FUNG
GLOBAL
INSTITUTE

Risk and Uncertainty in Financial Models – Implications for Asia

Michael Edesess

April 2014

Key Insights

- Overreliance on mathematical risk models, such as Value at Risk (VaR), contributed to misreading the risks and uncertainties leading to the global financial crisis (GFC). These models are built on unrealistic assumptions and are vulnerable to data errors. They also ignored radical uncertainty and are unreliable at best, almost meaningless at worst.

- Uncertainty is rooted in dynamic, non-linear feedback behavior between market participants that can lead to unpredictable (by current risk models) and sometimes chaotic outcomes. Radical uncertainty exists because participants cannot predict counterparties' reactions with precision, even in a simple two-party game. Systems theorists from other disciplines, such as engineering, physics and biology are aware that feedback loops can either be negative, meaning that the system will revert back to stable equilibrium, or positive, meaning that the system could explode and enter into chaos or breakdown.

- Dealing with uncertainty requires trust to create mutually predictable outcomes; however, the breakdown of trust creates uncertain behavior at the system level.

- True uncertainty cannot be quantified by a probability; hence the risk models failed to predict "black swan" events, such as the 'London Whale' incident, in which JP Morgan is reported to have incurred US$6 billion of losses (Bernanke 2004). Panic resulting in loss of trust in counterparties and in the system led to cascading bank runs.

- Post-GFC, **trust** is essential in stabilizing the financial system. In 2008, the U.S. and other governments took sudden and drastic measures to preserve trust and prevent panic from a wholesale bank run. The state restored trust in the system, but only by taking on high debt and other costs, including 'moral hazard'.

- The industry and regulators must adopt new quantitative and qualitative methods to model at the microcosmic level the 'ebb and flow' of collective trust in the macro-economy. These models should be embedded in cultural, traditional and ethical contexts to fully understand the financial system.

- 'Thinking the unthinkable' requires us to recognize that the free-market system may not always be able to address its endogenous problems. The state should have a backup plan to step in directly to shore up the financial supply chain to avoid a sudden seize-up of funding, as happened during the GFC. This is important for the non-financial sector, especially SMEs that generate the bulk of jobs, growth and innovation.

Introduction

Prior to the global financial crisis (GFC) that began in late 2007, many believed that the financial and economic problems of the world were better understood than they had ever been before and that they had largely been solved, or at least moderated.

The period from the mid-1980s until 2007 was commonly referred to as the "Great Moderation," notably in a speech with that title delivered in 2004 by U.S. Federal Reserve Board Governor, Ben S. Bernanke, who later became the Federal Reserve Chairman.

What defined the Great Moderation, according to Bernanke, was that "the variability of quarterly growth in real output (as measured by its standard deviation) has declined by half since the mid-1980s, while the variability of quarterly inflation has declined by about two thirds" (Protess and Silver-Greenberg 2013).

Before the 2007-2009 GFC, this moderation of the business cycle was widely seen as a permanent condition due to the mastery of the macroeconomic cycle by central banks, and lightly-regulated or unregulated financial innovations, which brought about widespread sharing and diversification of risk.

Enter the Crisis

In late 2007 and 2008 events unfolded rapidly that thoroughly undermined that euphoria. The origin of the crisis lay deeply within the financial system, where – as in virtually all banking – highly liquid short-term liabilities funded less liquid long-term assets. The interest rate spread between liabilities and assets paid as usual for bank intermediation. The only new developments were that the range and types of assets and liabilities were more complex than ever before, and bank profits from intermediation reached unprecedented heights. Furthermore, many of the intermediating entities did not fit the classical model of a deposit bank and therefore became known as "shadow banking" entities.

These were in fact the conditions that an emerging philosophy regarded as ideal for stabilizing the economic system. Deregulation and technological advances had unleashed financial innovation, allowing a finely-reticulated, self-organizing network to flourish and grow. Self-organizing networks, composed of a complex web of individual bilateral and multilateral interactions and agreements constructed from the bottom up, were believed to be robust – or antifragile, to borrow a term from Nassim N. Taleb. The model for such a self-organizing system was the internet, which had grown organically after being launched by a U.S. government program.

The belief that this finely-reticulated, self-organizing network that was the financial system was robust and antifragile turned out to be wrong. On the contrary, it is now conventional wisdom that it was the very definition of a fragile system.

What was wrong with that assessment?

The concept of a stable, self-organizing system had been derived in large part from the paradigm of a highly-evolved ecosystem.

There were, however, two key ways in which the financial system was not like an ecosystem:

1.   It depended on participants' trust to function; loss of trust can spread like an epidemic.

2.   Far from being completely self-organizing, it was tethered to a small set of centrally-defined, model-based prescriptions.

These two conditions, acting together, rendered the system highly fragile. The first condition is permanent. The financial and economic system will always depend on its participants' trust.

The second condition is modifiable. Modifying it effectively, however, may require a sea-change in economic perspectives. First, economic theory at the time had developed in such a way that it largely ignored the first condition. Second, both economic theory and practice were in thrall to a set of models that were applied in ways that were deeply flawed. The momentum in the discipline will not be easy to change.

The GFC and Asia

The last major financial crisis that affected a global region was the Asian financial crisis of 1997, involving several countries of Southeast and East Asia. Short-term dollar loans, so-called "hot money", were not rolled over when confidence fell, causing a panic. According to a 1998 paper by Steven Radelet and Jeffrey Sachs, "At the core of the crisis were large-scale foreign capital inflows into financial systems that became vulnerable to panic… [W]hile there were significant underlying problems and weak fundamentals besetting the Asian economies at both a macroeconomic and microeconomic level, the imbalances were not severe enough to warrant a financial crisis of the magnitude that took place in the latter half of 1997" (Radelet and Sachs 1998).

The International Monetary Fund (IMF) stepped in to rescue Thailand, Indonesia and South Korea, imposing austerity conditions in exchange for urgently-needed loans. These conditions included high interest rates and cutting back on programs that provided vital social services. The policies have since come under severe criticism, and the IMF has altered its policies in response (Asian Financial Crisis, 1997-98).

The recent GFC, ten years after the Asian financial crisis, was caused by circumstances arising within the United States' financial system. As will be explained in Section 4, these circumstances were shaped in part by the risk-weightings assigned by the U.S. in implementingBasel I rules, and the credit ratings on which they were based (Jablecki and Machaj 2011, 200-227). All of Europe has been severely affected by this crisis. Meanwhile, Asian countries have withstood the crisis comparatively well.

Prior to the GFC, the conventional wisdom was that the liberalized Western financial system had benefited from decades of theory and experimentation to reach a point of optimal refinement. It was assumed that financial systems in emerging economies like Asia's would gradually liberalize too, following a similar path that would enable them to mature and become less vulnerable to potential crises such as in 1997.

Post-GFC the view is more uncertain. Western financial system models appear to have failed, throwing much of the developed Western world into recession, while at the same time Asian economies have withstood the GFC much better. It may not be too much to say that there are murmurs that stronger

state control of the financial system, as practiced for example in China, might offer a more robust model after all. Additionally, it is possible that more restrained models may be preferable, such as that of Shariah law requiring more delimited relationships between borrowers and lenders.

In short, the entire concept of what a stable financial system should look like is up for grabs – at least it is widely perceived that it should. Meanwhile, at the heart of the coordinated response in the developed economies has been a tendency to maintain the same forms of control over the financial system that were in place before the crisis, but to tighten them.

It is appropriate to ask whether this is the best avenue to pursue. Decades and millions of hours of academic and practical effort have been devoted to the creation of models that have been shown to be in practice, and can be shown to be in theory, on the wrong track. Is it right to continue to deploy them? Will their continued application subject the financial system to continued risk?

China, meanwhile, has been diligently attempting to adopt the Basel III rules, which are under development to tighten capital and liquidity requirements (Cousin 2012). In many cases, adoption of the rules can be salutary for the health of China's banks, and will ease these banks' acceptance and entry into European markets. Nevertheless, implementation of the Basel rules will pose a challenge for China's banks (Chan, Wan, and Yang 2011).

It is worth asking if Asia should not seek a stronger role in the setting of international financial standards, and in the moderation of practices that have been shown to and may again invite crises. Emerging market economies (EMEs), the bulk of which are in Asia, are expected to account for 39 per cent of world financial assets and 45 per cent of GDP by 2030 (Sheng 2013). They have withstood resiliently the economic crisis that developed economies caused, having learned from their own crisis in 1997, which was treated with a flawed remedy that was prescribed by developed economies.

The EMEs of Asia have learned well from valuable economic precepts developed in the West, but they have not gone overboard in implementing those precepts as many believe developed economies have done. Thus, they may be in a position to assume a new authority and to use it to mount alternative proposals for the regulation and oversight of finance that are less mired in the framework of current imperfect models.

Systemic Financial Risk

The focus of public concern about financial risk is systemic risk. Systemic risk has been defined as "any set of circumstances that threatens the stability of or public confidence in the financial system" (Bisias et al. 2012). A loss of confidence in the financial system can cause a reluctance to engage in financial dealings and hence a slowdown in these activities, which in turn can cause a slowdown in economic transactions and activity and hence a recession or at worst a depression.

There are two major concerns about the implications of such a slowdown, although concerns vary among schools of economics. For the Austrian school, periodic slowdowns and recoveries are regarded not only as part of the inevitable pulsation of a vibrant economy, but even as necessary for the process of "creative destruction." For other schools of economics, however, such as the Keynesian school, prolonged recessions and particularly depressions are undesirable not only because of the suffering they incur contemporaneously but because of the political and long-term economic

destabilizations they can cause with potential dire consequences. For example, some view the Second World War as a product of the Great Depression.

In the extreme case and of greatest concern would be a situation in which bank depositors could not withdraw their deposits on demand at a bank or ATM. This could occur if banks could not borrow or lend to each other or sell their assets or borrow from any other entity, that is, if the flow of funds seized up or if customer records were lost. The flow of funds seizes up for an individual bank when there is a bank run. Systemic risk, however, is not concerned with an individual bank run, but with the equivalent of a run on the entire financial system – on all, or many banks at once.

If a natural disaster occurs that disrupts the delivery of subsistence needs like food, such as hurricane Sandy in the eastern U.S. in 2012 or the 2004 Indian Ocean tsunami, there exist organizations that stand ready to replace the supply chain that has been disrupted. In the case of Sandy, the U.S. Federal Emergency Management Agency (FEMA) stepped in, while in the case of the tsunami non-governmental organizations like Oxfam, Mercy Corps and World Vision intervened.

There is no similar organization with a mandate to directly replace the supply chain of financial funds in case of disruption. When the threat of such a disruption, or a panic that could cause such a disruption, loomed in the U.S. in 2008, the government response was to save the banks from bankruptcy so that they could continue to carry out their essential functions. When the U.S. government decided to let Lehman Brothers go bankrupt, it discovered that Lehman Brothers was so interconnected with financial counterparties that its bankruptcy threatened the entire financial supply chain.

Indebtedness is fundamental to a financial system, since the most basic financial activity, the deposit of funds in a bank, incurs a debt on the part of the bank. The bank in turn, to fund its service, must either charge the depositor a fee, invest or loan the funds to other entities. Hence, the problem of avoiding systemic financial risk is the problem of avoiding a breakdown in the supply chain or web of indebtedness.

Modeling in Finance

This paper will address these issues with a special focus on the use of models in finance. The mathematical modeling of risk in finance, its historical development, uses and deficiencies will be the subject of Section 2.

How trust and stability in finance can be enhanced will be the subject of Section 3. A key concern is that trust invites increased risk, which engenders a decline in trust. Economics has largely ignored this common psychosocial conundrum in its modeling work, tending to obscure the basic problem. Section 3 will cover briefly the broad range of topics related to this subject.

Section 4 will address financial regulation, its challenges and pitfalls, the options being presented and possible alternatives. It will go beyond this to discuss "thinking the unthinkable," that is, worst-case scenarios and unthinkable policies such as a FEMA for the financial supply chain to address extreme financial risk and instability.

Section 5 concludes with a summary, conclusions, recommendations and points for discussion.

2.   Mathematical Modeling of Risk

This section begins by explaining the fundamental building blocks of what later came to be known as modern portfolio theory (MPT). The first part of this section could be regarded as a parable because it exemplifies the promise and disappointment of nearly every advanced mathematical model that has been developed in modern finance.

Harry Markowitz's Nobel Prize-winning Breakthrough

Prior to the publication of Harry Markowitz's paper on "Portfolio Selection" in The Journal of Finance in 1952, investment risk was perceived as a property of an individual asset. A broker or financial advisor who recommended a stock to a client was held by law to be in breach of his duty if the stock was too risky for the particular client.

Markowitz's most valuable contribution was to shift that view to the risk of a whole portfolio of investments. Because individual investments can be less than perfectly correlated, adding a risky stock to a portfolio might not add to the risk of the whole portfolio, and it could even reduce it. That is, it might diversify the risk of the portfolio and thereby mitigate that risk.

In his 1952 article Markowitz stated this in a mathematical form, using three assumed properties of investment securities. The first property is each security's future expected return on investment, the second is the variability of that future return (as measured by standard deviation) and the third is the correlation between returns on two securities.

This fundamental return-risk-correlation structure became permanently embedded in mathematical finance and is at the core of virtually every development since.

Markowitz's Efficient Frontier

Using these three properties of securities, Markowitz constructed in theory what he called the efficient frontier of portfolios. A portfolio of securities contains some number of securities in percentage weightings adding up to 100 per cent. An expected return can be calculated for the whole portfolio in terms of the expected returns of the individual securities, and a standard deviation can be calculated in terms of the individual standard deviations and the correlations.

The standard deviation of the portfolio's return was interpreted by Markowitz as the portfolio's risk. Thus, each portfolio has an expected return and a number representing its risk. Plotting all portfolios on a graph of risk versus return results in an envelope that is convex upward – it is a curve above which no portfolios lie (see Figure 2.1).

Figure 2.1. Markowitz's Efficient Frontier
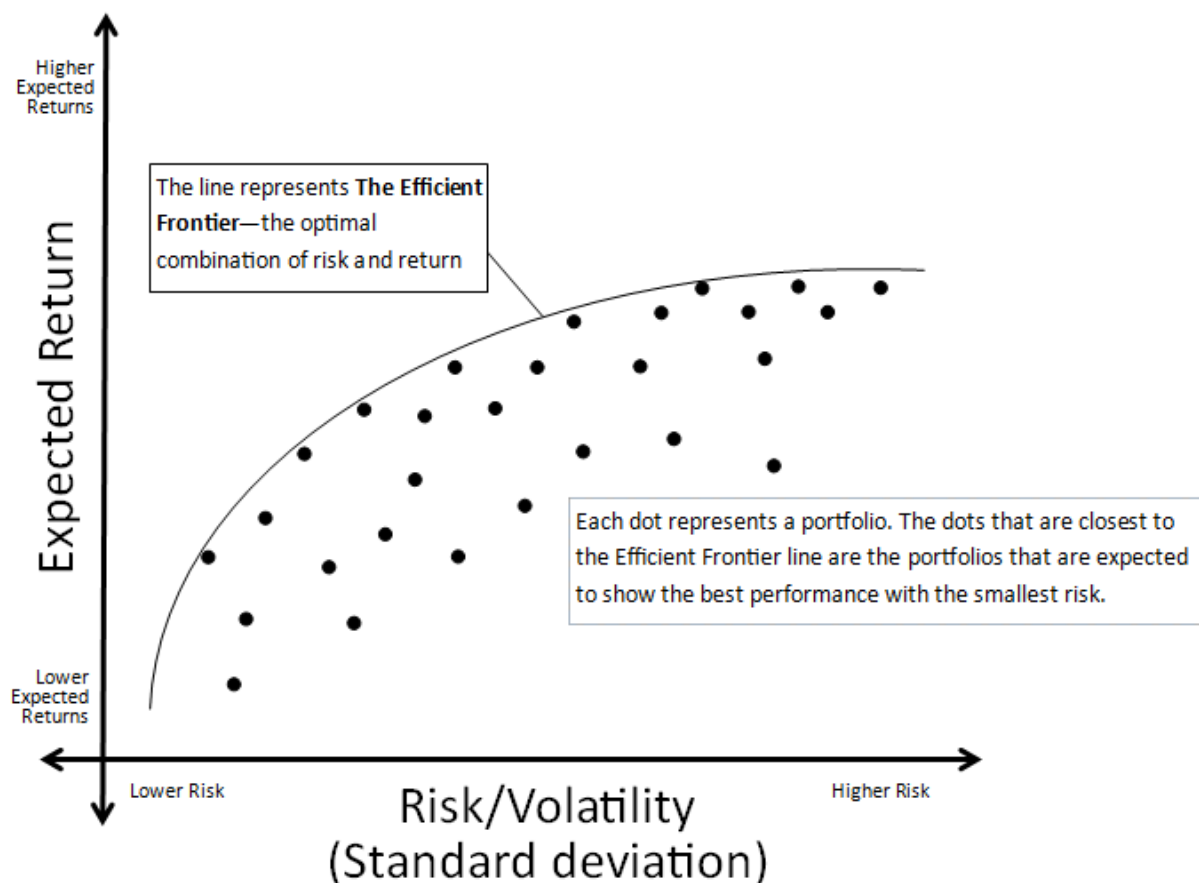
Markowitz concluded that the best, and the most "efficient", portfolios were those on that envelope, which he called the efficient frontier. Any portfolio on the efficient frontier had a higher expected return than any other portfolio with the same level of risk, and it had a lower risk than any other portfolio with the same expected return. A portfolio on the efficient frontier is "Pareto-optimal": any reduction in its risk must be accompanied by a decrease in its expected return, and any improvement in its expected return must be accompanied by an increase in its risk. Hence, no other portfolio is superior to it in both risk and return.

Attempts to Apply Markowitz's Mean-variance Optimization

Markowitz is a mathematician who was employed at the time at the Rand Corporation. Rand had been the innovator of important applied mathematics techniques in the field of operations research. Markowitz was now confronted with the mathematical challenge of finding out exactly which portfolios lay on the efficient frontier.

The solution method he devised involved a mathematical technique called quadratic programming (not to be confused with computer programming).

Given a universe of investment securities (in the initial applications, it was the universe of listed stocks), the Markowitz algorithm calculated all of the weighted combinations of these securities that lie on the efficient frontier.

This methodology excited people in the field, especially academics and newly-minted finance graduates. They attempted to apply the Markowitz quadratic programming algorithm to create the efficient frontier of portfolios of public stocks.

The first thing they noted was the sheer magnitude of the number of inputs that the method required. There were thousands of listed stocks. Markowitz's algorithm required an expected return and standard deviation for each of them, with a correlation coefficient for each pair. If the portfolios were to be constructed from 1000 stocks, the algorithm would require 501,500 inputs.

Since the model seemed to represent a major breakthrough, analysts and academics tried not to be discouraged and find ways to reduce the number of inputs, for example, by placing stocks into groupings and assuming that the correlation coefficients between stocks in two groupings were all the same.

The second problem was that they needed to obtain numbers for the inputs. The obvious approach was to use historical data to estimate the inputs. So they tried using historical average returns for the expected returns, historical standard deviations for the standard deviations and historical correlations for the correlation coefficients.

When those numbers were fed into the model the results that came out were absurd. A model that was expected to reduce risk through diversification might instead say that the efficient frontier consisted of combinations of only two very risky stocks. The reason was that those stocks were the ones that had the highest historical returns, and therefore had been assigned the highest expected returns; or they might be the ones with the lowest historical standard deviations, and therefore were assigned the lowest future standard deviations.

No matter what they did, the modelers found that the outputs of the model made no sense. Eventually, they employed trial and error to see what inputs would produce acceptable outputs.

In the end, this is precisely how the Markowitz mean-variance model is used. Its inputs are laboriously rigged to produce outputs that the modeler or the modeler's employer or client feels are acceptable. If it is found to be too difficult to rig the inputs to produce the desired outputs, then constraints on the output variables – portfolio weightings – are imposed to ensure that they come out as desired. In other words, the desired outputs are specified, then the inputs and constraints are created in such a way as to produce them.

Asset Allocation

Because running the Markowitz mean-variance optimizer on a large number of individual securities is so cumbersome, and constraining it to produce acceptable results is so difficult, it is now used only for "asset allocation." That is, the number of assets in a portfolio is reduced to a small number, such as ten broad asset classes. For example, the asset classes may include categories like small value

stocks, small growth stocks, large value stocks, large growth stocks, perhaps mid-cap stocks (stocks of companies with mid-sized capitalizations), and short-term, intermediate, and long-term bonds.

This still requires rigging the inputs, and usually, constraining the outputs to be within acceptable ranges. Running an asset allocation on ten asset classes still requires values for 65 input parameters.

The end result has been the complete Orwellian perversion of the Markowitz mean-variance computational model (see Box 2.1). Although it is often boasted that it is used, it is not really used – its implementers only go through the motions, using it chiefly as a sales tool. The model has proven useful only for its conceptual insights, not for its application in practice as a computational tool. This is a fact that is not recognized widely enough.

Box 2.1. How Mean-variance Portfolio Optimization (Asset Allocation) is Actually used in Practice

Most financial firms that offer financial advisory services perform an "asset allocation" for their clients. This is often represented as a very important part of their service. They let the client know that a sophisticated computer model is used to create the asset allocation. Frequently they point out that the technology used is Nobel Prize-winning technology.

Through trial and error the firm has created (or outsourced the creation of) inputs to the model that produce an acceptable efficient frontier of outputs. They then ask the client for her current portfolio. From that they construct her current asset allocation. Invariably, an asset allocation on the firm's efficient frontier will have a higher expected return for the same level of risk. This is necessarily the case "by construction," since the firm has manipulated the outputs of the model so that they have higher expected returns at each level of risk than any other portfolio. The result, however, is a function of the arbitrary inputs they have created, not of reality. The model is used as a sales device, not to solve a real-world problem. It has not been used in any other way.

Adding the Time Dimension

Although the Markowitz construct proposes a measure of risk standard deviation of investment returns), it has no time dimension. (We shall explore the adequacy of the risk measure later.) The model is a "one-period" model, that is, it applies only over a single time period of unspecified duration.

In the early 1960s an extension of the model added the time dimension, making the model mathematically much richer. It was noted that price movements of frequently traded securities could not be distinguished from the mathematical pattern called a random walk. Then it was discovered that a French mathematician named Louis Bachelier had used a similar model to describe securities price changes in his 1900 PhD dissertation.

Once the components were pieced together, the standard mathematical model of securities prices had been created – multivariate geometric Brownian motion. Multivariate Brownian motion had already been thoroughly analyzed in the mathematical field of stochastic processes.

The Black-Scholes option pricing model can be derived from this body of mathematical theory. The Black-Scholes model has the same problem that the Markowitz mean-variance optimization model has,

but it is much simpler. Like the Markowitz model, it needs inputs to produce an output, but it needs many fewer inputs.

The principal input required is the volatility of the option's underlying security. If this number is derived from the history of the security's price, the model will generally produce an output that is not credible. Hence, the input must be altered in such a way as to make the output believable.

The chronicler of the "quant" (quantitative analyst) culture Donald MacKenzie suggests that mathematical models are used in finance more as a means of technical communication among quants than for obtaining results (MacKenzie and Spears 2012). ("Quant" is short for quantitative analyst.) This is not, however, the way the models are generally understood by non-quants – by the general public, the creators of regulations, or even, in most cases, by the quants' superiors at their firms. They regard these models as means to provide hard facts and answers.

A Fundamental Problem with the Options Pricing Model and its Derivation

All economic and financial models, as is well known, rest on a series of assumptions. Most of these assumptions are not in keeping with real-world conditions. The incorrectness of some of these assumptions, including that stock price changes have a geometric normal (Gaussian) distribution, has been pointed out many times. This erroneous assumption is, in fact, one of the main reasons why the volatility calculated from historical returns will produce incorrect results when plugged into the Black-Scholes formula.

However, little has been said about the importance of one key assumption, which, having been mistaken for reality, has already contributed to a financial crisis (see Box 2.2).

The Black-Scholes pricing model rests on an approximation that is readily apparent in its original derivation. It assumes that a perfect arbitrage – i.e., a perfect hedge – can be set up between a stock and an option on that stock. That is, it assumes that a portfolio holding a combination of the right amount of the stock with the right amount of the option will be riskless.

Such a "riskless" arbitrage can be constructed, but it is riskless only over an infinitesimal time period. There is no discrete (or real) time period over which the hedge is riskless. Thus, the riskless hedge is only theoretical.

It is common, nonetheless, for people in the financial modeling field, including quants, to infer that results that hold theoretically over infinitesimal time periods will hold over discrete time periods as well. Another way of saying it is that they assume that discrete stochastic processes can be very closely approximated by continuous processes. Box 2.2 shows how that assumption played out in the "portfolio insurance" crash of October 19, 1987.

As often as it is repeated that the obviously incorrect assumptions that are made to enable theoretical concepts in economics and finance to be translated into mathematical models are only assumptions, they nevertheless creep into the belief systems of the modelers, and are forgotten or neglected by everyone else.

The corrective measures that are applied to hold this "model culture" in check are inadequate. Only moving the models far from center-stage, and redeploying the modelers to other tasks in which their

skills will be much more useful (this will be discussed below) will help to reduce the distortions in financial regulation and practice that are caused by the use of these models.

Additionally, more credence and a greater role must be assigned to less quantitative and non-quantitative methods for risk control that recognize the role of uncertainty (or unquantifiable risk), and which firmly subordinate the results of quantitative modeling to a sound process of judgment.

Box 2.2. The Portfolio Insurance Crash of 1987 and the Assumption of Continuous Processes

A portfolio invested in a broadly-diversified equity fund such as a Standard and Poor's 500 (S&P 500) index fund could, in theory, be protected over a three or five-year horizon against the downside of a negative return by purchasing a three or five-year put option. That is what portfolio insurance, devised by Berkeley professors Hayne Leland and Mark Rubinstein in the early 1980s, was intended to do. It would enable institutional investors like pension funds to increase their portfolio allocations to equities, taking advantage of equities' superior expected returns, while protecting their portfolios on the downside.

Leland and Rubinstein teamed up with John O'Brien to form LOR, Inc., which sold what they called portfolio insurance. The intent was to provide the equivalent of three-year or five-year put options as an overlay to an equity portfolio. Put options with such a long expiry were not generally available in the market, so LOR used the strategy of "option replication."

The principle of option replication comes from the derivation of the Black-Scholes option pricing model. It is that at any moment, an option is perfectly hedged – or equivalently, perfectly replicated – by a calculable amount of the underlying security. Hence, one can replicate at any moment the owning of an option by owning that exact amount of the underlying security, which perfectly replicates the option at that moment. This amount changes constantly over time; hence, the replication strategy constantly varies the quantity of the security owned. In practice, LOR did not buy the underlying security, the S&P 500 index, but S&P 500 futures.

This was not however a foolproof strategy because securities prices did not really move continuously, as they did in the theories. Hence, unlike the put option itself, if it existed, the replication strategy would not provide the same guarantee, or the "insurance", that the option would provide. Nevertheless it would work in normal markets, which are approximately continuous. It was possible to make this assumption with impunity because it was what was commonly assumed in virtually all theoretical work, and was simply part of the culture's implicit worldview to which all participants tacitly assented.

Portfolio insurance, once it caught on, rapidly became very popular among institutional investors because it enabled them to increase their commitment to equities without fear of a multi-year fall in prices.

The replication strategy, however, entailed selling S&P futures whenever their prices fell, to maintain the calculated relationship to the option. On October 19, 1987, the selling of S&P futures by portfolio insurance providers created a positive feedback loop in which selling led to falling prices, which led to more selling, creating a cascade of discontinuously falling prices. As a result, it was not possible to sell at, or close to, the price at the last market tick. The sale could only be completed at a much lower price.

The cascade accelerated toward the end of the day. By the day's close, it had seen the largest percentage drop ever in the Dow Jones stock index in a single day, -22.6 per cent. The chief cause was the belief entrenched by the assumptions of the Black-Scholes model and other academic financial models, that stock prices move continuously.

The VaR Model

Value at Risk (or VaR) is the most frequently used model for measuring financial risk. It is actually an umbrella model, one that can subsume a variety of sub-models. VaR was constructed by quants at JPMorgan in the late 1980s and early 1990s at the behest of then-CEO Dennis Weatherstone, who wished to have a way to measure firm-wide risk after the 1987 crash.

JPMorgan's quant team chose the only model they probably could have chosen, given that their quantitative finance training focused on geometric Brownian motion.

Economist Barry Eichengreen described the adoption of VaR in a 2009 article (Eichengreen 2010):

In the wake of the 1987 stock-market crash, Morgan's chairman, Dennis Weatherstone, started calling for a daily "4:15 Report", summarizing how much his firm would lose if tomorrow turned out to be a bad day. His counterparts at other firms then adopted the practice. Soon after, business schools jumped to supply graduates to write those reports. Value at Risk, as that number and the process for calculating it came to be known, quickly gained a place in the business-school curriculum.

VaR's key feature is that it assumes that each individual asset's short-term return on investment can be described by a Gaussian probability distribution, and that the collection of assets in a portfolio can be described by those individual distributions with pairwise correlations between the assets.

If those assumptions are correct, then the probability of any specific portfolio return can be calculated. In particular, VaR can answer the following question: "I want to be 99 per cent (or possibly 99.9 per cent) certain that our bank will not lose more than X dollars tomorrow. What is X?"

X would be the answer, in the 4:15 Report that Weatherstone requested, to the question "how much would the firm lose on a bad day." If the number were US$50 million, it would mean there was a 99 per cent chance that JPMorgan would lose no more than US$50 million tomorrow. If the number suddenly went up to US$75 million the following day, management could ask what had happened to change it.

The VaR model has many flaws. Defenders of the model make the reasonable argument that despite its flaws, it can be a useful tool for monitoring risk when understood in light of the approximations it makes, and when combined with other risk indicators. The flaws that cause it to measure risk only with great imprecision do not detract from the fact that when properly applied, it can detect changes in risk, which can serve as warnings that the situation requires more careful scrutiny.

These are good arguments for the judicious use of the measure, but the emphasis must be on the word "judicious." If specific VaR limits are embedded in corporate policy or regulatory code, they can limit risk managers' ability to exercise judgment in using the model, inviting manipulation and gaming of model inputs and of the model itself and creating excessive faith in its outputs.

Some of the key criticisms of the VaR model are:

1. It relies on inputs that can be estimated only very poorly. In this, it resembles the Markowitz mean-variance optimizer. The same inputs are required for the VaR model as for the mean-variance optimization model.

2. It does not account for fat tails (see Box 2.3) and hence, the assumption of a normal distribution underestimates probabilities of extreme events.

3. It does not estimate the amount that is actually at risk of being lost, but only the value that is not at risk. For example, if the VaR is US$50 million, it means that there is a 99 per cent chance that less than US$50 million will be lost. It does not say how much might be lost in the 1 per cent of scenarios in which more than US$50 million can be lost. It does not tell you, for example, that someone is taking a gigantic risk with a relatively small percentage of the firm's capital and with a low probability of failure, but a risk that will lose billions if it turns out badly.

> **Box 2.3. The "Fat Tails" Problem**
>
> The problem of "fat tails" was recognized almost at the same time that the Brownian motion model of securities prices was being constructed. The 1964 book "The Random Character of Stock Market Prices," edited by Paul Cootner, was the first compendium of articles to expound on the random walk model of stock market prices. (The Brownian motion process is the continuous counterpart of the discrete random walk process.) While Bachelier assumed price changes were normally distributed in his 1900 doctoral dissertation which is reproduced in that book, the book also includes an article by Benoit Mandelbrot in which he argues that the observed distribution of price changes resembles a more fat-tailed distribution, the "stable Paretian" distribution. Hence, the assumption of a normal distribution will underestimate the probability of extreme events.

4. While a 99 per cent certainty that the loss will be no more than the VaR tomorrow may sound like a very high level of certainty, it means that on one per cent of all trading days – one in one hundred, or about two per year – the loss will be greater than the VaR, perhaps incalculably so (see 3 above).

5. VaR measures only the risk that can be modeled by market fluctuations or "volatility." This is an inadequate measure of risk; it may not detect deeply-embedded risks of untoward events that could emerge suddenly without having been reflected in market volatility. Such a risk, for example, was the one that emerged with the subprime crisis of 2007.

6. A more recent criticism is that true uncertainty – the inability to predict "black swan" events – cannot be quantified by a probability. This argues for building resilience into the portfolio, if possible, or better yet, as Nassim Taleb states it, an antifragility that grows even more resilient with the experience of risk or failure.

The VaR model can be implemented either by model parameterization or by simulation. In model parameterization, the multivariate normal distribution is assumed with parameters derived from history and/or other means and with the same difficulty as with the Markowitz model.

In simulation, the standard method is to run the current portfolio through some number of days of recent market history (commonly 501, as per a standard risk management text (Hull 2012), and plug in the values of relevant market variables on those days to see what would have happened to the portfolio. This has been rightly criticized because in the mid-2000s, these were generally days of historic market stability. This is part of the reason why VaR failed so dismally

in the GFC. Since the GFC, it has become common to use "stress VaRs," deploying a period in history of sequential days during which markets were unusually stressed.

Over-reliance on the VaR model has been identified as one of the errors that was commonly made by financial institutions leading up to the GFC. Box 2.4 shows how over-reliance on VaR contributed to very severe write-downs on financial assets at UBS, a large Swiss bank, during the GFC in 2007-2008.

Box 2.4. UBS's Write-Downs

On April 18, 2008 the large Swiss bank UBS produced a 50-page document, "Shareholder Report on UBS's Write-Downs" (UBS 2008) describing the bank's processes and errors that led to a very large write-down of assets during the 2007-2008 GFC. Under the heading of "Risk Control Measurement and Monitoring Tools," the report listed the errors committed due to uncritical over-reliance on models and the concurrent inadequate attention paid to other risk monitoring procedures. These errors included:

•"…The historical time series used to drive VaR and Stress are based on five years of data, whereby the data was sourced from a period of relatively positive growth… When updates to methodologies were presented to Group and IB [Investment Bank] Senior Management, hindsight suggests that these updates did not attribute adequate weight to the significant growth in the US housing market and especially the Subprime market. The Market Risk function did not develop scenarios that were based on more fundamental attributes of the US housing market."

• "Lack of Housing Market Risk Factor Loss [RFL] limits: In a similar vein, it appears that no attempt was made to develop an RFL structure that captured more meaningful attributes related to the US housing market generally, such as defaults, loan to value ratios or other similar attributes to statistically shock the existing portfolio."

• "Over-reliance on VaR and Stress: MRC [Market Risk Control] relied on VaR and Stress numbers, even though delinquency rates were increasing and origination standards were falling in the US mortgage market. It continued to do so throughout the build-up of significant positions in Subprime assets that were only partially hedged. Presentations of MRC representatives to UBS's senior governance bodies did not provide adequate granularity of Subprime positions UBS held in its various businesses. No warnings were given to Group Senior Management about the limitations of the presented numbers or the need to look at the broader contextual framework and the findings were not challenged with perseverance."

• "Over-reliance on ratings: MRC relied on the AAA rating of certain Subprime positions, although the CDOs [collateralized debt obligations] were built from lower rated tranches of RMBS [residential mortgage-backed securities]. This appears to have been common across the industry. There is no indication that MRC sought to review the quality of existing portfolios as questions were being raised in relation to the Subprime sector more generally. A comprehensive analysis of the portfolios may have indicated that the positions would not necessarily perform consistent with their ratings."

Estimation of Probability Distributions of Individual Assets under VaR

The VaR model for a bank portfolio is only as good as the probability estimates for each of the assets in the portfolio, as well as the estimates of the correlations between assets. Probability estimates for two asset classes in particular are highly questionable: asset-backed debt securities and derivatives.

Correlations between assets are themselves difficult to determine, especially for an entire bank portfolio. An investment made in one corner of the bank could hedge or cancel out another investment in another corner of the bank, meaning that there could be a strong negative correlation, or one investment could correlate strongly with and exacerbate the risk of another investment. It may be very challenging to determine which.

The difficulty of determining the risk of a bank portfolio is illustrated by a January 2013 JPMorgan Chase report on the bank's multi-billion-dollar losses on the trades of the "London Whale" in 2012 (JPMorgan chase 2013). The report's "Appendix A, VaR Modeling," offers some insights (Box 2.5).

Box 2.5. The "London Whale"

Appendix A of the JPMorgan Chase report on the "London Whale" incident in which JPMorgan lost more than US$6 billion shows how *ad hoc* and unreliable the process of computing VaR is, even at a bank as sophisticated as JPMorgan. The appendix reveals that:

- "the model was limited in the manner in which it estimated correlation risk";
- "the model operated through a series of Excel spreadsheets, which had to be completed manually by a process of copying and pasting data from one spreadsheet to another";
- "many of the tranches were less liquid, and therefore, the same price was given for those tranches on multiple consecutive days, leading the model to convey a lack of volatility";
- It was not clear to the Model Review Group which model was being used in a pricing step in the VaR model, an approved vendor model or one developed by the modeler;
- "a spreadsheet error caused the VaR for April 10 to fail to reflect the day's US$400 million loss in the Synthetic Credit Portfolio";
- "further errors were discovered … including, most significantly, an operational error in the calculation of the relative changes in hazard rates and correlation estimates. Specifically, after subtracting the old rate from the new rate, the spreadsheet divided by their sum instead of their average, as the modeler had intended. This error likely had the effect of muting volatility by a factor of two and of lowering the VaR."

These errors make it sound like control over this program had run off the rails at JPMorgan. But in fact such problems are par for the course even at a large and sophisticated bank. It is impossible for a Model Review Group to adequately evaluate and control the activities of a programmer/modeler, either one who is operating a model by cutting and pasting across Excel spreadsheets, or one who has written a fully-automated computer program. It is common knowledge at such institutions, not openly voiced, and perhaps not common knowledge at higher levels of management, that model outputs cannot be taken very seriously.

The Credit Risk Modeling Failure

Much has been written about the travesty of ratings agencies assigning senior tranches of subprime CDOs triple-A ratings, implying nearly zero probability of losses, even though these securities were backed by "liar loans," "NINJA loans" (no income, no job, no assets) and other similarly low-quality loans to borrowers with very questionable repayment capacities.

Most of what has been written has focused on the perverse incentives that presumably caused the ratings agencies to produce these ratings, namely the fact that they were paid very handsomely by the packagers of the securities to assign high ratings. , Agencies even earned high consulting fees for helping the investment banks that packaged the securities to design them so that they would receive high ratings.

Some articles have assigned the blame on the models being used, as well as the perverse incentives. But very little has been written for public consumption about how the mathematical models actually produced these ratings. After all, the ratings agencies really did use mathematical models, they did not just pretend to perform analysis. Agencies would not have gone as far as to fraudulently alter the models so that they could produce the desired ratings, especially because such practices would expose them to law suits for fraud, which plaintiffs would have won handily.

In fact, the ratings agencies actually did run their models mostly in accordance with their own ingrained and widely accepted practices.

How, then, could the results of the models have been so wrong? To answer this question, we have to look both beyond the "model culture" to the "data culture," and to the compartmentalization that impaired the ability to take a broad overview of risk assessment.

The Model Culture and the Data Culture

The danger of the "model culture" is that the output of models will be believed, even when it can or should be known that the models are highly unreliable either because the inputs to the models are arbitrary or poorly estimated, their assumptions are wrong or because they are extremely vulnerable to the likelihood of coding errors, human error and other mishaps that will with high frequency cause them to produce incorrect or meaningless outputs.

The danger of the "data culture," on the other hand, is that it assumes an "objective" posture by relying only on data that is certifiable, documentable and quantifiable, even when it should be transparently obvious that the data is unreliable, wrong or does not apply to the current situation. The certifications may emanate from persons or entities that, whatever their credentials, may be negligent or corrupt. The documentation may be tampered with or certified without being read, and the quantification may be slapdash, miscalculated or misrepresented. Nevertheless, the data culture of the ratings agencies demands that they take no subjective view and only rely on putatively objective information derived from historical data.

In fact, in the U.S. subprime mortgage crisis that triggered the GFC, most of these flaws could be found in the data that were used as inputs in ratings agencies' and investment banks' models. Information on mortgage loan applications was deliberately misrepresented. Appraisals by certified

appraisers were overstated. Borrower credit scores were manipulated, and in particular, historical data on default rates were not representative of likely default rates on the loans that were actually being processed.

Nevertheless, because these facts could only be regarded as "soft information," ratings agencies and investment banks ignored those facts when they prepared individual default rate inputs to the CDO risk models. Box 2.6 elucidates the crucial role that bad data on individual loan default rates played in producing triple-A ratings.

Compartmentalization and the Absence of an Adequate Risk Modeling Overview

The second factor in the failure of the credit rating models was the low correlation assumed between mortgage default rates. That is, it was assumed that the likelihood that one borrower would default and another would default were nearly independent of each other.

As is now known, the default probabilities were in fact highly interdependent because the lack of default depended on rising housing prices. When housing prices ceased to increase, a large number of mortgages went into default at once.

The error was exacerbated by practices in risk assessment modeling, one in particular, namely the construction of so-called mezzanine CDOs, stands out as being especially egregious. Mezzanine CDOs, which were among the worst performers in the crisis, were developed in a two-step process, the first of which was the creation of asset-backed securities (ABS), which were collateralized by mortgages and other loans. These ABS were tranched to create triple-A, mezzanine and equity tranches, the least risky of which were the triple-A and the most risky of which were the equity tranches. The default risk of each was assigned with a formula that assumed a correlation between loan default probabilities.

Often, the mezzanine tranches of ABSs were then repackaged to form CDOs that were similarly tranched. When the model was run on the packaging of mezzanine ABS tranches, a low correlation among their default probabilities was assumed, similar to that between individual mortgage loans. As sociologist Donald MacKenzie, who studies the culture of financial modeling, has observed, in rating ABS CDOs in a two-step process, the "free lunch" of diversification was being consumed twice: first, implicitly, in the rating of the ABSs, then, explicitly, in the rating of the CDOs (MacKenzie 2011).

Box 2.6. How Ratings Agencies Used Bad Input Data to Certify Subprime Securities as Triple-A

The formulas commonly used to rate CDOs required only two inputs: (1) the probability that any individual loan would default; and (2) the correlation between loans.

In the case of subprime CDOs, both of these inputs were wrong, but attention has focused on correlation because of the role thought to have been played by the mysterious (to non-mathematicians) Gaussian copula formula.

The low probability of default on any individual loan that the ratings agencies used, however, may have had more impact on the results than the correlation. For the individual default rate, the agencies used – in fact believed they were obliged to use – only the historic mortgage default rate, which was low.

As is now well known, the quality of mortgage loans was dropping fast in the mid-2000s, driven by the fact that mortgage lenders could sell the loans to investment banking firms, which would then repackage them in CDOs that the ratings agencies would rate triple-A. So the ratings agencies were driving a race to the bottom. Even if they knew this – which they may not have – they still had to use the historical default rates to estimate the probability of default of an individual loan because that was more "objective" than using their subjective impression of how likely it was that the current borrowers would default.

The journalist Roger Lowenstein wrote, in an article titled "Triple-A Failure" in the April 27, 2008 *New York Times*: "Moody's did not have access to the individual loan files, much less did it communicate with the borrowers or try to verify the information they provided in their loan applications. "We aren't loan officers,"Claire Robinson, a 20-year veteran who is in charge of asset-backed finance for Moody's, told me. "Our expertise is as statisticians on an aggregate basis. We want to know, of 1,000 individuals, based on historical performance, what percent will pay their loans?"

If the ratings agencies had conducted a simple random sampling and interviews of borrowers in one of the subprime CDOs, it would have revealed that they were unlikely to repay unless the price of their house went up substantially and they could refinance. But that would require a subjective judgment on the part of the ratings agencies, a risk they were unwilling to take. It was safer for them to base their ratings on the historic data.

One might ask, if the ratings agencies were paid by the investors to evaluate the debt securities instead of by the issuers of the securities, would the agencies have done it differently and tried harder to render an honest opinion, even if it required subjective judgments? It is possible, given the prevalence of the "data culture," that they would not have.

The two-step process was carried out by separate risk modeling groups. Apparently no integrating intelligence was imposed to correct for the fact that low correlation between individual default rates was assumed in creating the now-diversified ABS mezzanine tranches, which were then assumed again to have low correlation to each other. This negligence was probably partly a result of the

incentives presented to the ratings agencies to produce high ratings, but it was facilitated by a segmented and inadequately self-aware modeling process.

VaR's Problem with Derivatives

Because VaR reduces the probability distribution of returns for all asset classes to a normal distribution, some way has to be found to represent the probability distribution of price changes for various derivatives as normal distributions. That is easy if one assumes that the only parameters needed are the expected return and standard deviation. Those can be calculated (or rather, estimated) for any option from the delta of the option (the rate of its price change relative to the price change in the underlying security). Sometimes the calculation goes one step further and uses the option's "gamma" (the rate of change of the delta). In any case, it provides a normal distribution as VaR requires. Box 2.7 shows how wrong this can get the probability distribution of the option's possible price changes.

Box 2.7. An Option whose Probability Distribution is far from Normal (Gaussian)

VaR's assumption that the percentage price changes (i.e., returns) for all assets follows a normal distribution is clearly wrong, but it also wrongly assumes that it is not far from wrong for a short time period. Suppose a naked call option costing US$1 has been written on a stock whose price is US$100 at a strike price of US$100, and the option expires tomorrow. At the 99.9 per cent probability level the stock might gain US$10, causing the option writer to lose US$10 – a loss of US$10 on a gain of US$1. However, there is also a 50-50 chance the option writer will lose nothing. This is a highly asymmetric probability distribution that is badly modeled by any symmetric distribution, including the normal distribution. The example may seem extreme, but with the variety of option and derivative contracts that are available and can be constructed, in addition to the incentive of risk traders to outsmart VaR, it is not far-fetched at all.

The Neglected Role of Uncertainty

In the face of recent "black swan" events such as the GFC, a fundamental verbal distinction made in 1921 by economist Frank Knight has come into wide currency (Box 2.8). Knight distinguished uncertainties that could be measured quantitatively with probabilities from uncertainties that could not, by designating the former "risks" and the latter true uncertainties.

This construct is useful because it is a reminder that not everything can – nor should – be quantified. Information quantified in terms of probabilities using sophisticated models can provide useful input to risk management, but should not overwhelm the process of judgment required to integrate these assessments with other factors.

Box 2.8. Risk and Uncertainty

It is useful to clarify the distinction between "risk" and "uncertainty". (Economist Martin Weitzman points out that while economists use the terms "risk" and "uncertainty" to describe this dichotomy, scientists use the terms "uncertainty" and "structural uncertainty," respectively, to characterize the same dichotomy (Weitzman 2007).)

The distinction, made originally by economist Frank Knight in 1921 (Knight 1921), came to the fore in 2002 because of a much-mocked statement by then-U.S. Secretary of Defense Donald Rumsfeld. Rumsfeld, referring to the difficulty of determining whether Iraq under Saddam Hussein had weapons of mass destruction, said:

"Reports that say that something hasn't happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns - the ones we don't know we don't know."

This was widely regarded as a "bizarre remark" (BBC News 2003) (compounded by his statement the following day: "I believe what I said yesterday. I don't know what I said, but I know what I think... and I assume it's what I said."); however, it became recognized later (Blanchard 2009) as a possibly inadvertent reference to "Knightian uncertainty," that is, uncertainty that cannot be quantified with a probability.

Simply put, in the economist's distinction, "risk" refers to uncertain (that is, not deterministic) outcomes to which numerical probabilities *can* be assigned. "Uncertainty," on the other hand, refers to uncertain outcomes to which numerical probabilities *cannot* be assigned.

by Many key risk-avoidance decisions have historically been made based on more than the modeling process. For example, financial industry mogul Jamie Dimon has acquired fame and respect in part because of having minimized or avoided the impacts of risks on the financial institutions he managed on two occasions, the Russian debt crisis of August 1998 and the GFC of 2007-2009.

Richard Bookstaber, the author of the prescient 2007 book "A Demon of Our Own Design: Markets, Hedge Funds, and the Perils of Financial Innovation," was the Managing Director in charge of firm-wide risk management at Salomon Smith Barney in 1998, where Dimon was co-CEO. In his book, Bookstaber describes how Dimon's vehement insistence on reducing the firm's exposure to Russian debt in June of 1998 enabled the firm to avoid the worst of the effects of Russia's debt default in August of that year.

Meanwhile, another firm, Long-Term Capital Management (LTCM), known for sophisticated quantitative models and employing two Nobel Prize winners in economics, suffered disastrously from the default. When the Federal Reserve became concerned that its failure could cause a broad financial crisis, it organized a meeting to force LTCM's creditors to bail it out (Edwards 1999).

The Role of Regulators

How could so many abuses have occurred when there were regulatory agencies overseeing the process? It is widely known that the prevailing set of economic beliefs before the GFC were responsible for creating an atmosphere of light and lax regulation. Nevertheless, there were still regulators whose job it was to apply regulations.

The answer to the question is in two parts. The first applies to practices that were – and still are – harmful to individuals, for example those who invest individually or as part of an investment pool such as a pension fund. The second applies to practices that pose a danger to the financial system as a whole.

In the first case, the protection of investors against financial harm, the underlying philosophy of financial regulators is very different from the philosophy that motivates other agencies, such as the U.S. Food and Drug Administration (FDA). The FDA, which oversees foods and pharmaceutical drugs, must approve all new drugs and monitor food processing plants for food safety. The FDA diligently prohibits inaccurate or misleading statements in advertising or marketing materials about the benefits of foods or drugs.

Unlike the FDA, the regulators of financial products do not have to approve all new financial products before they are released, and they are relatively indifferent to the verifiability of the advertising and marketing claims made for those products. As a result, financial products and services are advertised and marketed in ways not far removed from the sales and marketing of patent medicines in the 19th century. Instead of having to pass claims about the benefits of financial products and services through diligent regulatory agencies for approval, financial industry marketers and innovators are able to design and market products within a very loose regulatory framework. Where the FDA considers its job to be guarding the physical health of the consumers of food and drug products, financial regulators do not consider it their job to guard the financial health of consumers of financial products and services, except against blatant fraud.

Some proposals have been mounted to make the regulation of financial products and services more like the regulation of food and drugs by the FDA (Posner 2013), but they are unlikely to go far given the continuing anti-regulatory mood in the field of economics, in government, and in academia.

On the other hand, enhanced regulation to protect the financial system itself from systemic risk is being seriously considered, and some has been passed. Prior to the GFC, regulators might occasionally challenge the risk management assumptions used by banks, for example in their VaR models; but the regulators were at a disadvantage in the ensuing debates because the banks had more knowledge of their systems and investments than the regulators did. Besides, there was a philosophical disposition to believe that financial institutions were in the best position, and had incentives, such as protecting their reputations, to regulate themselves. Since the GFC, however, the U.S. Federal Reserve has required banks to perform regular stress tests. It recently vetoed capital-depleting measures such as stock buybacks and dividend payouts when it believed they would result in capital inadequacy (Hall, Chon, and Arnold 2014).

It remains to be seen whether the new regulatory measures that are being considered and that have been passed will mitigate systemic risk. This is discussed further in Section 4. So far, however, it

appears unlikely that protections for individual consumers of financial products and services against deceptive and misleading advertising and marketing will be substantially strengthened.

Summary and Conclusion

This section has shown that the outputs of mathematical risk models are unreliable at best and can be almost meaningless at worst. The more they are used and mandated, the more they will fail or produce erroneous results, be manipulated and gamed and even cause financial disruptions as the ratings agencies' triple-A ratings did.

Nevertheless, the insights of mathematical modeling are valuable and the knowledge gained by modeling and by the modelers themselves can be put to good use. The process of risk management should, however, be the result of the use of a diversity of risk management tools, some quantitative and some qualitative. Qualitative tools should include planning and stress-testing under various potentially threatening scenarios. The ultimate deciding factor should be judgment, informed by a variety of information including model outputs.

Modelers themselves might be better used if they were not confined to a laboratory where they spend their time implementing and trying to perfect mathematical risk models. Instead, they should be mingling with and monitoring the activities of traders, with a mandate to report potentially excessive risks when they see them. Ideally, they could engage with the traders and with each other and discuss the issues on a daily basis, reporting their findings in a concise memo to top management. This method would prove to be more effective than using a grandiose, but extremely flawed, mathematical model to report their findings, which are usually reduced to a single number.

3. Financial Risk and Uncertainty in the Macroeconomy: Trust and Stability

The subject of trust and stability covers a multitude of issues, both long-term and short-term. A key conclusion of this section will be that there is a large hole in economic theory. An area that has not been explored enough and needs to be given much more attention is the study of rational competitive behavior leading to undesirable equilibria or disequilibrium. The assumption that a benign "invisible hand" always rules has crowded this field of study out of mainstream economics, moving it to a back burner. This topic needs to be moved front and center.

The Necessity of Trust

Until the early 20th century, means of enforcing a debt contract, or sanctions if the contract were breached included imprisonment of the debtor or worse. Now, the worst the debtor can experience is bankruptcy from which he can emerge and restart.

The reasons for the elimination of debtors' prisons were complicated. But a debtor must trust that if he is unable to repay, the consequences will not be too dire or he will not borrow. A creditor must trust that if the debtor is in default he will not be incapacitated, or it will be impossible to renegotiate and eventually be repaid.

When a depositor deposits money in a bank, the depositor must not only trust the bank, but he must also trust that the other depositors will not suddenly withdraw their funds all at once. That is, the depositor must trust the other depositors to maintain their trust in the bank, and in each other.

In U.S. history, the government has often stepped in to shore up confidence when a breakdown in trust has occurred in the financial system. After the bank runs and bank failures of the 1920s and 1930s, the government established Federal Deposit Insurance to insure bank accounts up to US$100,000. In 2008, the government temporarily raised that guarantee to US$250,000 in response to the financial crisis.

In September 2008, a large money market fund, the Reserve Primary Fund, "broke the buck" when the value of its shares fell to US$0.97 the day after Lehman Brothers filed for bankruptcy protection, jeopardizing the fund's short-term Lehman loans.

There had never been any assurance given to investors that the value of a money market fund would not fall below the purchase price. Investors should have understood that, but the U.S. Department of the Treasury established the Exchange Stabilization fund anyway to ensure that the net asset value of money market funds would never fall below US$1 (SIFMA Money Market Reform Resource Center, n.d.). It did this to ensure that the key economic requisite of trust did not erode too far because investors in the fund may have had the impression that it could never lose money, and when it did, panic could have resulted.

After the stock market crashed on October 19, 1987, U.S. Federal Reserve Chairman Alan Greenspan pumped money into the economy to provide reassurance that the economy would keep moving and not collapse into recession. When the Federal National Mortgage Association ("Fannie Mae") and the Federal Home Loan Mortgage Corporation ("Freddie Mac") became insolvent, the U.S. government rescued them even though they were only "quasi-governmental" entities (supposedly privatized, but

previously owned by the government). When most of the large banks in the U.S., Ireland and Spain were threatened with bankruptcy in 2008 their governments rescued them.

All of these government backstopping measures were taken to preserve trust in the financial system and to prevent sudden panic. A certain level of trust in the financial system and in its individual institutions is necessary for it to function. Trust is difficult to establish and can deteriorate very rapidly. Not only does it depend on how trustworthy the system or an individual institution within the system is, but how trustworthy others believe it to be. When the trust of individual participants in a collective begins to erode, it can spread like an epidemic.

Most of the problems addressed in theories of finance, financial regulation and the macroeconomy concern themselves with how collective trust plays out over time, and how government should intervene to regulate and moderate the ebb and flow of collective trust. What is not studied adequately is the interaction between trust and the participants in a collective, what is necessary to maintain trust, when a collective reaches a tipping point where trust collapses and other related problems.

When trust is at a high level, expectations of risk-mitigation are high. It is a time when the consequences of risk are asymmetric. Rewards for success are high but an anticipated safety net of trust in case of failure makes expectations of ill consequences of failure low. It is a time of high "moral hazard" – the danger that excessive risks will be taken because the consequences of failure are expected to be shared with the collective. When trust is at a low level, anticipation of risk-mitigation is low and expectations of ill consequences of failure are high. It is a time when an injection of moral hazard might be a good thing, because it will help to restore trust.

Macroeconomic Theories, Trust, Moral Hazard and Expectations

Macroeconomic theories differ in their views of the way in which these characteristics of an economy – trust, moral hazard and expectations – ebb and flow, and what effect government actions will have on that ebb and flow. The theories, though they prescribe government actions that they argue will have specific effects on that ebb and flow, do not, oddly enough, occupy themselves greatly with stating the ultimate long-term objectives of shaping that ebb and flow.

Current macroeconomic theories can be roughly divided into two groups according to their beliefs about the ebb and flow of trust in the collective. These two groups can be identified with Keynesianism on the one hand, and the rational expectations hypothesis on the other.

Keynes believed that expectations about the future flowed in "waves of optimism and pessimism" that helped determine the level of economic activity (Sargent 2008). Hence, there is a natural pro-cyclicality to expectations. Optimism entails trust in the collective and positive expectations, which Keynes believed begets more trust and positive expectations. On the reverse side, pessimism begets pessimism. Natural sentiments tend to intensify a wave of optimism or pessimism. Many economists who hold this view believe that government needs to step in to exert a countercyclical force on the natural pro-cyclicality of economic bubbles and crashes.

Adherents of the rational expectations hypothesis believe, on the other hand, that the ebb and flow of trust and expectations tends to be countercyclical. As the economist and Nobelist, Thomas J. Sargent, states it, "in recurrent situations the way the future unfolds from the past tends to be stable, and people

adjust their forecasts to conform to this stable pattern." Thus, when a perturbation from the normal pattern begins, they rationally see it as only a perturbation and do not let it change their expectations. Those stable expectations will work naturally to suppress the perturbation.

These are obviously very different, almost polar opposite, theories of macroeconomics. We now delve a little further into the implications that have been derived from each.

Rational Expectations and the Dynamic Stochastic General Equilibrium Model

There has been an ebb and flow also in the fortunes of the two warring versions of economics, depending in large part on how well they seemed to have been modeling current economic conditions. After the Great Depression, Keynesian economics came to be dominant. Enthusiasm for it waned in the 1970s when one of the corollaries that had been inferred by Keynesian disciples, that higher inflation would correlate with lower unemployment and vice versa, was controverted by the advent of "stagflation" – inflation combined with high unemployment. That paved the way for the ascendancy of the rational expectations hypothesis and its offshoots beginning in the 1980s.

A large part of the recent efforts of academic economists has gone into development and refinement of a mathematical model spawned by the rational expectations hypothesis, the Dynamic Stochastic General Equilibrium (DSGE) model. Box 3.1 provides more information about this model and a comment by Nobelist economist Robert Solow, but the key feature is that it reduces the economy to a single "representative agent" with rational expectations, who stands in for all of the individual actors in the economy collectively.

Box 3.1. Dynamic Stochastic General Equilibrium

In an introduction to a U.S. House of Representatives hearing on economic models (Building a science of Economics 2010), it was stated that "The dominant macro model has for some time been the Dynamic Stochastic General Equilibrium model, or DSGE." The introduction then goes on to state:

"The agents populating DSGE models, functioning as individuals or firms, are endowed with a kind of clairvoyance. Immortal, they see to the end of time and are aware of anything that might possibly ever occur, as well as the likelihood of its occurring; their decisions are always instantaneous yet never in error, and no decision depends on a previous decision or influences a subsequent decision. Also assumed in the core DSGE model is that all agents of the same type—that is, individuals or firms—have identical needs and identical tastes, which, as 'optimizers,' they pursue with unbounded self-interest and full knowledge of what their wants are. By employing what is called the 'representative agent' and assigning it these standardized features, the DSGE model excludes from the model economy almost all consequential diversity and uncertainty—characteristics that in many ways make the actual economy what it is. The DSGE universe makes no distinction between system equilibrium, in which balancing agent-level disequilibrium forces maintains the macro-economy in equilibrium, and full agent equilibrium, in which every individual in the economy is in equilibrium. In so doing, it assumes away phenomena that are commonplace in the economy: involuntary unemployment and the failure of prices or wages to adjust instantaneously to changes in the relation of supply and demand. These phenomena are seen as exceptional and call for special explanation."

In his testimony at the hearing, economics Nobelist and MIT professor emeritus Robert M. Solow said of the DSGE model, "Here we are near the bottom, as the chairman said, of a deep and prolonged recession and the immediate future is very uncertain. We are in desperate need of jobs, and the approach to macroeconomics that dominates the elite universities of the country and many central banks and other influential policy circles, that approach seems to have essentially nothing to say about the problem. It doesn't offer any guidance or insight and it really seems to have nothing useful to say."

This is the end result of taking the invisible hand hypothesis to extremes – that the economy acts as if it were one single individual maximizing its utility. No wonder that the model neither predicted, nor has anything to say about the financial crisis of 2007-2009. That crisis was the result of the interaction of individual agents within the economic collective. It could not be explained by regarding the collective as if it were a single individual.

Resurgence of Keynesianism after the GFC

After the GFC, many of the tenets of rational expectations seemed to have failed because they were unable to either predict or explain the crisis. Instead, the theories of a Keynesian disciple named Hyman Minsky, specifically his "financial instability hypothesis," appeared to describe what had happened. In his paper on the financial instability hypothesis Minsky states, "from time to time,

capitalist economies exhibit inflations and debt deflations which seem to have the potential to spin out of control. In such processes, the economic system's reactions to a movement of the economy amplify the movement – inflation feeds upon inflation and debt-deflation feeds upon debt deflation (Minsky 1992)."

Similar views had been expressed earlier by other Keynesian economists such as John Kenneth Galbraith, whose statement of the hypothesis in a more tongue-in-cheek vein is in Box 3.2. However, since the Keynesian view had gone out of fashion in academia while rational expectations and the DSGE model were in fashion, the Keynesian view had to be revived.

Box 3.2. The Bezzle

From *The Great Crash of 1929*, a 1954 book by John Kenneth Galbraith:

"To the economist embezzlement is the most interesting of crimes. Alone among the various forms of larceny it has a time parameter. Weeks, months, or years may elapse between the commission of the crime and its discovery. (This is a period, incidentally, when the embezzler has his gain and the man who has been embezzled, oddly enough, feels no loss. There is a net increase in psychic wealth.) At any given time there exists an inventory of undiscovered embezzlement in — or more precisely not in — the country's businesses and banks. This inventory — it should perhaps be called the bezzle — amounts at any moment to many millions of dollars. It also varies in size with the business cycle. In good times people are relaxed, trusting and money is plentiful. But even though money is plentiful, there are always many people who need more. Under these circumstances the rate of embezzlement grows, the rate of discovery falls off and the bezzle increases rapidly. In depression all of this is reversed. Money is watched with a narrow, suspicious eye. The man who handles it is assumed to be dishonest until he proves himself otherwise. Audits are penetrating and meticulous. Commercial morality is enormously improved. The bezzle shrinks…"

Alternatives to the Current Mainstream Academic Models

The DSGE is currently the mainstream academic macroeconomic model, but as Box 3.1 makes clear it is purely an academic exercise and not applicable to – and perhaps not even intended to be applicable to – the real world. Such purely theoretical exercises are not uncommon in academia. The tradition of the field of pure mathematics is that its findings are intended to be contributions to abstract mathematics and not to applied mathematics, which is often found in an entirely different department at a university. The current fashionable model of physics, string theory, occupies much of physicists' efforts but is criticized for being not only untested against actual physical phenomena but untestable. Economist David Colander says in his textbook Economics, "Modern macroeconomists see their models as only indirectly relevant for policy. For example, when Robert Lucas, a Nobel Prize-winning modern macroeconomist at the University of Chicago, was asked what he would do if he were appointed to the Council of Economic Advisers, he said that he would resign" (Colander 2010).

Nevertheless, applied economists, like applied mathematicians, seek to develop theories that can directly provide answers to real-world problems. Since the prevailing theory of macroeconomics has

failed to do this, economists are considering what other avenues to pursue. Two of these are agent-based models, or what we will call "rational irrationality" models, after a term popularized by economics journalist John Cassidy.

Agent-based Modeling

Agent-based models (The Economist 2013) are computer simulations of an economy in which the economy is modeled from the bottom up by creating a large number of simulated economic entities or actors, called agents, and assigning to each of them interactive characteristics whereby they act together with other agents in the economy. The simulation is then run to see what results it brings. No specific macroeconomic assumptions are made. Features and patterns of the macro-economy are "emergent properties" observed only when the simulation is run. This approach has the advantage of not being bound by assumptions about economic aggregates that restrict macroeconomic theories.

Rational Irrationality Models

We have chosen to label this class of models "rational irrationality" models, but they also go by a number of other names. One of them is "Tragedy of the Commons" – a process in which a group of people each responding rationally to their incentives winds up creating a situation in which their welfares are all reduced. In his famous 1968 article "The Tragedy of the Commons" (Hardin 1968), sociologist Garrett Hardin presented the case of a commons on which numerous cattle owners graze their herds. Each owner has an incentive at any moment to add one more cow to the herd being grazed, because the benefit will be greater than the cost to the owner from the incremental damage done to the commons. And yet, eventually the commons becomes overgrazed and all cattle owners suffer losses exceeding their benefits. This is related to the classic environmental economics problem of "externalities," in which an action by agreement between private parties inflicts damage, such as pollution, upon uncompensated third parties.

It is also closely related to a problem of game theory known as the "Prisoner's Dilemma," in which two prisoners are separately interrogated and asked to confess and incriminate the other. The rewards or punishments are different if one prisoner confesses, neither does or both do. The equilibrium turns out to be a result that is sub-optimal for both; if they could meet and agree, that is, if they could play a "cooperative game" instead of the "non-cooperative game" that has been set up for them, they could have arrived at a much better result for both.

A bank run is a typical example in finance of a non-cooperative game. Bank runs have been studied in theory. The classic article on the subject was published in 1983 by Douglas W. Diamond and Philip H. Dybvig (Box 3.3). It concludes unsurprisingly that the result of demand deposits at a bank can be multiple equilibria, one of which is a "tragedy of the commons" in which depositors, suspecting that others will do so, run on the bank to withdraw their deposits all at once and are therefore unable to do so because the bank only keeps enough in reserve for the normal flow of withdrawals.

Many phenomena in finance could be modeled as similar to bank runs, including asset bubbles and crashes. In these cases, the actions of individual participants in the market are conditioned on what they think other participants will do. If participants believe that other participants will all buy, they will try

to buy before them, while if they believe that other participants will all sell, they will try to sell before them.

The solution in the case of Garrett Hardin's Tragedy of the Commons is to partition and enclose – privatize – the commons. It is, however, less clear what the solution is in the cases of financial tragedies of the commons.

What is clear is that such phenomena have not been studied enough in light of their tragedy of the commons, rational irrationality nature, in large part due to the dominance of a single-equilibrium invisible hand optimality fixation.

> **Box 3.3. The Theory of Bank Runs**
> The classic Diamond-Dybvig paper on bank runs says in its abstract: "Traditional demand deposit contracts which provide liquidity have multiple equilibria, one of which is a bank run. Bank runs in the model cause real economic damage, rather than simply reflecting other problems. Contracts which can prevent runs are studied, and the analysis shows that there are circumstances when government provision of deposit insurance can produce superior contracts."

The Effects of and Remedies for Moral Hazard

We have earlier described the various measures that have been taken in the United States to avert financial panics, including Federal Deposit Insurance, insurance against the possibility that the net asset value of money market mutual funds will fall below its purchase price, bailouts and other government actions. It is hotly debated how many of these measures the government really should take and should have taken to avert the potential for panic. Austrian economists would argue for less or none, while other economists would argue that they are needed. It is hoped that intensified study of the "rational irrationality" characteristics of the actions of individuals comprising a collective, perhaps aided by the use of agent-based modeling, can help to clarify when and how many of these measures should be taken.

When these measures to shore up trust are taken, however, they create "moral hazard" – they encourage economic actors to take more risk because they anticipate that the collective will backstop them if they fail. However, the more risk they take the more the burden falls on the collective to backstop them, which can encourage more risk-taking. It is this self-reinforcing cycle that caused banks before the GFC to be highly-leveraged, maintaining as little capital as possible while borrowing as much as possible to invest in risky credit derivatives.

The main problem of financial regulation is how to contain this propensity to take more risk that is then socialized to the collective once the inclination of the collective to absorb this risk has been announced by its past actions to forestall financial panic.

In the absence of moral hazard, the Miller-Modigliani theorem says that a corporation, including a bank, should be indifferent as to how much of its capital it obtains from equity and how much from debt. But in the presence of moral hazard, banks leverage up to get as much of their capital as possible from debt. In Section 4 we shall take up the possible and proposed regulatory measures that can be taken as a response to moral hazard to contain the risk assumed by financial institutions, which can propagate and become systemic risk.

Section 4. Financial Regulation and the Basel Accords

The Basel rules were a major cause of the GFC. In combination with credit agency ratings, whose central position was embedded in the Basel rules, the Basel rules were a necessary, if not sufficient, condition for the crisis to occur.

The Basel-induced process that led to the GFC began in 2001, when the Recourse Rule, an amendment to the 1988 Basel accords, was adopted in the United States. It conferred Basel I's 20 per cent risk weight on privately issued asset-backed securities, including mortgage-backed securities (MBSs), that had received a AA or AAA rating from a Nationally Recognized Statistical Rating Organization (NRSRO or credit rating agency). NRSROs are credit rating agencies such as Moody's, S&P and Fitch that have been approved for registration by the U.S. Securities and Exchange Commission (SEC) (Friedman 2011, 26). Basel II later extended this rule outside of the United States.

Shortly after adoption of the Recourse Rule, issuance of CDOs soared (Figure 4.1).

Figure 4.1. Global CDO Issuance (Bernstein and Eisinger 2010)
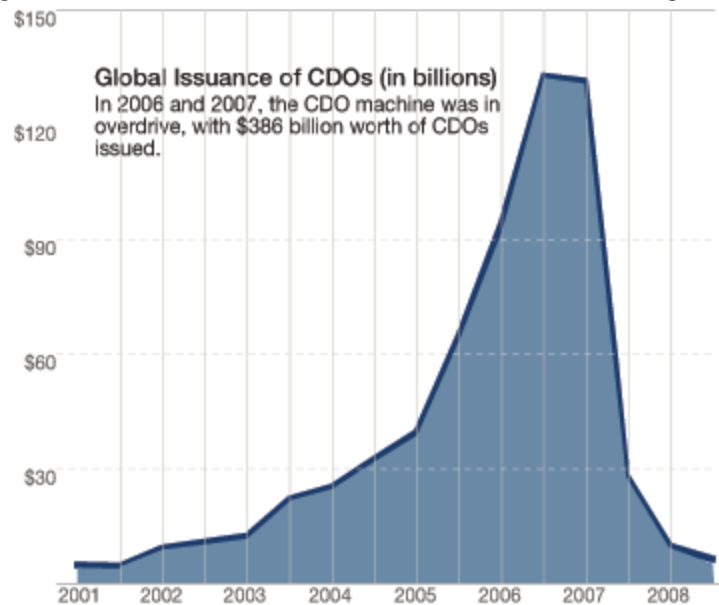
Source: Asset-Backed Alert

The mechanism that brought this about was as follows. In the United States, Basel I capital adequacy requirements were implemented to oblige "adequately capitalized" banking institutions to hold at least 8 per cent of their assets in "Tier 1" and "Tier 2" capital (equity plus some subordinated debt), and required "well-capitalized" institutions to hold 10 per cent.

However, this percentage to be held in Tier 1 and Tier 2 capital varied according to how risky the institution's assets were. Hence, the capital requirement came to be known as a requirement to hold in capital a certain percentage of "risk-weighted assets" (RWA).

For example, no risk was assigned to cash, gold and sovereign bonds issued by OECD governments, or commitments with an original maturity of up to one year. Therefore, if all of the assets of the bank were invested in these assets then no Tier 1 and Tier 2 capital would be required.

A 100 per cent risk weight was assigned to commercial real-estate loans and corporate bonds, among other investments. Hence, if all of the assets of the bank were held in these assets then its minimum

capital adequacy requirement to be "well-capitalized" – the amount to be held in Tier 1 and Tier 2 capital – would be 10 per cent of total assets (100 per cent of the 10 per cent requirement).

A 50 per cent risk weight was assigned to mortgage loans. Hence, their capital requirement was 5 per cent (50 per cent of 10 per cent).

However, the Recourse Rule adopted in 2001 (and later Basel II) assigned only a 20 per cent risk weight to AAA- and AA-rated asset-backed securities, such as CDOs. Therefore, the capital requirement for these assets was only 2 per cent.

Banks had a preference for holding as little of their assets as possible in Tier 1 and Tier 2 capital. They preferred to be as highly leveraged as possible because their risks were largely socialized due to Federal Deposit Insurance and other bank safety nets provided by the federal government.

Because the capital adequacy requirement for mortgage-backed securities, such as CDOs of MBSs, was only two-fifths as much as the capital requirement for individual mortgages, banks had an incentive to sell their individual mortgages to packagers (i.e., securitizers) and use the proceeds to purchase CDOs of mortgage-backed securities back from the packagers instead. That way they could leverage up and receive a higher return on equity than by investing in individual unpackaged mortgage loans.

This incentive drove the mortgage securitization mill. It thus drove the alienation of banks from the actual borrowers and the insertion of a chain of intermediaries instead. It also drove securitized mortgages to fall in quality because in the securitization chain the investor was alienated from the borrowers and had less information about them.

Nevertheless, the ratings agencies, which had been placed on the process's critical path by the Basel rules and the Recourse Rule, continued to assign AAA ratings to CDOs of deteriorating quality. They did this, as noted earlier, because of their commitment to use mathematical models, only "hard" historical default rate data and the low default rate correlations that they assumed in the models.

Economist Gary Gorton attributed the crisis to the fact that "the interlinked or nested unique security designs that were necessary to make the subprime market function resulted in a loss of information to investors as the chain of structures, including securities and special purpose vehicles, stretched longer and longer" (Gorton 2009).

That is, the loss of information at the investor end of the chain that occurred because mortgages had been separated by such a distance through a long chain of intermediaries from the actual borrowers made it difficult (or nearly impossible) for investors to evaluate the investments properly. This was apparently also the case for ratings agencies, in addition to problems associated with their calcified analytic traditions, as well as their conflicts of interest.

The regulatory arbitrage that was a response to the Basel I rules and the Recourse Rule propelled the replacement of traditional mortgage lending with this chain of intermediaries.

Structured Investment Vehicles and the Basel Rules

Foreign banks purchased almost as many CDOs as U.S. banks, even though they were not subject, until later, to the risk-weightings of the Recourse Rule. Why was that? The reason was that the Basel rules gave them another way they could leverage up to buy the higher-yielding CDOs.

Recall that a zero risk weighting was assigned to commitments with an original maturity of up to one year. European, as well as American, banks found that they could set up shadow banking entities called structured investment vehicles (SIVs) off their balance sheets to invest in CDOs.

These SIVs, based on standard banking practice, were funded with short-term obligations for long-term investments. Their bank sponsors invested or raised the funds by issuing asset-backed commercial paper with less than a one-year maturity. In this way, they could almost completely avoid the Basel capital requirements. The bank sponsor only needed to provide a credit or liquidity enhancement of at most 0.8 per cent of assets (Jablecki and Machaj 2011, 213-214).

Thus, the Basel rules and their implementation in the United States caused a complex web of regulatory arbitrage devices to be constructed, each of which made financial markets more opaque and convoluted. Far from a bottom-up, self-organizing system, the structure of finance was the organic offshoot of a particular set of regulations and the modeling processes that they required, like the marine colony that grows around a sunken ship.

The Endogenous Unintended Consequences of Model-based Rules

The hedge fund manager and philanthropist George Soros has advocated a philosophy of "reflexivity," which says that by appraising the fundamentals of an investment, an analyst can change those fundamentals. A forecast of a bright future for a company, for example, can enhance the company's ability to raise capital, while a dismal forecast can impair it. Therefore, as has been discovered in other fields such as physics, there is no such thing as a wholly detached observer describing a situation objectively.

A similar observation has been made about economics by sociologists Donald MacKenzie (2005) and Michel Callon (2006). They say that in some cases, economic theories, instead of being descriptive, can cause the economy to behave in a way that they predict. For example, MacKenzie noted that option pricing theory, for a while, caused option prices to behave in a way that the theory predicted because everyone was using it. They call this the "performativity" of economics.

However, MacKenzie also observed that economics can be "counter-performative" (Box 4.1). In these cases it can cause the opposite of what models predict to occur.

This appears to be what has happened as a result of the Basel rules and their implementation. When a flexible, malleable and organic process is required to touch down at a certain limited number of fixed points– like a tree whose roots are cemented over – it will adjust in a devious and not wholly predictable manner.

> **Box 4.1. Counter-performativity (Goldman, the SEC and performativity 2010)**
>
> "MacKenzie (2009) focuses on the use of the rating system to value credit derivatives such as collateralized debt obligations. The use by the credit rating agencies of a fixed set of characteristics, clearly specified and communicated to investors through software packages such as S&P's Evaluator, made it possible for investors to game the system. It worked as follows: every security has multiple qualities and characteristics. Knowing that only a few of them are quantified and used to value the credit derivatives, the arrangers of these derivatives took bonds of mortgages that were on the whole close to worthless but that looked good from the narrow standpoint of the credit rating. Investors could naturally choose to evaluate these mortgage derivatives themselves. But given their sheer complexity, the speed at which they had to respond to bid-lists (a few hours), and the fact that credit rating had worked perfectly well in the past, few of them chose to do this."

Should the Model-based Framework be Reconsidered?

The response to the numerous signal failures of modeling, regulations based on modeling and internal risk-management systems based on modeling, before and during the GFC, has been to tinker with the models to close the loopholes that evasive action drove trucks through the last few times they were used.

It might be more useful, however, to step back from modeling and consider whether there is something inherent about its use in finance and economics, or something about the mainstream modeling framework, that makes its application ineffective at best and counterproductive at worst. It is worth considering whether a different approach might serve better to achieve the ultimate goal.

The Purpose of Capital Adequacy Requirements

Non-financial companies do not need government regulation to ensure that their ratio of debt funding to equity funding is not too high. It is assumed that they do not have an incentive to seek the highest leverage possible.

The theory advanced by Merton Miller and Franco Modigliani in 1958 suggests why. Their work concluded that the value of a firm is unaffected by its ratio of debt to equity funding as laid out in the following reasoning. Assume the firm has a value at a certain debt to equity ratio. The equity funding is more costly to the firm per unit than the debt funding because investors perceive equity as more risky. Then is there an incentive to the firm to increase debt and reduce equity?

If debt were increased however, it would make both debt and equity more risky because of the smaller cushion against bankruptcy. Hence, increasing debt would reduce capital costs if all else were equal, but would raise them because it would increase the riskiness of both debt and equity.

The Miller-Modigliani theorem argues that under certain assumptions, these two effects wash, so that the aggregate cost of capital and the value of the firm are insensitive to the debt to equity ratio.

This argument is invalidated, however, if the risk is curtailed. This is the case in the financial industry, in which the government has felt impelled, at least since the bank runs of the 1920s and 1930s, to provide government guarantees to buffer financial institutions against panic. These guarantees have broadened since that time, reaching their zenith (some might say their nadir) with the extensive bank bailouts of 2008.

The resulting moral hazard induces financial institutions to leverage up and take more risks. The response of government has been to establish regulations requiring minimum levels of equity and near-equity-like capital to provide a buffer against loss.

The task of the Basel Committee on Banking Supervision (BCBS) in creating the Basel Accords (Basel I, II, and III) is primarily to make recommendations to countries for capital adequacy requirements that can be implemented in their national banking regulations. The purpose is to harmonize these requirements across countries so that there will not be a bottom-seeking process in which financial institutions look for a country with the lowest capital requirements to base their operations.

What is the ultimate purpose of government support for financial institutions?

The ultimate purpose of this whole process, however, is not merely to prevent financial companies from taking as much risk as they want. The purpose is to minimize the burden on the government – that is, on the taxpayer –in preventing some of the consequences of that risk.

Hence, ideally, government regulation and supervision should separate financial institution risks whose consequences are socialized from those that are privatized. If that can be done, government can seek to abstain from undertaking any risk-mitigation in those areas where the consequences of risk can be privatized.

Government should then focus on the risks whose consequences are unavoidably socialized. Determining which ones fall into this category may not be an easy task.

Risk Separation and Structural Reforms

The Basel capital adequacy requirements do not make a distinction between these two forms of risk. Government guarantee programs such as federal deposit insurance do not adequately make the distinction either. That is why programs that insure banks against both public and private consequences of their risk-taking have become the norm.

"Structural reforms" being proposed in the U.S., UK and EU may go some distance to rectify this situation. The original structural reform was the Glass-Steagall Act of 1933 which, in response to the 1929 crash and the Great Depression, required investment banking and deposit banking institutions to be separate. That Act was repealed in 1999 in an atmosphere of belief in a lightly-regulated financial industry and its powers of risk diversification and reduction.

Since the GFC, in addition to the negotiation of the Basel III rules – which still use risk-weighted assets determined by fallible models as the criteria for capital adequacy and do not explicitly separate socialized risks from privatized risks – structural reforms are being proposed that attempt to isolate socialized risks. Structural reforms have been proposed by the Liikanen Report in the EU chaired by former Bank of Finland governor Erkki Liikanen, the Independent Commission on Banking in the UK chaired by former Bank of England chief economist John Vickers, and former U.S. Federal Reserve Chair Paul Volcker.

These proposals differ in detail, but all propose institutional separations between risks that could cause socialized consequences such as a panic and inability to withdraw or borrow funds, and those whose consequences are privatized. The differences have to do, in part, with differing views on the possibility that institutional separations may reduce institutions' abilities to diversify and thus reduce risks.

Recovery, Resolution and Living Wills

Yet more directly aimed at ameliorating socialized risk are requirements, embedded for example in the Dodd-Frank Act of 2010 in the U.S., that financial institutions document in detail how they will be wound down in case of failure. Such documents could enable the government to launch a process called recovery and resolution in case of bank failure, in which it recovers what value it can from bank assets by selling them to value-seeking investors.

This would have been of limited usefulness in the 2008 crisis because potential buyers were having difficulty determining the value of the assets. The bank run in this case took the form of the refusal of lenders to routinely roll over short-term loans (many of them overnight loans called repos) that were used to fund the shadow banks' long-term investments. The reason for this run on shadow banks was that an uncertain percentage of their mortgage loans were discovered to be of very low quality. However, it was not known which and what part of their assets were invested in these low-quality loans. Since no one could differentiate the good assets from the bad, all of the banks' collateral that supported their short-term borrowing became suspect.

Had living wills been prepared by the financial institutions with close monitoring and supervision by regulatory authorities, it might have been easier to determine the value of the assets. In fact, under such monitoring, supervision and requirements for documentation of securitization processes and structures, it may have been too difficult for the financial industry to create as many of these financial instruments as indiscriminately as they did.

In his preface to the republished version of his 2007 book "A Demon of Our Own Design," author and former hedge fund risk manager Richard Bookstaber writes about the CDO crisis, "it is hard to understand how this elephant in the room was missed. How can a risk manager at a place like Citigroup see inventory grow from a few billion dollars to ten billion dollars and then to thirty or forty billion dollars and not react by forcing that inventory to be brought down?"

He concludes it is likely that "The risk managers did not have the courage of their conviction to insist on the reduction of this inventory, or the senior management was not willing to heed their demands… Even a risk manager who got it right might not have been able to carry the day against the traders."

Internal financial institution oversight, monitoring and risk management are inevitably compromised, conflicted and repressed, but diligent external oversight and monitoring by government authorities that is done clearly on behalf of protecting taxpayers' ability to unwind the institution before disaster hit might have uncovered the problems.

Thinking the Unthinkable

The government has no grand plan to step in to replace the national food supply chain in case it somehow breaks down. It does not even have a plan to intervene in the event of a failure of the electricity supply chain, even though that supply chain does break down massively from time to time even in the most developed countries, usually with only brief and non-catastrophic effect.

Is it necessary, nevertheless, for the government to create a plan to step in if the financial supply chain breaks down and people and businesses cannot spend their own money? This may not seem necessary now, but it may become necessary if another, larger financial crisis hits in the future bringing the economy to a complete standstill. In fact, should that happen, it may seem that the government should have already created such a plan. Consequently, the government should at least, as an exercise if not as a precaution, consider as a research project how it might intervene and take over the supply chain from a bleeding and supine financial industry.

This should be contemplated not only because the financial supply chain could break down if borrowing and lending were to freeze. In the event of the GFC there were signs that the banking

system could even take risks with the ability to maintain adequate depositor, borrower and investor records. If the information systems for recording in detail individual ownership of deposits, debts and investments were to break down, it could cause even more financial havoc than a temporary inability to supply funds. Nevertheless, in their haste to package CDOs, banks were negligent in keeping proper records of home ownership, a negligence that has continued to plague them and their borrowers long after the GFC (Paltrow 2011).

Even if no plan were actually intended to be implemented, the exercise of creating a plan in principle for the government to step in directly to shore up the financial supply chain could help determine what financial institutions should include in their living wills, and how the remnants of financial institutions and government would work together to salvage the financial supply chain in case of disaster. It would also be an integral part of scenario planning and stress-testing of financial supply chains under the various scenarios.

This is merely an extension to the limit of the planning and implementation of the living will provisions of the Dodd-Frank Act now taking place. In the United States, such a "thinking about the unthinkable" exercise could be attacked by anti-government groups as a conspiracy to take over the banking system. In the 1950s and 1960s however, the original "thinking about the unthinkable" exercises that were carried out to consider how a nuclear war between the United States and the Soviet Union might play out were attacked by anti-war groups as more likely to cause a war than to deter one. Such exercises may nevertheless have been instrumental in averting war.

The Cultural, Traditional and Ethical Context

There is no such thing as a society, market or business environment determined solely by the structure of a country's laws, regulations or even institutions. Those well-defined features are undergirded by innumerable tacit cultural and societal understandings concerning both interpersonal and business interactions without which laws, regulations and institutions would not work. Those understandings cannot all be explicitly written into laws and regulations. Even though they share some universal characteristics, they are to some degree culture-specific. They can also change subtly over time, sometimes making it more difficult (or easier) for laws or regulations to work or be enforced.

These understandings are sometimes conditioned by intellectual antecedents. That is why Keynes could say that "Practical men, who believe themselves to be quite exempt from any intellectual influence, are usually the slaves of some defunct economist."

Beliefs about what constitutes the most advanced form of a financial system or of financial regulation are formed in part by the prevailing body of intellectual beliefs. That body of beliefs in developed economies has been revealed to have been suffering from a form of hubris. Nevertheless, even industry and government leaders in developing economies, which one would think might be experiencing a bit of schadenfreude after the GFC, have bought into them because many of them have been educated in developed economies.

This may be a good time for those leaders to dig into their roots as a means of suspending judgment from the intellectual influence they have acquired. Other models besides those perfected in developed

economies may acquire new interest. Since the intellectual upheaval in economics occasioned by the GFC, alternative economic views may be more readily examined than before.

The relevance of the Shariah perspective on the securitization of debt, for example, may become apparent, though it would seem highly unlikely to those in developed economies of America and Europe. About one third of Asians are Muslim. One paper, "The Relevance of Islamic Finance Principles to the Global Financial Crisis" says (Rehman 2009):

> The Shariah perspective on the securitization of debt is compelling: the majority of Islamic scholars deem the sale of debt impermissible… The implications of the ban on selling debt … are sweeping. First, the separation between origination and bearing credit risk is removed, and the originator is forced to take on the credit risk. This will naturally lead to more careful lending. Second, the lack of a transfer of the debt would mean that the holders of credit risk would retain a "line of sight" to the actual loans, making them better aware of the risks. Rather than relying solely on credit ratings, lenders would have access to information specific to the underlying transactions.

The economist Paul Romer has said that "a crisis is a terrible thing to waste." In the aftermath of the GFC, it would be a good thing to reexamine what had become monolithic and almost universally accepted economic precepts – and they are indeed being reexamined. Assumptions that unfettered economic forces always lead to a beneficial equilibrium and that a proliferation of innovative financial instruments reduce risk, have both come into question.

This is a good time to examine and reconsider other models, including some that may have been discarded or assumed to be uncompetitive, outmoded or unsophisticated. The assumption that the Western economic model is supreme has been severely undermined. In the ensuing debate, other philosophical approaches have earned the right to be considered on equal footing.

4. Summary and Conclusions

This paper has shown that the process of risk management can become distorted if it relies only on mathematical risk models. These models are extremely vulnerable to a host of sources of error. However elegant or seemingly sophisticated the models may be as abstracted from the real world, they frequently require inputs in order to be used in practice that are either unobtainable, very poorly estimated or subject to manipulation. Although computer programs can be relied on to do exactly what they are told, programmer and data input errors are much more prevalent than the vast majority of those who use their outputs realize. Assumptions made at different stages of a complex and compound risk assessment process can clash, or accumulate to comprise a set of assumptions that are in aggregate very different from anything that can be considered reasonable in the real world.

Nevertheless, like all attempts to abstract and represent real-world processes in theoretical models, mathematical risk models can have their beneficial uses if we understand and account for their weaknesses and potential dangers. When models are subordinated to a qualitative and judgmental process of managing risk, they can frequently lend insight, and occasionally provide alternative viewpoints that might not otherwise be considered.

In view of the dangers of mathematical risk modeling, a great deal of caution should attend any procedure that tends to substitute the results of mathematical risk models for other risk management measures. Modeling should play the subordinate role. Any attempt to subordinate other risk management judgments to the results of modeling should be looked upon with great suspicion.

This suggests that procedures that lock in place mathematical risk model outputs in rules and regulations either imposed by government or by corporate oversight bodies should be enacted only with great hesitation and caution, if at all. Uncertainties associated with unquantifiable risk are present not only in the processes that mathematical risk models attempt to replicate, but also in the modeling processes, inputs and assumptions themselves. Mathematical models may in principle represent risk in a probabilistic form, but then the models require inputs to create their outputs, and those inputs themselves are subject to high levels of unquantifiable uncertainty. The notion that everything can be assessed in a quantitative manner is an illusion.

This has implications both for the modeling of financial risk at individual institutions and modeling of the entire financial system and the economy itself. In modern times, before the global financial crisis, an impoverished and narrow range of financial and economic models had gained a monopoly on intellectual inquiry in the financial and economic fields. This monopoly is still largely in place due to momentum, but it is at last being widely questioned.

In both the fields of financial modeling and modeling of the ebb and flow of collective trust in the macroeconomy and its effects, new and different or formerly ignored models should be introduced or reinvigorated.

In the financial modeling field, less emphasis should be placed on volatility-based models. For too long mean-variance-covariance constructs have completely dominated financial risk modeling. A naive newcomer to the field would wonder why risk is being modeled solely as the level of variation of market prices when presumably true financial risk is more akin to the likelihood of running out of money.

Financial risk modeling, as currently practiced, is similar to modeling the risk of air travel only by measuring the vibration of the plane without assessing the probability that it will crash. If the MPT model can keep from crowding out other, entirely different models, theoretical innovation in the financial field might be revitalized.

In macroeconomic modeling too much attention has been paid to models that assume the macro-economy can be modeled as if it were one person or group of persons with a coherent flow instead of as a large group of separate individuals engaged in numerous negotiations or games with each other, both cooperative and non-cooperative. A number of interesting alternative macroeconomic modeling procedures have begun to be explored that have the potential to bear fruit, and could be useful in practice in the future.

In short, far more diversification of risk models is needed than has prevailed in recent years – both in quantitative and qualitative models and in risk management systems. Ultimately, quantitative risk modeling must be subordinated to qualitative risk judgment, until such a time, which seems unlikely to arrive, when quantitative risk management systems have not only been so reliably perfected that they can take over qualitative management, but it can be proven conclusively that they can. Merely stating this as a possible prospect makes it clear how unlikely it is to come to pass, and yet much risk modeling has been implemented as if that state of affairs had already been realized.

Bibliography

Bernanke,Ben. 2004. "Remarks by Governor Ben S. Bernanke At the meetings of the Eastern Economic Association, Washington, DC: The Great Moderation." The Federal Reserve Board, February 20. http://www.federalreserve.gov/BOARDDOCS/SPEECHES/2004/20040220/default.htm.

Bernstein, Jake, and Jesse Eisinger. 2010. "Banks' Self-Dealing Super-Charged Financial Crisis." ProPublica, August 26.

Bisias, D., M. Flood, A. W. Lo, and S. Valavanis. 2012. "A Survey of Systemic Risk Analytics." Office of Financial Research, U.S. Department of the Treasury, Working Paper #0001.

Blanchard, Olivier. 2009. "(Nearly) nothing to fear but fear itself." The Economist, January 29. http://www.economist.com/node/13021961.

"Building a Science of Economics for the Real World." 2010. Hearing before the Subcommittee on Investigations and Oversight, Committee on Science and Technology, House of Representatives, One Hundred Eleventh Congress, July 20. http://www.gpo.gov/fdsys/pkg/CHRG-111hhrg57604/pdf/CHRG-111hhrg57604.pdf.

Callon,   Michel. 2006. "What Does it Mean to Say that Economics is Performative?" In Do Economists Make Markets? On the Performativity of Economics, edited by Donald MacKenzie, Fabian Muniesa, and Lucia Siu. Princeton Press. http://halshs.archives-ouvertes.fr/docs/00/09/15/96/PDF/WP_CSI_005.pdf.

Chan, Albert, Ying Wan, and Jacky Yang. 2011. "Beyond Basel III: The future of high performance in Chinese banks." Accenture. http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-Beyond-Basel-III.pdf#zoom=50.

Colander, David C. 2010. Economics, 8th Edition. McGraw-Hill.

Cousin, Violaine A. 2012. "Basel I, II, III – We Want it All at Once." World Financial Review, Forthcoming, May 25. http://ssrn.com/abstract=2066465.

Edwards, Franklin R. 1999. "Hedge Funds and the Collapse of Long Term Capital Management." Journal of Economic Perspectives 13 (2): 189-210. http://www0.gsb.columbia.edu/faculty/fedwards/papers/Hedge_Funds_&_the_Collapse_of_LTCM.pdf.

Eichengreen, Barry. 2010. "The Last Temptation of Risk." The National Interest, March 26. http://nationalinterest.org/article/the-last-temptation-of-risk-3091.

Friedman, Jeffrey. 2011. "Capitalism and the Crisis: Bankers, Bonuses, Ideology, and Ignorance." In What Caused the Financial Crisis, edited by Jeffrey Friedman. University of Pennsylvania Press.

"Goldman, the SEC and performativity." 2010. Socializing Finance blog, April 26. http://socfinance.wordpress.com/2010/04/26/goldman-the-sec-and-performativity/.

Gorton, Gary. 2009. "The Subprime Panic." European Financial Management, European Financial Management Association 15(1): 10-46. http://ideas.repec.org/p/nbr/nberwo/14398.html.

Hall, Camilla, Gina Chon, and Martin Arnold. 2014. "Fed Rejects Citi's 2014 Dividend and Stock Buyback Plans." *Financial Times,* March 27. http://www.ft.com/intl/cms/s/0/71683e54-b50d-11e3-a746-00144feabdc0.html?siteedition=intl#axzz2x7F4p9GW.

Hardin, Garrett. 1968. "The Tragedy of the Commons." Science, 13 December: 162 (3859): 1243-1248. http://www.sciencemag.org/content/162/3859/1243.full.

Hull, John C. 2012. Risk Management and Financial Institutions Third Edition. New Jersey: John Wiley & Sons, Inc.

Jablecki, Juliusz, and Mateusz Machaj. 2011. "A Regulated Meltdown: The Basel Rules and Banks' Leverage." In What Caused the Financial Crisis, edited by Jeffrey Friedman. University of Pennsylvania Press.

Knight, Frank H. 1921. Risk, Uncertainty and Profit. Cornell University Library.

MacKenzie, Donald. 2005. "Is Economics Performative? Option Theory and the Construction of Derivatives Markets." School of Social and Political Studies, University of Edinburgh. http://www.espanet2012.info/__data/assets/pdf_file/0017/3419/is_economics_performative.pdf.

MacKenzie, Donald. N.d. "The Credit Crisis as a Problem in the Sociology of Knowledge." http://www.york.ac.uk/media/satsu/events-docs-images-media/york-final-mckenzie.pdf.

MacKenzie, Donald, and Taylor Spears. 2012. " 'The Formula That Killed Wall Street'? The Gaussian Copula and the Material Cultures of Modeling." School of Social and Political Studies, University of Edinburgh. http://www.sps.ed.ac.uk/__data/assets/pdf_file/0003/84243/Gaussian14.pdf.

Markowitz, Harry. 1952. "Portfolio Selection." The Journal of Finance: 7(1), March.

Minsky, Hyman P. 1992. "The Financial Instability Hypothesis." Working Paper No. 74, Levy Economics Institute of Bard College, May. http://www.levyinstitute.org/pubs/wp74.pdf.

"New Model Army." 2013. The Economist, January 19.

Paltrow, Scot J. 2011. "Special report: Banks continue robo-signing." Reuters, July 18. http://www.reuters.com/article/2011/07/18/us-foreclosure-banks-idUSTRE76H5XX20110718.

Posner, Eric A., and E. Glen Weyl. 2013. "An FDA for Financial Innovation: Applying the Insurable Interest Doctrine to Twenty-First-Century Financial Markets." Northwestern University Law Review: 107(3). https://www.law.northwestern.edu/lawreview/v107/n3/1307/LR107n3Posner.pdf.

Protess, Ben, and Jessica Silver-Greenberg. 2013. "JPMorgan Expected to Admit Fault in 'London Whale' Trading Loss." New York Times, October 15. http://dealbook.nytimes.com/2013/10/15/jpmorgan-said-to-reach-deal-with-trading-regulator/.

Radelet, Steven, and Jeffrey Sachs. 1998. "The Onset of the East Asian Financial Crisis." Harvard Institute for International Development. http://www.earth.columbia.edu/sitefiles/file/about/director/pubs/paper27.pdf.

Rehman, Aamir A. 2009. "The Relevance of Islamic Finance Principles to The Global Financial Crisis." Panel Discussion on the Evolution of the Global Financial Crisis from the Current Crisis. Harvard Law School Islamic Legal Studies Program, March 16. http://ifp.law.harvard.edu/login/view_pdf/?file=Aamir_Rehman.pdf&type=workshops.

"Report of JPMorgan Chase & Co. Management Task Force Regarding 2012 CIO Losses." 2013. JPMorgan Chase & Co., January 16. http://files.shareholder.com/downloads/ONE/2272984969x0x628656/4cb574a0-0bf5-4728-9582-625e4519b5ab/Task_Force_Report.pdf.

"Rum remark wins Rumsfeld an award." 2003. BBC News, 2 December. http://news.bbc.co.uk/2/hi/3254852.stm.

Sargent, Thomas J. 2008. "Rational Expectations." The Concise Encyclopedia of Economics. http://www.econlib.org/library/Enc/RationalExpectations.html.

"Shareholder Report on UBS's Write-Downs." 2008. UBS, 18 April. http://www.propublica.org/documents/item/shareholder-report-on-ubss-write-downs.

Sheng, Andrew. 2013. "Outlook for Global Development Finance – Excess or Shortage?" Background Research Paper Submitted to the High Level Panel on the Post-2015 Development Agenda, United Nations Post-2015 Millennium Development Goals Development Agenda, May. http://www.post2015hlp.org/wp-content/uploads/2013/06/Sheng_Outlook-for-Global-Development-Finance-Excess-or-Shortage.pdf.

SIFMA Money Market Reform Resource Center. N.d. http://www.sifma.org/issues/regulatory-reform/money-market-reform/overview/.

"The Asian Financial Crisis, 1997-98." N.d. National Graduate Institute for Policy Studies, Course lecture. http://www.grips.ac.jp/teacher/oono/hp/lecture_F/lec11.htm.

Weitzman, Martin. 2007. "Role of Uncertainty in the Economics of Catastrophic Climate Change." AEI-Brookings Joint Center Working Paper No. 07-11, May. http://ssrn.com/abstract=992873 or http://dx.doi.org/10.2139/ssrn.992873

Acknowledgement

This report reflects work in progress and the views expressed therein are those of the authors and do not necessarily reflect those of the Fung Global Institute or Chatham House. The authors are solely responsible for any errors or omissions.