and respond to new types of conduct risk that come with these changes.

(f) The elephant in the room, so to speak, is the risk of broader societal disfunction, e.g. sovereign default. There is a tendency to put this sort of risk into the too difficult to handle category (or too politically sensitive to handle). Unfortunately, if we catalogue a broad range of historic systemic risk events rather than just focusing on the 2007–09 Credit Crisis then we find that the most serious ones often contribute to or feed off broader societal challenges. We might hope that we can limit our study of systemic risk just to things that financial systems have some clear input into and control over. However, this is not the way the financial system interacts with the rest of the economy and society more broadly.

My hope is that by articulating how systemic risk may be managed, measured and analysed, this book can help practitioners play a role in helping the financial system foster a robust economy and society. These ultimately provide the key ingredients for longer-term financial stability.

References

Almgren, R. and Chriss, N. (2001). Optimal Execution of Portfolio Transactions. *Journal of Risk*, 3, 5–39.

BCBS (2012). Consultative Document: Fundamental Review of the Trading Book. *Basel Committee on Banking Supervision.*

DTCC (2015). Understanding Interconnectedness Risks. *DTCC.*

FCA (2014a). One Minute Guide – EU Regulation on OTC derivatives (EMIR). *UK Financial Conduct Authority*, 12 September.

FCA (2014b). Capital Requirements Directive (CRD IV) FAQs. *UK Financial Conduct Authority*, viewed 18 September.

Fouque, J-P. and Langsam, J.A. (2013). *Handbook of Systemic Risk*. Cambridge University Press, Cambridge, UK.

FRC (2014). The UK Corporate Governance Code (September 2014). *Financial Reporting Council.*

Fruth, A., Schöneborn, T. and Urusov, M. (2011). Optimal Trade Execution and Price Manipulation in Order Books with Time-Varying Liquidity. *SSRN.*

Gale, D. and Yorulmazer, T. (2011). Liquidity Hoarding. *SSRN.*

HM Treasury (2009). A review of corporate governance in UK banks and other financial industry entities – Final Recommendations. *UK HM Treasury*, November.

IAA (2009). Note on Enterprise Risk Management for Capital and Solvency Purposes in the Insurance Industry. *International Actuarial Association.*

IAA (2013). Stress Testing and Scenario Analysis. *International Actuarial Association Insurance Regulation Committee.*

Khwaja, A. (2016). USD OIS Swap Volumes Surge. *Clarus Financial Technology.*

Reinhart, C.M. and Rogoff, K.S. (2009). *This Time is Different: Eight Centuries of Financial Folly*. Princeton University Press, Princeton, New Jersey.

Index

M.H.D. Kemp, *Systemic Risk*, DOI 10.1057/978-1-137-56587-7

J

K

T