**Statistical, Financial, and Institutional Risk**

In this paper, I wish to explore and compare three different approaches to the management of risk: a statistical approach, a financial approach, and an institutional approach. This paper can be read as a stand-alone document, but it can also serve as a less formal introduction to my book Financial Risk Management. Except where explicitly noted, all references in this paper are to the Second Edition of that book.

One of the primary objectives of my book, beginning with the first edition written in 2002, was to make clear the distinction between traditional risk management approaches based on statistical analysis (sometimes labelled **actuarial risk**, because of its association with insurance) and financial risk management approaches that take advantage of liquid instruments as a primary tool. The global financial crisis of 2007 -2008 reinforced my view of the importance of this distinction, as a large part of the problem could be traced to managing very illiquid positions with risk management methods specifically designed for liquid instruments. But this crisis also convinced me that more attention needs to be paid to risks that can only be controlled through proper institutional design.

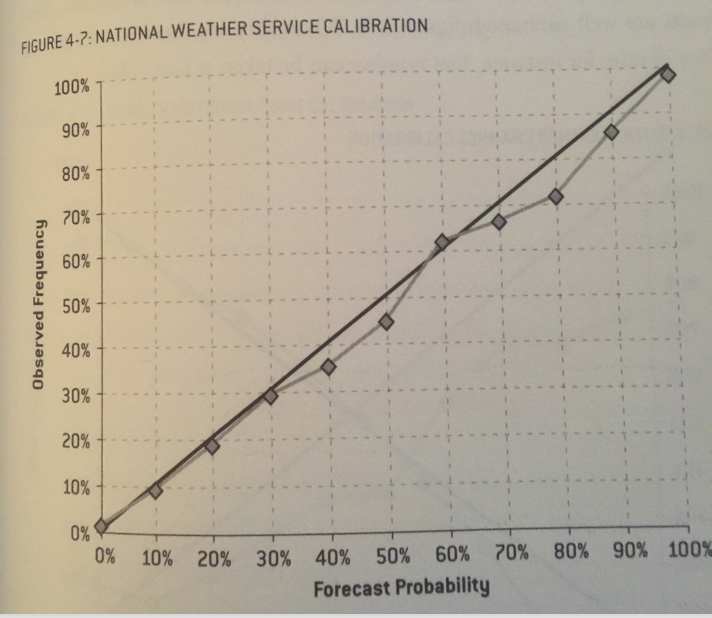
I’ll begin by defining these three different types of risk management methods – statistical, financial, and institutional. I will then discuss the application of these ideas to a number of areas: (1) trading, (2) bank lending, (3) retirement investing, (4) insurance, and (5) credit risk on derivatives. For each of these applications, I will offer an extended discussion of the techniques of financial risk management, the primary focus of my career and of my teaching, and a briefer discussion of risk management through institutional design.

**Statistical Risk Management**. The statistical approach to risk management has been actively employed ever since the 17th Century. While many accounts of the foundation of probability theory in the 17th Century by Pascal, Fermat, and the Bernoullis focus on the applications to gambling problems, a more thorough history, such as The Empire of Chance by Gigerenzer, et al, shows that at least as important a motivation was the application to pricing of early insurance contracts and annuities. From then on, probability and statistics developed hand-in-hand with applications to practical problems of decision making and risk assessment. During the Second World War, Abraham Wald and others developed a practical approach to using statistics and probability theory to guide decisions on allocation of resources and battles between ship convoys and submarines. After the war, the field of operations research applied these insights to a wide variety of problems in government, industry, and insurance, as can be seen by the prominent place of statistical decision theory in classical texts such as Hiller and Lieberman’s Operations Research. Early applications to banking and investment, such as Harry Markowitz’s portfolio theory, were closely based on this statistical decision theory paradigm, with both expected return and probability of loss of long-term investments judged by statistical models.

**Financial Risk Management**. A prominent issue for the statistical risk management paradigm was that the statistical models often depended heavily on subjective judgement. Sometimes this was explicit, with models derived from Bayesian or personalist theories of statistical reasoning; sometimes it was implicit, when historical statistical frequencies used as model inputs could vary significantly based on choices of relevant data. In any case, there was often suspicion of decisions being manipulated to meet the needs of those who had something to gain, whether business managers personally invested in a project being approved or risk controllers who had more to lose from a bad result than from a choice not to go ahead with an investment. Senior managers were often forced to get closely involved in the process to assure that decisions were being made in the best interests of the firm, with a resulting loss in speed and nimbleness of decisions-making.

An alternative approach, in which reliance was placed on the ability to quickly exit positions in liquid instruments, gave rise to a greater ability to decentralize the decision making process, with the primary monitoring being independent tracking of liquidation prices. This could be performed at a lower organizational level, with senior managers only needing to be sporadically involved, in setting and reviewing loss limits. While some use of statistics could not be avoided, since some added loss could take place in the time between when the decision to close a position was made and when it could be executed, these were mostly statistical models of shorter-term market moves, which could be frequently tested and improved.

As an example of how statistical models of short-term movements that allow for frequent testing and improvement can achieve very stable results, look at the chart just below for the close match between frequency of precipitation forecast by the National Weather Service and actual frequencies (this example and chart come from Nate Silver’s The Signal and the Noise). Contrast this with wide-spread skepticism about forecasts such as probabilities of winning elections or the impact of tax cuts on the economy or long-term climate change where models can only be tested very infrequently (for more on how statistical models can be judged in economics, Dani Rodrik’s recent Economic Rules is an excellent source).



**Risk Management Through Institutional Design**. It is usually recognized that institutional design is one way of mitigating statistical risk – the larger the aggregation of imperfectly correlated risks, the greater the greater the ability to offset losses in one segment of the portfolio with gains in another segment. This is the classical argument for having individuals and companies buy protection from an insurance company that can pool risks.

Less often recognized but even more important is the role of institutional design in providing a large base to share losses that are difficult to anticipate. I like the statement by Aaron Brown in his book Red Blooded Risk: “The first principle of modern quantitative risk management is to split the analysis into two parts. For normal, common events you have plenty of data to make reliable quantitative conclusions using historical statistics. But long-term success requires at least surviving the abnormal, unexpected, uncommon events you know will also occur.” At the most extreme tails, with little historical experience on which to base projections, statistics won’t provide assurance, and liquidity can’t be counted on. Only institutional arrangements that can spread losses broadly can assure survival – insurance, government guarantees, mutualization of losses.

This applies to all risk management, not just financial risk, but it is particularly relevant for financial risk. When you are estimating the extreme tails of a physical process, you can at least make use of knowledge regarding the underlying physical process. But with financial risk, you must remember that as Nassim Taleb says in The Black Swan: “Money in a bank account is something important, but certainly *not physical*. As such it can take any value without necessitating the expenditure of energy. It is just a number!”

But every institutional arrangement that spreads losses broadly brings with it a fear of moral hazard – “socialization of losses and privatization of gains.” Section 2.1 discusses the potential impact of moral hazard anytime there is an asymmetry between information and reward, with better-informed insiders, such as firm managers, owning more of the upside and less of the downside than less-well-informed outsiders, such as regulators. Section 2.2 discusses the closely related issue of Ponzi schemes. Well-designed institutional arrangements need to address issues of moral hazard as well as issues of assuring survival.

Unfortunately, arrangements that are advantageous for reducing statistical risk tend to exacerbate issues of moral hazard. For example, one large insurance company will achieve a larger aggregation of risks and hence a greater mitigation of statistical risk than many small insurance company. But a single large insurance company will know that it is “too big to fail,” i.e., that the government will need to come to its rescue if it is facing bankruptcy. This knowledge may lead to the taking of larger risks in pursuit of profits, figuring that profits will be shared between managers and stockholders while large losses will be mostly absorbed by the government. Possible remedies are trying to assure that managers who place the firm in such a position receive large retrospective penalties, such as “claw-backs” of past compensation, and close monitoring of positions by regulatory agencies acting for the government. But such close monitoring tends to mitigate the advantages of decentralizing decision making that financial risk management has been able to achieve.

Now let’s see how these concepts can be applied to specific areas of risk-taking.

**(1) Trading**

Let’s start by considering an idealized paradigm for managing trading risk utilizing liquidity – one that actually closely resembles a good deal of real trading situations. We can then look at how to deal with deviations from this idealized paradigm.

In our paradigm (described in more detail in Section 6.1.1), senior management allocates to a trader (or trading team) a **stop-loss limit**, a dollar amount of losses that would trigger liquidation of the trader’s positions. Unless and until this limit is reached, the trader is afforded great latitude in creating positions and in making changes to those positions with little interaction or oversight by senior managers. The positions can vary from trades that seek to benefit from very short-term market moves to ones that are based on predictions about events years in the future. The only requirement is that all positions must be taken using instruments for which a reasonably liquid market exists. Management’s willingness to grant the trader autonomy is based on the confidence that large losses can be avoided by liquidating positions.

A critical requirement for this paradigm to work is that the liquidation value of positions can be obtained frequently (probably at least daily), reliably, and independently of the trader. This process, called **mark-to-market**, is at the heart of any stop-loss system. It is only by having reliable independent valuation that management can be assured that stop-loss triggers will be effective. Key issues in marking-to-market are addressed in Section 6.1.3.

But it is not sufficient to just know when a stop-loss has been triggered – it is also necessary to have a good estimate of losses that may be incurred in the event of liquidation. Any liquidation takes time to complete and prices may move adversely during the liquidation process. If management wants to limit a trader’s losses to $30 million and estimates that there is a chance that $10 million could be lost in the liquidation process, the stop-loss limit needs to be set at $20 million. Estimates of potential liquidation cost need to be based both on statistical analysis of historical price moves, called **value-at-risk**, and economic analysis of potential price moves in the event of a temporary liquidity crisis, called **stress tests**. While this part of financial risk management still requires the same sort of probability judgments that are required in traditional risk analysis, it has the advantage of only needing to forecast for brief time horizons (the period of time over which liquidation will take place). Key issues in using value-at-risk and stress testing to estimate potential liquidation cost are addressed in Chapter 7.

Stop-loss limits do need to be negotiated between managers and traders in a sensible way. If a trader is pursuing a longer-term strategy that may incur substantial losses before paying off, it would be foolish to attempt this with a tight stop-loss limit that has a good likelihood of pulling the plug before the strategy has been given a real chance. If this requires a loss limit that is higher than what management is comfortable with, they should recognize up front that this is not a strategy they should pursue. And when a trader is close to their loss limit and (almost always) presents good reasons why the market will soon move in their favor and they just need to be granted a small increase in their loss limit, a meaningful dialogue between the trader and the managers to arrive at a reasoned judgment is needed; managers who treat requests for loss limit increases as bureaucratic box-ticking are asking for trouble.

Given my claim that this is a sensible paradigm, how can we explain the trading disasters that have all too frequently made headlines? One common theme behind many of these incidents (e.g., Barings Bank, Allied Irish Bank, Societie Generale) has been fraudulent behavior that has resulted in hiding the size of positions (detailed in Section 4.1). Another theme is the failure to recognize the illiquidity of positions, resulting in massive underestimation of the cost of position liquidation. The most prominent cases of this have been:

* Long Term Capital Management, which failed to adequately address the impact of the size of its positions relative to the depth of the market (detailed in Section 4.2.1). This threatened to have a severe impact on many large banks that were its trading partners, due to inadequate collateralization of derivatives trades, a point that we will return to in a later section on the role of collateral in financial risk management of the credit risk of derivatives.
* The JPMorgan “London Whale” loss, which resulted from a combination of traders disguising the illiquidity of their positions and management failing to recognize clear signs of this illiquidity (detailed in an accompanying paper “London Whale”).
* The catastrophic losses on collateralized debt obligations backed by subprime mortgages that caused several major firms to go bankrupt or require government rescue in 2008. There is no doubt that the primary cause of these losses was applying a paradigm designed for management of liquid positions to a set of positions that possessed no liquidity at all, as analyzed in Chapter 5.

In none of these cases is the primary culprit inadequacy of value-of-risk measurements, which seems to be a popular scapegoat for financial losses in the press and among some academics (and even some opinion pieces written by members of the financial industry, who should know better). Why does value-at-risk (VaR) seem to get such a grossly disproportionate share of the blame:

* There is often a fundamental misunderstanding of the role of VaR. It is only designed to measure potential losses from large market movements on positions that are sufficiently liquid to be traded at market prices. It is not designed to measure potential losses on positions that are too large to be liquidated in a reasonably short time period or for which no liquid market exists. We will address the tools for managing illiquid positions shortly – VaR plays no part in this.
* It should be clear from the description I have given of the paradigm for managing liquid position risk that the most important indicator of potential loss would be the size of stop-loss limits. But stop-loss limits are not shared with the public, and VaR measures are. There is a tendency for journalists and bank analysts to focus on the data they possess, even if it is of limited relevance.
* Many of the criticisms of VaR are based on misinformation as to how it is calculated. One often sees references to the inadequacy of assuming normal distributions or Gaussian copulas, even though every major bank for the past 20 years has used simulation techniques for measuring VaR that makes **no** use of these assumptions. We’ll have more to say about industry standards for estimating probability distributions as we next address how to deal with less liquid positions.

There are a wide variety of factors that can cause trading positions to be illiquid. Fortunately, analysis is simplified by a single methodology that can be applied to all of them that is a reasonably well-accepted industry standard. This **liquid proxy** methodology consists of creating a basket of liquid positions that can serve as a good proxy for an illiquid position and then performing statistical analysis of the potential distribution of differences between the cash flows of the actual position and of this liquid proxy. (This methodology is detailed in Sections 6.1.2, 6.1.4, 8.2.6 and 8.4). The liquid proxy is used to represent the illiquid position in all trading reports of the firm, including daily mark-to-market, VaR, and stress tests, with the statistical analysis of differences used to create valuation reserves.

For some types of illiquidity, this statistical analysis is fairly easy to design. If the illiquidity stems from infrequent trading, when independent quotes as to where trading is taking place may only be available sporadically, it is a reasonably straightforward task to create a basket of liquid instruments whose price movement closely follows the price movement of the less liquid instrument. Statistical analysis of the distribution of changes in this relationship is then easy, since every time a market price is available for the less liquid instrument, a data point is created. If the illiquidity stems from taking a large position in what would otherwise be a liquid instrument, the instrument itself becomes the liquid proxy and an estimate is needed for how much the proceeds of liquidation will differ from the price of the instrument at the time that liquidation was decided on. A simple estimation technique that can use a lot of readily available data is to estimate the time that an “orderly” liquidation would take (i.e., a liquidation that does not impact market prices) and then do a VaR-type analysis on the distribution of price moves over this time period. Estimation of the time period for orderly liquidation starts with an estimate of the size of position that can be traded in any one day without an impact on the market and then dividing this into the size of position held.

For other types of illiquidity, the statistical analysis is harder to design. Positions may be taken in instruments that are rarely traded or have such unique design that no independent pricing data is available. In such cases, a liquid proxy still serves the useful purpose of making sure that any element of the position that can be managed using liquid markets is so managed. But it is challenging to create a model for the illiquid position that can identify the best liquid proxy and can be used as a starting point for analysis of the differences between cash flows of the liquid proxy and cash flows of the illiquid position.

An excellent principle for designing such models comes from Emmanuel Derman’s 2001 paper on “The Principles and Practice of Verifying Derivative Prices.” To quote Derman: “Therefore, for illiquid positions, it is important to estimate the adjustments to conventional marked values that can occur as a result of long-term hedging. One should build Monte Carlo models that simulate both underlyer behavior and a trader’s hedging strategy to create distributions of the resultant profit or loss of the whole portfolio. These distributions can be used to determine a realistic adjustment to the trading desk’s conventional marks that can be withheld until the trade is unwound and their realized profit or loss determined. … Monte Carlo analysis provides a good sense of the variation in portfolio value that will be exhibited over the life of the trade due to transactions costs, hedging error and model risk.”

The advantage of using Monte Carlo simulation based on long-term hedging strategies is that it is based on cash flow projections that go all the way through to expiration of the trade and is not forced to make any assumptions about prices at which a non-liquid position might be liquidated in the future. Simulation allows a full calculation of all factors that might influence cash flows, including complex waterfall conditions that can be part of highly structured products such as collateralized mortgage obligations or collateralized debt obligations.

The principles for designing distribution assumptions for input variables to any model used in financial risk management, including those used in VaR and stress testing analyses as well as those used in analyzing cash flow differences from liquid proxies are (a more detailed discussion can be found in Sections 7.1.1, 7.1.2 and 7.2.3):

* The shape of distributions (e.g., fat tails, skew, degree of clustering of large moves) is best estimated using the longest time series possible, since extreme moves may only occur rarely. It is therefore advisable to start with an empirical joint distribution of variables utilizing as much historical data as is available. Only time series that have reliable market quotes for common times of day such be used in this part of the analysis; other series can have their relationship to these core series estimated statistically.
* In contrast to higher moments of the distribution, mean and standard deviation should not necessarily be estimated based on a long historical period. For each variable, mean and standard deviation should be estimated using the statistical method though most suitable for that series, which may include different choices of historical period, weighted moving averages, or GARCH.
* The distribution shape determined by longer time periods can be combined with means and standard deviations derived by other methodologies by simply multiplying each of the daily observations by a constant factor to achieve the desired standard deviation and then adding a common factor to each observation to achieve the desired mean.
* Risk managers are concerned with the tails of probability distributions and so cannot afford to ignore subjective judgments that might indicate a shift in economic regime from past experience (see Section 1.2 for a more detailed justification); Dimson, Marsh and Staunton’s 2013 paper “The Low Return World” is an illustrative example of experienced analysts making a sustained argument in favor of future expectations that differ from historical experience. Chapter 10 of Ricardo Rebonato’s Plight of the Fortune Tellers contains an excellent discussion of the need for blending subjective judgment with historical probability distributions in estimating risk.
* When statistical analysis or subjective judgment indicates the need to assume different correlation assumptions from historical experience, the empirical distributions derived from history may need to be mixed with distributions manufactured to move correlations in a desired direction.

Let’s examine a good prototype of how Derman’s principle can work. Consider a portfolio of options positions for most of which reliable external market quotes cannot be determined, due to a wide range of tenors and strikes. Mark-to-market would usually be performed through interpolation from available market quotes, using the Black-Scholes model as the key tool in interpolation (see Section 11.1 for details). In effect, we are using positions for which market quotes are available as a liquid proxy and using Black-Scholes to model the relationship between the actual portfolio and this liquid proxy. How can we estimate the degree of risk that this entails? Fortunately, the Black-Scholes model makes it very easy to compute delta hedging positions in the underlying instrument for the difference between the actual portfolio and the liquid proxy and the size of these hedges is not very sensitive to inaccuracies in the implied volatility used in their computation. It is therefore quite easy to perform a Monte Carlo simulation of the impact of following a dynamic delta hedging strategy with rehedging performed not too frequently, to keep transaction costs within reasonable bounds (See Section 11.3 for details of this simulation). Many risk management departments have long experience with such simulations and they have reliably produced reasonable distributions of eventual profits and losses from hedging until option expiration.

Trying to use this example as a prototype for other illiquid instruments can be challenging. Determining the size of rehedges can often be far more computationally intense than was true in the Black-Scholes case, often prohibitively so – rehedges need to be computed at each time step on each of thousands of Monte Carlo simulation paths. Methods to get computationally reasonable Monte Carlo simulations have to be addressed on a case-by-case basis. Here are some examples:

* Interest rate swaps that are illiquid due to having a longer tenor than those for which liquid quotes are available can utilize the longest tenor liquid swap as a liquid proxy. Simple strategies can be worked out for later switching from one liquid proxy to another (for example, for a 5 year swap with an initial 3 year liquid proxy, at the end of two years, the then 1 year liquid proxy can be switched to a 3 year liquid proxy that exactly matches the real position). Computation of probability distributions of gains and losses from such a strategy are easy to perform (in this example, historical data on the differences in market yields of 3 year and 1 year swaps). See Section 10.2.2.
* There is a theoretical hedge based on Black-Scholes pricing theory that can be used to create a portfolio of liquid option positions that closely replicate a swap based on realized variance and somewhat less closely replicate a swap based on realized volatility. Probability distributions can be easily generated for the differences (Section 12.1.1).
* A call at one strike and a put at a different strike can be used to create a reasonable liquid proxy for a digital option and probability distributions easily generated for the differences (Section 12.1.4). Many exotics contain digital features (see for example Sections 12.1.5, 12.1.6, and 12.3.4), so it is particularly important to develop good methodology for handling digitals.
* Models of a fully realistic hedging strategy for barrier options are extremely, probably prohibitively, costly to simulate. But a reasonable approximation can be based on Peter Carr’s theory of put-call symmetry, which creates a position that only needs to be rebalanced once, if and when the barrier is breached. A probability distribution of differences can be generated based on historical experience with the shape of the volatility surface for standard puts and calls (Sections 12.3.2 and 12.3.3). Jim Gatheral’s The Volatility Surface and Emanuel Derman’s The Volatility Smile both give excellent discussions of the impact of differing modeling assumptions on the hedging of barriers and on static hedge portfolios that can be used as reasonable approximations (see particularly Gatheral Chapter 9 and Derman p. 32-44, Chapter 12, and p. 292-296).
* Options on baskets of assets can be hedged with positions in options on the individual assets and a probability distribution generated for the difference by simulation (Section 12.4.2).
* Bermudean swaptions can be hedged with a basket of European swaptions and the probability distribution of the difference generated by simulation (section 12.5.2).
* The tremendously illiquid CDOs that were at the heart of the 2008 banking crisis should have been analyzed using simulations that looked at ultimate default loss with the only mitigation being gains from defaults in short positions (Section 13.4.3). Had such an approach been taken, it would have become necessary to consider some probability of long-term scenarios in which there was a substantial decline over several years in housing prices (Section 5.2.5.7).

Note that in many of these cases we attempt to estimate probability distributions based on data which can frequently be updated – for example, a probability distribution based on the shape of the volatility surface, for which there is a new observation each day, is favored over a probability distribution of the multi-year evolution of prices.

Even when trading risk disasters occur, they are generally too idiosyncratic and firm-specific to lead to major bankruptcies or to require government rescue. It is bank lending that almost always is the culprit in triggering government rescues and raising moral hazard issues. The one large exception was the 2007-2008 CDO crisis, which was really a case of illiquid lending positions being misidentified as liquid trading positions. Since bank lending does lead to a significant need for the government to assure against extreme tail risks, the close government regulation needed of banks will also apply to the trading positions of those banks. The institutional consequences of both are addressed at the end of the next section, on Bank Lending.

**(2) Bank Lending**

The risk of large losses of a size that threatens a bank’s viability has been due to loan losses far more often to trading losses. The banking industry across many countries and over many years has been subject to periodic crises due to loan losses (a good history of these crises and their doleful impact on national economies is This Time is Different by Reinhart and Rogoff). Over most of this history, the tools for utilizing liquid instruments to manage this risk has been quite limited – some ability to sell loans. Risk management of bank lending was treated as one of the classic cases of statistical risk management. The advent of the Credit Default Swap (CDS) derivative instrument in the late 1990s began to provide a wider possibility of offsetting loan exposure with liquid instruments for certain borrowers. The CDS offered a form of insurance against default without requiring the insurance provider to come up with the large amount of cash that would be needed for an outright loan sale.

For those borrowers for whom there is a reasonably liquid CDS market, the financial risk tools of marking loan position to the price at which CDS insurance can be purchased, and of exiting positions that have come to be seen as too risky, are now available. But the CDS market is not large enough or diverse enough (in terms of number of borrowers covered) to manage the bulk of bank loan risk. Attempts to design instruments that could transfer the credit risk of an entire portfolio of bank loans with one transaction, the Collateralized Debt Obligation (CDO), fell far short of the task in the 2007-2008 crisis, as documented in Sections 5.1 and 5.2. Not all of the risk got transferred, with originating banks still liable on losses above a certain level, and the problem was exacerbated by the fundamental error of treating all of the loan portfolio as being manageable with liquid instruments when only part of it was. The completely inappropriate use of financial risk management tools such as mark-to-market for illiquid positions not only blinded senior managers to adequate assessment of the size of risks, but caused accounting treatment of loans that triggered the immediate need to raise capital against projected loan losses; in previous credit downturns, the proper accounting treatment of loan portfolios meant that capital could be raised over more reasonable time periods (see Section 5.3.2). When a truly liquid position suffers a mark-to-market loss that requires an increase in capital, there is a readily available remedy: sell some of the liquid position to reduce the need for capital and also reduce a source of further losses requiring capital. Illiquid positions don’t have this option, so the proper approach is to build up a capital buffer in advance, when raising capital is less expensive.

Ultimately, the size of potential losses on bank loans, and the importance of bank lending to a smoothly functioning economy, necessitates an institutional risk solution through government action, such as deposit insurance and rescue arrangements by central banks. The resulting moral hazard issues of “too big to fail” are well documented (see Section 5.2.5.9). Regulators have responded by tightened controls in the form of monitoring compensation policy, stricter capital requirements, and limitations on size and allowable activities (Section 5.5.5). Particularly significant, in my view, is the use of capital requirements based on stress-tests mandated by the regulatory authorities with specified levels of longer-term macroeconomic stresses, such as drops in GDP, home prices, stock markets indices, and a rise in unemployment, accompanied by regulator-mandated formulas for calculating the impact of these macroeconomic variables on loan losses. I do not think the risk managers of a firm have any specialized regarding the probability of macroeconomic events and they can be subject to internal political pressures within the firm to hold down the impact of stress tests on required capital, due to moral hazard considerations.

**(3) Retirement Investment**

Investing for retirement, whether by an individual or by a pension fund on behalf of a group of individuals, is a challenging long-range planning project -- a classic example of a difficult problem for statistical risk. Proper risk assessment requires forecasting decades into the future what goods will need to be purchased, projected costs of a wide variety of items, life expectancy, prices and returns of financial assets, etc.

Providing an absolute floor protecting retirement income requires an institutional risk solution, one that has been provided by businesses or governments through some form of inflation-adjusted defined benefit pension. But the difficulties of risk management of these pension liabilities has led many businesses and some municipal governments to switch to defined contribution pension plans, which eliminates the absolute floor protection, or has led to actual or threatened bankruptcy of pension plans. Ultimately, it is only a national government with its broad power of taxation that, through social security plans, Medicare-like guarantees of medical care, and pension guarantees, can provide the institutional structure to provide absolute floor protection.

But individuals will also seek to protect retirement income above the absolute floor protection that national governments are willing to provide. This requires a long-range investment strategy, which once again produces a difficult problem of statistical risk. National governments can provide an important tool that can make this problem more manageable through the investment alternative of long-term government bonds that provide government-guaranteed payments, combined with sensible policies of central bank currency management that assures these investments a reasonable inflation-adjusted return. Most developed nations have provided a long-run real (i.e., inflation-adjusted) return over long time periods of between 2% and 3% on their government-guaranteed bonds, though there have been some significant shorter periods where inflation reduced this return. More recently, many of these nations have provided investors with a new tool, government-guaranteed bonds that provide a return expressed as an assured spread above the actual inflation rates (in the US, these are called Treasury Inflation Protected Securities, TIPS for short). To make this work, the institutional structure must include a true commitment to resisting the temptation to meddle with payment formulas in eras of high inflation.

Up until 2005, these securities generally provided guaranteed real returns of between 2% and 3%, but these then began to decline precipitously in all countries offering this type of investment, with guaranteed returns dipping as much as 1% below zero and only subsequently recovering to a level of about 1%. “The low-return world” by Dimson, Marsh, and Staunton provide most of the data utilized in this analysis (see particularly their Figure 3 on the pattern of guaranteed real returns) along with well-reasoned arguments why these lower returns may persist for some time.

Several economists argue strongly that all but the most wealthy individuals saving for retirement should invest almost exclusively in TIPS-like inflation-guaranteed government bonds or in inflation-protected annuities (a good summary of their arguments is in Robert Merton’s “The Crisis in Retirement Planning” in the July-August 2014 issue of the Harvard Business Review and in Bodie & Taqqu, Risk Less and Prosper). But many investors are extremely reluctant to lock-in the current low levels of guaranteed return. Must such investors be totally reliant on statistical risk management? Here, financial risk management can come to the rescue with a strategy called contingent immunization, created by Liebowitz and Weinberger (see Section 6.1.7). With this strategy, investors may freely choose any mix of investments (analogously to how a trading desk is free to pursue any strategy it chooses subject to a stop-loss limit) but must constantly monitor the liquidation value of their portfolio relative to the amount needed to be invested in inflation-guaranteed government bonds or annuities to assure an acceptable level of retirement income. Calculation of this liquidation value requires all of the machinery of financial risk management: accurate, independent, and continuous marking-to-market, VaR and stress test computations to assess potential loss in liquidations, and valuation reserves against less liquid positions.

**(4) Insurance**

Insurance contracts are one of the foundational paradigms for statistical risk analysis – indeed, it is from insurance actuarial analysis that the alternate name of actuarial risk analysis is derived. Once a firm takes on insurance risk, it is difficult to liquidate (there is a limited market for reinsurance to share some of the risk with other insurance firms). So there is limited scope for financial risk management to impact the insurance market.

There is one trend in finance that could alter this state. Investment banks have created derivatives and securities that supply risk protection similar to that offered by insurance firms – caps and floors on interest rates, foreign exchange rates, stock indices, and energy prices that offer insurance against financial risks faced by business firms and investors; credit derivatives that offer insurance against defaults; catastrophe bonds that offer insurance against major damage from natural disasters such as hurricanes and floods; weather derivatives that offer insurance against temperature levels or precipitation levels that may negatively impact a business.

Some of these new instruments offer an insurance against events that were not within the traditional realm of insurance company products – caps and floors and credit derivatives – although in some cases insurance companies followed the introduction of these derivative products with competing insurance contracts. By contrast, catastrophe bonds were a direct attempt to introduce a product that could provide insurance companies with a means of liquidating existing risk. Weather derivatives, stand somewhere in between – they could be used to liquidate the risk on existing contracts offered by insurance firms or they could be used to offer a competitive product to insurance contracts.

In all cases, the treatment of insurance-like derivatives in financial risk should always depend on the nature of the risk protection and not the legal form of the instrument. For example, protection against the default of a debtor is exactly the same whether issued in the form of a credit default swap or an insurance contract with identical terms that pays off on default. Either should be treated as analyzable by financial risk methodology to the extent there is a liquid market to eliminate the risk and treated as analyzable only by statistical risk methodology to the extent there is no such liquid market.

As we have already noted in the section on trading desks, caps and floors are very often manageable by financial risk management techniques that employ options models to determine the size of dynamic hedges in the underlying instrument. But there are exceptions – if the underlying instrument is not sufficiently liquid to allow dynamic hedging or if the caps or floors are sufficiently complex to make it very difficult to create effective dynamic hedges, more reliance will need to be placed on statistical risk management.

Weather options provide an interesting study in when financial risk techniques can be utilized and when all reliance must be placed on statistical risk management. A good statement of the distinction comes from Jewson, Brix, and Ziehmann, Weather Derivatives Valuation. When weather swaps are traded for the location to which the option is tied, or on locations whose weather is closely correlated with the weather of this location, valuation should be based on market prices and arbitrage-pricing models (e.g., Black-Scholes). By contrast, “for locations where the [underlying] weather [swap] is not traded, and which are not highly correlated with locations on which swaps are traded, actuarial valuation of the options is the only choice, with actuarial valuation based on “historical meteorological data and meteorological forecasts to predict the distribution of possible outcomes.”

Since it is only large firms, insurance companies and banks, that can provide credible institutional protection against large tail risks in insurance, the worry is that they become “too big to fail,” ultimately relying upon government guarantees and government rescue and hence raising significant moral hazard issues. We already looked at government actions to deal with these issues for large banks at the end of the Bank Lending section. Insurance companies have also always been subject to close government regulation and there are actions being taken to intensify this regulation, such as applying to large insurance companies the government-mandated stress tests being applied to large banks.

**(5) Credit risk on derivatives**

The rapid growth of derivative markets over the past forty years has provided a much greater degree of flexibility to firms and individuals who wish to take on or modify their exposure to different economic factors (e.g., foreign exchange rates, interest rates, commodity prices, stock market indices). Positions can be taken without large initial investments, can be taken in either direction, and can usually be liquidated easily. But this flexibility comes at a cost, since derivatives positions are contracts between two counterparties and each then has credit risk exposure to the other counterparty’s failure to live up to its contractual obligations.

The use of statistical risk management for the credit risk of derivatives contracts is even more challenging than its use for bank loans, since you now need a statistical estimate of the size of the loss in the event of default as well as a statistical estimate of the probability of default, along with the correlation between them. It is not surprising that considerable effort has been devoted to applying financial risk management to this issue.

As with the application of liquid instruments to managing trading risk, it is helpful to begin with an idealized paradigm of how to completely manage credit risk on derivatives using liquid instruments and then to see how this paradigm needs to be modified in certain circumstances. Throughout our discussion, we will try to utilize the close parallel between this paradigm and the paradigm for managing trading risk to illuminate similarities and differences between the two. Many references throughout this section will be to Jon Gregory’s up-to-date and thorough The xVA Challenge.

In this paradigm, credit risk on a derivative contract is minimized by a contract provision that calls for continuous transfer of cash as collateral against any change in the value of the derivative. So that if one party to the derivative contract declares bankruptcy, the other party is already holding sufficient cash to be able to purchase an identical replacement contract without loss. This paradigm requires continuous (daily or more frequently) calculation of the replacement value of the contract (parallel to the continuous calculation of liquidation value for trading risk). It also requires VaR and stress test calculations to determine potential loss that may occur during the time period in which replacement takes place (in any case, it will require some time to find a replacement counterparty, but there also could be a period of time before declaring bankruptcy when the current counterparty has stooped meeting its margin calls; to minimize this, the paradigm allows a contract to be terminated in the event margin calls are not met, but even then there is some time delay in recognizing that a failure to post margin is not just an operational error – see Gregory Section 6.6.2). This potential loss, estimated by VaR and stress test methodology, is covered by an **initial margin** amount posted by the counterparty which is separate from and additive to any of the continuous transfer of cash, which is called **variation margin**. (Highly liquid securities, such as government bonds, are sometimes used instead of cash for margin, particularly for initial margin. In this case, more margin would need to be posted to cover potential loss on the securities in the event of liquidation. This topic is covered in Gregory Sections 6.1, 6.2.4, and 6.2.6).

Just as the trading risk paradigm allows great latitude to traders without needing to check continuously with senior managers, the counterparty credit risk paradigm allows firms or individuals who are willing to post margin to be free of close monitoring of their credit-worthiness – it is the collateral posted against the liquid position that provides assurance against loss.

Can this paradigm work in practice? For many liquid instruments it does, in the form of exchange-traded derivatives (see Section 14.2 of my book and Section 6.7 of Gregory for details). Derivatives traded on exchanges offer the added benefit that all contracts between counterparties are immediately replaced by two contracts between the exchange and each of the counterparties. This means that counterparties can focus on market risk and positioning and leave all considerations of credit risk to the exchange, which takes on no market risk but specializes in managing credit risk through rules for initial and variation margin. It also offers the advantage of very easy unwind of existing positions – since all trades are booked with the exchange as a counterparty, once an offsetting trade is made with any counterparty, the positions with the exchange cancel and no longer exist. And, since an exchange is the counterparty for a significant volume of any customer’s trades, the exchange is in a good position to see if a customer is building up an illiquid concentration in a particular trade and can then either limit this customer’s trading or impose higher initial margin requirements to cover the longer time period over which liquidation would take place (see Gregory Section 9.3.6).

But exchanges must be able to quickly offset positions they acquire when a counterparty fails to make a margin call. Since exchanges only acquire market positions through a failure to meet margin call, they will not be experienced in managing trading positions and must rely heavily on liquidity to offset positions. This requires that exchanges deal only in the most liquid contracts, restricting contract terms to only a few standardized dates and standardized terms to promote liquidity. In the area of risk mitigation through institutional design, exchanges manage extreme tail risk by mutually sharing residual losses across a wide base of member firms.

For less standardized but still liquid products, centralized counterparties can take on much of the structure and function of exchanges, and offer much of the advantages, including limiting all credit exposure to the centralized counterparty, easy unwinds, enhanced ability to detect illiquid concentrations, and mutual sharing of residual losses. The major difference with exchanges is that positions on which margin calls have not been met are closed out not by the exchange itself but by auctioning positions off to exchange members.

But derivatives which are not exchange-traded or centrally cleared (called **over-the-counter** or **OTC derivatives**) remain very popular (see Gregory’s Figure 3.1). Why? Some positions are too illiquid to be handled in any sort of standardized way. But, probably more importantly, many counterparties, particularly non-financial firms and individuals, prefer to avoid the operating costs and collateral expenses of posting margin. Investment banks serve as dealers in OTC derivatives both as a way of satisfying customer needs and as a major source of revenue. While OTC derivatives dealers may utilize some of the initial margin and variation margin techniques of exchanges and centralized counterparties to reduce credit exposure, they frequently reduce the use of margin to accommodate customers who wish to avoid posting margin. Conceptually, this is equivalent to lending the required margin to the customer and hence exposes the dealer to credit risk. The calculation and management of credit exposure on the portion not covered by margin is a major issue for dealers; see Section 14.3 of my book for an overview and Gregory Chapters 7, 10, 11, 12, and 17 for more detail. Calculation of credit risk on less liquid instruments should reflect the uncertainty of liquidating such positions, parallel to the discussion of illiquid instruments in the trading risk section of this paper.

Given the importance of the OTC derivatives product to investment bank dealers and the amount of effort they devote to calculating and managing the resulting credit risk, it has been most disappointing to see the major risk management failures that have occurred in this area. To focus on the major three failures:

* In the LTCM crisis of 1998 (Section 4.2.1 of my book), virtually every major investment bank was found to have required grossly inadequate initial margin from LTCM on its OTC trades. LTCM’s excellent reputation for its investment management skills, its track record of excellent trading results, and its lack of transparency in not disclosing to its dealers the extremely illiquid size of its positions all contributed to this lack of vigilance. It required a joint effort by 14 of the major investment banks to take over LTCM’s positions to liquidate them in an orderly way to avoid a truly major negative impact on global markets. This incident triggered a major industry study of credit practices relating to trading counterparties, the Counterparty Risk Management Policy Group report of 1999.
* In the period leading up to the 2007-2008 crisis, many investment banks purchased from insurance companies derivatives protection against credit losses on large holdings of super-senior tranches of CDOs (Section 5.2.5.1 of my book). These derivatives either had no margining provisions or extremely inadequate margining provisions. Given that the only time a payoff on the derivative would be due would be a time of major financial crisis, there was almost no chance that the insurance companies would be capable of making these payments, so it would only be through margining that they would provide any real protection. In 2008, the payments that were required caused these insurance firms to seek bankruptcy protection and the contracts either proved of little value or forced the government to come to the rescue (as in the case of derivative contracts of the AIG insurance company). This represented an extreme form of wrong-way risk, as discussed in Sections 5.4.5.1 and 14.3.4 of my book, and should have been recognized by senior managers, risk managers, and regulators as violating any reasonable norms of financial risk management. For at least some of the investment banks involved, this was probably just a cynical attempt to create an appearance of protection to mislead government regulators, an extreme example of moral hazard, as discussed in Section 5.2.5.9, 5.4.5.1, and 14.3.4 of my book.
* The 2007-2008 crisis also revealed serious flaws in the assumptions behind the management of credit risk on all OTC derivatives. The bankruptcy of a major firm, Lehman Brothers, and threatened bankruptcy of several other major firms, caused severe disruptions in the settlement process for OTC derivatives on which margin calls could not be met (Section 5.3 of my book). The response from regulators was a concerted effort to force greater use of exchange-traded and centrally-cleared derivatives, reducing the use of OTC derivatives, and to require a much higher level of capital protection to be held against OTC derivatives (Section 5.5.7 of my book). But as more trades get directed onto exchanges and central clearing houses, these become new potential sources of “too big to fail” institutions that may require government rescue, see Gregory 9.3.6 for further discussion.

The higher level of capital that regulators are requiring against OTC derivatives has resulted in higher prices being charged to customers who still wish to use OTC derivatives. The determination of these higher prices, and disputes over proper methodology and accounting for them, have been a major topic in the risk management literature over the past few years. The principal components of these price increases (known as valuation adjustments, or xVAs in the shorthand of Wall Street) are:

* Credit valuation adjustment (**CVA**) has been around for a long time, reflecting the expected loss component of credit exposure calculations. What has changed is that the more stringent regulatory requirements for keeping capital against potential losses above the expected level has led to more explicit pricing for this capital requirement. This is called **KVA** and is covered in Section 16.3 of Gregory.
* The lowered credit ratings of many investment banks in the wake of the 2007-2008 crisis has led to greater scrutiny of the funding costs of collateral posted as margin, offset by the funding benefit from collateral received as margin. The funding valuation adjustment (**FVA**) takes this into account in pricing derivatives. Some treatments distinguish between the funding cost of the variation margin, called FVA, and of the initial margin, called **MVA**. See Chapter 15 and Section 16.2 of Gregory.