Max Weber Lecture Series

THE GREAT MODERATION, THE GREAT PANIC AND THE GREAT CONTRACTION

Charles Bean
The Great Moderation, the Great Panic and the Great Contraction

CHARLES BEAN
Abstract
This lecture examines the causes of the recent financial crisis and subsequent recession. On the macroeconomic side, the Great Moderation encouraged an overly optimistic assessment of risk. Combined with low interest rates, reflecting both loose monetary policy and relatively high Asian savings rates, that encouraged a build-up of excessive leverage in the banking system. On the microeconomic side, distorted incentives led to a concentration and mispricing of risk. Informational complexities associated with new financial assets and the interconnectedness of financial institutions then resulted in the closure of funding markets and a flight to safety when loan defaults rose unexpectedly. The episode indicates the need to focus on agency and information problems between banks and their funders in addition to those between the banks and their borrowers. It also suggests that the process of financial intermediation should play a more prominent role in macroeconomic models.

Keywords
Great Moderation; global imbalances; financial crisis; leverage; credit booms; macroprudential policy.

JEL codes: E32, E44, E52, E58, F32, G01, G21

The lecture was delivered for the Max Weber Programme on Wednesday 20 January 2010 and was first delivered as the Schumpeter Lecture at the Annual Congress of the European Economics Association, Barcelona, 25th August 2009.

Deputy Governor for Monetary Policy and Member of the Monetary Policy Committee, Bank of England. I am grateful for helpful comments from Viral Acharya, Tobias Adrian, Tim Besley, Claudio Borio, Steve Cecchetti, Charles Goodhart, Paul Klemperer, Paul McCulley, Ken Rogoff, Hyun Shin, John Taylor, Bill White, Fabrizio Zilibotti and from numerous colleagues at the Bank of England. I am especially grateful for the expert assistance of Adrian Penalver. The views expressed are those of the author and do not necessarily reflect those of either the Bank of England or the Monetary Policy Committee.
1. Introduction

It is a great pleasure to deliver this Max Weber lecture. It is a nice historical touch that Ramon Marimon invited me to come here after I delivered my talk today as the Schumpeter lecture at the European Economic Association meetings last August. Weber and Schumpeter were roughly a generation apart in age. Though I can find no evidence that they actually met, it is clear that they were familiar with each other’s work on the dynamics of economic development and the sources of economic change.

Weber famously argued that it was only the ideology of Protestantism which supported the accumulation of capital, sparking the Industrial Revolution. For Weber, ideas drove reality. This was in stark contrast to Marx, who believed that ideas and social relations were just a superstructure imposed over an underlying materialist dynamic. And Schumpeter rejected both of these characterisations of economic development. For him, the hero of the development story was the entrepreneur who spots an opportunity and has the will to exploit it.

What has this got to do with the current financial crisis you may ask? Well, echoing Weber, some see the near-failure of the banking system in many advanced economies and the subsequent global recession as evidence of the bankruptcy of capitalism; new ideas will be needed to bring about a new world. At the other end of the spectrum, there are those who argue that the failure of the economics profession to foresee the crisis demonstrates the inadequacy of neo-classical economic theory; our ideas must change if we are to understand the world. Finally, some see this crisis as a failure of entrepreneurship. Powerful bankers had a vision to reshape the financial world and the will to bring it about. Unfortunately that vision turned out to be destructive rather than constructive.

In this lecture, I will examine the origins of the crisis and why it has proved so virulent. But rather than giving a blow-by-blow account of events, which have been well documented elsewhere, I focus instead on the underlying economic forces. How could what initially appeared to be little more than an over-extension of loans to a high-risk subset of US households end up generating global financial sector losses estimated by the IMF in October 2009 to be in the region of $3.4 trillion (IMF 2009a) and an output gap that is expected to be of the order of 5% of GDP in 2010 (IMF 2009b)? It is tempting to look for single causes, but a variety of factors – both macroeconomic and microeconomic – have contributed to the genesis of the crisis. While each might have been of limited impact in isolation, collectively they constituted a potent cocktail.

On the macroeconomic side, the widespread and prolonged experience of stable growth and low inflation during the period of the Great Moderation resulted in an overly optimistic assessment of likely macroeconomic risks. Moreover, loose monetary policy, particularly in the United States, resulted in low short-term interest rates, while high Asian savings rates were associated with low longer-term interest rates and substantial international capital flows into the United States and other advanced countries. Taken together, these factors encouraged banks and other financial institutions to leverage up their balance sheets so as to raise the return on capital.

At the microeconomic level, a variety of distorted incentives resulted in the concentration, rather than the diversification, of risk as well as the under-pricing of that risk. Moreover, the opaque structure of some of the new asset-backed securities and the high degree of interconnectedness between financial institutions meant that the information necessary to evaluate exposures accurately was missing when conditions deteriorated. The consequence was the closure of funding markets and a flight to safe assets. A number of dynamic processes then amplified the shock within financial markets and propagated it into the real economy. Aggressive, and in some cases unconventional, policy actions have helped to limit the damage to the real economy. But it is likely that loss in output associated with the crisis will be both substantial and persistent.
Understanding the causes of the financial crisis and the associated sharp contraction in output represents a central challenge for the economics profession, especially in view of its collective failure to anticipate the building risks ahead of time. Some have suggested this demonstrates the bankruptcy of the conventional economic wisdom and the need to develop completely new approaches. My belief, though, is that much of what went wrong can be adequately explained utilising our standard economic toolkit, especially in understanding the role played by distorted incentives, information problems and co-ordination failures. But reconstructing macroeconomics so that it provides a useful tool for policymakers will require the profession to move beyond the paradigm of frictionless capital markets to incorporate the central role played by financial intermediaries in a meaningful way. In particular, recent experience suggests the need to focus on agency and information problems that arise at the interface between banks and their funders, in addition to those between banks and those to whom they lend.

The rest of the paper is structured as follows. The next section briefly sets the scene by summarising some of the key features of the period of financial innovation and rapid credit growth, mainly for the purchase of financial and real assets, which preceded the crisis. Section 3 then explores the macroeconomic antecedents of the crisis: the impact of the Great Moderation on perceptions of risk; the role of loose monetary policy; and the causes and consequences of the global imbalances. Section 4 shifts the focus to the various incentive and information problems arising within financial markets. Section 5 then explores the amplification and propagation mechanisms that made the financial crisis so severe and its impact on the real economy so great. Section 6 briefly considers the policy response, particularly that of central banks. Finally, Section 7 considers some of the lessons for the future development of the discipline.

2. Setting the scene: The credit boom

During the decade or so up to 2007, there was a substantial increase in the ratio of household and business debt to GDP in a number of the advanced economies. For instance, between 1995 and 2007, the ratio of household and business financial liabilities to GDP rose from 123% to 179% in the United States and from 128% to 213% in the United Kingdom. The increase in Spain was even more marked: from 77% to 214% A striking feature of this period was also the increased reliance of financial institutions on the credit markets to fund their activities, rather than relying on the traditional medium of deposits. For instance, according to the US Flow of Funds, the total wholesale credit market liabilities of the US financial sector doubled from around 55% of GDP in 1995 to around 110% in June 2007, just before the start of the crisis.

This build-up in debt was accompanied by a wave of financial innovation. Between 2000 and 2007, there was a marked expansion in the issuance of asset-backed securities (ABS): securities backed by bundles of residential (RMBS) and commercial (CMBS) mortgages, other consumer credit, and loans to businesses. By selling these securities to investors, the loan originator could thus shift the associated risks off his or her balance sheet. And a totally new class of asset was born in the shape of Collateralised Debt Obligations (CDOs). A CDO re-packages the payoffs in a bundle of ABS into an ordered hierarchy of other securities, with each tranche only bearing losses once tranches senior to it have been wiped out. A bundle of risky assets could thus be re-packaged into other assets with different risk characteristics. Those offering bond-like returns in most states of nature were suitable for investors seeking low-risk assets, such as pension funds, while those tranches more likely to suffer losses could be sold to investors willing to bear more exposure to risk. This was a particularly attractive way of treating riskier loans, such as those to US sub-prime households. The increased importance of securitisation is illustrated in Figure 1, with the total value of new ABS issued globally rising from less than $1 trillion in 2000 to more than $3 trillion at the peak of the credit boom in 2005 and 2006.
Alongside the increased use of securitisation, there was also an explosion in the issuance of Credit Default Swaps (CDS). These allow the holder of a bond or other security to buy insurance against default on the part of the issuer of the security. As with CDOs, this allowed the risk in an asset to be unpacked, with risk redistributed to those more willing to bear it. At the end of 2004, the Bank for International Settlements estimated the global value of the notional principal of outstanding CDS at less than $10 trillion. By the end of 2007, this had risen to almost $60 trillion, roughly the same as annual global GDP.

By allowing specific risks to be split and then traded, this financial innovation appeared to facilitate the redistribution and dispersion of those risks (whether this was indeed the case is a matter I return to in Section 4). As the set of state-contingent securities became more complete, so financial markets appeared to draw ever closer to an idealised Arrow-Debreu world. Moreover, it seemed natural to think that this dispersion of risk should make the financial system more, rather than less, stable. That view is neatly encapsulated in the following remark of the then Chairman of the Federal Reserve:

> These increasingly complex financial instruments have been especial contributors, particularly over the past couple of stressful years, to the development of a far more flexible, efficient, and resilient financial system than existed just a quarter-century ago. (Greenspan, 2002a).

Significantly, the expansion in credit was not associated with a marked expansion either in the demand for goods and services, or in consumer price inflation. But it was associated with rises in the prices of certain assets, particularly those of housing. The rise in house prices was especially marked in those countries where credit rose most sharply, notably the Anglophone countries and Spain (see Figure 2), though some countries, such as Germany and Switzerland, avoided a rise in house prices altogether. In passing, it is also worth noting that the run-up in house prices in the United States was actually less marked than in the other Anglophone countries and Spain, despite being at the epicentre of the crisis. That reflects a more elastic response in the construction of new homes in the United States than elsewhere. Consequently, although the overvaluation of house prices might have been lower there than in some other countries, the overhang in the housing stock at the time of the bust was correspondingly greater.
The point to stress is that there was nothing self-evidently wrong with this expansion in balance sheets and the associated rise in asset prices. There is some evidence from developing economies suggesting that increased development in financial markets helps foster growth and catch-up (e.g. Levine, 2005). And if financial innovation allows households to gain easier access to mortgage finance, then one should expect that to lead to a rise in the real price of houses. But just as there is evidence that increased financial intermediation is often a Good Thing, so there is also evidence from numerous past credit/asset price boom-bust episodes that a juxtaposition of rapid credit expansion and sharp asset price increases is often a harbinger of a painful bust to come (e.g. Allen and Gale, 2007; Borio and Lowe, 2002). So everything hinges on whether the underlying developments are sustainable or not. The mistake of (most of) the economics profession, along with almost all bankers and market participants, was a failure to see what was going on beneath the surface and to understand how disruptively things could unwind.

3. Macroeconomic antecedents

The proximate causes of the present crisis undoubtedly lie within the financial markets. But it is worth dwelling on the macroeconomic environment in the years preceding the crisis, as this created fertile ground for the formation of what proved to be an unsustainable boom in credit.

3.1 The Great Moderation

To begin with, the decade and a half preceding the crisis was characterised by an unusually high degree of macroeconomic stability, with steady growth and low and stable inflation in most of the advanced economies. This is documented in Figure 3, which shows the average volatility in output growth and inflation in each of the G5 countries before and after 1992 (in the United States, the Great Moderation actually appears to have started a little earlier, in the mid-1980s). All five countries exhibit lower volatility in the post-1992 period, though in the Japanese case the lower output volatility was relative to a sub-par growth trend. This reduction in macroeconomic volatility is discussed at more length in Blanchard and Simon (2001) and Stock and Watson (2003); the latter appear to have been the first to describe the phenomenon as a “Great Moderation”. Bernanke (2004) lays out three broad explanations for the Great Moderation.
First, it could just be a happy accident, if this was a period characterised by unusually small or benign shocks. In particular, the emergence of China and other emerging market economies as a source of cheap manufactured goods generated both an increase in global competition and a terms of trade gain for the advanced economies, resulting in a beneficial “tailwind”, only mutating into a “headwind” towards the very end of the period as rapid global growth raised oil and other commodity prices (White, 2008).

Second, changes in economic structure could have reduced the macroeconomic impact of a given shock to aggregate demand or supply. Candidate structural changes include: the shift from manufacturing to services; tighter control of inventories leading to an attenuation of the inventory cycle (McConnell and Gabriel Perez-Quiros, 2000); increased competition in product and labour markets; and innovation in financial markets facilitating the dispersal of risk and an enhanced ability to smooth consumption (Dynan, Elmendorf and Sichel, 2006).

A final possibility is that better macroeconomic policies, including improved policy frameworks, had moved economies closer to the eponymous Taylor frontier tracing out the lowest achievable inflation volatility for a given volatility of the output gap (e.g. Romer and Romer, 2002; Taylor, 2008). In particular, earlier periods were bedevilled by monetary policies that had failed to anchor inflation expectations. But the pursuit of inflation targets and similar stability-oriented monetary policies by independent central banks seemed to have been more successful in anchoring inflation expectations.

Now if the Great Moderation was down to good luck, then improved macroeconomic performance was more likely to prove ephemeral. But if it was down to structural changes or better policy, there was more chance of it proving permanent. Moreover, if it was permanent, then part of the dividend from this reduced macroeconomic volatility should be a greater willingness to take on risk.

As it happens, the academic literature has so far failed to provide a clear answer on the relative contributions of the various explanations for the Great Moderation. Thus Cogley and Sargent (2005), Sims and Zha (2006), and Stock and Watson (2003) all ascribe a sizeable portion of the improved performance to good luck, while Clarida, Gali and Gertler (2000) and Lubik and Schorfheide (2004) put more weight on better policies. This inconclusiveness of the empirical literature in part stems from the fact that shocks are not measured directly, only their consequences. Thus what appears to be
good luck may in fact reflect the beneficial influence of structural changes or better policy in ameliorating their impact (Bernanke, 2004). See Benati and Goodhart (2009) for further discussion of this rather fundamental identification problem.

In any case, what matters is not what academics thought, but how market participants responded to these benign conditions. Market participants were faced with what is, in essence, a complex signal-extraction problem: How much of the reduction in economic volatility was likely to be permanent? But whereas many such problems involve learning about the first moments of a distribution, this involved making inferences about higher moments. The longer such a period of low volatility lasts, the more reasonable it is to assume that it is permanent. But as tail events are necessarily rarely observed, there is always a danger of underestimating tail risks (see Barro, 2009, for a discussion of the macroeconomic implications of tail risks).

Certainly, measures of volatility implied from options prices suggest that the perceived risks in financial markets had shrunk to extremely low levels by 2006. Figure 4 plots the implied volatilities for US equities (VIX) and treasuries (MOVE); both exhibit a marked decline after 2003-4. There are two ways to rationalise this. One possibility is straightforward extrapolation into the future of the benign experience of the preceding few years. An alternative and complementary explanation is excessive faith in the ability of policy makers to maintain steady growth in the real economy. The FOMC’s willingness to relax monetary policy aggressively in the wake of sharp asset price falls, coupled with its reluctance to raise interest rates to moderate asset price rises, may also have helped place a floor under expected asset prices (the so-called Greenspan Put; see Miller, Weller and Zhang, 2002).

A tendency of people to underestimate future risks during periods of good economic performance is a recurring theme in the history of financial markets. In part, this seems to be a mixture of a willingness or desire to believe that the world has changed for the better, and moreover to find ways to rationalise that belief – in other words, “this time is different”. To that must be coupled a limited imagination – shared in this instance with regulators and central bankers – to see quite how badly things could go wrong. But when David Viniar, the Chief Financial Officer of Goldman Sachs, noted in August 2007 that they “were seeing things that were 25-standard deviation moves, several days in a row” it was a
The Great Moderation, the Great Panic, and the Great Contraction

sure sign that something was badly awry. Wrong, that is, not with the world, but rather with market participants’ models of risk.

A particular feature of this underestimation of risk lay in the prevalent assumption that bad shocks were likely to be uncorrelated with each other. Thus the management of Northern Rock probably believed that if they were unable to roll over their short-term wholesale funding in the Asset-Backed Commercial Paper (ABCP) market, then they would still be able to find an alternative source of funds in the inter-bank market. As it happened, the same forces that led to a shutting down of the former in August 2007 also led to the closure of the latter at all but very short maturities. By the same token, many investors thought that they had purchased relatively safe securities because there were insured against default by the monoline insurers or through the purchase of CDS contracts. But that insurance was potentially worthless when large correlated losses occurred and the insurers proved to have insufficiently deep pockets. A systemic financial crisis is apt to generate the dysfunction of several markets simultaneously, making bad tail outcomes much more likely to occur than they appear when risks are treated as largely independent. That is a lesson thrown up repeatedly by financial crises, including most recently by the Tequila (1994), Asia (1997) and LTCM (1998) Crises. But it is a lesson that market participants seem to keep on forgetting.

3.2 Monetary policy

Other macroeconomic features probably also played into the substantial growth in credit preceding the crisis. In particular, it has been argued that monetary policy was too loose over this period, especially in the United States (see Taylor, 2007). Figure 5 shows the difference between the policy rate and benchmark estimates of an appropriate policy rate derived from a Taylor rule. Notwithstanding the usual caveats about the construction of such benchmarks, the very low level of the Federal Funds rate relative to the Taylor rule from 2001 to 2005 in the aftermath of the collapse of the dot-com bubble is striking. Of course, the Federal Open Market Committee held rates this low for a reason, namely concerns that the United States might go on to experience a deflation of the sort experienced by Japan during its “Lost Decade”. In order to mitigate that risk, it therefore not only cut the target Federal Funds rate to very low levels, but also committed to holding the policy rate at a low level for a sustained period so as to put further downward pressure on longer-term market interest rates.

But this is not the end of the story. A number of countries, most conspicuously China, chose to peg their currencies to the dollar. Consequently, monetary laxity in the United States was transmitted overseas, as those countries chose either to relax their own monetary policies to prevent a surge in capital inflows, or else to accumulate extra foreign reserves, which were then recycled into the US capital market, primarily as increased holdings of US treasuries. Moreover, highly elastic global supply associated with the availability of a large reserve army of labour from the rural sectors of China and India meant that global inflationary pressures remained subdued until constraints on the supply of commodities began to bite. Instead, much of the increase in liquidity showed up in global asset price inflation. Of course, this extra leg could have been avoided if the pegging countries had allowed their currencies to appreciate, or else pegged against a basket of currencies.

---

1 Dowd, Cotter, Humphrey and Woods (2008) calculate that, with a normal distribution, a 25-sigma daily event should occur about once in every $1.31 \times 10^{35}$ years; to put this in context, the universe is estimated to be just 12-14 billion years old!
Figure 5 does not, however, suggest that policy rates were noticeably out of line with the Taylor benchmark in the euro area and the United Kingdom. And even in the United States, where more of a case can be made that policy rates were too low, inflation remained low well into the recovery from the downturn in the early part of the decade. So there was very little evidence of conventional overheating. Credit growth was nevertheless relatively rapid and asset price inflation – especially for real estate – elevated in all three jurisdictions. That led some – most notably White and colleagues at the Bank for International Settlements (Borio and White, 2003; White 2006) – to argue that, despite the stability of inflation, monetary policy was nevertheless too loose. Moreover, the credit/asset-price boom was storing up trouble for the future. Central banks should, they argued, therefore “lean against the wind” by holding policy rates higher so as to mitigate these risks, even though their inflation objectives might not be threatened. I return to this issue below in Section 6.

3.3 Current account imbalances

A third feature of the macroeconomic landscape was the counterintuitive pattern of current account deficits and surpluses and the associated flows of international capital. The rapid development and integration into the global market economy of China, India, Brazil and other emerging market economies might have been expected to lead to capital flows from the advanced countries into these economies in order to provide the investment to equip the extra workers being brought on to the global labour market. Instead, we saw China and some other emerging market economies running large current account surpluses and exporting capital, while countries such as the United States and the United Kingdom ran current account deficits and imported capital. These payment flows are documented in Figure 6, where the data have been smoothed by taking moving five-year averages in order to make the underlying trends clearer. In 2007, for instance, the US current account deficit stood at $730 billion, while China’s surplus stood at no less than $372 billion.
A number of competing explanations have been advanced for this counterintuitive pattern of flows. First, Dooley, Folkerts-Landau and Garber (2004) argue that it simply reflects the juxtaposition of strong demand in the United States, coupled with China’s adoption of an undervalued exchange rate in order to accelerate the development of its traded goods sector and to soak up labour from the rural areas; in other words, that it was an equilibrium outcome reflecting policy choices. Second, Bernanke (2005) lays the blame on a “savings glut” in the surplus countries resulting from the lack of an adequate household safety net in China, the accumulation of precautionary holdings of international reserves by several emerging market countries after the Asia crisis, and rational savings of the windfall gains from higher oil prices on the part of oil producers. Finally, a number of authors link the imbalances to imperfections in financial markets. Caballero, Farhi and Gourinchas (2008) argue that the United States holds a comparative advantage in the creation of “high quality” financial assets from real investments, leading to a direction of savings into the US financial markets. Mendoza, Quadrini and Rios-Rull (2009) suggest that the relatively greater depth of the US financial markets has resulted in a comparatively lower volume of savings there. And Song, Storesletten and Zilibotti (2009) propose that China’s high savings and accumulation of foreign assets is a by-product of more productive enterprises facing restricted access to capital and needing to rely instead on internally generated finance.

Now while these flows are certainly likely to have added to the growth in credit, on the face of it they could only be a part of the explanation. For instance, the cumulative US current account deficit over the 2000-2007 period was $4.7 trillion. Over the same period, the stock of household, corporate and government debt outstanding rose by $14.4 trillion, more than three times as much. But it is possible that the global imbalances may have played into the credit boom through another route. Bernanke’s savings glut hypothesis was advanced primarily as an explanation for the decline in long-term real interest rates observed during this decade, as shown in Figure 7. And that low level of real returns on government bonds in turn encouraged financial institutions to shift into other, riskier, assets that offered a higher return, including the burgeoning class of asset-backed securities – what came to be known as the “search for yield”.

---

**Figure 6. Current account balances**

(\% of GDP, 5-year rolling average)

Source: IMF and national sources

![Current account balances graph](image-url)
There is relatively little research that sheds light on the precise contribution of the global imbalances to the credit/asset-price boom. As it happens, there is a noticeable inverse correlation across countries between house price inflation and current account positions (see Figure 8). But the causality could plausibly run from house prices, via demand, to the current account, as well as in the other direction. In fact, Granger-causality tests point to bi-directional causality, running both from current accounts to house prices and vice versa. That is consistent with some influence of the imbalances on asset prices. But the results could equally indicate that both are driven by a common underlying factor, such as the level of demand or the availability of credit.

4. Financial market failings

While the macroeconomic environment may have provided fertile ground for the credit/asset price boom to develop, to explain both the extent of the growth in credit and the magnitude of the subsequent crisis one also needs to dig down into what was happening inside financial markets. There, one finds a variety of incentive distortions and information problems that appear to have played a central role. (For much more on these and other failings that originated within the financial markets, see the contributions in Acharya and Richardson, 2009a.)

4.1 Distorted incentives

Banks are vulnerable institutions because they make long-term loans with uncertain returns that are financed predominantly by short-term debt instruments and callable deposits. If creditors think that the borrowers will be unable to repay the loans, then there is every incentive for them to run. That is why banks are required to hold a buffer of capital and reserves large enough to absorb losses in most feasible states of nature.

---

2 With four lags of each variable on the right-hand side and a full set of time and country fixed effect dummies for sixteen countries over the period 1975-2007, the $F$-statistic for the exclusion of house prices in the equation for the current account is 8.2, while the $F$-statistic for the exclusion of the current account in the house price equation is 4.2.
Consider for a moment a very simple financial intermediary (a “bank”) that can finance risky lending with either equity or debt, which here encompasses both deposits and credit market instruments. Assume that the bank pays \( R_D \) on its debt and earns (a stochastic) \( R_L \) on its loans, net of intermediation and monitoring costs. Then the return on its capital, \( R_K \), is given by:

\[
R_K = (1 + \lambda)R_L - \lambda R_D,
\]

where \( \lambda \) is the leverage (debt-to-equity) ratio. In the normal course of events, one would expect the return on debt to be an increasing function of leverage, as the greater the leverage, the greater the chance that the bank’s capital will be wiped out in the event of low returns on the loans and it will be forced to default on part or all of the debt. Indeed, with symmetric information between bank and creditor, there should be no gain from raising leverage at all by virtue of the Modigliani-Miller theorem.

Aside from the privileged tax status of debt, there are a number of reasons why the Modigliani-Miller theorem does not apply to banks in practice. First, creditors – especially if they are households rather than sophisticated financial market participants – may not even factor in the implications of higher leverage for the possibility of default. Second, even those creditors who do understand the consequences of higher leverage may expect the debt to be partially or wholly underwritten by the state in the event of loss. That has, of course, been exactly what has happened during this financial crisis, with many governments extending the protection afforded to retail deposits, and in some cases also indemnifying wholesale creditors from any losses too. This would not matter if the cost of the implicit insurance provided by the state had been properly priced and passed back to the bank as an up-front charge varying with the degree of leverage. But that was not the case in practice. Third, and relatedly, creditors may believe that a bank will be deemed by the authorities to be too important to be allowed to fail, in which case the state may be forced to inject additional capital, ensuring that creditors are still repaid. (The recent proposal for contingent convertible bank bonds (CoCos), which convert to equity before the existing shareholders are wiped out, is an attempt to address some of these issues.)
For all these reasons, there is an incentive for the bank to raise its leverage, because it will not suffer an offsetting penalty on its cost of funding. Moreover, the lower the perceived uncertainty associated with the loans is, the more the bank can afford to leverage up, while maintaining the same uncertainty over the return on its capital. The stability of the Great Moderation thus further encouraged financial intermediaries to increase leverage.

These distortions are characteristic of banking systems everywhere. In addition, a number of distortions relating specifically to securitisation have figured in the crisis. Securitisation is attractive because once a bank has sold on its loans, and the attendant risks, it can then leverage fresh loans off its capital. And, as noted earlier, such distribution of risks should in principle be beneficial for the economy. Unfortunately, at least in its application over the past few years, securitisation also introduced a number of distortions.

First, a loan originator who can sell a loan on at par has less incentive to take care who he lends to. The much higher default rates on cohorts of US sub-prime loans originated in 2006 and 2007 compared with earlier years is a symptom of the reduced incentive for care in the extension of loans (see Figure 9). The originator needs to have enough “skin in the game” if he is going to exercise due diligence in loan origination.

Second, the crisis revealed that many of the risks that securitisation was supposed to disperse had, in fact, remained either within the banking system, or else could be re-intermediated back into it during a crisis. Thus, the debt securities issued by one bank were frequently not sold on to investors outside the banking system, but were instead bought by the proprietary trading desks of other banks, leading to a network of cross-holdings of the underlying loans. That may make sense when the main risks are idiosyncratic to the loans originated by particular banks, but looks rather foolish when the risks are macroeconomic in nature and the returns on such securities are correlated.

Moreover, a considerable amount of the remaining risk was contained in institutions which, while not formally recognised as banks, engaged in exactly the same sort of maturity transformation, financing long-term assets by short-term debt instruments – the “shadow banking system”. These included
entities such as conduits, which housed the securitised loans and then financed them by selling short-term paper. But in many cases such entities held back-up credit lines with the originating bank, so that when funding difficulties arose, the securitised loans were in effect returned to the originating bank’s balance sheet. And even where there was no formal obligation to act as a lender of last resort to an off-balance-sheet vehicle, originators frequently chose to provide back-up finance in order to protect their reputation in funding markets. For more on the failure of securitisation to transfer risks and its consequences, see Acharya, Schnabl and Suarez (2009).

The motive for setting up these off-balance-sheet entities was entirely one of regulatory arbitrage. In most jurisdictions, off-balance-sheet vehicles were not required to hold capital in the same way as a bank would if the loans were on their balance sheet. So it appeared to be a neat way to boost profits without having to raise more capital (Acharya and Richardson, 2009b). The Banco d’España, the Spanish banking supervisor, insisted however that Spanish banks would have to treat conduits and the like as on balance sheet for capital purposes. As a result, Spain did not see the mushrooming of off-balance-sheet vehicles that was seen in the United States and the United Kingdom. A good question is why other supervisors were not equally aware of the regulatory arbitrage taking place.

A third distortion lies in the nature of remuneration contracts. Incentive contracts that offer potentially unlimited upside rewards, but cap the downside losses, encourage traders to take on excessive risk. One unintended consequence of the financial innovation was that it enabled clever traders to create positions with considerable embedded leverage – that is, portfolios requiring little payment up front, but whose returns amplified changes in the value of the underlying assets. Traders then had a natural incentive to gravitate towards these types of highly risky instruments. A puzzle is why those who framed the contracts were not more aware that this might happen.

A related problem is that it is extremely difficult for management to observe the risk being taken on by their traders, particularly when innovative financial instruments have unusual return distributions. Take, for example, a deeply out-of-the-money option. This pays a steady income premium and has little variation in value when the underlying instrument is a long way from the strike price, but generates rapidly escalating losses in bad states of the world. In good times this looks like a high-return, low-risk instrument. Only in very bad states of the world do the true risks being run become apparent.

### 4.2 Information problems

Information problems have also been central to the crisis. Though ABS may sound simple, in practice they are anything but. A plain vanilla RMBS typically contains many thousands of underlying mortgages and even the documentation on them can run to a couple of hundred pages. Such instruments provide diversification, but unless the mortgages are fairly homogenous in nature, it may be difficult for an individual investor to monitor the evolution of the underlying risk exposure precisely. The problem is compounded in the Master Trust model of securitisation, under which the mortgage pool must continually be topped up to replace maturing mortgages, since the characteristics of new mortgage cohorts may differ from those of existing ones.

In addition, US RMBS contain embedded options that further complicate evaluation of their worth. Ordinary US mortgage loans are usually non-recourse in nature and offer the borrower the option both of early repayment and of default if the price of the house falls below the value of the loan, leaving the lender to recover what they can from re-sale of the house at an uncertain price. US sub-prime mortgages also embed an option for the lender. Essentially sub-prime mortgages are structured as a rolling sequence of short-term loans. After an initial two or three-year period, the interest rate typically rises sharply, so that the borrower has a very strong incentive to re-finance. The lender then has discretion over whether to offer such a new and more affordable loan, depending on whether house prices have risen or not; see Gorton (2008) for further details. As a consequence of these
embedded options, RMBS, particularly those comprising sub-prime mortgages, have a rather complex pay-off structure that depends on the evolution of house prices, and are both tricky to price and difficult to monitor.

CDOs have yet more complex pay-off structures. As noted earlier, these re-package the payoffs in bundles of MBS into an ordered hierarchy of other securities, with each tranche only bearing losses once tranches above it have been wiped out. The pay-offs on these assets are therefore highly non-linear in the underlying risks – they behave very much like catastrophe bonds. As a consequence, the prices of these securities are apt to drop sharply once the expected losses on the underlying mortgages reach a certain point. This has indeed been a feature of the crisis, as illustrated in Figure 10, which shows the values of the ABX indices for differently-rated tranches of such MBS. The lowest rated tranches (BBB- and BBB) rapidly lost almost all their value during the early stages of the crisis as conditions worsened, followed subsequently by the higher rated tranches (A and AA). Even the AAA-rated tranche, which investors presumably bought on the premise that it offered pretty safe returns, lost a fifth of its value in the period following the collapse of Lehman Brothers.

![Figure 10. Prices on CDO tranches](source: ABX.HE indices from JP Morgan Chase and Co)

A typical CDO comprises a large number of RMBS from a variety of originators. On the face of it, this might seem like a good thing as it creates diversification. However, it is even harder to monitor the underlying risks than with a plain vanilla RMBS. That may not matter too much when defaults are low and only the holders of the first, equity, tranche suffer any losses. But once defaults rise materially, it matters a lot what such a security contains. And with highly non-linear payoffs, returns can be extremely sensitive to small changes in underlying conditions.

When defaults on some US sub-prime mortgages originated in 2006 and 2007 started turning out much higher than expected, there was a realisation that losses on some of these securities could be much greater than previously believed. And a growing realisation of the informational complexity of these securities made them difficult to price in an objective sense. Essentially, investors switched from believing that returns behaved according to a tight and well-behaved distribution to having very little idea about the likely distribution of returns – in effect, a state of Knightian uncertainty (Caballero and Krishnamurthy, 2008).

A further information problem arose out of the interconnected nature of modern financial markets. Banks, when they appear at all in economic models, usually just exist to transfer funds from savers to
borrowers and to monitor the latter. But the reality is a high degree of exposure of financial institutions to each other as a result of interbank loans, holdings of CDS written by other parties, holdings of securities issued by other parties, and so on. In normal times, one does not need to know much about a counterparties’ risk exposure. But, when there are doubts about the solvency of a major institution, a creditor needs to know not only his own exposure to that institution, but also about the exposure of his other counterparties. And the information problems multiply when more than one institution is at risk of insolvency.

The collapse of Lehman Brothers was critical. Prior to that – and particularly in the light of the earlier rescue of Bear Stearns – such an institution was seen as being too big and interconnected to fail. But the inability of the US authorities to put together a viable rescue plan, together with growing doubts about the effectiveness of the Troubled Asset Relief Program (see Taylor, 2009), sharply raised the likelihood that a whole range of other financial institutions might fail too. Evidence of this can be seen in the behaviour of major bank CDS spreads, shown in Figure 11. These rose especially sharply after the collapse of Lehman’s; note also the rise and subsequent fall earlier in the year, around the time of the rescue of Bear Stearns. Moreover, the prospect of having to make significant default pay outs called into question the solvency of a major provider of that insurance, American International Group (AIG).

Figure 11. Five-year senior CDS spreads for major banks

The bottom line of all this is that the information requirements for managing risk effectively escalated sharply at the same time as investors also realised that the distribution of prospective returns on a range of ABS was a great deal worse than previously believed. So it was not surprising that there was a rush for the exits and a drying up of wholesale funding to institutions that were thought to be at all vulnerable; in other words, a wholesale run took place (see Uhlig, 2009, for a model of such a run). Moreover, attempts to cope with that by selling off assets ran into an adverse selection problem: if someone was selling an asset, then perhaps that indicated that they knew it was a “lemon”. And even the revelation that the central bank was acting as a Lender of Last Resort by providing liquidity support, far from solving the problem, could be interpreted as a negative signal about an institution. That was the case in September 2007 with Northern Rock, when the public revelation of a Bank of England credit line triggered the first significant retail run on a UK bank for more than a century.

As a result, at times such as the “Little Panic” of August 2007 as well as the “Great Panic” of September-October 2008, the markets for a whole range of securities closed completely and investors
sought safety in liquid government assets. An indication of these pressures is provided by the spread between bank funding costs and equivalent maturity government debt. Figure 12 shows the three-month US Treasury Bill and three-month Eurodollar rates, together with the implied spread between them. Periods of heightened funding pressures on banks include: the outbreak of the crisis in August 2007; the scramble for funding as the subsequent year-end approached and banks sought to dress their balance sheets; around the time of the rescue of Bear Stearns in spring 2008; and especially after the collapse of Lehman’s in October 2008. And even when confidence appeared to improve somewhat, ABS markets generally remained dysfunctional.

Figure 12. US Treasuries-Eurodollar (TED) spread

5. Amplification and propagation

The presence of incentive distortions and information problems helps explain why the crisis happened. But why has the impact been so great? Back in 2007, the US Federal Reserve estimated the unexpected losses associated with the sub-prime mortgage market would be of the order of $50 billion. By October 2009, the IMF were putting the global write-downs in banks and other financial institutions on credit originated in mature markets at no less than $3.4 trillion over 2007-2010 (IMF, 2009a)! Alongside the escalation of expected losses, there has also been a sharp and internationally synchronised slowdown in global growth – the “Great Contraction” – that will almost certainly turn out to have been the sharpest in the post-war period, together with an unprecedented contraction in world trade (see Figure 13). The IMF has put the shortfall in advanced country GDP, relative to potential, to be of the order of around $2.2 trillion by 2010. How could this come about?

The first thing to be said is that much more was involved than the imprudent extension of loans to a few poor US households. Rather, the unexpected losses on sub-prime mortgages and assets backed by them acted as a trigger for generalised doubts about the viability of a whole class of assets and a

---

3 Testimony by Ben Bernanke to the US Senate Banking Committee, 19 July 2007.
A number of amplification and propagation mechanisms have, however, also kicked in. Some of these feedback loops have operated entirely within the financial sector. The most obvious is that the withdrawal of funding and/or the need to maintain adequate capital ratios prompted fire sales of assets. But if there is great uncertainty about the fundamental value of an asset, or there is a suspicion that the seller may have chosen to sell it because they know it is a lemon, then prices are likely to have to fall a long way before buyers step forward. That is particularly so if potential buyers are concerned about their own access to funding and therefore wish to remain in safe and liquid assets. But given the use of mark-to-market accounting to value trading books, the falls in asset prices then worsens the balance sheets of other institutions. That in turn reduces their capital ratios, aggravating their problems in raising funds and generating a further round of asset sales.

Further amplifying mechanisms arise as a consequence of leveraged positions. The net worth of a leveraged investor deteriorates much faster than his gross worth when asset prices fall, as all of the loss is scored against capital. Thus even to hold leverage constant requires substantial asset sales, further depressing asset prices. For instance, an investor with $100 million of assets financed by $10 million of capital and $90 million of debt who experiences a $5 million loss in asset value would have to sell $45 million of assets and redeem an equivalent amount of the debt to restore his original leverage ratio. Brunnermeier (2009) refers to this as a “loss spiral”. In fact, things are even worse than this because lenders typically expect borrowers to fund a larger fraction of an investment themselves in such circumstances, reinforcing the downward spiral. Adrian and Shin (2008) report empirical evidence to support this, finding that the leverage ratio of the five large US investment banks is strongly correlated with the size of their balance sheets.

Now it was noted above that once counterparty risk becomes a material concern, it is not only the state of an intermediary’s immediate counterparties’ balance sheets that matter, but also the state of those counterparties’ counterparties balance sheets, and so on. In practice, what used to be a very simple process of intermediating funds between savers and borrowers has evolved into a highly complex network, which is both difficult to comprehend and a major source of uncertainty. This complexity is illustrated in Figure 14, which shows the exposures between the 24 largest UK banks at the end of
Charles Bean

2009 Q2. (An equivalent network map for the major global banking entities would be even more revealing, but unfortunately the necessary data is not readily available.)

**Figure 14. Network of large exposures between UK banks**

The raw data underlying this map, drawn from regulatory returns, is the value of “large” exposures of bank A to bank B and vice versa, where “large” is defined as exceeding 10% of the debtor bank’s eligible capital. Each node represents a bank, with the size scaled in proportion to the sum of the total value of exposures to that bank and the total value of exposures of that bank to others in the network. The thickness of the line connecting two nodes is proportional to the value of the bilateral exposures between that pair of banks. Data source: FSA returns for 2009Q2

The analysis of the behaviour of such financial networks is still very much in its infancy (for more, see Haldane, 2009). But a key point material to regulatory design is that some types of network will be inherently more stable than others, depending on both the nature of the network and the obligations on its members. For instance, a network where all institutions are of similar importance and exposures are evenly spread is likely to be relatively stable, as the consequences of an adverse shock to any one institution will be spread widely and thinly. In contrast, as we saw with the collapse of Lehman’s, networks in which there are a relatively small number of key players are potentially very susceptible to the failure of a key player. By the same token, a network in which exposures are collateralised or can be netted across the system will be more stable than one where gross exposures are forfeited in the
event of default, as in the latter case the cascade through the network of the losses is likely to generate more failures on the way.

Everything so far relates to amplification within the financial sector. But what about the substantial impact on the real economy? The IMF’s estimate of total financial sector write-downs is around $3.4 trillion. If we assume an annual marginal propensity to consume out of wealth of 5%, that suggests a fall in demand of around $200 billion. But the loss in advanced country GDP relative to potential is projected to be more than 10 times bigger than that. So we cannot come near to explaining the Great Contraction purely in terms of wealth effects.

The most obvious additional mechanism for the propagation of the financial crisis into the real economy is through a “credit crunch”, as banks tighten lending standards and with it the supply of credit to businesses and households. That could arise because lenders seek to correct a previous under-pricing of risk and the associated lending to unsuitable borrowers. But it could also arise because of the banks’ need to reduce leverage and repair their balance sheets. Both the bank lending data and credit conditions surveys in the United States, the United Kingdom and the euro area appear consistent with some restriction in the supply of credit from the middle of 2007 onwards. But there was little evidence after the collapse of Lehman’s of a further pronounced tightening in the supply of credit large enough to account for the sudden sharp contraction in activity in 2008Q4 and 2009Q1.

There are two candidate, though not mutually exclusive, explanations for the large impact on output. Expectations of future credit availability may have worsened sharply. Or, more plausibly, there was a more general increase in uncertainty as a result of the near-meltdown in the financial sector, leading to an increase in precautionary savings, the postponement of planned investment projects and a running down of inventories.

The sharp contraction in world trade is particularly notable (see Figure 13). In the years leading up the crisis, world trade was growing at around 8% a year. In the two quarters following the collapse of Lehman’s, world trade was almost 20% lower than its level a year earlier. Nothing like this has been seen in peacetime. While the ultimate magnitude of the contraction in trade was greater during the Great Depression, that took nearly three years to come about and was in part associated with the introduction of protectionist measures. Trade fell much more sharply on this occasion.

A number of factors appear to have played into the collapse in trade. First, the downturn in global demand was unusually synchronised. In part, that reflected the global impact of the near-meltdown of the world’s financial system. And, in part, it reflected the rapid international transmission of the shock to aggregate demand down globalised supply chains. Second, the downturn in demand was concentrated in consumer and producer durables, which are more highly traded than consumer and business services. Third, the globalisation of supply chains and the associated unbundling of production mean that value added early in the production process typically crosses frontiers a number of times as intermediate inputs are exported and then incorporated into other goods and then re-exported. The rapid expansion of world trade during the decade or so before the crisis in part reflects this lengthening of supply chains. Consequently, when GDP – a value-added concept – falls, it is possible for world trade to fall more than proportionately. Fourth, there are suggestions that the turmoil in financial markets may have particularly impacted on the availability of trade finance. Baldwin (2009) contains more on the causes and consequences of the collapse in international trade.

History suggests that downturns associated with financial crises tend to be both deeper and last longer than normal demand-driven downturns. Reinhart and Rogoff (2008, 2009a, 2009b) have mapped the macroeconomic characteristics of the present financial crisis against preceding episodes in both advanced and emerging market economies. Their results can be summarised as follows: asset price collapses tend to be deep and prolonged; the aftermath of banking crises tend to be associated with substantial and prolonged falls in output; the downturns typically lead to a substantial build-up in
public debt; and the antecedents of the present crisis appear remarkably similar to those in previous episodes. Similar results reported by the IMF (2008) suggest that the cumulative output loss in outright recessions, as well as in milder slowdowns, on average tends to be almost twice as much if preceded by a financial crisis.

Moreover, the results of Cerra and Saxena (2008), across a large number of countries and a variety of types of crises, suggest that even in the long run the losses from crises may never be fully recovered. IMF analysis (IMF, 2009b) of the aftermath of no less than 88 banking crises suggests that, although growth recovers, there is a persistent loss in output of, on average, about 10% relative to the pre-crisis trend, though there is quite a lot of heterogeneity in country experience. Qualitatively similar results are reported by Furceri and Mourougane (2009), who find that the typical OECD financial crisis knocks around 2% off potential output, which rises to 4% for deep crises.

These findings suggest that financial crises have a material impact on both the dynamics of the business cycle and on the underlying equilibrium growth path. The literature on the “broad credit channel” and the “financial accelerator” points to an amplifying effect of changes in firms’ financial health on their cost of capital. In particular, when borrowers have less collateral they face a higher external finance premium to compensate lenders for the increased monitoring costs (see e.g. Bernanke, Gertler and Gilchrist, 1999). But the quantitative magnitude of the effect in this type of model is usually relatively modest. And the banks in this class of model are simple or non-existent, and any incentive or information problems relate to the borrower, rather than the financial intermediary.

Instead, it seems likely that the adequacy of bank capital and the associated phenomenon of financial sector de-leveraging play a central role. Capital is essential because it is the means by which intermediaries persuade creditors that they will almost surely get their money back. Increasing capital is one way to get leverage down; that is why governments in the United States and other affected countries forced banks to raise more capital privately or else accept injections of extra capital financed by the taxpayer. Extending guarantees on funding, whether retail or wholesale, is another way to make creditors believe that they will get their money back, as are measures to insure or quarantine the risky bad assets.

An open question is therefore whether the banking system will be able to support an adequate volume of credit as the extensive network of public support is removed. If not, then further attempts to de-leverage are possible. And de-leveraging through the restriction of new loans generates a feedback loop analogous to that through asset sales. De-leveraging by restricting credit growth may be privately rational, but if every intermediary tries to do the same, the result will be lower activity and higher defaults and little improvement in leverage ratios – what, to paraphrase Keynes, might be described as the “paradox of de-leveraging” (McCulley, 2008). A recent contribution from Gertler and Karadi (2009) explicitly captures the role that de-leveraging can play in slowing a recovery when financial intermediaries need to hold capital in order to convince creditors that they will be repaid.

6. Monetary policy during and after the crisis

The severity of the financial crisis and the sharpness of the contraction in activity have prompted an aggressive response by policy makers. These include measures to support the banking system: capital injections; funding guarantees; underwriting of losses on toxic assets; and central bank liquidity support. Fiscal and monetary policies have also been highly stimulatory in order to counteract the associated shock to aggregate demand. Lack of space precludes lengthy discussion of the measures taken by governments to deal with the stress in the banking system (see Swagel, 2009, for an insider’s view of US policies and IMF, 2009a, for a recent international analysis), but some brief commentary on the unorthodox monetary policies adopted by central banks is appropriate here.
The immediate response was to offer liquidity support, particularly after the intensification of the crisis in October 2008. Following Bagehot (1873), the classic dictum in the face of a financial crisis is for central banks to lend freely against good collateral to solvent but temporarily illiquid institutions, though at a penalty rate to discourage moral hazard. Although the details of the operations undertaken differ, the central banks in each of the main jurisdictions widened both the range of the collateral they were prepared to lend against and extended the tenor of the loans, though in some facilities the penalty appears to have been on the nugatory side. In the United States, because of the critical importance of non-bank financial intermediaries in US capital markets, the Federal Reserve also extended the range of counterparties that it was prepared to deal with. As a result, central bank balance sheets expanded sharply in the wake of the collapse of Lehman’s (and therefore well before policy rates reached their near-zero lower bounds); see Figure 15.

**Figure 15. Central bank balance sheets**

Central Bank total liabilities (index: Aug. 2007 = 100)

All the main central banks also cut their policy rates aggressively, so that both policy rates and short-term market rates came close to their effective floor of zero. Before the crisis, and with policy rates averaging somewhere in the 4%-5% range, most central bankers believed they had plenty of room for manoeuvre to offset adverse demand shocks. But the transmission mechanism of monetary policy is weakened when the financial system is de-leveraging. Hence central banks found themselves needing to inject even more stimulus, taking them into unknown territory.

There are two primary options confronting a central bank that has reached the zero interest rate lower bound, both of which seek to influence a wider range of asset prices (Buiter and Panigirtzoglou, 1999, discuss these and other, more exotic, approaches). The first is to provide a commitment to keep the policy rate low for a significant period into the future. This approach had already been followed by the Federal Reserve earlier in the decade, when it set interest rates much lower than suggested by a
Taylor rule. At the present juncture, the Federal Reserve, the Bank of Canada and the Riksbank have all made explicit statements that imply policy rates will remain low.

The idea behind providing such commitments is that they both pull down market interest rates further out along the yield curve and raise expected future inflation. Indeed, in the canonical New Keynesian/New Classical DSGE model, this is the only way of stimulating the economy at the zero interest rate lower bound, as the impact of monetary policy is completely summarised by the current and future path of the policy rate. A variation on this theme, in essence adopted by the European Central Bank, is to provide unlimited financing to the banking system at the policy rate at longer maturities than usual.

The other option involves the outright purchases of assets, rather than the normal repurchase agreements (i.e. collateralised lending) undertaken by central banks. This may involve the purchase of either government debt or corporate assets. And they may be financed by the issuance of extra central bank reserves, raising the monetary base. Or they may be financed by the sale of other assets in which case there will be no effect on the monetary base. The Federal Reserve and the Bank of England have bought both government and private assets, though with different emphases, in part reflecting the different financial market structures in the two countries. In particular, the Federal Reserve has bought significant quantities of a variety of corporate assets, whereas the Bank’s purchases have been more heavily concentrated on purchases of government debt.

The aim of purchases of corporate assets is to enhance market liquidity by having the central bank stand ready to act as a backstop buyer. In so doing, it should bring down spreads over government bond yields by reducing liquidity premia, encourage issuance and generally improve the functioning of capital markets. Importantly, the central bank may not need to make a large volume of purchases to make a difference. The mere fact of credibly standing ready to buy may be enough to have the desired impact. However, the central bank will be taking private credit risk onto its balance sheet, potentially exposing itself – and indirectly the taxpayer – to the risk of credit losses. So ultimately, this is better thought of as a fiscal action.

Purchases of government debt are instead intended to have a more general impact by pushing down bond yields. That, in turn, can be expected to push up a whole range of asset prices as the sellers subsequently re-balance their portfolios. Rather than being considered unconventional, purchases of government debt financed by the issuance of extra central bank money is really just a return to the classic monetary policy operation of the textbook: an open market operation. The only things that distinguish the present operations from a traditional open market operation are the circumstances under which they are taking place and their scale.

The mechanism whereby asset purchases are supposed to influence the economy are absent in the canonical New Keynesian/New Classical macroeconomic model, because of the assumption that the returns on interest-bearing assets all move in tandem and that any risk or term premia are constant. The rationale for asset purchases – that relative asset supplies affect asset prices and returns – instead lies in an older strand of the literature, dating back to Tobin (1969) and Brunner and Meltzer (1972).

The initial responses in the United Kingdom to these measures have been moderately encouraging. Government bond yields fell significantly on the commencement of the programme of asset purchases, and yields appear to be some 50-75 basis points lower than they would otherwise be. And there are also signs of beneficial effects on conditions in the relevant corporate credit markets. Meier (2009) provides a fuller assessment. But it would be premature to draw conclusions about the efficacy of such measures, as the transmission lags through to nominal spending are likely to be quite long. Moreover, even after the event, it will be difficult to draw firm conclusions, as we can never know with certainty what would have happened in the absence of such policies.
Let me now turn to the lessons from the crisis for the future conduct of monetary policy. As noted in Section 3.2, in the years preceding the crisis, there was a lively debate amongst central bankers as to the right way to respond to rapid credit growth and rising asset prices. On the one hand, the Bank for International Settlements and others (e.g. Borio and White, 2003; Cecchetti, Genberg and Wadhawni, 2002; White, 2009) advocated “leaning against the wind” by raising interest rates above the level warranted by the immediate outlook for inflation and activity. On the other hand, the Federal Reserve espoused a policy of “benign neglect” coupled with aggressive relaxation when asset prices fell sharply (see e.g. Greenspan, 2002b; Bernanke and Gertler, 2001), on the grounds that identifying an asset price bubble in its early stages was difficult and trying to deflate an established one was dangerous.

The events of the past couple of years have clearly tipped the balance in favour of taking some sort of pre-emptive action. And in a second-best world, where monetary policy is the only instrument available to cool a credit/asset-price boom, a degree of leaning against the wind makes sense. Indeed, an inflation-targeting central bank ought to be willing to undershoot its target in the medium term, if it thereby improves its chance of meeting the target further out through the avoidance of a disruptive bust (Bean, 2003). But monetary policy is a somewhat blunt weapon to use to control credit and asset prices because it also has a wider impact on aggregate demand. And raising interest rates enough to cool a credit/asset-price boom in full swing is likely to require a substantial increase in interest rates and thus involve substantial collateral damage to real activity (though for a dissenting view, see Adrian and Shin, 2008, who point out that even small rate changes can have large impacts when institutions are highly leveraged).

Ideally therefore, with an additional objective in the shape of managing credit growth and asset prices, one really wants another instrument that has a more targeted impact on credit growth and asset price inflation than do interest rates. That is what “macro-prudential regulation” is supposed to achieve.

Most discussion in policy circles has so far revolved around introducing pro-cyclical capital requirements, or its close cousin dynamic provisioning. Thus banks would be required to build up extra capital/reserves during a credit/asset-price boom, which can then be run down in the event of a bust. This should reduce the incentive for banks to leverage up in a boom, as well as making the financial system more robust in a bust. However, other instruments could also be deployed to this end. For instance, credit/asset-price booms seem to be characterised by an excessive shift into riskier forms of lending. In that case, an instrument more directly targeted at the microeconomic distortion would be to increase the risk-weights that are attached to such lending when computing banks’ required capital. And, as we have seen in the present crisis, much of the action may take place outside the regulated banking system in the wider credit markets. In that case, varying margin requirements might be a more appropriate instrument for dealing with vulnerabilities building up in the capital markets more generally. The optimum approach seems likely to involve a portfolio of instruments.

Once potential instruments have been identified, a host of other questions then need to be addressed. Precisely which variables should the instruments vary with and how are the variations supposed to be calibrated? To do all this properly, one needs a clear understanding of the nature of both the market and organisational failures. And once that has been answered, policy makers need to understand both the macroeconomic impact of the instruments and how their operation interacts with monetary and fiscal policies. Moreover, everything so far has been couched in terms of moderating the impact of the credit cycle over time. But the earlier discussion of network effects highlighted the potential role played by large and highly interconnected financial intermediaries, graphically illustrated in the aftermath of the collapse of Lehman’s. That points to another dimension for macro-prudential regulation, namely to increase the robustness of the financial system by requiring systemically important intermediaries to carry heavier capital requirements.
There is clearly much work to be done in developing and calibrating a suitable macro-prudential toolkit. Borio and Shim (2007) and Brunnermeier, Crockett, Goodhart, Persaud and Shin (2009) provide fuller discussions of the issues. This is an area where future academic research can potentially be of great assistance to policymakers and regulators.

7. Some lessons for the discipline

I conclude with some observations regarding the implications of the crisis for the future development of our subject. In all probability, the Great Panic and the Great Contraction of 2008 will join the Great Depression of the 1930s and the Great Inflation of the 1970s as discipline-defining events, especially for macroeconomics. The struggle to understand and deal with the Great Depression led to the invention of macroeconomics as a distinct branch of the subject. In the case of the Great Inflation, economic theory for once led events, with the development of the natural rate hypothesis foreshadowing the subsequent take-off in inflation. But the subsequent decades were spent refining our understanding of the inflationary process. What will be the legacy of the Great Panic and the Great Contraction?

First, it would be a mistake to rely on monocausal explanations. Underestimation of risk born of the Great Moderation, loose monetary policy in the United States and a perverse pattern of international capital flows together provided fertile territory for the emergence of a credit/asset-price bubble. The creation of an array of complex new assets that were supposed to spread risk more widely ended up destroying information about the scale and location of losses, which proved to be crucial when the market turned. And an array of distorted incentives led the financial system to build up excessive leverage, increasing the vulnerabilities when asset prices began to fall.

The multiplicity of factors that played into the crisis help to explain why the profession failed to foresee the full extent of the risks that were building up in the system. To be sure, a few wise souls can justifiably claim that they saw at least part of the developing threat. Shiller and others drew early attention to the incipient bubble in the US housing market (e.g. Case and Shiller, 2003). White and colleagues at the Bank for International Settlements highlighted the dangers of lax monetary policy and rapid credit growth even when inflation might seem to be under control (e.g. White, 2006). Numerous authors, from Obstfeld and Rogoff (2001) onwards, noted the threat posed by the pattern of international capital flows, though usually the expectation was that it would result in a large and disruptive depreciation in the dollar rather than a banking crisis. Some central banks highlighted the threats associated with the general compression in risk premia (e.g. Bank of England, 2005). And shortly after the crisis began, Reinhart and Rogoff (2008) correctly gauged the macroeconomic significance of the event. But no-one can really claim to have come near to seeing the whole picture in advance.

Second, the economics profession has oversold the virtues of unfettered financial markets. We usually start from a presumption that markets work best when they are left to themselves, unless there are obvious market failures present. By the same token, even though not strictly the case in theory, we usually start from a presumption that expanding the range of available securities is beneficial. Yet that has resulted in a deeply unsatisfactory outcome.

One view is that we need to bring psychology and the insights of behavioural economics into the study of financial markets (e.g. Akerlof and Shiller, 2009). That may indeed be part of producing a richer and more accurate description of financial market behaviour, particularly in regard to expectations formation. But distorted incentives, information failings, collective action problems and network effects have all played a large part in recent events. Our standard analytic toolkit can get us a long way in understanding what has happened.
Third, we should pay more heed to the lessons of history. Financial booms and busts have occurred with some regularity ever since the Tulip Mania of 1636-7. Yet macroeconomists tend to treat them as pathologies that happen at other times or in other places. Their frequency suggests that we would be better advised to think of them as a central feature of capitalist economies that our standard models should be able to capture. And once we have learned the lessons from this episode, we need to be sure that we (and our successors) do not then forget them.

Fourth, policymakers grappling with the crisis have found little to help them in the DSGE analysis that has dominated macroeconomics over the past couple of decades. Financial intermediaries are conspicuous by their absence in the workhorse New Keynesian/New Classical DSGE model. The focus is rather on intrinsic dynamics resulting from inter-temporal decision-making in the face of a variety of adjustment costs and impediments to price adjustment; there are no financial frictions to speak of. That such a framework has developed may be unsurprising in the light of the Great Inflation and its subsequent Great Moderation. But the fact that financial intermediation plays a negligible role in Woodford’s (2003) magisterial state-of-the-art opus, *Interest and Prices*, speaks volumes. To provide a more useful macroeconomics, we need to put credit back into it in a meaningful way.

Of course, a number of authors have tried to introduce banks into the standard model. Some recent contributions include: Christiano, Motto and Rostagno (2008); Curdia and Woodford (2009); Gerali, Neri, Sessa and Signoretti (2008); Goodfriend and McCallum (2007). But these build on the earlier strand of research relating to the broad credit channel that stems from the work of Bernanke, Gertler and Gilchrist (1999) and Kiyotaki and Moore (1997), and which focuses on agency or information problems relating to the borrower.

By contrast, frictions originating within the financial sector itself played a central role in recent events. In particular, the agency and information problems lay one step further back in the financial intermediation process, namely at the interface between the banks and their funders. Some recent contributions have begun the task of introducing such considerations into the standard framework; see, for instance, Aikman and Paustian (2006); Chen (2001); Gertler and Karadi (2009); Meh and Moran (2004); and Van der Heuvel (2008). Goodhart, Sunirand and Tsomocos (2005, 2006) provide an even richer model of the intermediation process, though at the cost of greater complexity and with a simpler representation model of the real economy. And Adrian and Shin (2009) and Cecchetti (2008) have begun exploring the implications of the behaviour of such financial intermediaries for the monetary transmission mechanism. But there is still a long way to go in the development of rich, tractable and useful macroeconomic models containing financial intermediaries.

Reintroducing the financial intermediation sector as a source of shocks, rather than merely acting as an amplifying mechanism as in the broad credit channel, represents something of a return to older models of the business cycle, in which credit creation and destruction played a central role. Recent events have, in particular, stimulated renewed interest in the ideas of Hyman Minsky (see e.g. Minsky, 1982). In Minsky’s theory of endogenous financial booms and busts, prosperous times lead to an excess availability of credit and the gradual development of credit-fuelled asset-price bubbles. Financial crises result when those debt levels become excessive – a so-called “Minsky” moment. A credit crunch then ensues leading to a downturn in the real economy. Much of what is contained in this lecture echoes some of Minsky’s ideas.

As a student at Harvard, Minsky was a protégé of Schumpeter, whose teaching influenced him greatly. Schumpeter viewed credit as an essential ingredient of the capitalist machine and he would surely have regretted the downplaying of financial intermediation in macroeconomics over the past forty years. And he would surely endorse the need to rectify that omission now. So the final words go to him:

Credit creation [is] the monetary complement of innovation. This relation…is at the bottom of all the problems of money and credit. (Schumpeter, 1939, p.111)
Bibliography


