

## Reforming finance: are we being radical enough?

**2011 Clare Distinguished Lecture in Economics and Public Policy**

**18 February 2011**

Between 2007 and 2009 the financial systems of the developed world suffered a major crisis, the after-shocks of which we are still seeking to manage. As a financial crisis, it was as big as anything in 75 years; in the UK you have to go back to before the First World War to find an equivalent scale of bank losses or liquidity runs. And the crisis has had major macroeconomic and human consequences – unemployment, real income loss, some house owners in negative equity, and taxpayers burdened for a decade or more with dramatically increased government debts.

So, not surprisingly, the crisis provoked much discussion of the need for radical reform. And not just technical reform of financial regulations, but also, many argued, a need:

- to challenge the Anglo-Saxon model of financial capitalism;
- to halt the rising scale of financial activity within the economy and the apparently related rise of inequality;
- and to reconstruct economics as a discipline, challenging the free-market simplicities of rational expectations and efficient-market hypotheses.

Has this challenge actually occurred? Or have we, in the wake of the crisis, patched up the existing system, reformed the technical details, but left our model of capitalism unchanged and existing economic theories largely unchallenged? And if radical challenge has not occurred, should it? Have we been radical enough?

Different protagonists in the debate focus on different dimensions of radicalism. For much of the press, and probably much of the populace, the touchstone of radicalism is whether bankers' bonuses will be reduced. For some policy experts, the touchstone is structural reform of banks, breaking up the 'too big to fail' banks, rather than simply increasing their capital requirements. For the French government, in its presidency of the G7 and G20, the measure of success in 2011 will be whether we proceed beyond the reform of financial systems to reform the global monetary system, tackling the perceived problems of speculation in commodity markets, harmful volatility in short-term capital flows, and the supposed inadequacies of a system still dominated by one reserve currency.

So I will aim in this lecture, not only to ask are we being radical enough, but also what should we mean by radicalism, which problems do we need to fix, as well as how do we fix them?

One lecture cannot, however, cover all the issues and potential problems. So one set of topics I'll exclude for now are the French questions for this year – does effective reform have to imply reform of the international monetary system as well as the financial system?

And I will therefore concentrate on three issues, three possible dimensions of radicalism: (see Slide 1 – there is a link to the PowerPoint presentation on the top right of [this page](#))

- Radicalism in pursuit of financial stability. Have the technicians of financial regulation been radical enough in the reform of capital and liquidity? Have we significantly reduced the probability and severity of future financial crisis?
- Second, issues relating to financial intensity, the social value of financial activity and inequality. Have we thought deeply enough about why finance has grown so large relative to the real economy, why pay in finance is so high, whether this is a problem, and if so whether we can do anything about it?
- Third, radical challenges to dominant economic theory. Was this crisis, as some have argued, a crisis not just of specific institutions and regulations, nor even just a crisis of markets in general, but also of an entire economic theory? And if so has that theory been appropriately de-throned?

Those three dimensions may seem somewhat disparate. The first is technical; the second raises wider social issues; the third, issues of theory. But there is, I hope to show, a linking theme, which is the need to look at the total system and the macro-effects, recognising the deep roots of instability and excessive financial activity, and avoiding the delusion that if we can only fix a few specific incentive and market discipline problems, all the problems will disappear.

And across each of these dimensions, I will argue that there is a danger, as the crisis recedes, that we will not be adequately radical in our challenge to pre-crisis assumptions, or recognise how deep are the problems we may face.

But I will also suggest that one particular way of defining radicalism may be too simple.

For some commentators it is clear that dealing with 'too big to fail' banks is the policy which unlocks everything – that if we fix 'too big to fail' we would not only have financial stability, but would also address the problems of distributional fairness – the two issues linked because it is 'too big to fail' status that enables bankers to gamble with other people's money.

Now I want to be clear that I believe fixing 'too big to fail' is **very** important: if I didn't believe that, I would be seriously worried about how large a proportion of my working

hours and my mental effort this year are going to be devoted precisely to that subject. But I do not believe it is the sole and sufficient key to progress.

I think we could face an unstable system even if we had perfect mechanisms for allowing banks, however large, to fail without taxpayer support. And that we could have growing inequality, driven by growing financialisation of our economy, even if all banks could fail or banks were smaller. And there is I believe some danger that an **exclusive** focus on the 'too big to fail' debate, far from being theoretically radical, can reflect a new variant of the simplistic conventional wisdom which provided the theoretical underpinning for past policy mistakes.

That conventional wisdom assumed that individuals operating within the financial system act in a fully rational fashion. While it recognised that markets can be imperfect, it therefore tended to assume that these imperfections arose solely because of imperfect incentive structures and principal-agent relationships. As a result, fixing incentive problems, such as those created by 'too big to fail' status, would, under this past conventional wisdom, be a **sufficient** as well as necessary response.

There are, however, strong reasons for believing that financial market instability and inefficiency can arise not only because of poor incentives but also because individuals do not always act in fully rational ways, and because indeed an assumption of perfect forward-looking rationality is at odds with the reality of inherent irreducible uncertainty. Myopia, irrational exuberance and herd effects matter as much as poor incentives. As a result the financial systems and markets could remain deeply unstable even if obvious incentive problems, such as those created by 'too big to fail' status, were fixed. Other policy responses, focussed on the stability of the total system, are therefore essential alongside necessary but not sufficient policy measures to address the problems of 'too big to fail' institutions.

In summer last year, I discussed some of the issues covered in this lecture with Tommaso Padoa-Schioppa, the much respected former Italian finance minister, central banker and economist who sadly died late last year. '*The question we have not yet answered*' he said '*was whether this was a crisis of specific institutions or a crisis of markets and systems*'. A key conclusion of this lecture is that we must understand it as the latter.

## 1. Financial Stability

First, then, financial stability.

Are the reforms which the global financial regulators have now agreed – Basel III as they are now known – radical enough? Have we found an answer to the 'too big to fail' problem? And are there other aspects of regulatory reform on which we have failed to focus amid our concentration on Basel III and large systemically important banks?

To answer that we must first be clear about the essential problem we need to solve. The bank solvency and liquidity crisis of Autumn 2008 led public authorities to

provide taxpayer support to prevent harmful failure: government equity injections, government guarantees of senior term debt, and exceptional central bank liquidity support. And public debate therefore often focuses on the need to avoid any taxpayer support in future – indeed the Dodd Frank Bill in the US makes it legally impossible to provide such support on a bank-specific basis.

But it is a striking fact that the total direct cost of such direct support is typically very small compared with the macroeconomic harm wrought by financial crisis.

- Thus, the International Monetary Fund's (IMF) estimates of the total direct cost of public support during the crisis, published in June last year, suggest that on average it might amount to less than 3% of GDP. (Slide 1) And latest estimates for the US suggest that it could still be less, indeed it could be negative, with the public authorities making a profit, certainly in relation to the commercial banks, if not in relation to Fannie Mae, Freddie Mac and AIG. That reflects the fact that equity stakes are often subsequently sold at a profit, that direct guarantees often generate fee income but no losses, and that central bank liquidity operations are often profitable.
- But these direct support costs are swamped by the macroeconomic harm produced by the financial crisis. US public debt to GDP will increase by at least 50% in this recession, even if the direct cost of support ends up as zero. UK public debt similarly will increase by at least 50% of GDP, even though the direct costs may not exceed 5%.

And that illustrates that to think radically we must focus on the total system, not on individual institutions, and on the drivers and costs of aggregate instability, not solely on the causes and costs of individual institution failure. And the core of the problem we face is instability in the supply and demand for credit – first too exuberantly supplied in the upswing, then suddenly curtailed in crisis, inducing a credit crunch which, in turn, becomes a self-reinforcing recession. (Slide 2)

Credit cycles are subject to inherent self-reinforcing dynamics – (Slide 3) – with easy credit driving asset price increases, which confirm in the minds of both lenders and borrowers the wisdom of yet further credit extension, until the cycle breaks and moves into a harmful reverse. A touchstone of radicalism in financial stability reform should therefore be whether we have taken action which will reduce the severity of these cycles.

The core of the global regulatory response has been the new Basel III rules on capital and liquidity. On the capital side, we have increased the required ratio of equity capital to risk weighted assets (RWA) from 2% to effectively 7% – quite an increase. And, in fact, we have been more radical than even those figures suggests, because we have tightened the definition of what counts as equity; and we have tightened some definitions of risk weights, particularly in respect to bank trading activities. We have changed the numerator, ratio and denominator; and the combined effect is large.

Ideal capital ratios well above Basel III levels

But in an ideal world free from transitional concerns, optimal equity capital ratios would almost certainly be much higher still.

The case for believing that has been made by several economists: articles by David Miles (2010) of the Bank of England Monetary Policy Committee (MPC), and by Martin Hellwig (2010) and others are particularly valuable statements of the case. Crucial to their arguments are the need to distinguish between social and private costs, to focus on small probability but potentially very harmful events, and to focus on the total system rather than on specific institutions.

Those three considerations together suggest that the costs of higher bank equity requirements are lower and the economic benefits considerably higher than many participants in this debate – from the banking industry but also from some public authorities – have in the past assumed.

On the cost side, the insights of the Modigliani and Miller theorem (1958) are vital. If we increase the amount of equity which banks are required to hold, there must in the long run be some offsetting decrease in the cost of equity, the return which investors demand on a now lower risk investment. David Miles finds empirical evidence of this effect over the period 1992 to 2010: as bank leverage increased, bank equity betas and thus the cost of equity increased also. Credit Suisse, meanwhile, faced with the higher equity ratios which the Swiss authorities are imposing, has cut its return on equity target from 18% to 15%, signalling to investors that in future investment in bank equity should be seen as both lower risk and lower return than in the pre-crisis period.

And while at the private level this lower pre-tax cost of equity is offset by the fact that returns on equity capital are not tax-deductable, while returns on debt are, it is vital to grasp that that tax treatment creates a private cost of higher equity capital but not a social cost. If the tax-take from higher equity capital increases, governments will as a result receive larger tax revenues which they could, if they wished, use to offset any adverse effects arising from the higher cost of private credit intermediation.<sup>1</sup> That simple point is all too often ignored. It is indeed striking that the public debate on optimum equity ratios has, even among some public authorities, been shot through with a profound confusion between the quite distinct considerations of private cost and social optimality. There is no general social interest in ‘economising in the use of equity capital’ (i.e. having higher leverage), but that was an aim overtly discussed in the design of the Basel II regime.

In assessing the benefits of higher equity ratios, meanwhile, three insights are essential.

- The first is the importance of considering the impact of low probability but extreme events: David Miles, argues persuasively that the adverse costs of even very rare banking crises are so great as to outweigh any marginal growth penalty resulting from higher equity ratios.

---

<sup>1</sup> Whether or not such adverse effects actually exist is however itself highly contentious and depends on issues relating to the functions which credit performs within the economy. (See Adair Turner, *What do banks do, what should they do?*, lecture at Cass Business School, March 2010).

- The second, is the fact that the losses which the banking system stress tests may face are not simply exogenous shocks, but deeply endogenous to the system itself, and dependent on the subtle interplay of confidence and contagion. If confidence is lost, the system will suffer liquidity crunches resulting in a reduced credit supply which in itself creates credit losses.<sup>2</sup>
- Thirdly and crucially, it is vital to recognise that an undercapitalised banking system could impose a macro-volatility penalty even if no bank failures ever occur. If banks are so lightly capitalised that in the face of losses they constrain lending sharply, that can itself induce harmful macro-volatility, even if there is no danger of actual insolvency. Private sector arguments that our Basel III capital requirements are too high because they are in excess of any reasonable estimate of the losses which a bank might face, therefore simply miss the point. What matters is the macro-systemic stability of credit supply, not just the risk of individual failure.

These theoretical arguments are, I believe, compelling. They suggest that if global regulators were benevolent dictators designing regulations for a banking system to service a greenfield market economy, they would be wise to choose capital ratios far above even Basel III levels, something more like the 15% to 20% of risk-weighted assets which David Miles illustrates in his recent paper. And the empirical evidence is as compelling as the theoretical. Before the last 40 years or so, banking systems ran with much higher equity capital ratios, much lower leverage and yet economic growth was as high as it is today and investment as a percentage of GDP as high if not higher. (Slide 4)

There is simply no good theoretical argument or empirical evidence that we need to run banking systems with anything like as highly leverage as over recent decades. And today's regulators are, in a sense, the inheritors of a half-century long policy error, in which we have allowed private sector banks to pursue their private interest in maximising bank leverage, at times influenced by a deep intellectual confusion between private cost and social optimality.<sup>3</sup>

In an ideal world where we could choose to 'not start from here' much higher equity ratios would be optimal. But we are 'starting from here', and optimal policy needs therefore to take account not simply of the ideal end point, but of the transition. And while much higher equity ratios will not in the long run carry an economic penalty, the transition from a starting point of sub-optimally high leverage could, unless we manage it carefully, slow recovery from the crisis induced recession.<sup>4</sup> The analysis

---

<sup>2</sup> It is for instance notable that several countries where confidence in the banking system was lost or came under challenge during the financial crisis, have already chosen to introduce capital standards which go beyond Basel III standards. These include Switzerland, Spain and Ireland.

<sup>3</sup> This private benefit derives both from the tax-deductibility of debt, and from the put option of limited liability.

<sup>4</sup> There are two key ways in which the transitional economical consequences could be different from those which would apply in an ideal, comparative statics, analysis.

- The first relates to the cost of equity. The theory of Modigliani and Miller, supported by David Miles' empirical analysis, suggests that as equity ratios are increased the cost of equity will fall. The private sector argues, however, that if banks are forced over a short time period to raise additional equity, the cost will increase, given an upward sloping supply curve for new

by the BIS Macroeconomic Assessment Group (MAG) of the transition dynamics was therefore an essential input to the decisions of the Basel Committee and Financial Stability Board in the design of Basel III. And Basel III is, I believe, best understood as an extremely valuable step towards greater financial stability in a world where today's policymakers must operate within the context they inherit. Optimal policy, like economic structure, is to a degree path-dependent.

But it is still important to know that ideal equity ratios would be considerably higher than those which we agreed in Basel III, and that the system will therefore remain more vulnerable to instability than is ideal. Other aspects of policy are therefore extremely important. And, in particular, the policies for systemically important financial institutions (SIFIs as they have become known), i.e, with the set of banks which in Autumn 2008 were treated as 'too big to fail'.

### Fixing 'too big to fail': best and second best solutions

The clearly important principle is that all banks, however large, should be able to 'fail'. But a wider definition of the objective is that our largest banks, because they play such significant roles in the economy – in particular through their role within credit extension – should be so regulated as to ensure that their failure or near failure produces **neither** the need for public direct support **nor** a harmful disruption to their own or to the wider system's credit extension capability.

To achieve that objective, it is essential that private fund providers to banks (whether equity or debt) can absorb losses without this triggering either the fire sale losses which can accompany standard insolvency procedures, or knock-on effects among the fund providers (depositors, institutional investors, or indeed other banks) which could, in turn, lead to a reduction in credit supply.

These conditions were not met before the crisis. In Autumn 2008, faced with the potential failure of several large banks in the UK and the US and elsewhere, we were not confident that we had mechanisms to impose losses on deposit holders, senior debt holders or even subordinated debt holders, without that producing large macro-disruptive effects, and we therefore supported those large banks via capital injections, imposing dilution losses on existing equity holders, but no losses on any other category of fund providers.

---

bank equity capital. Both assertions could be correct: one over the long term, the other over the short term. The appropriate policy response is a transition path to higher capital requirements which enables most banks to meet them out of retained profit.

- The second relates to levels of leverage in the real economy. There is a reasonable case that, in some countries and sectors, real economy leverage had become sub-optimally high in the pre-crisis period, with credit intermediation costs sub-optimally low. In these cases, an increase in the cost of credit intermediation and a reduction in the credit supply could in the long run be positive. Even in such circumstances, however, the process of deleveraging from a sub-optimally high level can create deflationary economic effects (for the original and classic description of these see Irving Fisher, *The debt deflating theory of great depression*, *Econometrica*, 1933).

In future we must be able to impose losses on debt providers as well as equity investors – both in order to avoid taxpayer costs and in order to re-introduce *ex-ante* market discipline – with fund providers aware in advance that they could suffer loss.

To achieve those objectives, two ways forward are possible. (Slide 5)

- One is to require SIFIs to have 'higher loss absorbency', either more equity capital or more debt capital, with clear mechanisms specified for converting that debt capital to equity if needed to absorb losses and to maintain adequate equity capitalisation.
- The other is to create 'resolvability', by which we mean mechanisms which enable the authorities to impose losses on debt providers, through either write down or conversion to equity, in such a way that the bank can be recapitalised and maintain its operations without disruption and without public support.

Both policy options should be pursued: they are complements, not alternatives. Indeed at a conceptual level they are not distinct but simply different ways of achieving smooth loss absorbency. That becomes clear when we think about the potential role for what has been called 'bail-inable' debt (Slide 6), senior debt which can be written down or converted to equity at the point when the bank would otherwise fail, with this convertibility achieved either via contractual provisions or through statutory requirements. We can think of such bail-inable debt as being either a form of 'higher loss absorbency', a reserve army of potential capital, or as a means of ensuring that firms can be resolved without the complexities and disruptions which might arise when losses are imposed on ten or hundreds of thousands of depositors, rather than hundreds of senior debt securities. Either way, bail-inable debt could be very valuable regulatory tool.

But it can never be as certain a way of increasing the stability of the system as more equity, for reasons which take us back to the centrality of macro-system concerns.

Much of the debate about contractually bail-inable debt and resolvability focuses on the legal mechanisms: do we have the legal processes to resolve banks? Do we know enough in advance about internal organisation and legal bank structures and inter-dependencies to make rapid resolution possible? And can we overcome the complexities of multiple legal systems, property right conventions and insolvency regimes which would at present make smooth resolution of a large cross-border bank extremely difficult?

But let us assume for now that we can and will fix all of these problems. Would we then have fixed the problem of 'too big to fail'?

- Clearly we would, if we are talking about the idiosyncratic failure of a large bank – a failure like Continental Illinois in 1984, or Barings in 1995, unrelated to a wider collapse in asset values and confidence. And that in itself would be very valuable, helping to re-introduce some *ex-ante* market discipline.

- But the problem is that large systemically important banks are most likely to fail amid general systemic stress, when many other banks, big or small, are also under stress, and when the failure of any one might radically increase the stress faced by the others.

It is therefore crucial that our answers to the SIFI problem cover also the more difficult but more likely scenario of multiple bank systemic stress. And in such conditions, bail-inable bonds will only enable us to avoid the dilemma of Autumn 2008, if the following vital conditions are met:

- If regulators could be confident that those bonds are held outside the banking system; and
- in addition, confident that the bonds are held by investors who have so arranged their assets and liabilities that they could face the imposed losses without that in turn inducing systemic effects.

And it may be very difficult to be confident that those conditions we met.

There are two ways to gain that confidence – the first relies on empirical observation, the second on an assumption of fully informed investor rationality. Neither route may be entirely robust.

- The first way to seek such confidence, would be for regulators to understand, or to regulate, which investors hold bank medium-term debt. Our information on this today is imperfect. We believe a significant proportion is initially held by other banks, and a larger proportion still by a broadly defined group of ‘fund managers’. (Slide 7). But ownership after secondary market trading could be significantly different. And some of these ‘fund managers’ may be in turn financed by banks (e.g. hedge funds by prime brokers), or linked to the banking system by complex repo and derivative relationship so that losses suffered by one bank, could indirectly impose losses or confidence shocks on others. And our ability to track these complex inter-connections, and as a result to predict the knock-on consequences of initial losses in conditions of systemic fragility is imperfect today and likely to remain so. We need to improve our understanding of the complex interconnections of our financial system: but it is unclear that understanding will ever be good enough for us confidently to impose large losses simultaneously on the senior debt of multiple large banks (or indeed multiple small banks), in conditions of macro-systemic stress.
- The other route to confidence, would be based on faith in market and investor rationality, assuming axiomatically that investors who buy bail-inable bonds will only do so on the basis of rational assessments of their ability to absorb risks in all possible future states of the world, including those of macroeconomic stress. As Section 3 will discuss, this axiomatic assumption was at the core of the pre-crisis conventional wisdom, the reason why public authorities thought they could sleep easy in the face of an explosive growth in financial scale, complexity and interconnectedness. But it relies on an assumption of fully informed rationality, which may be simply untrue, and

indeed impossible. For as Andrei Shleifer et al (2010) have argued in an extremely perceptive recent paper, it may be inherent to human nature that in the good times investors systematically fail to take rational account of the tail of low probability adverse events.

A bail-inable bond will have a highly skewed probability distribution of pay-outs. (Slide 8) Over a long period of time, only the zero-loss segment of the distribution will be observed. A low probability of significant loss continues to exist, but Gennaioli, Shleifer and Vishny argue that that low probability will be wholly discounted through a behavioural process which they label 'local thinking' – the reality, deeply rooted in human nature, that not all contingencies are represented in decision makers' thought processes. After a period of good times, investors will assume that senior bank debt is effectively risk-free: as indeed they did, in the years before the crisis (Slide 9). Regulators cannot therefore rely on free-market discipline to ensure that the debt is only held by investors who can suffer loss without that causing knock-on systemic disruption.

If therefore we can neither perfectly and continuously monitor or regulate who owns bail-inable debt, **nor** rely on free-market discipline to ensure that it is always appropriately held, contractually bail-inable debt and technical resolvability will be valuable but still imperfect solutions to the 'too big to fail' problem. We can only be sure that losses can be smoothly absorbed if we are sure that the investors who provide funds do not suffer from 'local thinking' but remain perpetually aware of the full distribution of possible results. Subordinated debt which can convert to equity well before potential failure ('early trigger CoCos') may approach what is required since the price will presumably vary with probabilistic expectations of future conversion. But only with pure equity can we be fully confident that the dangers of 'local thinking' will not creep in over time, and that investors, facing day-by-day price movements up and down will remain continually aware that they hold a potentially loss absorbing instrument. The implication of Shleifer's 'local thinking' theory is that if investors are to remain continuously aware of the full frequency distribution of objectively possible results the **observed** frequency distribution of returns needs to include negatives and well as positives. This is achieved by equity returns but not by low risk debt.

While therefore the debate about 'too big to fail' banks often seems to imply that we will only have been truly radical when we have seen a major bank fail with losses imposed on debt holders, the truly radical and ideal solution remains one in which there is enough equity or close to equity instruments in the funding mix as to reduce to a minutely low level the probability of us ever having to impose losses on the debt holders of large banks.

So I think we need to be clear: in an ideal world we would increase equity requirements for all banks well above Basel III levels. And to make large systemically important banks safer in a world where overall equity standards are suboptimal, resolvability and/or bail-inable debt are valuable tools, but more equity (or close to equity instruments such as early trigger CoCos) is the best solution.

We need to allow for non-rational decision making, for myopia, as well as fix bad incentives. And we need to keep our focus on systemic stability and macroeconomic

consequences, not just on whether we can avoid taxpayer costs and smoothly resolve specific institutions. That focus implies that equity ratios are of central importance.

### Markets as important as institutions: shadow banking as important as banks

But it also implies that we must not focus exclusively on specific institutions, such as banks, but on total financial systems and markets. In the two years since the crisis, global regulators have focused primarily on the capital and liquidity regimes for banks, both in general (Basel III) and for big banks in particular (the SIFI agenda). But what is striking when one looks back at the events of the initial year of the crisis, 2007 to 2008, is that it did not look at all like a familiar banking crisis, but something entirely new, a crisis of 'shadow banking'.

Among the key events which marked the development of the crisis were:

- In June 2007, liquidity pressures at two hedge funds sponsored by Bear Stearns Asset Management led to the imposition of gates on investor redemptions, sudden increases in margin calls, and sudden drops in asset prices.
- In August, major losses were incurred by hedge funds which the market had thought were following low-risk market-neutral strategies, as a result of knock-on consequences from margin calls in structured credit portfolios.
- The closure in February 2008 of hedge funds Carlyle Capital and Peloton in the face of additional collateral calls on mortgage backed securities.
- Gradually growing problems throughout 2007 to 2008 in the liquidity and solvency position of off-balance sheets structured investment vehicles (SIVs) and conduits which had taken leveraged positions in structured credit products, and had funded those with liabilities far shorter than the contractual maturity of the assets, many of these liabilities (ABCP) bought in turn by money market mutual funds.
- The rescue of Bear Stearns in March 2008 and the failure of Lehman Brothers in September, the latter the key trigger for the dramatic intensification of the crisis. Both of them broker dealers/investment banks rather than commercial banks.
- The emergence in Summer 2008 of major stresses among money market mutual funds, which had previously seemed to promise investors an attractive combination of enhanced return, immediate fund access, and capital certainty, and with Reserve Primary Fund 'breaking the buck', on 16 September 2008.
- The development between August to October 2008, of a new form of liquidity run: a run as much in the secured lending markets (such as repo) as in unsecured funding.

- And, throughout late Autumn 2008, significant deleveraging by hedge funds, whose sales of credit securities into a falling market, helped drive the downward spiral of trading book asset values, which in turn undermined confidence in the solvency of major banks.

This therefore seemed at the time a new form of financial crisis, different in nature from some of the classic bank failures and bank runs of the past. And it occurred within a financial system which in the 20 years before the crisis, had, at least in the US, seen dramatic growth in a complex system of non-bank credit intermediation (Slide 10).<sup>5</sup>

This entailed:

- Money Market Mutual Funds (MMMFs) growing at the expense of bank deposits;
- credit increasingly extended via the purchase of credit securities rather than through traditional bank loans;
- an explosive growth in complexity, with derivatives and securities tranching giving us the alphabet soup of CDS and CDO and CDO squared; and
- with the most rapidly growing institutions being not traditional commercial banks, but broker dealers, i.e, the investment banks (Slide 11).

Given this history, it might indeed seem odd that so much of the regulatory reform focus over the last two years has been on the capital and liquidity position of commercial banks. In fact, that focus can be justified: banks, as leveraged, maturity transforming and credit providing institutions, play an absolutely central role in the system, and it was when the crisis spread from 'shadow banking' to the core banking system in Autumn 2008 that it threatened major macroeconomic harm. But we certainly need to understand the nature and consequences of these shadow bank developments, and identify the fundamental drivers which could in future lead to the re-emergence of 'shadow banking' or the risks it created, in new forms.

Shadow banking can be understood as a new variant, and both a turbo charged variant, of non-bank credit intervention. (Slide 20). Credit can flow from ultimate fund providers (households or corporates) to ultimate users (households, corporates or governance) through the banking system, or through non-banking routes. Both routes have always existed, with for instance, the individual able to buy a corporate or government bond, directly or via a insurance company or pension fund intermediation, as well as able to deposit money at a bank which in turn lends it on. But securitisation and shadow banking dramatically changed the scale of such non-bank credit intermediation, and changed its nature and its riskiness in two crucial ways.

---

<sup>5</sup> While this growth was most evident in the US, with bank credit intermediation remaining more dominant in other countries, the banking systems of other countries were deeply interconnected with the US shadow banking system, whether via trading operations based in the US, as purchasers of credit securities, as counterparties in derivative and repo markets, and because US shadow banking institutions such as money market funds were (and still are) key providers of short-term funding.

- It introduced leverage and maturity transformation, the classic functions of banks, into the non-bank intermediation channel, with MMMFs performing large scale maturity transformation, with hedge funds employing leverage, and with SIVs both leveraged and maturity transforming.
- And while it was in part a parallel system of credit, it was also deeply entwined with the classic banking system – with MMMFs providing funds into the banking system and with the banking system via repo and other secured finance markets providing funds to SIVs, conduits, hedge funds and other investors.

But it is important for us to understand not just what occurred but why. And two recent papers argue persuasively that among the most fundamental drivers was investor demand for very low-risk debt instruments; a demand, however, which exceeded the quantity of truly low-risk instruments which could objectively exist.

Thus:

- Gorton and Metrick's analysis (2010) focuses on the dramatic development of the repo market from 1990 to 2010, with huge increases in total values transacted, but also a huge widening of the classes of collateral used in repo transactions. They see these markets as essentially deriving from a demand for 'money equivalent' assets – immediately liquid, and of certain or close to certain capital value, but delivering attractive returns above those available on pure risk-free T bills.
- Gennaioli, Shleifer and Vishny's analysis meanwhile, illustrates how the 'tranching' of credit securities met investor demands for an apparently low or zero risk debt instrument (the triple A tranches of structured credit products) delivering attractive yield uplift versus pure risk free Treasury bonds.

These developments were, however, fundamentally unstable, because they were based on assumptions about the available risk and return which were, at the macro-systemic level, objectively impossible. Beyond the financial system there exists a real economy with corporates, households and governments whose cash flows available to service debt are subject to a combination of both objectively modelable risk and inherent irreducible uncertainty. The financial system can divide, repackage, and distribute those risks, but only to a limited extent can it reduce them. But the complexity of a large financial system, combined with 'local thinking' of the sort which Shleifer et al have described, can result in assessments of risks which are, in aggregate, impossible given the objective reality of the non-financial real economy.

Subject to 'local thinking', investors in the good times assume that objectively risky instruments are close to risk free but then rapidly, and in a herd, revise their estimates of riskiness when the first evidence of arising defaults emerges, bringing back into their consciousness, as it were, the previously ignored downward tail of the distribution.

Such a description of investor decision making clearly breaks with the assumptions of rational expectations which have dominated much of economics for the last several decades: and I return in Section 3 to the importance of challenging that assumption. Its implications for appropriate regulatory reform are profound, as Gennaioli, Shleifer and Vishny themselves highlight. It means that many credit securities *'owe their very existence to neglected risk'*, and thus that the total amount of credit extended to the real economy could be larger than optimal. It implies that *'it is not just leverage but the scale of new claims itself, which might require regulatory attention'*, and that *'recent policy proposals, while desirable in terms of their intent to limit leverage and fire sales, do not go far enough'*.

Risks in the financial system thus derive not only from the fact that the incentives faced by individual institutions might encourage them to act in socially harmful ways (e.g. by increasing leverage above socially optimal levels) but also from inherently imperfect investor evaluations of risk. And that in turn implies that:

- aggregate levels of debt and leverage in the real economy, and trends in those levels, are key determinants of financial stability;<sup>6</sup>
- risks can exist in interconnected markets as much as in specific institutions; and
- that those risks could exist even if we were able to resolve all banks, or even indeed if we broke up large banks into smaller ones. A system of multiple interconnected players could be as risky as one with large specific institutions.

Given that in addition the very process of regulating banks more effectively will create incentives for activities and risks to move to the non bank arena, this implies that looking forward we will need both to monitor and perhaps regulate overall developments in markets and credit supply, as well as the activities of specific institutions.

- In particular we may need to regulate the level of collateral haircuts/margins in the repo and other secured finance markets. Which is to say, to regulate leverage at the contract specific level within the market, rather than at the institutional bank level.<sup>7</sup> The central question again being how much equity and how much leverage, but now within markets rather than within specific institutions.
- And we will need to monitor closely overall developments in credit volumes and credit product innovations whether or not financed directly by banks.

### Complexity and interconnectedness as drivers of risk

---

<sup>6</sup> Luigi Einaudi, 'Debts' (1934) for a perceptive discussion of the rigidities which debt rather than equity constraints introduce into an economy, building on Irving Fisher, *The Debt Deflation Theory of Great Depressions*, 1933.

<sup>7</sup> CGFS paper No 36, *The Role of Margin Requirements and Haircuts in Procyclicality* (March 2010), discuss the possible case for such requirements.

Slide 10 showed the rapid increase in the US over the last 30 years of the shadow banking system. Equally striking is the growth in the size of financial sector balance sheets above and beyond the increase which can be explained by growing real economy leverage. Slide 15 shows debt as a percentage of GDP in the US: for the corporate sector this increased slightly: for the household sector very significantly; but what is most striking is the increase in debt claims by the financial sector on the financial sector, i.e. claims internal to the financial sector itself.

This growth of intra-financial system assets and liabilities is partially related to the growth of shadow banking, which, as per Slide 14, has produced a proliferation of claims between MMMFs and banks, between banks and off-balance sheet vehicles, banks and hedge funds. The multi-step credit intermediation of shadow banking creates a higher gross value of claims than the single step process of bank intermediation. But there has also been rapid growth of interbank claims, both via repo and derivative markets, and via interbank markets in unsecured deposits.

The overall effect of these developments is that major bank balance sheets are now dominated by claims against other banks and other financial institutions.

This development is in part a by-product of the development of securitised lending. But it also reflects a crucial shift in philosophy towards the role of bank treasury departments and interbank markets. Whereas 40 years ago, bank treasuries were seen as service functions and interbank markets as the means by which banks with surplus funds placed then at banks in deficit, increasingly treasury and trading activities were organised as business units, charged with making a profit through effective 'position taking'.

The resulting explosion of trading and position taking activity raises important questions relating to the social value added and the distributional consequences of increased financial intensity. Except to the extent that increased position taking delivers allocative efficiency benefits for the whole economy, it must by definition be zero sum in total, but could be rent extracting for specific successful players.

But whatever the answers to those questions (which are considered in Section 2), the increased complexity and interconnectedness of the financial system has created increased financial stability risks. In Autumn 2008, it was sudden awareness of downside risks (the sudden end of the risk neglect described by Shleifer et al) combined with market awareness that banks and shadow banks were linked in a complex web of imperfectly understood interconnections, which drove the rapid drying up of liquidity and self-reinforcing asset price declines.

There is therefore a strong case that regulators should focus on the interconnectedness of the financial system as a crucial issue in itself, separate from and in addition to the soundness and resolvability of individual institutions. And a strong case also that policy tools (such as counterparty capital requirements or tax) should be used to lean against the proliferation of complex interconnectedness and the externality it creates. How far it should do so, however, should ideally reflect an evaluation of whether this interconnectedness and complexity does indeed deliver a positive economic value added.

## Macro-prudential through-the-cycle policies

The fundamental macroeconomic problem is instability in the supply of credit, first too exuberantly supplied, then restricted. Shadow banking illustrates the need continually to analyse the totality of the financial system, monitoring credit extension, maturity transformation and leverage whether arising outside banks or within. And interconnectedness between banks itself creates risk, which need to be subject to overt focus and possibly to policy interventions. Together these findings illustrate the central importance of what has become known as a 'macro-prudential' approach.

Within the new UK regulatory structure, the responsibility for taking that macro prudential approach will reside with the Financial Policy Committee. Its remit will include keeping the entire system under review, and proposing policy responses. These could include responses which change the scope of regulation (e.g. to cover shadow banking institutions and markets) or which address complex risks arising from interconnectedness between firms rather than solely within firms. But they will also need to include the possibility of discretionary policies, varying through the cycle, which could lean against self-reinforcing cycles of credit extension and related asset prices. Such policies could involve the application of counter-cyclical capital requirements on an across the board or sectoral specific basis, and/or the application of constraints on borrowers, though for instance minimum loan-to-value ratios.

The conventional wisdom of the pre-crisis years was that financial stability was assured by monetary stability together with the non-discretionary application of an unchanging set of prudential rules. The Bank of England focused on using the interest rate instrument to achieve low inflation; the FSA focused on the regulation of individual institutions. It is clear now that this left a harmful underlap in macro-prudential analysis and policy. The creation of the Financial Policy Committee is therefore, along with Basel III capital and liquidity changes, a vitally important element of the policy response.

## Radical thinking on financial stability

So, have we been radical enough in our actions to ensure financial stability, to reduce the probability and severity of financial crises?

The answer is that we have done a lot, but that we still have more to do.

- Basel III is a major step forward, but in an ideal world, equity ratios would be set much higher.
- The SIFI agenda is important: we need to ensure resolvability; but the best way forward would also involve higher equity ratios (or close to equity instruments, such as early trigger CoCos).
- We must ensure we understand the drivers of shadow banking and guard against the re-emergence of new risks in new financial mutations.

- Complexity and interconnectedness are important in themselves, and specific regulatory action to offset the externalities created may be required.
- And we need to make a reality of macro-prudential oversight and policy response.

But as important as checking off the things that we still need to do, is the need to keep thinking deeply about the fundamental drivers of financial instability. Some commentators see the fundamental problem as taxpayer support for ‘too big to fail’ banks; and the key driver of instability as the lack of market discipline faced by bank executive and traders who knew in advance that they would be rescued. So, it seems, if we create adequate mechanisms to resolve big banks, the key problem will be overcome and the key driver of instability removed.

- But the implications of Shleifer’s analysis is that market discipline via debt instruments is very difficult to achieve. ‘Local thinking’ means that investors may over time come to assume that their debt holdings are close to risk free. And local thinking affects bank executives as much as bank investors. It is not clear that excessive risk-taking by banks in the years before the crisis was actually driven by an explicit or implicit assumption that they were ‘too big to fail’ and therefore did not need to worry about the risks they were taking. Rather it was driven by an irrational exuberance in which they truly but erroneously imagined that the risks they were taking were small, well understood and well contained. Incentives matter but so too does myopia.
- And the implication of the shadow banking story, of Gorton and Metrick’s analysis as well as Shleifer’s, and of the fundamental nature of credit and asset price cycles, is that we face deep and complex drivers of instability, which could exist even in a world where we had the operational capability to resolve all banks, and even in a world of multiple small banks rather than large banks. As a result, no single unchanging set of rules is likely to be sufficient to ensure stability: discretionary tools to lean against emerging excess credit extension will also need to be part of the policy armoury.

Fixing the ‘too big to fail’ problem by making it possible to resolve banks smoothly is an absolutely essential element of our policies to improve financial stability. But if we think it is a sufficient response, we will be failing to be adequately radical either in our analysis or in our policy prescription.

## **2. Financial intensity, inequality, and social value**

Technicians are necessarily focused on financial stability – on reducing the potential of the financial system, whether banks or shadow banks – to be autonomous drivers of instability in the economy.

But much popular focus is on bonuses, and on the perception that bankers who played a major role in creating the problem were hugely highly paid, and still are, for performing activities whose value is to many ordinary people unclear.

A key popular issue therefore, but also I will suggest one which carries some implications for the appropriate balance of financial stability regulation, is whether high compensation in financial markets results from activities which, directly or indirectly, deliver economic value added or whether it has been swollen by rent extraction, i.e. by activities which essentially transfer income from the rest of economy to the financial sector.

The background of these issues is the quite startling increase which has occurred over the last 30 years in the relative scale of financial activity within developed economies, an increase which has been among the factors driving increased inequality of income at the top end of the income distribution.

A wide range of indicators illustrate the increasing 'financialisation' of our economy, an increase in the scale of financial activity relative to real GDP. (Slide 16).

- Trading activity has massively increased. Daily FX trading volumes have increased 234 times between 1977 and 2010, a period during which nominal GDP increased seven times. Daily oil futures trading has gone from a one time multiple of underlying physical flows in 1980 to ten times today.
- Innovation has created whole new investment categories, particularly in securitised credit, with volumes growing from zero in 1980 to many trillions today.
- Financial sector balance sheets as already shown, have increased dramatically relative to real GDP, the growth of intra-financial system claims greatly exceeding even the rapid growth of credit extended by the financial system to the real economy.
- And while it would have been theoretically possible for this growth to occur without financial sector compensation and profits growing as percent of GDP, in fact there has been a dramatic increase in that percentage, the latter half of the 20<sup>th</sup> Century seeing the US financial sector regaining the relative weight it enjoyed in the 1920s and then going on to still higher levels (Slide 17) and UK financial firms accounting for a dramatically increased share of total private corporation gross operating surplus (Slide 18).

Alongside this growth in the financial sector, income inequality has increased significantly in rich developed societies. This has involved the lowest income deciles falling further behind the median. But also a quite startling increase in inequality at the top end of the income distribution – the top decile pulling away from the median, the top percentile from the top decile, and the top tenth of the top percentile from the not-quite-so-rich rest of the top 1%.

And one driver, though not the only driver of that increased inequality at the top end appears to be the growth of the financial sector and the often huge earnings made in it.<sup>8</sup>

---

<sup>8</sup> For other drivers of inequality, see Adair Turner, Lionel Robbins Memorial Lectures 2010, Lecture 1.

Philippon and Reshef (2009) for instance illustrate (Slide 19) how financial pay relative to average pay for similar levels of skill soared in the period of 1920s financial intensity, fell away entirely in the period of 'financial repression' 1935-75, but then grew again (and indeed exceeded 1920s levels) from 1980 onwards.

That increase in 'excess wage', as Philippon and Reshef see it, reflects moreover, not just a small number of bank CEOs and other senior executives but a pervasive tendency for key employees at many levels and in many financial industry sub-sectors to be very highly paid. Asset management has driven high remuneration quite as much as banking: alternative asset management (i.e. hedge funds) has seen the highest remuneration of all: and within banks and investment banks, star traders are often more highly paid than senior executives. As a result, finance has become the first choice career for talented young people wishing to make serious money not at the end of a sustained entrepreneurial or business management career, but early in life.

This phenomenon has been apparent for many years and a focus of increasing academic attention. But popular awareness and resentment only became acute in the wake of the financial crisis, and of the taxpayer-financed bank rescues. Some financial activities it became clear, far from adding value in some complex though difficult to understand fashion, in fact created financial instability and produced economic harm. And these activities have ended up being subsidised by taxpayers: the gains were privatised, the losses socialised. The enduring resentment against higher bankers' bonuses is therefore easy to understand.

But to ensure that we properly understand the role which finance has played in increasing inequality, and what if anything we should and can do about it, we need to place a focus on 'bankers' bonuses' within a wider analysis, posing four questions.

- Whether and how far the huge expansion of financial activity has indeed delivered economic value: whether some of it instead reflected rent extraction.
- Whether rent extraction opportunities derive primarily from the ability to privatise gains but socialise losses in 'too big to fail' banks, or whether from a wider and deeper set of factors.
- Whether those measures we are taking for financial stability reasons, both Basel III and the 'too big to fail' agenda, will as a by-product address any legitimate issues of excessive financial activity and resulting income inequality.
- And whether more generally alternative possible conclusions on the social value added of financial activity carry any implications for the financial stability agenda.

The answers to these questions are not clear and they are worthy of far greater attention from economists than they have so far received.

But let me suggest some reasons for believing that the factors at work here are deeper than those implied by a sole focus on bankers' bonuses in rescued banks.

In some ways, it is likely that our financial stability reforms will impact the scale of financial activity and, as a result, remuneration in some segments of finance. Higher capital requirements overall, and in particular higher capital against trading activities, where the pre-crisis capital regime was most woefully inadequate, may result in less trading activity within banks, and as a result in fewer highly paid traders.<sup>9</sup> But expectations of a significant impact on aggregate remuneration from other aspects of the reform agenda may turn out to be exaggerated. In particular:

- Regulation of the structure of bankers' pay, while an important element within our financial stability agenda, is unlikely in itself significantly to reduce total remuneration and might even under some circumstances increase it. Poor remuneration structures before the crisis certainly played a role in creating incentives for excessive risk taking: traders were often rewarded with large cash bonuses on the basis of one year mark-to-market profits, long before it was clear whether the profits had been generated at the expense of a trail of toxic losses in later years. New regulations which require a significant proportion of bonuses to be deferred, to be paid in non-cash form, and to be subject to claw back in the light of subsequent profit and loss, will play a useful role in reducing incentives for excessive risk taking.

But it would be possible for the market reaction to these structural changes, while compatible with reduced financial instability, to result in **increased** total remuneration. Reductions in bonuses have already been offset by higher salaries: and the total size of bonuses deferred might be higher than those which were previously paid up front.<sup>10</sup>

The crucial determinant of whether there are in Philippon and Reshef terms 'excess wages' will therefore be not the structure of remuneration, but whether across the financial system, within banks and beyond them, pre-compensation profitability is excessive relative to economic value added. If it is, either shareholders or employees will enjoy the benefit: and if the skills which determine the relative allocation of this rent between different firms are at all scarce, a high proportion of the benefit will accrue to skilled employees.

If economic activities earn super normal pre-compensation returns, attempting to stop those returns reaching employees by regulating pay structures or levels of pay, will be as ineffective as attempting to contain inflation via prices and incomes policies in conditions where nominal demand is increasing at an excessive rate.<sup>11</sup>

---

<sup>9</sup> The reform of the trading book capital regime is a key agenda item for the Basel Committee and the Financial Stability Board during 2011.

<sup>10</sup> This would occur for instance if the total pay of affected staff were determined by market competition by other sectors of the financial industry (e.g. hedge funds) so that the market required present value of total remuneration is unchanged. In this case the gross value of deferred bonuses might rise to reflect decreased certainty of receipt.

<sup>11</sup> See Philip Augar, *Britain's Bank Bonus Embarrassment*, Financial Times, 12 January 2011 for a clear statement of this argument.

- And while ensuring that all banks are resolvable is, as Section 1 argued, vitally important, we should be wary of assuming that this will in itself radically transform risk taking behaviours. Post facto, some traders in rescued banks were effectively ‘gambling with other people’s money’, but it is not at all clear that they knew that at the time, nor that ensuring that they have more personal ‘skin in the game’ can significantly change risk perceptions and risk taking behaviours. Dick Fuld, the CEO of Lehman Brothers, had huge equity at stake in that firm, but still allowed massive risks to develop. The tendency to ignore the down-tail of the distribution (Shleifer’s ‘local thinking’) is inherent to human beings, and affected traders and bankers as much as investors and users of financial products. Myopia is important as well as poor incentives.
- Constraints on pay within banks, moreover, even if effective, will not in themselves constrain the level of pay in other segments of the financial system, including in the shadow banking system, to which financial activities currently performed by banks could in future migrate.

Overall, therefore, while measures to make the banking and financial systems safer may have some impact on the ‘excess wage’ phenomenon illustrated by Philippon and Reshef, it seems unlikely that they will be transformational.

The central issue in any fundamental analysis of inequality and financialisation therefore remains whether increased financialisation has delivered commensurate added value. If it has, the increase in inequality has at least been accompanied by an increase in the size of the economic cake. If it has not, then we need to recognise that fact, even if there is no straightforward policy response.

So has increased financial intensity delivered economic value added?

To a true believer in the efficiency of free financial (and other) markets, the answer is, axiomatically, yes it must have. If the measured contribution of financial services has increased as in Slide 17, that must be because financial services have been bought in a market by rational customers who value them in line with their market prices. If shareholders receive increased profits and employees increased earnings, that will, by national income accounting rules, count as ‘value added’ within measures of GDP: but in perfectly efficient markets it must also be economic value added, accounting reflecting economic truth.

And there are indeed some reasons to expect the relative role of financial services to increase as societies get richer. As households get richer, they tend to demand more lifestyle consumption smoothing products (saving at one point in their life, borrowing at others): as a result the aggregate scale of credit intermediation tends naturally to increase with income. And the global economy of capital flows and floating exchange rates, which emerged from 1970 onwards, does demand some complex services (such as the provision of forward FX market liquidity and hedging products), which generate increased financial activity and which deliver economic benefit.

But while some increase in the relative scale of financial activity may have been both inevitable and desirable, there are both empirical and theoretical reasons for

questioning whether it all was. Empirically it is worth noting that the period of post-war 'financial repression' from 1935 to 1975, apparent both in measures of financial activity and of financial sector remuneration, was one of rapid and steady economic growth which compares well with the subsequent 30 years. There is no aggregate level empirical evidence to support the belief that financial liberalisation and financial deepening has generated superior economic performance. In particular ways it may have done so, but the case needs to be proven at the level of specific value added functions, specific economic benefits.

And in theory there are several ways in which the potential for rent extraction, rather than true economic value added, is likely to be greater in financial services than in other sectors of the economy. At least four categories of effect can be identified.

- The first involves rent extraction from society in general via tax management activities, which benefit the financial service sector's clients, but also, through fees earned, the industry itself. And a depressingly large proportion of what is labelled 'innovative product structuring' is based on tax management activities.
- The second involves rent extraction via high but non transparent margins charged both to retail and to wholesale customers, exploiting deep asymmetries of information and expertise. One way in which these high margins are achieved is via the sale of products which appear to give superior return (yield uplift) but which have deeply embedded options, likely to cause large but very occasional losses, but with the potential for those losses poorly understood not only by the customer, but often also by the salesmen themselves.
- The third involves super normal returns for activities which are also to a degree value added. Market making, i.e. providing liquidity to enable customers to contract in large volumes at reasonable bid offer spreads is, at least up to a certain level of market liquidity, a value creative activity.<sup>12</sup> But market making is an activity likely to be characterised by naturally arising oligopolies and opportunities for super normal returns. Dominant market makers become more dominant because customer order flow goes to where the liquidity is greatest. And customer order flow provides subtle opportunities for position taking which exploits information advantages. Market makers therefore enjoy a systematic tendency to make position taking profits, while customers systemically lose, but in an opaque fashion which few customers understand. As a result, some of the most highly profitable activities in wholesale finance are not in fact excessively risky, and are likely therefore to remain highly profitable, and highly remunerated, even if we effectively reduce financial stability risks.

---

<sup>12</sup> Increased market liquidity is, however, likely beyond some point to be of declining marginal benefit. It is for instance difficult to imagine what possible economic value could arise from arbitrageurs being able to spot microscopic divergences in market prices a few seconds (or now with algorithmic flash trading a few milliseconds) before other arbitrageurs do the same. See Benjamin Friedman, *Overmighty Finance Levies a Tithe on Growth*, Financial Times, 26 August 2009.

- The fourth involves the encouragement by agents (fund managers or banks) of valueless increased trading activity which not only generates income directly, but which, by generating volatility, can itself stimulate demand for hedging services, an effect explored by Paul Woolley in his chapter in the LSE Finance Report (*Why Are Financial Markets So Inefficient and Exploitative?*).

The latter two factors together seem likely to be highly pertinent to the conundrum posed in Section 1, when considering the increase in financial trading activity and related interconnectedness within the financial system. In theory this activity must either be socially value added or zero sum at the social level, with the private gains of some participants offset by the private losses of others. But while the value added seems uncertain or partial, industry participants certainly believe they make money from these activities, and Slide 17 illustrated that total industry profits grew alongside with increased financial intensity. The explanation may be that beyond an inner core of banks and other financial institutions enjoying economies of scale and advantages of skill, there is a wider array of market participants (corporates, individuals, and indeed smaller banks) which, without being fully aware, are systematic losers within a zero sum game.

There are therefore good reasons for rejecting the axiomatic assumption that all of the increase in financial activity has delivered real economic value added, and for instead suspecting that it has arisen from a mix of value added and rent extracting activities. What the balance is deserves further detailed examination, but if we assume for now that the rent extracting element is at least significant, what consequences follow for financial regulation?

Let me suggest four:

- the first two implying caution about what public policy should and can achieve;
- the second two implying, however, the need to be at least open to radical ideas, and to recognise that decisions on appropriate trade offs within financial regulation cannot be entirely divorced from judgements about the value added of increased financial intensity.

(i) Creative and distributive activities: imperfect ability to discriminate

It seems likely that some (and perhaps much) financial trading and position taking activity is either zero sum or rent extracting in its impact, rather than value creative. But it does not follow that public policy can or should attempt to distinguish and extinguish the rent extracting element.

The existence of non-value added economic activities is not unique to finance. Roger Bootle indeed, in his recent book *The Trouble with Markets* (2009), makes the insightful though rarely made point that much economic activity is at least to a degree 'distributive' (i.e. rent extracting), rather than 'creative' (i.e. value added) in its primary economic effect. Thus, for instance, the sales person who wins an order for Company A against Company B's competition, does not as a direct effect increase the economic cake, he redistributes income from Company B to Company A, with economic value added arising only to the extent that this activity on average enables

the more efficient company to grow more rapidly than the less efficient. The case for the market economy, for capitalism, should never therefore be made on the grounds that all activities are axiomatically value added, but simply that it is a system somewhat more likely than the available alternatives to result in a somewhat higher proportion of creative rather than distributive activities.

And any attempt precisely to distinguish 'value creative' from 'rent extracting' activities in finance, at the level of detail which would be required to inform detailed regulation, would be extremely difficult. Market liquidity has a value, but a value subject probably to declining and perhaps at the limit to negative marginal benefit. But regulators and policymakers cannot judge precisely where that limit lies. Public policies are as imperfect as markets.

(ii) Competition based solutions less powerful than assumed?

If both policymakers and existing markets are imperfect, the appropriate response might seem to be to concentrate policy initiatives on making markets more competitive. It is unclear, however, that this will be as powerful a lever as often supposed.

- In retail financial services it is, for instance, notable that some of our greatest concerns about high distribution margins and inappropriate advice have arisen in activities (e.g. the sale of pension and other long-term investment products) characterised by huge numbers of competitive firms and relatively easy market entry.
- And in some wholesale financial services, for instance in market making activities, economic rents have often not been associated with overall market shares which would typically excite competition authority concerns, but from subtle opportunities arising from dominance of highly specific product/market segments or from large but still minority market shares.

It is therefore possible that we face in financial markets imperfections which are unamenable to our standard policy responses and which produce a mix of value added and rent extracting activities between which public policy cannot distinguish. We may face a divergence from social optimality which cannot fully offset, but which still leaves intact the case for a market economy and free financial markets as better than the alternatives.

But while we should recognise that the existence of 'distributive' rather than 'creative' activities is not unique to finance, it seems likely, as Roger Bootle argues, that financial services may be characterised by a greater prevalence of distributive activities than other sectors of the economy. And while we should not imagine that there is a perfect policy response, the probability that financial markets can entail very significant divergences from social optimality should at least make us open-minded about policy options previously rejected, and carries implications for how regulatory trade-offs should be struck.

(iii) Openness to radical policy options

If measures to make competition more effective are not always capable of generating optimal results, neither taxation nor state provision should be excluded from the policies considered.

- The potential existence of economic rent or of financial activity on a harmful rather than useful scale could justify financial transaction or other financial activity taxes, such as those discussed by the IMF in June 2010. If there are super normal returns which are unamenable to structural remedies, economics teaches that taxation is the appropriate response.
- And in retail financial services we should be open to the possibility that the state could sometimes be a more efficient provider of some services, removing the churn and excess cost which pure private competition can create. This was the logic of the UK Pensions Commission's recommendations, subsequently accepted by government, for a state role in providing a low cost mass market pension (the soon to be launched the National Employment Savings Trust (NEST)).

(iv) Trade-offs in financial stability policy

Finally, and most crucially from the point of view of the financial regulator, the realisation that not all financial activity is axiomatically beneficial, has important implications for the trade-offs which financial regulators have to strike.

Section 1 highlighted, for instance, the massive growth in financial sector interconnectedness which has occurred over the last 30 years, a growth closely related to increased trading activity and to an increasing focus on position taking as a potential profit source. It argued that policy may need to lean more aggressively against this interconnectedness, offsetting the externality which complexity in itself creates. But measures to reduce interconnectedness will tend to make some markets less liquid: so too will measures to regulate margins in repo markets, so too will further possible increases in trading book capital.

In the past, proposals for such policy measures were often countered by the argument that they would have a 'chilling effect' on liquidity, product innovation, price discovery and market efficiency. That argument had power because it was assumed by axiom that ever more liquid markets delivered economic benefits, and that increased financial activity always produced economic value added as well as private profit.

Those axiomatic assumptions are clearly invalid: not all financial activity is socially useful. The size of the financial sector, the volume of trading, and the pace of product innovation, should therefore never be seen, as they were before the crisis, as always positive indicators of policy success. And in deciding how far to go in tightening prudential requirements in pursuit of financial stability, regulators cannot entirely avoid making judgements on whether discouraged activity delivers economic value added or not.

The technical issues of financial regulations and the wider issue of whether and to what extent increased financial intensity delivers real value added cannot be entirely divorced.

### **3. Economic theory: fixable incentive structures versus inherent irreducible imperfection**

Which brings me to my third issue, that of economic theory. If this was, pace Tomasso Padoa-Schioppa, not merely a crisis of specific institutions but also of markets, was it also a crisis for economics.

I believe it was. Not so much because academic economics was monolithic – it never has been – but because in the translation of ideas from academia to public policy, a dominant, over-simplified and dangerous conventional wisdom developed. That dominant conventional wisdom needs to be dethroned if we are to think clearly about the drivers of financial instability and the required policy response.

Let me begin with a caricature of the dominant conventional wisdom.

For over half a century the dominant strain of academic economics has been concerned with exploring, through complex mathematics, how economically rational human beings interact in markets. And the conclusions reached have appeared optimistic, indeed at times panglossian. Kenneth Arrow and Gerard Debreu (1954) illustrated that a competitive market economy with a fully complete set of markets was Pareto efficient. New classical macroeconomists such as Robert Lucas illustrated that if human beings are not only rational in their preferences and choices but also in their expectations, then the macroeconomy will have a strong tendency towards equilibrium, with sustained involuntary unemployment a non-problem. And tests of the efficient market hypothesis appeared to illustrate that liquid financial markets are not driven by the patterns of chartist fantasy, but by the efficient processing of all available information, making the actual price of a security a good estimate of its intrinsic value.

As a result, a set of policy prescriptions appeared to follow:

- Macroeconomic policy – fiscal and monetary – was best left to simple, constant and clearly communicated rules, with no role for discretionary stabilisation.
- Deregulation was in general beneficial because it completed more markets and created better incentives.
- Financial innovation delivered economic value added because it completed more markets, and speculative trading was beneficial because it ensured efficient price discovery, offsetting any temporary divergences from rational equilibrium values.
- And complex and active financial markets, and increased financial intensity, not only improved efficiency but also system stability, since rationally self-

interested agents would disperse risk into the hands of those best placed to absorb and manage it.

Now of course, as a description of academic economics, this is not only a simplification but a caricature. Throughout the last half century much of academic economics has been devoted quite explicitly to understanding why and under what conditions these simplistic assumptions do not apply. Kenneth Arrow himself spent much of his career exploring the market imperfections which made his illustration of a Pareto efficient equilibrium inapplicable in the real world. Lancaster and Lipsey (1956) illustrated that if some markets were imperfect, then making other markets closer to perfect might not be welfare optimal. Stigler and others considered the costs of gathering the information required to make markets efficient. Many researchers, including Summers, Poterba and Shiller, found serial correlations and other patterns in share prices which contradicted simple efficient market hypotheses.

All economic textbooks included taxonomies of potential market failure, which might justify policy interventions such as pollution taxes and provision of public goods. And, more fundamentally, the work for which Nobel laureates James Mirrlees, Joe Stiglitz and George Akerlof were acclaimed, illustrated that once we really understand the implications of information economics, markets can settle far from an efficient equilibrium, and equilibria might be multiple and fragile. While the work of the behavioural economists – such as Daniel Kahneman, also awarded a Nobel Prize – has questioned the very assumption of rational choice, of a *homo economicus* driven solely by the parts of his brain devoted to rational information processing.

So academic economics has not been monolithic, it has explored complexities and made assumptions clear, it has produced multiple schools of thought and the most prestigious prizes have gone to people of strongly opposing views. And some of those views help us understand why the financial crisis occurred.

But despite this academic diversity, it is I think fair to say that in the translation of ideas into ideology, and ideology into policy and business practice, it was one oversimplified strain which dominated in the pre-crisis years.

Keynes famously wrote that *‘the ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than commonly understood. Practical men, who believe themselves to be quite exempt from any intellectual influences, are usually the slaves of some defunct economist’*. But I suspect the greater danger lies not with entirely practical men or women exempt from any intellectual influence, but with the reasonably intellectual men and women who are employed in the policymaking functions of central banks, regulators and governments and in the risk management departments of banks, who are aware of intellectual influences, but who tend to gravitate to simplified versions of the dominant beliefs of economists who are not yet defunct but still very much alive.

For it is striking in the pre-crisis years how dominant and how overconfident, at least in the arena of financial economics, was a simplified version of equilibrium theory which saw market completion as the cure to all problems, and mathematical

sophistication decoupled from philosophical understanding as the key to effective risk management.

Institutions such as the IMF, in its Global Financial Stability Reviews (GFSR), set out a confident story of a self-equilibrating system. Thus for instance the GFSR of April 2006, only 18 months before the crisis broke, stated that:

*'There is growing recognition that the dispersion of credit risk by banks to a broader and more diverse group of investors, rather than warehousing such risk on their balance sheets, has helped make the banking and overall financial system more resilient.*

*The improved resilience may be seen in fewer bank failures and more consistent credit provision. Consequently the commercial banks may be less vulnerable today to credit or economic shocks'.*

Market completion was thus explicitly seen as the key to a safer system, with it assessed by axiom that financial instruments would only be held by investors who rationally understood the specific risks involved, and who were in a position to absorb those risks.

Moreover, the GFSR continued:

*'Credit derivatives enhance the transparency of the markets' collective view of credit risks and thus provide valuable information about broad credit conditions and increasingly set the marginal price of credit'.*

Price discovery was thus the key to allocative efficiency: a liquid market in credit risk ensuring an appropriate and rational equilibrium price.

Meanwhile in the risk management departments of banks, and in regulation of market risk capital, assessment of trading risk was dominated by the use of 'Value At Risk' (VAR) techniques which assumed that one could infer the probability distribution of future potential movements of market prices from the observation of movements over the recent past. Market prices were assumed to be driven by the rational interaction of multiple independent agents, and market price risk was therefore mathematically modelable.

All of these assumptions turned out to be fatally flawed.

- Risks were often held not by fully rational and foresightful investors, but by investors who, as per Shleifer's local thinking, had assumed that objectively risky instruments were close to risk free.
- The market price of credit, as Slide 9 shows, provided neither forewarning of future risk nor efficient allocation of capital, but drove a cycle of over exuberant credit supply and then sudden credit withdrawal.

- And apparently sophisticated market risk models failed utterly to anticipate the scale of potential price movements, leaving capital allocated against trading books woefully inadequate.

Fatal policy mistakes were thus based on over confidence in the tendency of financial markets towards rational and efficient equilibria. And that over confidence had, in the years before the crisis, become a belief system. In regulators such as the FSA, the assumption that financial innovation and market liquidity were valuable because they completed markets and price discovery, was not just accepted, it was part of the institutional DNA.

The belief system did not, of course, exclude the possibility of market intervention. But it did determine assumptions about the appropriate nature and limits of intervention. Regulation to protect retail customers could be appropriate: requirements for information disclosure could help overcome asymmetries of information between businesses and consumers. Regulation and enforcement to prevent market abuse was justifiable, because rational agents can also be greedy, corrupt or criminal. And regulation to increase market transparency was not only acceptable, but a central tenet of the doctrine, since transparency completes markets and helps generate increased liquidity and price discovery. But the belief system of market regulators and financial policymakers in the most financially advanced centres tended to exclude the possibility that rational profit seeking by professional market participants might generate rent extraction and financial instability rather than social benefit – even though several economists had clearly shown why that could be the case.

What the dominant conventional wisdom of policymakers therefore reflected was not a belief that the market economy was actually at an Arrow-Debreu nirvana, but the belief that the only legitimate interventions were those which sought to identify and correct the specific market imperfections preventing the attainment of that nirvana.

An appropriately radical response to the financial crisis requires that we take into account explanations of financial market imperfection and instability which go beyond the identification of specific information asymmetries or incentive problems. Three challenges to the dominant conventional wisdom need to be recognised.

- The first, drawing on insights from evolutionary biology, behavioural analysis and neuroscience, recognises that human beings are not always fully rational in their decision making but driven also by rules of thumb (Kahneman et al 1982), and by instinctive reactions. In our financial decision making, as Andrew Haldane (2010) puts it, we are sometimes literally ‘in two minds’, because we have both a pre frontal cortex uniquely capable of ratiocination and a limbic system which can predispose us to myopia, inertia, irrational exuberance or herd effects.
- The second, clearly expressed by Joe Stiglitz in his Nobel Prize Lecture (2001), illustrates that information asymmetries and imperfections can be so inherent and pervasive as to be unamenable to any feasible policy of identifying and removing them to bring the system closer to efficiency and stability.

- The third draws on Frank Knight's (1921) crucial distinction between mathematically modelable risk and inherent irreducible uncertainty. It points out that the very idea that the expectations of economic agents are distributed around the objective probability distribution of future outcomes is a philosophical category error, since no such probability distribution of future outcomes objectively exists. As a result assessments of the financial market 'risks' which trading book capital needs to absorb, cannot be reduced to the precise mathematical modelling of VAR approaches.

Together, these insights amount to a profound challenge to the pre-crisis conventional wisdom. The latter two together imply, as Roman Frydman and Michael Goldberg (2011) argue, that financial markets could be susceptible to occasional large divergences from equilibrium values **even if** agents were, at the individual level wholly rational. The first introduces a major additional driver of instability, with myopia and other non-rational decision-making processes as important as poor incentives. In combination they mean that financial markets and systems which are as free and competitive as we can feasibly make them will sometimes deliver neither stability nor allocative efficiency.

As a result, fixing poor incentives – such as those created by 'too big to fail' banks or by perversely designed bonus arrangements – while a necessary part of the regulatory response, cannot be sufficient. Our policy response needs also to include policies which focus on the complex dynamics of the whole system, above all through higher equity capital requirements, and macro-prudential policies which can respond in a discretionary fashion to the credit cycles and instabilities which could arise even in a system where individual agents' incentives were always well designed.

## References

David Miles, Jing Yang and Gilberto Marcheggiano *Optimal Bank Capital*, External MPC Unit, Discussion Paper No 31, January 2011

Anat Admati, Peter Demarzo, Martin Hellwig and Paul Pfleiderer, *Fallacies, Irrelevant Facts and Myths in the Discussion of Capital Regulation, Why Bank Equity is Not Expensive*, Max Planck Society, 2010

Nicola Gennaioli, Andrei Shleifer and Robert Vishny, *Neglected Risks, Financial Innovation, and Financial Fragility*, September 2010

Gary Gorton and Andrew Metrick, *Regulating the Shadow Banking System*, September 2010

Committee on the Global Financial System, *The Role of Margin Requirements and Haircuts in Procyclicality*, CGFS Paper No 36, March 2010

Adair Turner, *Financial Markets: Efficiency, Stability and Income Distribution*, Lecture II of Lionel Robbins Memorial Lectures, October 2010

Paul Woolley, *Why Are Financial Markets so Inefficient and Exploitative – And a Suggested Remedy*, Chapter 3 of the Future of Finance, The LSE Report 2010

T. Philippon and A Reshef, *Wages and Human Capital in the US Financial Industry 1909-2006*, NBER Working Paper No 14055

Roger Bootle, *The Trouble with Markets*, Nicholas Brealey 2009

Andrew Haldane, *Patience and Finance*, Oxford India Business Forum, September 2010

Philip Augar, *Britain's Bank Bonus Embarrassment*, Financial Times, 12 January 2011.

Benjamin Friedman, *Overmighty Finance Levies a Tithe on Growth*, Financial Times, 26 August 2009.

Luigi Einaudi, *Debts', Selected Economic Essays'*, Macmillan 2006. First published as *'Debiti*, La Reforma Societe XLI, January 1934

Irving Fisher, *The Debt Deflation Theory of Great Depressions*, Econometrica, 1933

Roman Frydman and Michael Goldberg, *Beyond Mechanical Markets*, Princeton University Press, 2011

Modigliani, F: Miller, M (1958), *The Cost of Capital, Corporation Finance and the Theory of Investment*. American Economic Review 48 (3): 261-297

Joseph Stiglitz, *Information and the Change in the Paradigm of Economics*, Nobel Prize Lecture, December 2001

Frank Knight, *Uncertainty and Profit*, 1921

Kenneth Arrow and Gerard Debreu, *Existence of an Equilibrium for a Competitive Economy*, *Econometrica*, Vol 22, 1954

Richard Lipsey and Kelvin Lancaster, *The General Theory of the Second Best*, *Review of Economic Studies*, 1956