Endogenous and Systemic Risk

Jon Danielsson
London School of Economics

Hyun Song Shin
Princeton University

Jean–Pierre Zigrand
London School of Economics

This version August 2011
First version July 2010

Abstract

The risks impacting financial markets are attributable (at least in part) to the actions of market participants. In turn, market participants' actions depend on perceived risk. In equilibrium, risk is the fixed point of the mapping from perceived risk to actual risk. When market players believe trouble is ahead, they take actions that bring about realized volatility. This is "endogenous risk." A model of endogenous risk enables the study of the propagation of financial booms and distress. Among other things, we can make precise the notion that market participants appear to become "more risk–averse" in response to deteriorating market outcomes. For economists, preferences and beliefs would normally be considered independent of one another. We discuss modeling of endogenous risk and some of its distinctive features, both theoretical and empirical.

1 Introduction

Financial crises are often accompanied by sharp price changes. Commentators and journalists delight in attributing such unruly volatility to the herd mentality of the financial market participants, or to the fickleness and irrationality of speculators who seemingly switch between fear and overconfidence in a purely random fashion. Such crisis episodes lead to daily headlines in financial newspapers such as "Risk Aversion Rises" or "Risk Aversion Abates."

Such price swings would be consistent with price efficiency if they were entirely driven by payoff-relevant fundamental news. A large part of this volatility is however due to a number of feedback effects that are hard-wired into the system. We labelled this risk *endogenous risk* in Daníelsson and Shin (2003) to emphasize that while the seeds of the volatility are exogenous, a large part of its eventual realized magnitude is due to the amplification of the exogenous news within the system.

Endogenous risk is the additional risk and volatility that the financial system adds on top of the equilibrium risk and volatility as commonly understood. For this reason, in the formal modeling exercise we will assume that financial institutions are risk neutral. This has the advantage that any feedback effects must be due to the system itself, rather than due to the risk averse behavior of the financial institutions.

Once the financial system is fully modeled to take account of the hard-wiring of risk feedback, it has the potential to magnify risk considerably. However, by the same token, the system can dampen realized risks "artificially" thereby encouraging the build-up of potential vulnerabilities. Part of our task is to show under which circumstances the magnifications occur. For empirical evidence iAdrian and Shin (2010) who note that the major market–based financial intermediaries were deleveraging aggressively during the crisis, and that such financial intermediaries could be seen as the marginal investors for the determination of risk premiums.

In our model, endogenous risk and the inherent nonlinearities of the system are associated with fluctuations in the capitalization of the financial sector. As the capital of the financial sector fluctuates, so does realized risks. The balance sheet capacity of the financial sector fluctuates for both reasons. The risk exposure supported by each dollar of capital fluctuates due to shifts in measured risks, and so does the aggregate dollar capital of the financial sector itself.

The mutual dependence of realized risks and the willingness to bear risk means that the risk capacity of the financial system can undergo large changes over the cycle. Occasionally, short and violent bouts of risk shedding sweep the markets during which the financial institutions' apparent willingness to bear risk evaporates. Those are the episodes that are reported on in news under the headline "risk aversion." It is as if there was a latent risk aversion process that drives financial markets. Of course, the fluctuation in risk aversion is itself endogenous, and in this paper we sketch the mechanisms that drive the fluctuation.

By conceptualizing the problem in terms of constraints rather than preferences, we can address an apparent puzzle. How can it be that human beings are risk averse one day, in a perfectly coordinated fashion, selling their risky holdings across the board and reinforcing the crisis, only to become contagiously risk-loving not too long thereafter, pushing prices back to the pre–crisis levels? Surely they do not all together feel compelled to look right and left ten times before crossing the street one day while blindly crossing the next?

We then apply the results from the theoretical model to more practical questions of systemic risk. Empirical studies and financial history have taught us that financial markets go through long periods of tranquility interspersed by short episodes of instability, or even crises. Such behavior can be understood within our framework as periods where leverage growth and asset growth go together, leading financial institutions on a path driven by positive feedback between increasing leverage, purchases of risky assets and higher prices. During this period, greater willingness to take on risk dampens measured risks and tends to reinforce the dormant volatility. The active trading by financial institutions works to reduce volatility while thickening the tails of the outcome distribution, and increase the magnitudes of extreme events.

The amplification of risk over the cycle poses considerable challenges for bank capital regulation. A fundamental tenet of microprudential capital regulation is the idea that if every institution is individually safe, then so is the financial system itself. A surprising and counterintuitive result of analyzing prudential regulations is that the individually prudent behavior by a financial institutions causes an overall amplified crisis. This is an illustration of the *fallacy of composition*, as discussed by Daníelsson, Embrechts, Goodhart, Keating, Muennich, Renault, and Shin (2001), which criticized the Basel II capital rules on these grounds.

Our results also have implications for financial risk forecasting and recent models of empirical systemic risk forecasting. The vast majority of such models assumes financial risk is exogenous, which may be an innocuous assumption during times when the financial markets are calm, but not during market turmoil. In general, an assumption of exogenous risk is likely to lead to an underestimation of risk when things are calm and overestimation during crisis. This means that most extant financial risk forecast models and empirical systemic risk models might fail when needed the most, i.e., in the accurate forecasting of extreme risk.

There are also implications for regulation of over the counter (OTC) derivatives. The impact of moving OTC derivatives to central counterparties (CCPs) are analyzed by Zigrand (2010) in the light of endogenous risk. He notes that CCPs need to protect themselves from counterparty risk, implying institutionalized initial margin and maintenance margin rules based on continuous marking–to–market. Endogenous risk appears in several guises, to be elaborated below.

2 Endogenous Risk and Price Movements

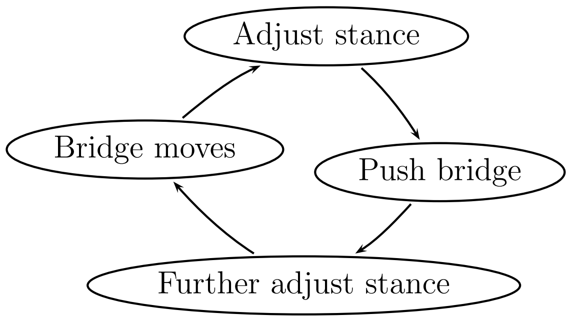
In the main, price movements have two components — a portion due to the incorporation of fundamentals news, and an endogenous feedback component due to the trading patterns of the market participants over and above the incorporation of fundamentals news.

Large price movements driven by fundamentals news occur often in financial markets, and do not constitute a crisis. Public announcements of important macroeconomic statistics are sometimes marked by large, discrete price changes at the time of announcement. These changes are arguably the signs of a smoothly functioning market that is able to incorporate new information quickly.

In contrast, the distinguishing feature of crisis episodes is that they seem to gather momentum from the endogenous responses of the market participants themselves. This is the second component, the portion associated with *endogenous risk* (see Daníelsson and Shin, 2003). We can draw an analogy with a tropical storm gathering force over a warm sea or with the wobbly Millennium bridge in London.

A small gust of wind produce a small sway in the Millennium bridge. Pedestrians crossing the bridge would then adjust their stance slightly as a response, pushing the bridge further in the same direction. Provided sufficiently many pedestrians found themselves in the same situation, they will find themselves coordinating spontaneously and unwittingly to move in lockstep, thereby reinforcing the swaying into a something much more violent. Even if the initial gust of wind is long gone, the bridge continues to wobble. Similarly, financial crises appear to gather more energy as they develop. And even if the initial shock is gone, volatility stays high. What would have been almost impossible if individual steps are independent becomes a sure thing given feedback between the movement of the bridge and the adjustment by pedestrians, see Figure 1.

Figure 1: Feedback Loop of the Millennium Bridge



By analogy, as financial conditions worsen, the willingness of market participants to bear risk seemingly evaporates even in the absence of any further hard news, which in turn worsens financial conditions, closing the loop. Any regulatory interventions might be best aimed at understanding and mitigating those negative spillover effects created purely within the financial system. If one cannot prevent gusts of wind, then at least one can make sure the pedestrians do not act in lockstep and cause the bridge to collapse by critically amplifying the initial swing.

The workings of endogenous risk can be sketched as follows. An initial negative piece of news, leading either to capital losses to the financial institutions (FI) or to an increase in market volatility, must be followed by a risk exposure reduction on behalf of many market participants (or capital raising, which are difficult to do pull off quickly, especially in the midst of a crisis). The reason for contagious behavior lies in the coordinated responses of market participants arising from the fact that market prices are imperatives for action through risk constraints imposed on individual traders or desks (such as Value-at-Risk (VaR) constraints[1]), or through the increase in haircuts and the implied curtailment of leverage by credit providers.

To the extent that such rules are applied continuously, the market participants are induced to behave in a short–termist manner. It follows that the initial wave of asset sales depresses prices further, increasing the perceived risk as well as reducing capitalization levels further, forcing a further round of fire sales, and so on. The fall in valuation levels is composed of a first chunk attributable to the initial piece of bad news, as well as to a second chunk entirely due to the non-information related feedback effects of market participants. In formal models of this phenomenon, the feedback effects can be many times larger than the initial seed of bad news.

2.1 Leading Model

We illustrate the ideas sketched above through the dynamic model of endogenous risk developed in Daníelsson, Shin, and Zigrand (2011). The model has the advantage that it leads to a rational expectations equilibrium that can be solved in closed form. Here, we give a thumbnail sketch of the workings of the model. The detailed solution and the properties of the model can be found in Daníelsson, Shin, and Zigrand (2011).

Time flows continuously in $[0, \infty)$. Active traders (financial institutions) maximize profit by investing in risky securities as well as the riskless security. The financial institutions are subject to a short-term Value–at–Risk (VaR) constraint stipulating that the Value-at-Risk is no higher than capital (tangible common equity), given by V_t. In order to emphasize the specific contribution of risk constraints to endogenous risk, all other channels are switched off. The short rate of interest r is determined exogenously.

[1] See Daníelsson and Zigrand (2008) where a VaR constraint lessens a free-riding externality in financial markets, and Adrian and Shin (2010) for a model whereby a VaR constraint is imposed in order to alleviate a moral hazard problem within a financial institution.

Given rational behavior, prices, quantities and expectations can be shown to be driven in equilibrium by a set of relevant aggregate variables, chiefly the (mark–to–market) capitalization level of the financial sector. The financial institutions are interacting with each other and with passive investors (the non–financial investors, including individual investors, pension funds and so forth).

The risky security has an (instantaneous) expected equilibrium return μ_t and volatility of σ_t. The equilibrium processes μ and σ are endogenous and forward looking in the sense that the beliefs $(\tilde{\mu}_t, \tilde{\sigma}_t)$ about actual moments (μ_t, σ_t) are confirmed in equilibrium. Financial institutions in equilibrium hold diversified portfolios commensurate with those beliefs, scaled down by their effective degree of relative risk aversion γ_t (solved in equilibrium) imposed upon them by the VaR constraints:

$$D_t = \frac{V_t}{\gamma_t} \Sigma_t^{-1} (\mu_t - r) \tag{1}$$

with D the monetary value of their holdings.

The model is closed by introducing value investors who supply downward-sloping demand curves for the risky asset. The value investors in aggregate have the exogenous demand schedule for the risky asset y_t where

$$y_t = \frac{\delta}{\sigma_t^2} \left(z_t - \ln P_t \right) \tag{2}$$

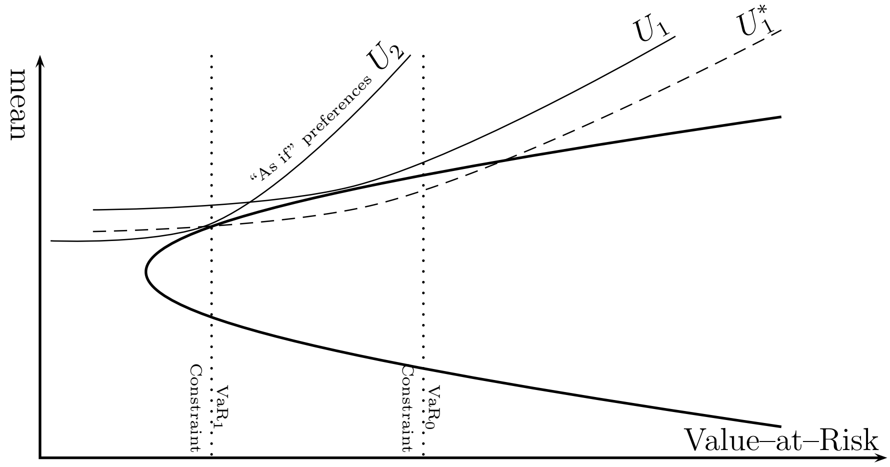
where P_t is the market price for risky asset and where dz_t is a (favorable) Itô demand shock to the demand of the risky asset. Each demand curve can be viewed as a downward sloping demand hit by demand shocks, with δ being a scaling parameter that determines the size of the value investor sector.

Even though the financial institutions are risk neutral, the VaR constraints imply that they are compelled to act "as if" they were risk averse and scale their risky holdings down if VaR is high compared to their capitalization level: [2]

$$\begin{aligned} &\text{coefficient of effective relative risk aversion} \\ &\qquad = \text{coeff. of relative risk aversion} \\ &\qquad + \text{Lagrange multiplier on the VaR constraint} \end{aligned}$$

[2]This goes back to Daníelsson and Zigrand (2008), first circulated as Daníelsson and Zigrand (2001).

Figure 2: Changing Risk Appetite



Thus, even if the traders were risk-neutral, they would act in a risk-averse way depending on how tightly the risk constraint is binding. Figure 2 illustrates the general intuition as to why risk aversion is effectively fluctuating randomly as a function of the tightness of the VaR constraints of financial institutions. For the purpose of illustration, we draw the indfference curves consistent with some degree of inherent risk aversion. Suppose the FI initially has sufficient capital so that its Value–at–Risk constraint is non–binding at VaR_0. In this case, the indifference curve is U_1. Suppose investment opportunities stay constant but capital is reduced, so that the VaR constraint becomes binding at VaR_1. Therefore, the optimal portfolio chosen is no longer a tangency point between the indifference curve (shifted down to U_1^*) and the efficient set. An outside observer might conclude that the VaR constrained portfolio choice actually was the choice of a more risk averse investor (steeper indifference curve U_2): "as if" risk aversion increased. In the dynamic model, investment opportunities change endogenously as well of course.

In a rational expectations equilibrium, the actual volatility of prices implicit in this equation, σ_t, and the beliefs about the volatility, $\tilde{\sigma}_t$, must coincide. To compute the actual volatility of returns, we resort to Itô's Lemma and get

$$\begin{aligned}\sigma_t &= \eta\sigma_z + \underbrace{\tilde{\sigma}_t \times (\text{diffusion of } V_t)}_{\text{vol due to FI's wealth-VaR effect}} + \underbrace{V_t \times (\text{diffusion of } \tilde{\sigma}_t)}_{\text{vol due to changing beliefs}} \\ &= \eta\sigma_z + V_t\left[\tilde{\sigma}_t + V_t\frac{\partial\tilde{\sigma}}{\partial V_t}\right].\end{aligned}$$

Equilibrium volatility is determined as the fixed point where $\sigma_t = \tilde{\sigma}_t$, which entails solving for the function $\sigma_t(V_t)$ from an ordinary differential equation. Daníelsson, Shin, and Zigrand (2011) show that there is a unique closed form solution given by

$$\sigma(V_t) = \eta\sigma_z\frac{\alpha^2\delta}{V_t}\exp\left\{-\frac{\alpha^2\delta}{V_t}\right\} \times \mathrm{Ei}\left(\frac{\alpha^2\delta}{V_t}\right) \tag{3}$$

where $\mathrm{Ei}(w)$ is the well-known exponential integral function[3]:

$$\mathrm{Ei}(w) \equiv -\int_{-w}^{\infty}\frac{e^{-u}}{u}du \tag{4}$$

The $\mathrm{Ei}(w)$ function is defined provided $w \neq 0$. The expression $\alpha^2\delta/V_t$ which appears prominently in the closed form solution (3) can be interpreted as the relative scale or size of the value investor sector (parameter δ) compared to the banking sector (total capital V_t normalized by VaR).

The closed form solution also reveals much about the basic shape of the volatility function $\sigma(V_t)$. Consider the limiting case when the banking sector is very small, that is, $V_t \to 0$. Then $\alpha^2\delta/V_t$ becomes large, but the exponential term $\exp\{-\alpha^2\delta/V_t\}$ dominates, and the product of the two goes to zero. However, since we have exogenous shocks to the value investor demands, there should still be non-zero volatility at the limit, given by the fundamental volatility $\eta\sigma_z$. The role of the $\mathrm{Ei}(w)$ term is to tie down the end point so that the limiting volatility is given by this fundamental volatility.

The endogenous term reduces the fundamental volatility if the FI are sufficiently capitalized (i.e. if V_t is large enough) and dramatically increases the volatility in a non-linear fashion as V drops, as depicted on Figure 3 where the properties of our model are illustrated graphically.

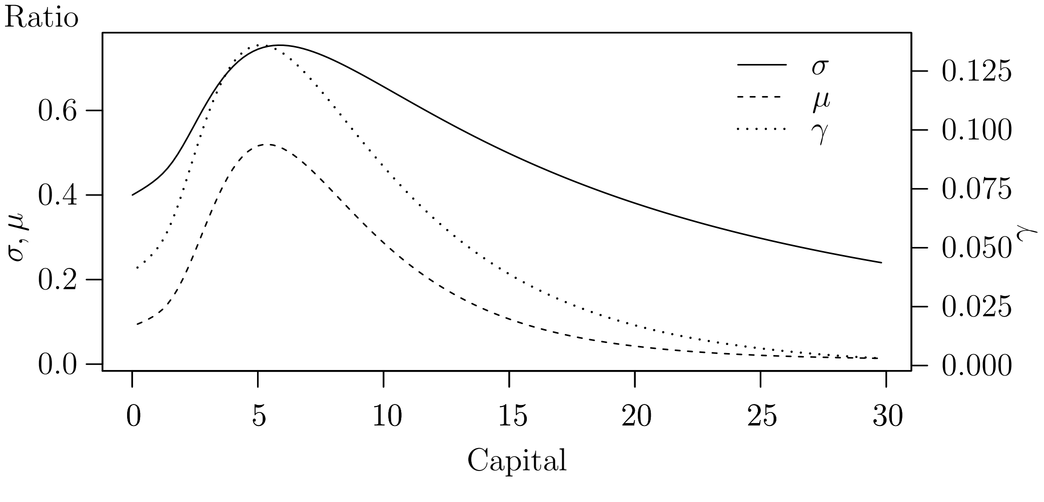
[3] http://mathworld.wolfram.com/ExponentialIntegral.html

The Figure plots the equilibrium diffusion σ_t, the drift (expected return) μ_t and risk–aversion γ_t as a function of the state variable V_t. The parameters chosen for all plots in this paper are $r = 0.01$, $\delta = 0.5$, $\alpha = 5$, $\sigma_z = 0.4$, $\eta = 1$ and $c = 10$. There is nothing special about this particular parameter constellation, almost any other combination of parameters would generate the same shapes of the plots and hence results. For this reason, the choice of parameters is not very important for understanding the basic intuition as a model. However, different parameters generate different magnitudes, and for this reason, we roughly calibrated the parameters so the outcome would correspond to daily returns for relatively high risk stocks. In future work we are planning to estimate the model, i.e., make the parameters be data driven.

σ is the equilibrium volatility and γ_t is the endogenous effective risk aversion. Higher levels of capital represent a well capitalized banking sector, where volatility is below the fundamental annual volatility of 40%. As capital is depleted, volatilities, risk premia and Sharpe ratios increase.

In the extreme case where capital gets fully depleted to zero, the economy has no financial institutions, and so volatility is equal to the fundamental volatility. With a well capitalized financial sector, variance is low as the financial sector absorbs risk.

Figure 3: Equilibrium Risk Premia, Volatility and Risk–Aversion/Sharpe Ratio

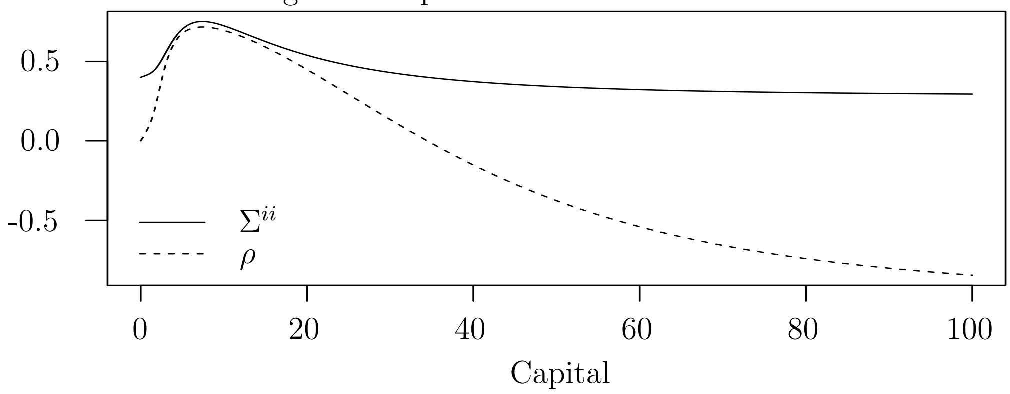


In the leading model, volatility, risk premia as well as generalized Sharpe ratios are all countercyclical, rising dramatically in a downturn, providing ex ante compensation for the risks taken as illustrated in Figure 3. These features align our model with available empirical evidence. As can be seen from the graphs, market volatility is a function of the state variable V_t and so the model generates stochastic volatility.

Volatility is lower than fundamental news–induced volatility in times when the financial sector is well–capitalized, when financial institutions play the role of a buffer that absorbs risks and thereby reduce the equilibrium volatility of financial markets. FI are able to perform this function because by having a sufficient capital level, their VaR constraints are binding less hard, allowing them to act as risk absorbers. However, as their capital is depleted due to negative shocks, their risk constraints bind harder inducing them to shed risk and amplify market distress.

A similar picture emerges in a multivariate version of the model when there is more than one risky security. The added dimension allows us to address the emergence of endogenous correlation in the returns of risky assets whose fundamentals are unrelated. We illustrate the properties of the bivariate case in Figure 4. Here, Σ^{ii} is the variance of the returns on security i and ρ^{ij} is the correlation coefficient between the returns on securities i and j, where securities i and j are intrinsically uncorrelated.

Figure 4: Equilibrium correlations



3 Feedback Effects and Empirical Predictions

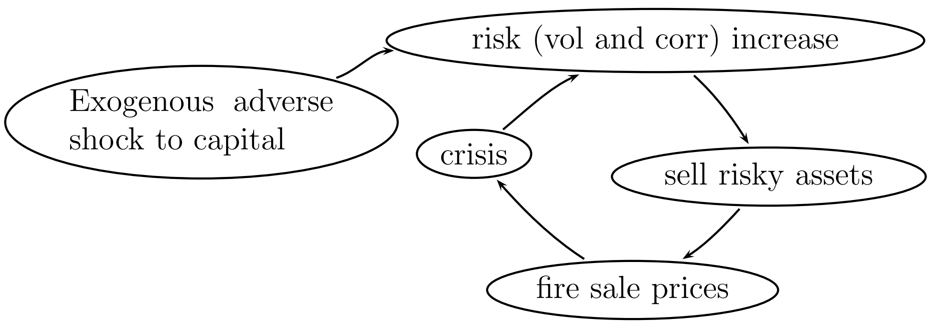
Some features of the model of endogenous risk can be presented under several sub-headings. We begin with the role of constraints in propagating feedback.

3.1 Constraints and Feedback

The main driver of the results in the leading model are feedback effects which increase in strength along with the homogeneity in behavior and beliefs amongst financial institutions (financial institutions), especially during crises. Just as in the example of the Millennium Bridge where an initial gust of wind eventually causes the pedestrians to react *identically* and *at the same time*, constraints on financial institutions together with marking–to–market can lead to synchronized institutional behavior in response to an external shock. The ultimate effect is to synchronize the behavior of all financial institutions, dampening risks in the up-turn and amplifying risks in the downturn.

For a well-capitalized financial sector, correlations between the various securities are reduced since the financial institutions have ample capacity to absorb risk. For low levels of capital, however, volatility increases as shown in Figure 4. This gives rise to an adverse feedback loop. When capital falls, financial institutions need to shed their risky exposures, reducing prices and raising volatility across all securities. This in turn forces financial institutions to engage in another round of fire sales, and so forth. This is illustrated in Figure 3. These effects are summarized in Figure 5, where an initial adverse shock to capital leads to an adverse feedback loop.

Figure 5: Feedback in Leading Model

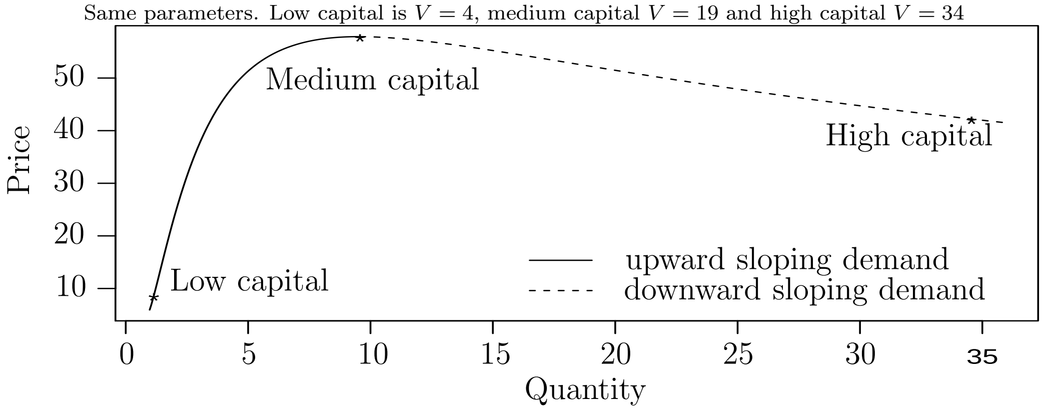


Within the leading model, the feedback effects can be understood in terms of the slope of the demand functions of the financial institutions. When the financial sector is undercapitalized, an adverse shock prompts the financial institutions to shed risky securities because risk constraints bind harder and because the price drop leads to a capital loss. So a lower price prompts a sale rather than a purchase. This sale in turn prompts a further fall in price and the loop closes.

This is demonstrated in Figure 6 which plots supply and demand responses. Note that Figure 6 charts total demand response taking account of changes in V and volatility in equilibrium, not the demand curve in a partial equilibrium sense at a given V for different prices. In other words, Figure 6 shows the continually evolving demand response as the FI continues buying or selling. The reduced-form demand function is upward sloping[4] for low levels of capital. As prices increase so does demand. This phenomenon is what gives rise to the amplification effects in the Monte Carlo simulations in Figures 7 and 8. As the FI becomes better capitalized, its equilibrium demand function assumes the typical downward shape. Instead, for small V, the FI increases endogenous risk, while for larger capital levels it decreases endogenous risk.

Figure 6: The Demand Function

Same parameters. Low capital is $V = 4$, medium capital $V = 19$ and high capital $V = 34$

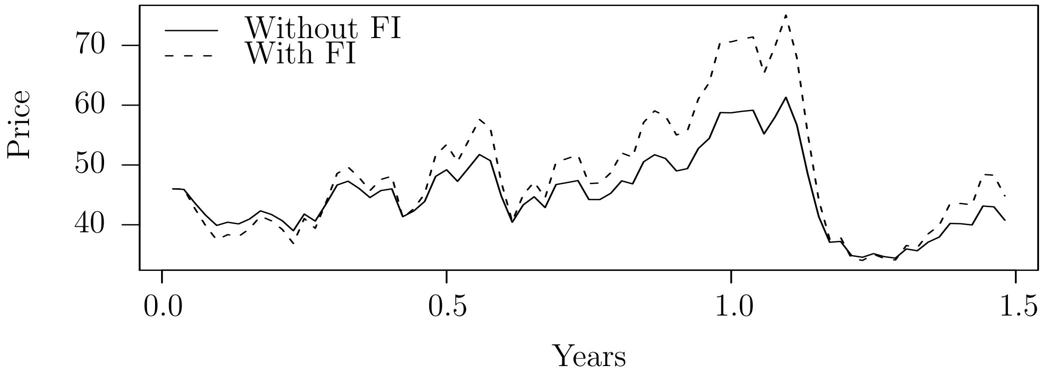


We further demonstrate this feature by means of simulations of price paths. Figure 7 shows a typical path with a year and a half worth of prices in a univariate model in the absence of risk-constrained traders (and hence where prices follow a geometric Brownian motion). The prices in the absence of risk constraints (P_a for autarchy) rise slowly (at a mean rate of return equal to the risk free rate), followed by a crash in the beginning of the second year. For the same sequence of fundamental shocks the prices when there

[4]An early example of an endogenous risk-type result with an upward sloping demand functions comes from Gennotte and Leland (1990). In their model of portfolio insurance, delta hedging of a synthetic put option requires the delta–hedger to sell a security into a falling market, magnifying the volatility.

Figure 7: Simulation of Price Paths

Start at $V = 12$. In first case σ and μ are constants, since the FI exerts no price impact when not present in market.



are risk-constrained FIs (P_{FI}) show a much bigger rise followed by a bigger crash.

3.2 Endogenous Risk and Comovements

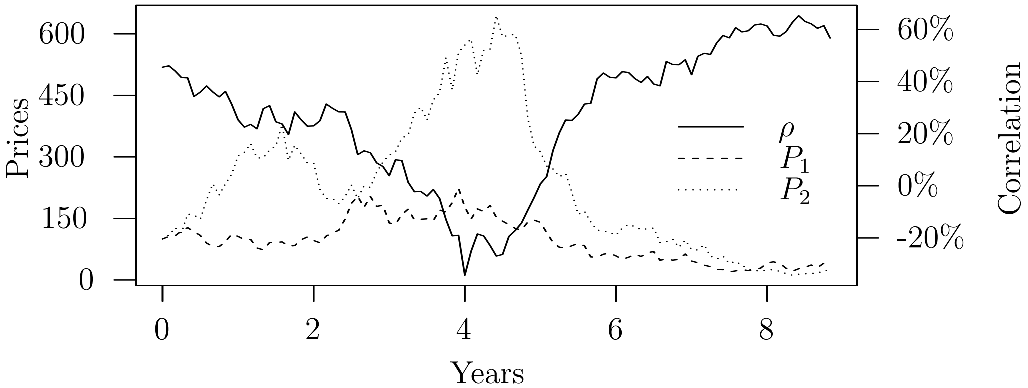
Correlations (or more generally dependence, linear or non-linear) between risky assets are of key importance in characterizing market returns. In the absence of correlations in the fundamentals, diversification can enable the mitigation of risk. However, endogenous risk and the associated risk constraints imply that assets whose fundamentals are unrelated may still give rise to correlations in market prices due to the fluctuations in risk constraints of the FIs. Since risk constraints give rise to "as if" risk aversion, the correlation in return is associated with fluctuations in the degree of as-if risk aversion. The sudden increase in correlations during the crisis is well documented and has repeatedly wrong-footed sophisticated proprietary trading desks in many banks that have attempted to exploit historical patterns in asset returns.[5] In crises, volatilities and implied volatilities shoot up at the same time, whether it be the implied volatility of S&P 500 options or of

[5]This occurs in equilibrium in our model, with the FI portfolio that gives rise to the described offloading itself chosen in equilibrium.

interest rate swaptions. Again, all those spikes in comovements are driven by the same unifying heightened effective risk aversion factor, itself driven by the capitalization level in the financial sector.

We illustrate this by simulating price paths for the bivariate model, shown in Figure 8. The correlations initially decline slowly, the price of the second asset increases sharply while the price of the first asset is steady. Then in year 4, an averse shock to its price leads to a sharp increase in correlations, causing the price of the first asset to fall as well.

Figure 8: Simulation of Prices and Correlations



As we see from Figure 4, variances move together, and so do variances with correlations. This feature is consistent with the empirical evidence in Andersen, Bollerslev, Diebold, and Ebens (2001) who show that

> "there is a systematic tendency for the variances to move together, and for the correlations among the different stocks to be high/low when the variances for the underlying stocks are high/low, and when the correlations among the other stocks are also high/low."

They conjecture that these co–movements occur in a manner broadly consistent with a latent factor structure. A good candidate for such a latent factor would be the tightness of the risk constraint implied by FIs' capitalization discussed above.

3.3 Endogenous Risk and the Implied Volatility Skew

Options markets offer a direct window displaying endogenous risk in simple graphical terms. Equity index options markets have, at least since 1987, consistently displayed a skew that is fanning–out over longer maturities. Out–of–the–money puts have much higher implied volatilities than out–of–the–money calls. Shorter dated options have a more pronounced skew compared to longer dated options. The fear in the market that drives such features in the options market seems to be of a latent violent downturn (against which the expensive out–of–the money puts are designed to protect), while strings of positive news over the longer term are expected to lead to less volatile returns over longer horizons, the great moderation. The sharp downturn is not expected to be permanent, hence the mean–reverting fanning–out of the skew. We find this result in our model, see Figure 9.

Our discussion of the way in which endogenous risk plays out in the market is a promising way to address the stylized empirical features in the option market. Endogenous risk embeds an asymmetry between the downside and the upside. Depletion of capital and endogenously increasing risks generate sharply higher volatility, while no such corresponding effects operate on the upside. The widely accepted version of the events of the stock market crash of October 1987 (see for instance the formulation of Gennotte and Leland (1990)) places at the center of the explanation the feedback effects from synthetic delta–hedged puts embedded in portfolio insurance mandates. The "flash crash" of May 6th 2010 almost certainly has more complex underpinnings, but it would be a reasonable conjecture that the program trades executed by algorithmic high frequency traders conspired in some way to create the amplifying feedback loop of the kind seen in October 1987.

As well as the omnipresent implied volatility skew at any given moment in time, our model also predicts that implied volatilities move together in a crisis, which has indeed occurred, across securities as well as across asset classes.

4 Implications for Financial Regulation

As we have seen, the financial system can go through long periods of relative tranquility, but once endogenous risk breaks out, it grips the entire financial

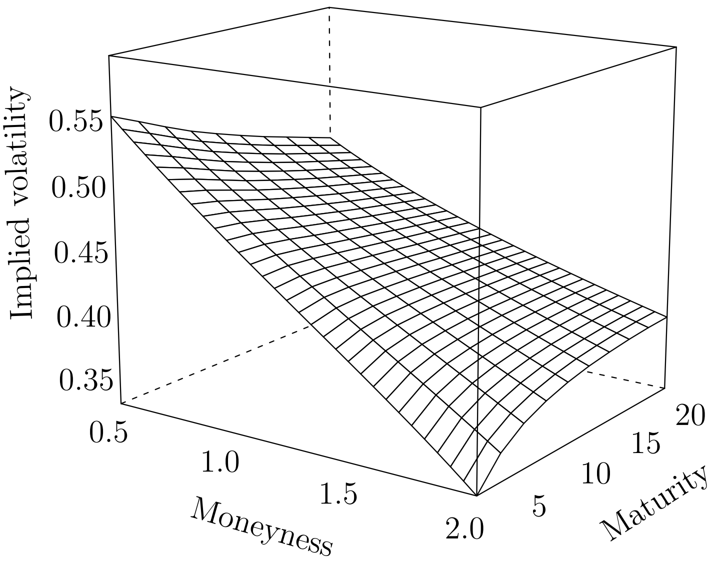


Figure 9: Implied Volatility Surface

markets. This happens because the balance sheets of large financial institutions link all securities. Our results hold important implications for financial regulation. Regulators will need to be prepared for prospect that once a storm hits, it has a significant probability of being a "perfect storm" where everything goes wrong at the same time. In the presence of endogenous risk, the focus of regulatory policy should be more towards the system, rather than individual institutions. Even if the economy starts out stable, continued prosperity makes way to an unstable system. An apposite comment is given in Crockett (2000):

> "The received wisdom is that risk increases in recessions and falls in booms. In contrast, it may be more helpful to think of risk as increasing during upswings, as financial imbalances build up, and materialising in recessions."

This reasoning is also consistent with Minsky's financial instability hypothesis. Stability can sow the seeds of future instability because financial institutions have a tendency to react to the tranquility by building up their risky asset holdings that increase the thickness of the left tail of the future outcome distribution, which ultimately undermines stability. At some point, a negative shock arrives, and markets go through an abrupt correction. The longer is the period of dormant volatility, the more abrupt and violent is the correction when it arrives.

While our model of endogenous risk has a single state variable (the FI capital level V), it would be possible to develop more complex versions where the history of the financial system affects future crisis dynamics. One way of doing so would be to posit market participants who extrapolate from last market outcomes in the manner recommended by standard risk management systems that use time series methods in forecasting volatility. One popular version of belief updating is the exponentially-weighted moving average (EWMA) method that forecasts future volatility as a function of last return realizations.

To the extent that volatility is simply dormant during upturns rather than being absent, there is a rationale for counter-cyclical tools that lean against the build-up of vulnerabilitis during upturns.

Our model of endogenous risk is consistent with leveraging and deleveraging of financial intermediaries as discussed by Adrian and Shin (2010) and Daníelsson, Shin, and Zigrand (2011). Credit increases rapidly during the boom but increases less rapidly (or even decreases) during the downturn, driven partly by shifts in the banks' willingness to take on risky positions over the cycle. The evidence that banks' willingness to take on risky exposures fluctuates over the cycle is especially clear for financial intermediaries that operate in the capital market.

Deleveraging causes risk aversion to curtail credit in the economy, leading to a downturn in economic activity. The role of a liquidity and capital provider of last resort can be important in dampening financial distress. While financial institutions may be overly leveraged going into a crisis, the endogenous feedback effects may lead to excessive deleveraging relative to the fundamentals of the economy, prompting institutions to curtail lending to the real economy.

4.1 Forecasting Risk

Our model of endogenous risk has direct implications for empirical risk forecasting. Almost every model used in practice for forecasting risk, assumes financial risk is exogenous. In other words, the financial institutions are price takers where their trading decisions do not affect price dynamics. So long as individual trading portfolios represent a relatively small part of overall market capitalization and financial institutions are different, an assumption of exogenous risk is relatively innocuous. This is likely to be the situation most of the time, perhaps 99.9% of all trading days.

It is however the other 0.1% that matter most for financial stability. That this when market turmoil becomes extreme, constraints become especially binding and financial institutions start acting in harmony, shedding the same risky assets and buying the same safe assets. At that time, financial risk becomes highly endogenous implying that financial risk forecast models based on an assumption of exogeneity of risk are likely to fail.

The underlying reason is the dual role of market prices. On the one hand, market prices reflect the current value of an asset, but on the other, they also reflect the constraints on financial institutions, and hence are an imperative to act. Constraints may not be binding tightly during calm times but may become highly restrictive during crisis, leading to adverse feedback between increasingly tight constraints and falling asset prices.

This suggests that market prices during periods of calm may be a poor input into forecast models, since any reliable empirical systemic risk model needs to address the transition from non–crisis to crisis. Market prices during calm times may not be informative about the distribution of prices that follow after a crisis is triggered. In addition, price dynamics during one crisis may be quite different in the next, limiting the ability to draw inference from crisis events.

Consequently, risk models are likely to underestimate risk during calm times and overestimate risk during crisis — they get it wrong in all states of the world.

4.2 Empirically Modelling Systemic Risk

The tendency of risk models to fail during crisis, as discussed above has particular implications for the the burgeoning field of empirical systemic risk modeling. Here, the question of interest is not the risk of financial institutions failing, but rather the risk of cascading failures. Consequently, the challenge for a reliable systemic risk model is to capture the risk of each systematically important institution, as well as their interactions. These models generally attempt to use observed market variables to provide an indication of the risk of some future systemic event. The current crop of systemic risk models is examined empirically by Danielsson, James, Valenzuela, and Zer (2011) who find that because of high model risk, such models are highly unreliable.

Systemic risk is concerned with events that happen during crisis conditions, looking far into the tails of distributions. This makes the paucity of relevant data a major concern. Over the last fifty or so years we have observed less than a dozen episodes of extreme international market turmoil. Each of these events is essentially unique, driven by different underlying causes. We should therefore expect that models that are fed with inputs from calm periods will perform much less well during periods of stress.

As a consequence, we feel that the current crop of systemic risk forecast models is unlikely to perform as expected. Instead, such models would need to incorporate endogenous risk explicitly if they are to capture the the underestimation of systemic risk prior to a crisis event, as well as the overestimation of systemic risk during the crisis event, both of which are damaging.

4.3 Leverage and Capital

Endogenous risk implies *non–linearities* due to the feedbacks that conspire to make the regulator's problem very difficult. Capital held by the FI is proportional to the risk–tolerance of the non-financial sector times the square of the tightness of the VaR constraint.[6] Leverage in the leading model is

$$\frac{\text{assets}}{\text{capital}} = \frac{1}{\text{VaR}_t}$$

[6]In Basel II, the level of tightness of the VaR constraints for market risk is three times the relevant quantile.

where VaR_t is proportional to volatility over short periods. In other words, the growth rate of the capital ratio is equal to the growth rate of volatility. Leverage is procyclical and builds up in quiet booms where VaR is low and unwinds in the crisis. In practice, deleveraging is exacerbated by increased haircuts, reinforcing the feedback loops further through this second channel of forced delevering, see Xiong (2001), Gromb and Vayanos (2002), Geanakoplos (2010), Adrian and Shin (2010) and Brunnermeier and Pedersen (2009). Of course, if capital requirements are not risk–based, for example by using the leverage ratio, procyclicality is not increased by the capital requirements.

Financial crises and strong destabilizing feedback effects naturally occur when capital levels are too low, as can be seen in the Figures above. When capitalization is adequate, financial institutions allow absorption and diffusion of risk, resulting in calmer and more liquid markets to prevail. But endogenous risk raises the fundamental level of volatility in the economy during periods of low capitalization and diminishes the fundamental level of volatility otherwise.

Low capitalization therefore go hand–in–hand with low liquidity. [7] The first effects of the recent crisis became visible through a liquidity crisis in the summer of 2007, where central bank interventions were crucial, but then the crisis quickly turned into a solvency crisis. The liquidity crisis was the harbinger of the later solvency crisis. The two must be linked in any account of the recent crisis.

Countercyclical measures that reduce the feedback loops can be one way to mitigate the boom bust cycle. Capital adequacy therefore has a major role to play. Since the strength of the adverse feedbacks is very sensitive to the procyclicality of capital adequacy rules, a sufficient capital buffer needs to be imposed in conjunction with countercyclical rules that lean against the build-up of vulnerabilities during the boom. A large capital buffer that either cannot be used, or that imposes positive feedback loops, is counterproductive exactly in those situations where it would be needed most. This is aptly demonstrated by Goodhart's metaphor of the weary traveller and the lone cab driver, (Goodhart, 2009, ch 8). A weary traveler arrives late at night by train to an unknown town. One taxi is waiting and the traveler goes to it requesting to be taken to his hotel. The taxi driver refuses and points to

[7] Recall the earlier discussion on the critical level of capital that would allow the financial system to perform its socially useful role.

a sign on the wall that says "local regulations stipulate that a taxi must be present at the taxi stand at all times." In addition, excessive bank capital tied up in government bonds is socially costly because it hampers the socially optimal activities of banks, to transform maturities and to take on risks by lending.

Risk builds up during the good times when perceived risk is low and imprudent leverage and complex financial networks build up quietly, perhaps aided by moral hazard (Altunbas, Gambacorta, and Marques-Ibanez (2010)). It is only in a crisis that this risk materializes and becomes plainly visible. A promising avenue to think about capital adequacy, based on an idea in chapters 10 and 11 in Goodhart (2009), that deserves further thought would be to require financial institutions to set aside an initial capital buffer, plus an additional variation capital buffer that is a function of the *growth rate* of various assets (both on and off balance sheet) as well as of the maturity mismatch (and of the probable liquidity in a crisis) imposed by those asset classes.

The variation buffer can then be naturally and countercyclically depleted in a downturn, provided the financial institutions do not feel compelled to take large amounts of hidden toxic assets back onto their balance sheets during the downturn. As far as we know, this idea has yet to be formally analyzed.

Note, however, that while countercyclical regulatory capital requirements are a step forward,[8] they are not sufficient to stem all procyclical forces in the markets. For instance, financial institutions will still allocate capital to traders according to a VaR type formula, forcing them to unwind risky positions if risk shoots up. Haircuts will always go up in a downturn. Central clearing houses will impose daily settlement and contribute to procyclicality. Net derivative positions will still be at least partly delta hedged, implying reinforcing feedback effects (on top of the VaR induced feedback effects) if delta hedgers are net short gamma.

In summary, the omnipresence and inevitability of adverse procyclical spillover effects in financial markets reinforces the need for countercyclical *regulatory* capital rules.

[8]Whereas regulators relaxed capital adequacy requirements during the S&L crisis, no such formal countercyclical regulatory forbearance seems to have been applied in this crisis.

4.4 Endogenous Risk and Central Clearing Counterparties (CCPs)

The volume of over the counter (OTC) derivatives exceeds the global annual GDP by some margin and such derivatives have been widely blamed for their contribution to systemic risk. In particular, the opaque nature of the OTC market, coupled with counterparty risk have been singled out as especially dangerous. Consequently, there are ongoing discussions about moving a non–negligible fraction of the OTC trade onto central counterparties (CCPs), with the expectation that the most dangerous systematic impacts of OTC would be mitigated if they were forced to be centrally cleared.

This directly related to the very recent development of credit value adjustment (CVA) desks in financial institutions which now are some of the largest desks in financial institutions.

The impact of moving OTC derivatives to CCPs are analyzed by Zigrand (2010) in the context of endogenous risk. He notes that CCPs need to protect themselves from counterparty risk, implying institutionalized initial margin and maintenance margin rules based on frequent marking–to–market. We have observed above that such margin calls bear the hidden risk of exasperating downward spirals. Endogenous risk appears in at least five guises.

First, an important question to ask is to what extent the current OTC markets resemble CCPs, i.e. how many feedback rules are embedded already in OTC? Daily collateral exchanges in the OTC market play the role of daily margin calls, and up–front collateral ("independent amount") plays the role of the initial margin. So some of the mechanisms to reduce counterparty risk are also applied in the OTC market, of course. Still, it seems that a sufficiently large part of the OTC exposures have not been dealt with in this way. ISDA states that 70% of OTC derivatives trades are collateralized, while a survey by the ECB (2009) estimated that EU bank exposures may be collateralized well below this. Singh (2010) estimates under collateralization is about 2 trillion dollars for residual derivative payables. This justifies our working hypothesis that should trade move onto CCPs, it is conceivable that feedback effects become stronger than they currently are.

The second appearance of endogenous risk is the fallacy of composition. It is not true that if all products are cleared, and therefore appear to be safe, that the system overall is safe. Indeed, it probably is safer to only require clearing of products that are mature and well understood, for the risk of CCP failure imposed by an immature contract is very costly.

The third aspect of endogenous risk arises in the way the *guarantee fund* of the CCP is replenished. CCPs provide very little guidance on how exactly they expect to manage the replenishment by member firms. It would appear natural to presume that the CCP would replenish through risk–sensitive (e.g. VaR) rules whereby in periods of higher risk or past capital losses, the CCP will ask for new capitalizations. Member firms being broker dealers, they may be forced to sell risky assets or increase haircuts to their debtors to raise the required capital, thereby contributing to procyclicality in the market place. Even the original move from OTC to CCP will require such a sale as there currently simply is not enough collateral (Singh, 2010).

The fourth aspect of endogenous risk has to do with the number of CCPs. Assume that one FI (call it FI1) currently trades with another one, FI2, in the OTC markets. Assume also, as occurs commonly, that the two financial institutions have two open exposures to each other that roughly net out. If both exposures were cleared by the same CCP, then a deterioration in the markets would have no effects on the variation margin calls (but may have an effect on the initial margin which we ignore for simplicity), and therefore will not create any feedback loops. If however both positions were cleared on two separate CCPs with no links between the two CCPs, or one position on one CCP and the other one stays bilaterally cleared, then an increase in volatility will lead, regardless of the direction of the markets, to margin calls and the selling of risk. Again, if capital is fixed in the short run, the individually prudent course of action is to shed risks. This affects prices, which in turn affects the mark–to–market capital and VaR of *all* financial institutions, not just of the two engaged in the original trades. All financial institutions start to act in lock step, and the bridge wobbles. The fallacy of composition appears again in the sense that even if every exposure is centrally cleared, the overall exposure is not centrally cleared when there are multiple CCPs. Cross–margining would mitigate this, as occurs for instance for options through the OCC hub between ICE Clear US and the CME.

The fifth endogenous risk feedback effect again has its origin in mark–to–market. Financial institutions mostly know when a contract is not liquid, and some financial institutions spend enormous amounts of resources on trying to properly value a derivatives position. If such a contract was centrally cleared and the price made available to the market, this mark may give the appearance of “officially correct audited market prices.”[9] But it is unavoidable that relatively illiquid products will get marks that will force all financial institutions, even the ones that have not traded that day and the ones whose accurate internal models predict better marks, to mark their books to these new CCP marks. Since by assumption this market is illiquid, the demand is inelastic and a big sale on one day will move prices and generate strong feedbacks through forced selling, leading to a quick drying–up of liquidity.

5 Conclusion

Each financial crisis has its own special features, but there are also some universal themes. In this paper, we have focused on the role of endogenous risk that propagates through increasingly tight risk constraints, reduction in risk-bearing capacity and increased volatility. Deleveraging and the shedding of risk imply that asset price movements increase manifold through the feedback effects that are hard-wired into the financial system itself. This paper has aimed at spelling out the precise mechanism through which endogenous risk manifests itself and has suggested ways of mitigating it.

[9] CCPs do have put in place mitigating procedures to try to make sure that the marks are actually prices at which clearing members would be willing to trade.

References

ADRIAN, T., AND H. S. SHIN (2010): "Liquidity and Leverage," *Journal of Financial Intermediation*, 19, 418–437.

ALTUNBAS, Y., L. GAMBACORTA, AND D. MARQUES-IBANEZ (2010): "Does monetary policy affect bank risk-taking?," Discussion paper, BIS, Working Paper 298.

ANDERSEN, T., T. BOLLERSLEV, F. DIEBOLD, AND H. EBENS (2001): "The distribution of realized stock return volatility," *Journal of Financial Economics*, 61, 43–76.

BRUNNERMEIER, M., AND L. H. PEDERSEN (2009): "Market Liquidity and Funding Liquidity," *Review of Financial Studies*, 22, 2201–2238.

CROCKETT, A. (2000): "Marrying the micro- and macro-prudential dimensions of financial stability," The General Manager of the Bank for International Settlements; `http://www.bis.org/review/rr000921b.pdf`.

DANÍELSSON, J., P. EMBRECHTS, C. GOODHART, C. KEATING, F. MUENNICH, O. RENAULT, AND H. S. SHIN (2001): "An Academic Response to Basel II," The New Basel Capital Accord: Comments received on the Second Consultative Package `http://www.bis.org/bcbs/ca/fmg.pdf`.

DANIELSSON, J., K. JAMES, M. VALENZUELA, AND I. ZER (2011): "Dealing with Systemic Risk When We Measure Systemic Risk Badly," mimeo, London School of Economics, `www.RiskResearch.org`.

DANÍELSSON, J., AND H. S. SHIN (2003): "Endogenous Risk," in *Modern Risk Management — A History*. Risk Books, `http://www.RiskResearch.org`.

DANÍELSSON, J., H. S. SHIN, AND J.-P. ZIGRAND (2011): "Balance Sheet Capacity and Endogenous Risk," `http://www.RiskResearch.org`.

DANÍELSSON, J., AND J.-P. ZIGRAND (2001): "What Happens When You Regulate Risk? Evidence from a Simple Equilibrium Model," Working paper, LSE, `www.RiskResearch.org`.

——— (2008): “Equilibrium asset pricing with systemic risk,” *Economic Theory*, 35, 293–319, `http://www.RiskResearch.org`.

Geanakoplos, J. (2010): “The Leverage Cycle,” *The 2009 NBER Macroeconomics Annual*, pp. 1–65.

Gennotte, G., and H. Leland (1990): “Market Liquidity, Hedging, and Crashes,” *American Economic Review*, pp. 999–1021.

Goodhart, C. (2009): *The Regulatory Response to the Financial Crisis*. Edward Elgar.

Gromb, D., and D. Vayanos (2002): “Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs,” *Journal of Financial Economics*, 66, 361–407.

Singh, M. (2010): “Collateral, Netting and Systemic Risk in the OTC Derivatives Market,” Discussion Paper WP/10/99, IMF.

Xiong, W. (2001): “Convergence Trading with Wealth Effects: An Amplification Mechanism in Financial Markets,” *Journal of Financial Economics*, 62, 247–292.

Zigrand, J.-P. (2010): “What do Network Theory and Endogenous Risk Theory have to say about the effects of CCPs on Systemic Stability?,” *Financial Stability Review, 14, Banque de France.*