Risk-Pricing and the Sub-Prime Crisis

Andrew G. Haldane

At the time of writing, the sub-prime crisis has just celebrated its first birthday. As Chinese leader Chou En-lai said of the French revolution, it is “too early to tell” what lessons ultimately will be reached about this crisis. Nonetheless, much has already been learnt, and even greater amounts said and written, over the past year on its causes and consequences. Without repeating that, I discuss here three aspects:

• The roots of the sub-prime crisis—how did we get here?
• Public policy tools for managing the transition path—how do we get out of here?, and
• Public policy tools for managing future crises—how do we prevent a return here?

(A) Roots of the crisis: risk in the wrong place at the wrong price

The origins of the sub-prime crisis are, in many respects, all too familiar. This is a story of classic informational and distorted incentive problems among market participants (Greenlaw et al., 2008). These combined to generate a broadly based under-pricing of risk within the system and its quantity counterpart—a credit boom. In many important respects, only the names have changed: this is the crisis of sub-prime, SIVs and CDOs rather than LDC debt, South Sea Companies and junk bonds. Plus ça change.
But even if the market failures themselves have been familiar, the virulence and duration of the accompanying problems has taken many by surprise. Changes in the structure of the financial system over the past decade provide a large part of the explanation. Crucial, here, has been the rapid emergence of instruments and vehicles which allowed risk to be disected, rebundled and then reallocated within the financial system. Such securitisation led to risk becoming an easily tradable commodity. As trading in this commodity grew rapidly in the early part of this century, the price (of risk) fell and the quantity (of risk-taking) rose in parallel (Figure 1).

Other things equal, improving market mechanisms for risk-sharing should boost welfare, by generating a fall in the equilibrium cost of risk capital. That was the prevailing orthodoxy as recently as 2007. Previously lumpy and concentrated risk exposures had been diversified to a wider clientele of investors supposedly better able to bear and manage this risk. This included insurance companies, pension funds and hedge funds, to whom $3 trillion of net credit exposure was transferred through the CDS market in 2006 alone. A risk shared was a risk halved—or perhaps even quartered, given the magic of diversification (Duffie, 2007). Or so ran the rhetoric.

Figure 1: Financial market liquidity

(a) The liquidity index shows the number of standard deviations from the mean. It is a simple unweighted average of nine liquidity measures, normalised on the period 1999–2004. Data shown are an exponentially weighted moving average. The indicator is more reliable after 1997 as it is based on a greater number of underlying measures.
So what went wrong and why? As illustrated in Figure 2, one consequence of securitisation was that the informational chain lengthened, with an alphabet soup of on and off-balance sheet conduits providing the additional links in this chain. Lengthened but, crucially, not strengthened. Because with every additional link in the chain, the scope for informational gaps and misaligned incentives increased. This was a party game of Chinese Whispers, but with thousands of participants and large quantities of real money at stake. The final outcome was, however, much the same: muddled messages at the end-point of the chain, which in turn resulted in distortions to risk perceptions, and hence risk-pricing, by end-investors. That is now the new conventional wisdom.

But what behavioural frictions allowed this elaborate party game to start and to continue for so long? In understanding those frictions, one revealing observation is that many of the instruments used to effect these risk transfers took a particular form; in particular, they embodied significant “tail risk” (Rajan, 2005). The most important of such tail risk instruments were Collateralised Debt and Loan Obligations (CDOs and CLOs), which tranche up the loss distribution on a pool of underlying assets. By construction, these instruments have a highly-non-linear, tail-heavy payoff distribution. Growth in these markets was exponential during the credit boom.

Pricing tail risk products, such as CDOs and CLOs, gives rise to two particular informational problems for investors. First, by definition, tail risk is realised very infrequently. So financial instruments embodying high degrees of tail risk are particularly sensitive to the small sample problems that face modellers of all financial time-series. This small sample problem was exacerbated by the high and enduring degree of macroeconomic stability experienced up until recently, often termed the “Great Moderation”.

*For more detail on the roles of participants in structured finance markets see Committee on the Global Financial System (2005), The role of ratings in structured finance: issues and implications, January, available at www.bis.org/publ/cgfs23.htm.
A second potential informational problem arises from investors using linear pricing models, such as CAPM, to price instruments with highly non-linear payoffs. One reason why this might arise is that many non-bank institutional investors based their investment decisions on simple summary statistics of risk, such as credit ratings. Credit ratings are based either on measures of expected loss or default probability. In both cases, they are measures of the central tendency of the loss distribution and provide no real indication of tail risk. As such, these measures are exactly analogous to the use of a linear pricing model.

Figure 3 illustrates these two potential informational frictions schematically. It shows the returns to a hypothetical CDO-like asset. These are highly non-linear and tail-heavy. If the price of the underlying asset remained in the range \( x-y \), a myopic investor would infer \( \alpha = \alpha^* \) and \( \beta = 0 \). That is, the asset would appear to offer the end-investor significant excess returns with no systematic risk. It is only once tail risk is revealed to the left of \( x \) that the true risk-return profile is revealed (\( \alpha = 0 \) and \( \beta = \beta^* \)). Even then, however, there is scope for mispricing if investors are fitting linear pricing models to this payoff distribution. These models would underprice risk because \( \beta (=\beta^*) \) is underestimated in the tails. So in both cases, informational frictions manifest themselves as investors mistaking beta (systematic risk) for alpha (excess returns) (Rajan, 2005).

How large quantitatively were these pricing mistakes? A formal model of CDO pricing can be used to assess this issue, calibrated to recent experience. The results are summarised in Table 1. It shows the excess returns (alpha) perceived by holders of different CDO tranches, using either a linear pricing model (shown in the first column), a restricted sample of observations (in the second) or both (in the third) (Haldane and Webber, 2008). The implied pricing errors are large, especially so for the highest-risk, equity tranche. It could be argued that holders of these high-risk tranches are likely to be sophisticated investors, less subject to informational
frictions. But even for the lower-risk tranches, the perceived excess returns are material, at around 0.5% per annum—a premium of around 25% over safe rates during the recent past.

Mispricing perceptions on this scale will have generated perverse incentives on the part of both originators and end-investors in structured credit products. On the demand side, perceived excess returns will have generated a ready demand from investors, at a time of low global real interest rates. And, faced with this demand, originators will have had strong incentives to manufacture these instruments for onward sale. These incentives will have roughly been in proportion to the perceived mispricing. In other words, price incentives to bear risk will have been strongest among those least, rather than best, able to price and manage it.

The bottom line from this analysis is an uncomfortable one from a systemic perspective. Informational frictions and financial engineering combined to deliver risk in the wrong hands at the wrong price. Contrary to the rhetoric ahead of crisis, risk gravitated to the unsuspecting and uninformed rather than the sophisticated and robust. And once risk is in the wrong hands it becomes difficult to move, whatever the price. The seizure in some securitisation markets over much of the past year is testimony to that.

(B) Easing the transition: three coordination failures

Summary measures of liquidity illustrate just how quickly this once gushing tap turned itself off during 2007 (Figure 1). The upshot was a logjam in money markets as banks hoarded liquidity. In consequence, money market yields spiked upwards sharply in September 2007. They have remained sharply higher since and market expectations are that they will

<table>
<thead>
<tr>
<th>Table 1: Overpricing of tranched derivatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived excess return per annum (%)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Tranche</td>
</tr>
<tr>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>Senior (22–100)</td>
</tr>
<tr>
<td>Mezzanine (9–22)</td>
</tr>
<tr>
<td>Equity (0–9)</td>
</tr>
</tbody>
</table>
remain at elevated levels in the period ahead (Figure 4). If LIBOR spreads have become the financial markets’ barometer of the health of the global financial system, the storm clouds show little sign of lifting in the immediate period ahead.

So are there ways in which this transition path could be eased or made shorter? To gauge that, it is important to recognise that the precise nature of the turbulence has altered subtly over the period since September. This can be seen by decomposing the money market spread into its liquidity and credit risk components, as shown in Figure 4. This suggests, broadly-speaking, three phases: September–December 2007, when illiquidity factors appeared to be dominant; December 2007–February 2008, when solvency concerns took centre stage; and March 2008 onwards, when solvency and liquidity concerns seemed to combine. All three phases can be thought to be rooted in coordination or collective action problems, albeit of different types.

The first—"illiquidity"—phase was brought to an end by the bout of coordinated central bank intervention by the world’s major central banks on 12 December 2007.¹ The scale of these interventions was modest

¹ Further coordinated central bank intervention took place in March 2008.
relative to the scale of global financial balance sheets. So why did these relatively modest (in financial terms) actions work in shifting illiquidity premia? And what lessons does this contain for dealing with future illiquidity crises?

Theory provides some of the answers. Theoretical models of liquidity are often founded on expectations or “higher order beliefs” (Allen, Morris and Shin, 2006). Liquidity is present when private sector agents each believe others will transact and expectations coordinate around this high-liquidity equilibrium. Conversely, when they believe others will not transact, expectations coordinate around a low-liquidity equilibrium. Liquidity equilibria are, in that sense, self-fulfilling and potentially multiple. Several implications follow from such a model of liquidity, each of which matches observed patterns of behaviour during the illiquidity phase of the present crisis.

First, liquidity is likely to adjust sharply, as expectations switch and coalesce around different equilibria. The spikes, both upwards and downwards, in Figure 4 are consistent with that multiple, knife-edge equilibria theory. Second, coordinated public sector liquidity actions are one means of coordinating private sector liquidity expectations. The December 2007 coordinated liquidity interventions by central banks are an example of such actions. Third, the success of such coordination will depend less on the money behind it than on the strength of the expectational effect. The surprise announcement of coordinated action in December would score highly on the latter criterion, but modestly on the first. And fourth, such announcement effects are, according to theory, likely to be particularly potent when public signals are relatively less noisy than private signals (Morris and Shin, 2002, 2007). In other words, talk by the public authorities is not cheap when private sector uncertainty is high, as it was in the midst of financial stress in December.

So the December and subsequent coordinated liquidity intervention by central banks underscore some important general points about dealing with crisis. These include the importance of clear, prompt and frequent communications, as a means of helping shift and coordinate private sector expectations. This role is especially valuable when private sector uncertainty is high and when actions are coordinated across jurisdictions. Theory explains why open mouths can ease the transition path for the

---

2 Even if both public and private signals are noisy in absolute terms.
financial system, perhaps at least as well as open wallets. Recent experience has tended to corroborate that theory.

In the second—“solvency”—phase of the crisis, perceptions of default probability rose for almost all financial institutions internationally and came to be the dominant determinant of interbank premia (Figure 4). A bank facing rising solvency fears can seek to strengthen its balance sheet in a number of ways. These include reducing dividends, raising new equity and shrinking asset growth. Coordination problems surround all three options.

From the perspective of an individual firm, there are clear disadvantages in the first two options. Cutting dividends and/or raising new capital serve as a visible signal of institutional stress and/or management failure. A number of the banks which have raised new equity as a result of the recent crisis have simultaneously seen a significant management shake-out. And most firms have similarly sought to insulate dividend payouts from losses, for fear of disturbing investor perceptions. Shrinking asset growth, in part because it is less visible, comes with less such stigma. So it is perhaps unsurprising that there has been a sharp and generalised contraction of banks’ credit supply internationally in response to recent events—so-called “deleveraging”.

While such a response is rational from the perspective of an individual firm, it may not be for the financial system as a whole. By reducing the supply of credit, depressing the macro-economy and hence potentially crystallising credit risk, actions to shrink balance sheets can become self-defeating. In effect, a credit crunch is a coordination failure. In these circumstances, there would be collective benefits from coordinated capital-raising and/or dividend-cutting by financial firms to defuse stigma risk, as an alternative to a concerted contraction of credit.

Regulatory intervention is one means of achieving this coordination. For example, regulators could require an across-the-board increase/decrease in new equity/dividend payouts at times of market stress. This could be justified on systemic insurance grounds and could, if necessary, be temporary. This would reduce the stigma otherwise attached to financial firms going down this road unilaterally. Having system-wide regulatory tools such as these to hand would be a useful addition to policymakers’

3 There are means of hitting two birds with one stone—for example, by paying dividends in stock rather than cash, as a number of firms have done over recent months.
Risk-Pricing and the Sub-Prime Crisis

toolkit in heading-off balance sheet co-ordination problems, such as the credit crunch currently afflicting many developed countries.

The third—“mixed solvency/liquidity”—phase of the crisis contained perhaps the most complex behavioural interactions. As asset prices fell, an increasing number of financial institutions found themselves marking asset portfolios down. This prompted some institutions to dispose of these assets in a fire sale which in turn fed back to a weakening of asset prices. This feedback loop between weakening asset prices and marking-to-market of balance sheets helps explain the correlation between liquidity and solvency problems in the period since March 2008 (Cifuentes, Ferrucci and Shin, 2005).

This market dynamic has had a number of systemic consequences. First, the increasing prevalence of marking-to-market has meant that institutions have been writing-down their balance sheets based on expectations of future losses rather than realised defaults. The first two bars in Figure 5 illustrate that write-downs have far exceeded realised losses on sub-prime loans during the current crisis.

![Figure 5: Losses on sub-prime asset-backed securities](image)


(a) Area below dotted line shows net write-downs by major UK banks and LCFIs since the start of 2007 to 22 April 2008, while total height of bar shows an S&P estimate (published on 13 March 2008) of write-downs by all investors.

(b) In the absence of data on realised losses, this estimate is derived from data on actual delinquency rates on outstanding mortgages by vintage and an assumption about the transition from delinquency to default, as described in Box 1, Bank of England, Financial Stability Report, April 2008, pp. 18–19.

(c) This estimate is derived in the same way as estimated credit losses, but assuming that serious delinquency rates in different vintages continue to rise at their average rate to date until the mortgages are four years old, when they are assumed to be plateau. See Box 1, Bank of England, Financial Stability Report, April 2008, pp. 18–19.
Second, fire-sale dynamics have contributed to an overshoot in the prices of a number of assets, to levels well below those justified by credit fundamentals. Figure 6 looks at market-implied loss rates on the ABX index—an index of credit default swaps written on US sub-prime mortgages. The modal loss rate in April 2008, of around 38%, is very difficult to reconcile with any plausible path for the real economy—for example, it would be consistent with a probability of default of 70% and a loss given default rate of over 50% (Bank of England, 2008). This suggests very substantial illiquidity and uncertainty premia in asset prices. The same effects, on a smaller scale, are evident across a number of other financial markets.

Third, given this overshoot in market prices, mark-to-market estimates of financial sector losses far exceed loss estimates built-up from an assessment of likely default paths. The third and fourth bars in Figure 5 illustrate that these two sets of loss estimate on sub-prime loans can differ significantly—in this case, by a factor of more than two. Pessimism about the scale of mark-to-market financial sector losses may itself have added to downward pressures on asset prices.

How is this adverse asset price feedback loop reversed? One alternative, mooted by some, would be to suspend mark-to-market accounting, to prevent financial firms marking down portfolios at firesale prices. But a shift to historic cost valuations would be a retrograde step at a time of acute uncertainty about the true balance sheet position of many financial firms; it would amplify uncertainties about losses rather than diminish them. Japanese experience during their “lost decade” is salutary in this respect.
An alternative approach, which defuses the feedback dynamic without dispensing with the fair-value principle, would be to seek greater coordination of marks made by financial firms, centring those around fundamental values. That coordination is not easily achieved by financial firms acting on their own, as can be seen from the degree of uncertainty around reported marks on assets. So a case can be made for some international body playing this coordinating role, offering more prescription on valuation methodologies and, more ambitiously, perhaps offering greater guidance on marks to ensure these are well aligned with fundamentals. This could have helped defuse some of the asset price feedback risk which has operated during the present crisis. And, in a world of increasingly widespread marking-to-market and hence greater asset price-sensitivity of balance sheets, it could play an even more valuable role looking into the future.

(C) Preventing a recurrence: curbing the credit cycle

A number of detailed recommendations have already been made to help address the fault lines exposed by the sub-prime crisis. In particular, the Financial Stability Forum issued a detailed action plan for the international authorities in April 2008, which was endorsed by the G7. In the UK, a consultation document was issued by the Tripartite authorities (HM Treasury, the Bank of England and Financial Services Authority) in January 2008 containing a number of proposals for improving financial stability and protecting depositors. Most of these proposals are aimed at minimising the collateral damage the next time crisis strikes.

An even more difficult task for public policy is, however, to prevent a next time. Of course, an almost riskless financial system could be made a reality tomorrow, provided the public were willing to pay the deadweight costs of acute financial inefficiency every period thereafter. At times of acute financial distress, there is often pressure to move in this direction, with Glass–Steagall and Sarbanes–Oxley being among the more memorable examples. This “Dangerous Dogs Act” risk—of an over-hasty legislative response—needs to be avoided because history suggests such mistakes are not easily rectified: it took the US 66 years to repeal the

---

4 Available at www.fsforum.org.
5 Available at www.hm-treasury.gov.uk. Two further consultation documents were issued in July 2008.
Glass–Steagall Act, which came as a direct response to the stock market crash of 1929.

The broader point is that at least some of the turbulence over recent months has simply been the weeding out of failed financial innovation. It is Schumpeterian “creative destruction” of the CDO-squareds and the SIVs. These are examples of markets failing—but not of market failures. In these circumstances, policymakers should stand aside. In other words, they should be seeking to curb (rather than eliminate) the credit cycle, finding a preferred (rather than corner) solution on the risk-return frontier. How? Two proposals, neither of them new, warrant detailed theoretical, empirical and practical investigation in the period ahead.

(a) Designing a counter-cyclical regulatory regime

Several features of the regulatory and risk management framework serve to amplify procyclicality of risk-taking and credit creation. From a risk management perspective, Value-at-risk (VAR) and stress-testing methodologies were lauded ahead of the crisis as a dramatic step forward in the technological frontier of risk management. That claim looks rather hollow in the light of subsequent events. At root, the problem with these approaches is their use of historical inputs, often drawn from a very small sample. VAR measures in practice simply mirrored the risk cycle, remaining low when historical risks were low and only signalling distress once it was upon us. Stress-tests were based on events which, with hindsight, were Himalayan foothills by comparison with the Everest of events banks have had to face over the past year. In this way, neither of these frontier risk management approaches served as an effective brake on excessive risk-taking during the upswing.

Improved regulation of banks’ capital and liquidity could, in principle and in practice, address this pro-cyclicality problem. Liquidity has been the poor relation of the regulatory world for the past several decades (Jenkinson, 2008). One consequence is that banks’ liquid asset holdings have been in secular decline for the better part of 40 years. For example, in the UK liquid asset ratios among banks have fallen by a factor of at least ten since the late 1960s. Over the same period, potential demands on firms’ liquid assets appear, if anything, to have increased—for example, because of contingent off-balance sheet exposures and the demands of real-time gross settlement payment systems. The failure of Bear Stearns
in March 2008 underscored just how quickly even sizable liquidity pots can be exhausted at times of acute stress.

Most countries already have some form of liquidity regulation in place. But these regulatory regimes differ greatly across countries. Crisis events have already led to the authorities revisiting best practice principles for liquidity regulation (Jenkinson, 2008). One potential direction of travel would be the design and implementation of an international regime for liquidity regulation, paralleling the one for capital regulation. A second, even more ambitious, proposal would be to design such a regime to lean explicitly against the wind, with regulatory liquid asset ratios rising during the upswing to help curb risk appetite. Such a liquidity regime could help serve as an automatic stabiliser for the credit cycle.

Capital regulation has of course already been in place for several years. It has plainly delivered some benefits. Capital ratios for US banks fell secularly from the middle of the 19th century up until the 1980s, by a factor of around ten. Basel I helped reverse that trend. But such regulation inevitably comes at a cost. The loopholes in Basel I were soon exposed by banks, to the point that it became part of the problem rather than the solution. And considerable concern was expressed ahead of the recent crisis about the potential pro-cyclicality of its supposedly superior successor, Basel II (Danielsson et al., 2001). The regime does have some built-in safeguards against such pro-cyclical effects. But it is fair to say there is a significant degree of uncertainty about whether these safeguards will prove adequate in a sharp downturn—we will probably not have long to wait to find out. More fundamentally, however, some have asked whether capital regulation could also be given an explicitly *counter-cyclical* role, with capital requirements rising during an upswing to choke off excessive risk-taking (Goodhart and Persaud, 2008).

No-one has as yet devised a fully coherent means of giving this idea practical effect. But the principles seem clear. Regulation is a tax and, within this, capital regulation is a tax on credit creation. Linking this tax to the cycle requires no more (and no less) a leap of faith than with any other macroeconomic policy tool, such as monetary or fiscal policy. The technical difficulties are daunting, such as defining the cycle and calibrating the tax to achieve the desired modulation in risk-taking. But these are the self-same calibration issues that have historically been faced by all other tools of macroeconomic management. With time and experience, these
problems were resolved satisfactorily and macroeconomic policy graduated to an explicitly counter-cyclical role. Just as monetary policy graduated from a Friedman fixed k% rule during the 1970s, the time may be ripe for capital regulation to evolve from the Basel fixed 8% rule that has governed since its inception.

(b) Redrawing the boundary for OTC instruments

One common factor distinguishes those markets and instruments that have been impaired most by the recent turmoil—they are almost all over-the-counter (OTC) markets. This includes not only complex securitised instruments, such as CDOs and CLOs, but also such plain-vanilla financial contracts as residential mortgage-backed securities and unsecured interbank loans. All of these markets have, for prolonged periods over the past year, been dislocated and, at times, have seized-up completely. In preventing a recurrence of such market failures, there is a case for rethinking the appropriate boundary between OTC-traded instruments and those traded, cleared and settled centrally.

The benefits of central trading, clearing and settlement of financial instruments are manifold (see, for example, Committee on Payment and Settlement Systems, 2007). At the trading end, they include the standardisation and simplification of financial contracts, which increases these contracts’ transparency and hence fungibility. This also simplifies significantly the closing out or hedging of positions in the market. In this way, price formation and liquidity in these markets is boosted. Lack of transparency, loss of liquidity and the accompanying interruption to price discovery has been a common feature within some OTC markets and instruments during the recent crisis. So too has been the difficulty of closing out or hedging such exposures, the latter which has exposed firms to significant basis risk.6

Centralised clearing of positions, for example through a central counterparty, offers further benefits. The most important of those is the elimination of counterparty credit risk, as the central counterparty interposes itself between each bilateral trade to act as guarantor. In the event of a party failing, the central counterparty takes on the resulting exposure and meets any loss from close-out using a central pool of funds drawn from

---

6 This also contributed to the emergence of serious back-office problems in matching and confirming certain OTC derivative trades.
participants (ex-ante or ex-post). This “ex-ante guarantee of trades and ex-post mutualisation of losses” model boosts market liquidity, by taking counterparty credit risk out of the equation. As illustrated in Figure 4, counterparty risk has been a key ingredient in the dislocation of a number of OTC markets in recent months, most notably the money markets.

Given these advantages, why are centralised arrangements for the trading, clearing and settlement of transactions not in place already? In part, this can be explained by the financial innovation cycle. Competition and diversity among financial products is desirable. Standardisation and centralisation would be neither cost-effective nor healthy from a competitive perspective at early stages of the product life-cycle. But beyond a certain size threshold, the benefits of standardisation and centralisation weigh more heavily, as risk exposures accumulate and network externalities intensify.

Recent experience suggests that those network thresholds may have been breached for a number of markets which, nonetheless, have remained resolutely OTC. This would include, for example, a number of securitisation markets, which have grown exponentially and had a combined outstanding stock of around $6.5 trillion at end-2006. It might also include OTC derivatives markets such as CDS, whose growth has exceeded Moore’s Law and whose amounts outstanding have reached almost £60 trillion.

Private participants in these markets may not have the individual incentives to create centralised infrastructure to contain such risks. Indeed, some of the larger firms are likely to earn rents from playing this *de facto* central bank and/or counterparty role in some markets. So there is scope for greater regulatory scrutiny of the appropriate systemic boundary for OTC markets. Having established those boundaries, there then is a case for concerted regulatory action to catalyse infrastructural change once these systemic thresholds have been breached. For example, the US President’s Working Group on Financial Markets and the Federal Reserve Bank of New York have recently set the financial industry some priority actions aimed at improving the infrastructure of the CDS market, including the creation of a central counterparty. Partly in response, a number of industry initiatives are underway to build these market infrastructures. There is further to go, both within the CDS market but more broadly across a wider array of large and systemically-important OTC markets.
Neither of these medium-term initiatives would prevent a next time with a 100% probability. Nor should they, for that would imply policymakers had overbalanced and fallen headfirst into a Dangerous Dogs trap. But, if implemented in practice, they might mean that more than just the names will have changed when the next thunderclap sounds.

References


