A REVIEW OF OPERATIONAL RISK IN BANKS AND ITS ROLE IN THE FINANCIAL CRISIS

Erika de Jongh, Dawie de Jongh and Riaan de Jongh
Centre for Business Mathematics and Informatics (BMI), North-West University

Gary van Vuuren
Fitch Ratings, United Kingdom and North-West University

Accepted: April 2013

Abstract

The role of operational risk in the 2007/2008 financial crisis is explored. The factors that gave rise to the crisis are examined and it is found that although the event is largely regarded as a credit crisis, operational risk factors played a significant role in fuelling its duration and severity. It is concluded that, from an operational risk perspective, 2008 was the worst on record. Considering the extensive role of operational risk in global financial calamities, suggestions are made to improve the management of this risk type.

Key words: operational risk, banks, credit crisis

JEL: G46, G21, 32

1 Introduction

The objective of this paper is to analyse the role of operational risk in the 2007/2008 financial crisis and to provide recommendations regarding the improvement of operational risk management to assist in the prevention of future crises. Several articles have covered the 2007/2008 financial crisis. Most of these focussed on credit risk (Kregel, 2008; Hellwig, 2009; Lo, 2012), but work that focussed on the crisis from an operational risk perspective have only appeared recently (e.g. Hess, 2011; Andersen et al., 2011; Cagan, 2009; Kirkpatrick, 2009; Robertson, 2011; Rose, 2009).

Operational risk events stem from varied causes, including transaction and execution errors, fraud, improper business practices, product flaws, technology failures, employment discrimination, natural disasters (or ‘acts of god’) and terrorism (Cruz, 2002:14). Operational risk measurement and management, therefore, should embrace a wide band of sources which should detail internal corporate weaknesses, well-defined losses and clear classification of these amounts, details of recovery procedures and more accurate definitions of the event commencement and termination dates. Accounting databases do not need to be nearly as comprehensive and detailed as operational loss databases: the latter require greater quantities and better qualities of loss data. The credit crisis of 2007/8, although widely expected, precipitated a severe, protracted reduction in credit availability which continues to affect the global economy participants (as witnessed in the on-going sovereign crises in Europe). Although there are numerous descriptions and explanations of the origins of the credit crisis, it is now generally accepted that principal causes were negligent lending practices by banks, low, protracted global interest rates which in turn initiated residential and commercial property price bubbles, high oil prices, historically low world-wide inflation, a ‘light-touch’ financial regulation environment and inappropriate assumptions made for the assignment of financial derivative credit ratings (Jobst, 2010; Tomasic, 2012). Operational failures have contributed to every catastrophic loss since 1990, including the 2007/8 crisis. The American Insurance Group (AIG) event – an example of principal-agent risk – represents the largest corporate loss yet recorded (Canadian Institute of Actuaries,
2011). Operational losses can be caused by all levels of staff – including Boards of Directors – whether intentional or not. Although they are often caused by individuals: many instances of fraud are affected by colluding groups of people. Whatever the magnitude of the collusion, the largest operational risk losses have historically occurred at the most senior levels of corporate governance (Canadian Institute of Actuaries, 2011). The combination of these toxic components threw the global economy into turmoil, but they present opportunities to assess previously untested claims that operational risk increases in times of financial turbulence. Operational loss characteristics have been explored before the crisis and during the crisis (the term post-crisis implies that the event has reached its end, an event that is widely disputed) to establish whether these events have altered (in frequency, severity or both (see e.g. Esterhuysen, Van Vuuren & Styger, 2010; Hess, 2011).

The paper is structured as follows: an overview of operational risk is provided in Section 2, including a review of the definition of operational risk, as well as types, measurement and management of operational risk. Section 3 provides an overview of the financial crisis, including a timeline of events and the major contributory factors. The role of operational risk in the financial crisis is undertaken in Section 4 which includes a discussion on lessons learnt and challenges for operational risk management. Section 5 concludes with some recommendations made for improving operational risk management so as to reduce the effects of future crises.

2 Overview of operational risk

2.1 Definition of operational risk
The Basel II definition of operational risk is the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events (BCBS, 2006). This definition excludes strategic and reputational risk, but includes legal risk. Note that operational risk typically deals with losses only, unlike market risk which consider the upside (profit) as well.

2.2 Measuring operational risk
Risk is the uncertainty associated with the outcomes of events. An operational risk event is typically modelled by a loss density which then provides a model of all possible outcomes of this loss event. The bulk of operational risk loss data occurs in close proximity to the density centre – usually referred to as the body of the distribution which comprises the expected losses i.e. those losses having a high probability of occurrence but with medium, or low, impact. Losses occurring away from the centre to the right hand side of the density are typically referred to as unexpected losses i.e. those losses having a low probability of occurrence but with high impact. Risk measures based on these distributions are defined in terms of the Value at Risk (VaR) – a quantile selected in the right tail of the loss density. A universal risk measure in common global use is economic capital, defined as the difference between the VaR (the 99.9% quantile as specified for operational risk by the regulator) and the expected loss as shown in Figure 1. Expected losses are usually covered by financial institutions through capital provision and pricing; economic capital is the capital retained to guard against unexpected losses.

It has become customary to model the above-mentioned loss distribution by assuming separate models for the body and tail of the distribution. A range of choices are possible, however a popular choice for the body of a distribution is the Burr distribution and, as motivated by extreme value theory (EVT), the generalised Pareto distribution (GPD) is a popular choice for the tail section. Recently, Ahn, Kim and Ramaswami (2012) have studied the class of Log phase-type (LogPH) distributions as a parametric alternative in fitting heavy tailed data. Ahn et al., (2012) analytically derive its tail related quantities including the conditional tail moments and the mean excess function, and also discuss its tail thickness in the context of extreme value theory. They argue that the LogPH can offer a rich class of heavy-tailed loss distributions without separate modelling for the tail side, which is the case for the GPD.
Nonetheless, the accuracy of the VaR estimate is very dependent on the number of data points used as well on the particular model assumed. Several authors have researched this issue such as Neslehova, Embrechts and Chavez-Demoulin (2006), Cope, Mignolia, Antonini and Ugoccioni (2009) and Dahen, Dionne & Zajdenweber (2010). One of the findings was that the VaR estimate is very sensitive to the extreme losses observed, especially if data are sparse, which is frequently encountered when only internal data are used. This gave rise to research into so-called robust methods that are resistant to outliers (see Horbenko, Ruckdeschel & Bae, 2011). In order to obtain better estimates for VaR, internal data are often augmented by external data and by expert opinion. The issue of scaling is a very important issue when incorporating external data and exactly how expert opinion should be incorporated remains a pertinent research issue. Recently Dahen and Dionne (2010) proposed scaling methods that they applied to both frequency and severity loss data and using credibility theory, Agostini, Talamo and Vecchione (2011) proposed an integration model that allows integrated parameter estimation through the use of historical loss events and expert opinion. The parameter integration is obtained by considering a compounded average of historical data and subjective parameter estimates whose weights express the credibility assigned to each source and are provided by the Bühlmann – Straub model for advanced credibility theory.

The accurate modelling of the tail of the loss distributions is of paramount importance in calculating economic capital since economic capital estimates are extremely sensitive to small changes in tail estimates (see e.g. Cope et al., 2009). The estimation of economic capital will be discussed in more detail in Section 2.5.

2.3 Operational risk types

Most of the operational losses encountered in practice are frequent and relatively small, however, of real concern to regulators and risk officers are the less frequent/high-impact losses. Examples of operational risk events that occur frequently are equipment failures, losses due to ineffective management processes, employee errors, internal and external fraud, IT system disruptions and natural disasters. An example of an unpredictable, considerable-impact operational risk event is the terrorist attacks in the US in September 2001. Such low probability/high impact events are referred to as black swan events, i.e. rare events but ones whose impact on financial markets can lead to extremely high losses. These losses place considerable emphasis on the effective determination of economic capital by financial companies and are of paramount concern in operational risk and regulators in their attempt to stabilise the international financial system.

Types of operational risks are discussed in most textbooks (see e.g. Chernobai, Rachev & Fabozzi, 2007; Bessis, 2010). A short review
of the most common types follows in the section below.

**Internal fraud**

Losses due to acts intended to defraud, misappropriate property or circumvent regulations, the law or company policy, which involves at least one internal party. An example of financial fraud is fund embezzlement by bank financial officers.

This was the case in the Daiwa Bank scandal of 1995 in which Iguchi Toshihide – simultaneously holding the position of bond trader and head of government bond trading in New York – answered only to himself. Toshihide's responsibilities never allowed him to take more than a two or three day vacation and his long stay in his positions ensured that his expertise regarding the vagaries of the US government bond market were matched by no others. In 1984 he misjudged interest rate movements and made a relatively small loss (about US$150 000). Embarrassed, Toshihide concealed these losses and continued to do so thereafter until his (unreported) losses reached US$1.1bn. Daiwa's customer accounts were raided by Toshihide to conceal these losses: he sold customer bonds and forged documents to give the appearance of authorisation. Daiwa’s internal audits failed to identify the fraud. A 1989 inspection by the New York State banking authorities (accompanied by a Fed examiner), found nothing, and two further inspections, (in 1992 – by examiners of the New York Federal Reserve and in 1994 – by auditors from Japan’s Ministry of Finance (MOF)) also detected nothing. US examiners eventually ordered Daiwa to end Toshihide’s dual capacity as head of trading and head of settlements, leading to Toshihide’s confession to the President of Daiwa Bank in 1996. Aware that they had failed to properly supervise Toshihide, Daiwa Bank’s management dithered and withheld the information from the Fed. In November 1995 Daiwa Bank was indicted on charges of conspiring to conceal trading losses and fined $340 million, the largest criminal fine ever at the time (see Tschoegl, 1999).

**External fraud**

Losses due to acts intended to defraud, misappropriate property or circumvent the law, by a third party. An example is credit card theft and subsequent usage. External fraud may be committed in collusion with company staff and, therefore, in some cases, internal and external fraud may coexist. Most often, however, fraud involves actions carried out independently by third parties, external to the institution but fraud detection systems have been used to great effect in the mitigation of operational risk (see Bolancé, Ayuso & Guillén, 2012).

**Rogue trading and self-dealing**

A rogue trader is defined as an individual who acts recklessly and independently of fellow employees – usually to the detriment of both the clients and the trader's employer. Rogue traders typically trade in high risk investments which cause considerable losses (usually preceded by large, but unsustainable, profits). Many such traders' actions have resulted in large losses: these often accumulate because of protracted disguise. Due to poor internal surveillance in Barings Bank (the UK's oldest merchant bank), a loss of US$1 billion resulted from rogue trading activities by Nick Leeson in February 1995 (see Leeson, 1996). The financial services industry has largely ignored this potentially catastrophic form of operational risk and far too few controls are in place to manage it. In 2008, a lone trader, Jerome Kerviel, lost US$7.2 billion in unauthorised European Index Future trades at the French bank Société Générale. Just three years later (early 2011) the Swiss Bank UBS suffered a US$2.3 billion loss on fraudulent Delta 1 and exchange-traded funds (ETF) transactions due to the actions of another lone trader, Kweku Adoboli. Since 2002 repeated cases of rogue trading have befallen the largest global financial institutions. Trading surveillance lapses were exploited by the rogue traders at several of these institutions (GARP Risk Professional, 2012).

**External robbery and theft**

The Enron scandal is an example of theft (Chernobai et al., 2007:8). This was the largest bankruptcy case in US history (excluding the credit crisis), with a loss of US$600 million. Fictitious income was used to create fictitious capital in order to fund high-risk – and
ultimately unprofitable – deals. This is sometimes referred to as a Ponzi scheme (Berkowitz, 2012). In this way risk was concealed from bondholders and investors. Investigators blamed this failure on a combination of poor accounting failures and management information: the company’s assets were fraudulently overstated by US$24 billion.

Errors in legal documents
The Irish Allied Bank provides an example of errors in legal documents (Chernobai et al., 2007:8). A loss of about US$700 million was experienced when a trader falsified bank statements to recoup losses. Careless legal wording in financial protection products – payment protection insurance (PPI) – cost UK banks £264 million in payouts to customers in the first half of 2011. Over £5 billion has been set aside by UK banks to cover potential future PPI-related compensation payments, but Canadian, US, Italian and Hong Kong banks have also been charged with abusing their positions by selling unsuitable products – highly complex or highly risky – to unsophisticated investors such as local government bodies (Campbell, 2011).

IT disruptions
IT systems are used to increase efficiency, simplify labour and improve the handling and flow of data. These systems sometime fail and typically result in high losses which can have a considerable impact on the particular institution or even the financial system. An example of IT disruptions is the MasterCard computer virus which involved a computer virus capturing customer data for fraudulent activities (Chernobai, et al., 2007:8). This loss could also be classified as external fraud. In November 2010, an extensive computer disruption occurred which affected the Swedish bank Swedbank’s systems (including branch and card systems, ATMs and its internet banking system). After the disruption, the bank’s crisis groups and backup routines were activated, customers were indemnified and subsequently, Swedbank made a thorough review, identifying and implementing improvements (Swedbank, 2010). McPhail (2003) identified several potential operational risk problems in the Canadian banking system such as the failure of time-sensitive payment requirements and the disruption and dislocation in payment systems which could contribute to severe liquidity shortfalls in financial institutions. A framework was identified which provided a unified and systemic perspective on operational risk. The implementation of the framework – which assisted in the assessment of operational risk management in relevant critical systems – promoted financial stability in the Canadian banking system (McPhail, 2003).

Principal-agent risk
One of the most important operational risks – this is the risk that arises from agents who act on behalf of the organisation but who pursue actions not in the best interest of the stakeholders, but rather their own. Many of the large losses in the financial crisis were driven by principal-agent risk (Lang & Jagtiani, 2010). Principal-agent risk was the underlying cause of two of the drivers of the 2008 global credit crisis: the sub-prime crisis and AIG’s credit default swaps debacle. Under normal operating circumstances, laws and regulations monitored through legitimate, transparent metrics generally prevent the exploitation of principals by agents. Where information asymmetries and flawed performance metrics exist, however, this is not necessarily true. In the period preceding the credit crisis, large – but ultimately spurious – profits were generously rewarded, while legitimate – but moderate in comparison – returns were criticised and in some cases penalised. In such situations, some agents engaged in business activities that created the appearance of profitability (while value was actually being destroyed) and even well-meaning management structures began to disregard fiduciary responsibilities. Irresponsible behaviour at just one firm very quickly replicated itself, eventually resulting in industry-wide trends. This operational failure was a key driver of systemic risk (Canadian Institute of Actuaries, 2011).

External (black swan) events
Extensive losses were made when four commercial aircraft were hijacked and used to
crash into the World Trade Centre in New York and the Pentagon in Washington in September 2001. The destruction resulted in billions in insured property losses, the single largest insurance hit in history (see Banham, 2002). This event – which caused considerable global economic and political impact – provides a compelling example of physical assets afflicted by external causes.

2.4 Enterprise-wide Risk Management
The management of operational risk is closely connected to the principles of Enterprise-wide Risk Management (ERM) as outlined by e.g. the ISO 31000: 2009 Risk Management Standard (ISO 31000, 2009). ERM embraces the following important steps for operational risk management:

- define the strategic goals of the company and translate these into operational risk types that must be managed;
- analyse risks by identifying, describing, estimating and evaluating each one;
- assess the likelihood and impact of the occurrence of events;
- explore ways in which the event occurrence probability might be reduced and how the impact could be reduced (risk mitigating strategies);
- institute risk thresholds, tolerances and controls to ensure that operational risk events are managed, monitored and controlled; and
- ensure that management processes (such as reporting and model validation processes and procedures) are in place.

Any breaches, gaps or inefficiencies in the ERM process could lead to higher-than-anticipated operational losses.

2.5 Calculating economic capital
Closely associated with the management and measurement of operational risk is the provision of sufficient economic capital to guide against unforeseen losses due to operational risk events. The determination and management of economic operational risk capital plays an important part in the assessment of operational risk. The Basel II Accord provides guidelines for the calculation options of economic operational risk capital for banks which are the Standard Approach, The Basic Indicator Approach and the Advanced Measurement Approach (AMA) (BCBS, 2011a). Of these most large banks employ the AMA and specifically the Loss Distribution Approach (LDA) (BCBS, 2011b).

The LDA requires banks to organise their operational loss data in units of measure or operational risk categories (ORCs). These categories are determined by a specific business line (e.g. retail bank) and event type (e.g. internal fraud) combination. An important assumption is that the ORCs must be selected in such a way that all loss data observed in an ORC may be considered from independent sources. The loss data are then modelled in each ORC by a frequency distribution (typically Poisson) and a severity distribution (typically a combination of a Burr for the bulk of the data and a Generalised Pareto for the distribution's tail). Using the random sums procedure (McNeil, Frey & Embrechts, 2005) the frequency and severity distributions are used to determine an aggregate loss distribution as well as the 99.9% VaR. This value is then used to determine the economic capital for each ORC. Assuming total dependence between ORCs, the individual economic capital figures may be added to obtain an overall economic capital figure for the bank.

As stated previously the economic capital estimates are very sensitive to many of the assumptions underlying the LDA approach. Recently Embrechts and Hofert (2011) gave an overview of observed practice and supervisory issues in operational risk and Cope et al., (2009) empirically analysed the sensitivity of economic capital estimates to various assumptions underlying the LDA. From these the following modelling issues are highlighted as most sensitive:

- Modelling of the severity distributions in each ORC and especially the accurate modelling of the tail of the loss distribution. (This entails augmenting internal data with external data and expert opinion information as well as the analysis of outliers as discussed in Section 2.2.)
- Modelling the aggregate loss distribution in each ORC. (This entails the establishment of the compound distribution of frequencies and severities and the use of Panjer
recursion or Monte Carlo simulation techniques.)

- Diversification assumptions and the modelling of dependence between ORCs in order to obtain the overall loss distribution. The internal data collated by banks seldom cover periods of more than ten years and typically five year data sets are the norm (see Cope et al., 2009). The determination of an accurate 99.9% VaR (i.e. a 1 in 1 000 year event) using 5 to 10 year data sets is dubious. To circumvent this problem external data banks (in which several banks pool loss data – such as the ORX data) have been compiled. The ORX operational risk database currently consists of 249 781 losses amounting to 107 billion euros (ORX, 2012). In practice, internal data are then augmented with external data and scenarios to improve economic capital estimates.

Estimation of the operational risk capital under the loss distribution approach requires evaluation of aggregate or compound loss distributions. Closed-form solutions are not available for the distributions typically used in operational risk; however, with modern computer processing power these distributions can be calculated almost exactly using numerical methods. Shevchenko (2010) reviews numerical algorithms that can be successfully used to calculate the aggregate loss distributions. In particular, Monte Carlo, Panjer recursion and Fourier transformation methods are presented and compared. Cope (2012) has recently proposed an alternative method for integrating information from loss data with that obtained from scenarios analyses. The stochastic process that generates losses within an ORC is modelled as a superposition of various sub-processes that characterize individual ‘loss-generating mechanisms’ (LGMs). Cope (2012) then provides an end-to-end method for identifying LGMs, performing scenario analysis and combining the outcomes with relevant historical loss data to compute an aggregate loss distribution for the ORC.

The assumption of total dependence is considered to be a conservative assumption since no provision is made for possible diversification. Copula models have been introduced to allow for modelling the dependence structure between ORCs. Böcker and Klüppelberg (2008) undertook the simultaneous modelling of operational risks occurring in different event type/business line cells. They found that this analysis posed serious challenges for operational risk quantification and they invoked Lévy copulas to model operational loss events dependence structures. The consequences of this dependence concept for both operational risk frequencies and severities were analysed and the authors argued that instead of estimating precise frequency correlations between different cells, more effort should be directed at the more accurate modelling of loss severity distributions. Cope (2012) the position of various sub-losses within an ORC data with that obtained from scenarios. Gourier, Farkas and Abbate (2009) performed an empirical study of the shortcomings of the standard methodologies for quantifying operational losses. Extreme value theory was used to model heavy-tailed data – characteristic of operational risk losses. It was found that using Value-at-Risk as a risk measure led to misestimations of capital requirements. By introducing dependence between the business lines through copulas, the authors explored stability and coherence and related these to the degree of heavy-tailedness of the operational loss data. Inanoglu and Ulman (2009) used aggregated weekly operational loss data to avoid synchronicity problems with sparse data and applied non-parametric estimation to operational loss sample losses. The authors used the empirical distribution function to build a pseudo-sample matrix of probabilities which emulate drawings from identical marginals required to fit a standard copula. The empirical distribution function matrix was used to fit standard Gaussian, t, and Gumbel copulas to operational loss data. Annual losses in each event-loss type and by business line were calculated by simple summation. The simulation results found substantially lower diversification ratios from all copula models at the 99.9th percentile than those found in in other studies. The authors proposed using distributional copula approaches (with larger numbers of parameters than the Gumbel) and the development of a Bayesian strategy.

Quantitative techniques are not only used for the calculation of economic capital, but also for assessing and threshold calculation of key risk indicators and for monitoring and controlling these key risk indicators. Loss distributions capture outcome severity together...
with the probability of frequency and impact components. This is based on the assumption that enough data are available for the particular event type. However, data are frequently not available and qualitative assessments must be made. This is usually done by gleaning information from risk experts and ascertaining their view on the likelihood of future events and associated impact. Key risk indicators are often defined and the likelihood and impact assessed. These values are then multiplied to obtain a risk rating so as to rank risk indicators. This rating, however, should not be regarded as a risk measure since the product of likelihood and severity estimates expected losses, while risk management is concerned with unexpected losses. The assessment of likelihood and impact is better viewed in a matrix framework (Jobst 2010).

3 Overview of financial crises

3.1 The definition of a financial crisis
There exists a substantial literature on financial crises and market crashes. Notable among these are the books by Reinhart and Rogoff, (2009) and Bielecki, Brigo and Patras (2011). The term financial crisis broadly refers to a variety of situations in which the value of financial institutions or assets reduces abruptly. Investors sell off assets or withdraw money from financial institutions with the expectation that the value of those assets will decrease further if they remain at the financial institution. When available money is withdrawn, the financial institution is forced to sell other assets to make up any shortfall. This frequently results in a ripple effect through the economy and in liquidity shortages (see the extensive descriptions in Reinhart & Rogoff, 2009).

3.2 Possible causes of a financial crises
The causes of financial crises are diverse and include shocks to inflation, currency, banking, external sovereign debt, domestic sovereign debt, serial defaults and asset price bubbles (Reinhart & Rogoff, 2009). Inflation shocks – for example – cause decreases in the real value of money and uncertainty regarding future inflation discourages investment and savings. High inflation leads to shortages of goods if consumers begin hoarding fearing future price increases. If elevated inflation levels continue, consumer confidence and economic growth declines, resulting in recessions. The severity of the crisis is determined by the severity of the rise in inflation. Reinhart and Rogoff (2009) define a crisis due to inflation as exceeding a threshold of 40% per month.

Asset price bubbles arise through different circumstances. If mortgage interest rates rise, home buying is discouraged and house prices decrease. Home owners struggle with higher interest payments leading to more defaults and banks owning these mortgages simultaneously face more defaults, lower value of the collateral and more bad debt. Depending on the size of the mortgage book, bad debt can increase considerably. This aspect is discussed in detail in the next section.

3.3 Background to 2007/8 financial crisis
The crisis originated in the US during 2007 and peaked in September 2008 with the failure of Lehman Brothers (McLean & Nocera, 2010). This event resulted in a lack of confidence in the financial system and plunging capital markets. At this stage, the global financial system was on the verge of collapsing. Investment banks began to collapse, including the largest global insurance company, AIG. The financial system was locked into its first systemic crisis of modern times (Bessis, 2010:4). Failures extended to all players, insurance companies and funds. The crisis manifested itself as a systemic one, involving the collapse of the global financial system, brought about by lack of confidence amongst financial institutions and investors concerning their financial stability. The crisis of confidence caused a credit crisis, as investors withdrew their funds from the markets and credit institutions drastically decreased lending to limit losses, producing a shortage of capital and effectively halting economic growth. It is interesting to note that although Basel II regulations for banking credit risk were enforced from 2008, the US banks refrained from full compliance to these new rules (Bessis, 2010:4) at the time.
Prior to June 2007, US house prices increased steadily. This increase was mainly attributed to a flourishing sub-prime mortgages industry. Sub-prime loan issuers argued that, should house prices rise, collateral would be more valuable so the sub-prime loans transform into prime mortgages.

At the same time banks were grouping these loans into Mortgage Backed Securities (MBS), which were bought by a variety of investment banks who then converted the MBS into Collateralised Debt Obligations (CDOs). The CDO owner is entitled to a part of the pool’s interest income and principal. Securitisation of mortgages allows distribution of the credit risk of lending activities to investors best equipped to bear it. Insurance companies and banks in turn issued credit default swaps (CDS) which meant that following a default on a loan the devaluated loan would be taken back into the balance sheet of the issuer of the swap at full value. Banks and mortgage brokers eagerly supplied clients with credit, even clients with dubious creditworthiness. These loans were readily bought by investment banks and other investors for the purpose of securitisation which in turn bought CDSs to cover their risks. Credit risk was therefore distributed widely over the financial system because, prior to 2008, these markets (mortgage, sub-prime, CDO and CDS) were highly profitable and resulted in large bonuses for entrepreneurs (Andersen et al., 2011). In mid-2007, several financial players were concerned about the house price bubble. House prices stopped rising and interest rates on the sub-prime loans increased. Although some financial institutions expected some difficulties, it was not generally expected to trigger a system-wide crisis. In the second half of 2007 a surge in mortgage defaults showed up and accelerated in subsequent months. This led to the devaluation of mortgage backed securities such as CDOs. The collapse of the US housing market together with the subsequent devaluation of mortgage backed securities constituted a causal mechanism to the financial crisis. The volatility in the US mortgage market then spilled over into stock, commodity, and derivatives markets worldwide, causing a crisis of systemic proportions (see Hellwig, 2009).

### 4 Role of operational risk in the financial crisis

In their studies of the financial crisis, Andersen et al., (2011:2) and Cagan (2009) ask some pertinent questions from an operational risk viewpoint, namely:

- Why were loans granted to individuals with limited ability to service these loans without proper documentation of income, wealth or employment status?
- Why have investment banks readily bought such loans for securitisation and further distribution?
- Why did the constructed securities receive investment grade ratings even when significant portions of under documented sub-prime loans were included in the underlying asset?
- How could insurance companies issue billions’ worth of credit default swaps without setting aside capital to cover potential claims?

In an attempt to answer these questions, Andersen et al., (2011) concluded that failure to manage operational risk in banks and mortgage brokers resulted in poorly documented loans contributing to erroneous or lacking assessment of borrowers’ credit-worthiness. This operational risk exposure was transferred into credit risk for the CDO owners. Some possible answers are considered below.

#### 4.1 Why were loans granted to individuals with limited ability to service these loans?

Access to loans by individuals with limited ability to service these loans has been shown to increase personal bankruptcy rates. For first-time applicants near the 20th percentile of the credit-score distribution, access to payday loans causes a doubling of bankruptcy filings over the next two years. The effects are statistically and economically larger in locations where the credit provider has fewer competitors (see Skiba & Tobacman, 2011).

Despite this research, banks were unconcerned because the risk had been passed on to investment banks through the sale of mortgage backed securities.
4.2 Why have investment banks bought such loans for securitisation and further distribution?

Investment banks both generated and invested heavily in CDOs. Citibank warehoused mortgages for future securitisation (Kregel, 2008), an element that added to the losses as the housing and CDO markets collapsed. The risk models of firms such as Citibank did not include scenarios in which real-estate values decreased sharply, which suggested that the risk of almost any mortgage was limited (Kolb, 2011). Investment banks who failed to set up appropriate risk management measures also faced challenges from the rapid development and increasing complexity of these products. Extraordinary profits generated by the market for securitised assets clouded the judgment of management and staff as salaries and bonuses skyrocketed in the years before the crisis. The fact that investment banks were confident buying under documented loans without requiring additional information from the loan originator, indicates that a risk management focus came second to profit generation. Whether or not a transaction was considered sound was less an issue for risk management and more of an issue to whom the transaction was presented within the organisation (Kolb, 2011).

Investment banks were highly leveraged as the opportunity to increase lending compared to equity provided by deregulation was fully exploited in an attempt to realise the full potential of the CDO market. The aggregated effect of the operational risk elements put the investment banks in a position where they could only withstand minor increases in default rates before the losses became critical. In fact, fully exploiting the 40 to 1 asset to equity ratio in practice meant that a reduction in asset values of less than 3% would result in the firm being eliminated, a case in point being the downfall of Lehman Brothers.

The investment banks’ failure to manage operational risk was transformed into shareholder risk as the investment banks were only capitalised to handle marginal losses. Moreover, the failure of investment banks to require thorough risk assessments and documentation from loan originators resulted in operational risk being transferred to credit risk for the CDO owners.

4.3 Why did the constructed securities receive good investment grade ratings?

Credit rating agencies assigned the same rating to derivatives compiled partly of sub-prime loans as those containing principally prime loans. These ratings became even more of a problem as sub-prime loans were usually under-documented making it nearly impossible to make any informed assessment of future default rates, and hence the riskiness of the securitised products. This led to a mis-representation of risk affecting the behaviour and decisions of financial institutions.

The post-crisis investigations into the practices of the credit rating agencies uncovered alarming results concerning how these institutions operated in the period of extreme growth in credit securitisation prior to the crisis (Andersen et al., 2011). A review carried out by the United States Senate Permanent Subcommittee on Investigations (US Senate, 2011) revealed a number of flaws and inconsistencies in the way ratings were generated for CDOs. The Senate revealed that the departments carrying out the CDO ratings were severely understaffed, and that the management system concerning how to conduct CDO ratings was lacking. For instance, it was discovered that none of the examined rating agencies had a documented procedure describing how to carry out CDO ratings. There was also a lack of written procedures for surveillance of accuracy of the ratings provided. This led to the staff at rating agencies being overworked and lacking proper guidance.

It should be noted that the high market demand for securitised assets led to increasingly complex CDOs, with increased fractions of sub-prime loans. Given the number of mortgages referenced in a single CDO, deriving a generalised model for assessing the credit risk of a CDO is difficult. A major problem with the models identified both by Rajan (2008) and Kregel (2008) was the reliance on historical default correlations between groups of borrowers as a predictor of future default rates. Subprime mortgages were at the turn of the century a fairly new invention, and had never previously been originated at the same rate and extent as during
2002 to 2006. Thus, the performance history of such loans was limited. It is not likely that the available history concerning default rates provided a remotely reliable predictor for how sub-prime loans would perform in the future. Despite apparent shortcomings in the models and severe organisational problems, the policy was that every deal should be rated, a policy that generated considerable income for the rating agencies. All of the identified issues within the credit rating businesses fall under the category of operational risk. The problems observed in the credit rating agencies gave rise to an undervaluing of risk through ratings that did not reflect the risk of the underlying assets (i.e. sub-prime loans). This overoptimistic assessment of risk, resulting from failed management of operational risk, was transferred into credit risk for the CDO holders.

4.4 How could insurance companies issue billions' worth of credit default swaps?

Several insurance companies and particularly a subsidiary of American International Group (AIG), issued so-called Credit Default Swaps (a form of debt insurance) for securitised assets. AIG alone was exposed to about US$500 billion worth of assets through the insurance of securitised loans. In 2007 the CEO of AIG Financial Products said: ‘It is hard for us, without being flippant, to even see a scenario within any kind of realm of reason that would see us losing one dollar in any of those transactions’ (Morgenson, 2008). He was referring to the CDS derivatives that would later inflict losses so great that only a government bailout could prevent AIG from going bankrupt. The belief in low future claims made the CDSs seem highly profitable, and for a while they were. In 2005 profit margins on CDS sales were as high as 83%. On average, CDS sales generated salaries and bonuses of more than US$1 million for each employee in AIG Financial Products. Because AIG Financial Products was not classified as an insurance company it was not subjected to requirements to report its activities to insurance regulators, and was allowed to conduct its business almost without oversight (Morgenson, 2008).

4.5 How is it possible that the crisis was not forecast?

Failures to properly assess the risk of the assets insured and failure to properly assess the need for collateral constitute the major operational failures concerning the practices for issuing CDSs. Based on available knowledge it seems that the insurance company AIG, represented by its subsidiary AIG Financial Products, did not carry out independent assessments of future default rates, and placed full confidence in the ratings provided by the credit rating agencies. The sentiment that default rates would remain low was reinforced by a strong belief that real estate values would continue to increase without significant variations in value (US Government, 2011).

The willingness of insurance companies to insure the debt contained in the CDOs contributed to escalating the market for these products by strengthening the illusion that CDOs represented a comparably low risk investment. Hence failure to manage operational risk on the part of the insurance companies was transferred into significant risk for the shareholders and, as it turned out in the case of AIG, for American taxpayers.

4.6 Concluding remarks

In the wake of the crisis much focus has been directed towards the remuneration practices within the financial industry, and several countries are currently implementing regulations restricting the bonus potential of employees within the financial industry. There is no denying that the potential for substantial bonus payments affected the actions and behaviour of central actors in the financial organisations. However, it is also possible to trace the frailty of the financial system to the failure to ensure quality throughout the supply chain for securitised assets. For instance, the rating agencies are not required to verify the information in the loan portfolio that was to be subjected to securitisation. There is also no requirement stating that the issuers of loans should perform due diligence. The fact that no one reacted to the extensive lack of documentation of, particularly, sub-prime loans is baffling to say the least.
Furthermore, investment banks spent vast amounts getting CDOs rated, but seemed to lack interest in whether the credit rating agencies possessed the necessary systems, tools and competence to provide reliable results. The disregard for supply chain management observed in the financial industry is uncommon in other industries, and regulators could possibly benefit from stricter regulations concerning the responsibilities to ensure that subcontractors and business partners run sound and sustainable operations (Andersen et al., 2011).

The 2008 financial crisis can be described as the worst crisis ever from an operational risk viewpoint. This is demonstrated by Cagan (2009) who shows, using Algorithmics’ FIRST database of risk case studies statistics, that 2008 was the most severe year in terms of the size and impact of the loss for all the events that involve financial institutions. The amount of operational risk losses observed in 2008 is almost four times greater than those observed in 2007. Hess (2011) analyses operational risk in the context of the 2008 financial crisis. The largest global repository of information on publicly reported operational losses, SAS OpRisk Global Data (SAS, 2007) was chosen as the underlying dataset. A significant impact on the riskiness of the loss severity was found for the trading and sales and retail brokerage business lines (BL) due to the financial crisis. Losses from investment banks caused by the market failure of auction rate securities are responsible for this result. A 150% higher VaR for the BL trading and sales and a 50% higher VaR for the BL retail brokerage was calculated using financial crisis data.

5 Improving operational risk management

It is important to manage operational risk effectively. Operational risk can lead to a financial crisis or, as it did in 2008, worsen a financial crisis through the supply chain. It may be deduced from Section 4 that severe failures to manage operational risk were present in all parts of the supply chain involved in generating and distributing the securitised assets known as CDOs. An absence of supply chain management, greed, lack of competency, and a naive belief that past history is the best predictor for the future are all ingredients that resulted in a financial crisis not seen since the early 1930s (US Senate, 2011).

The crisis exposed a financial system with a special ability to socialise losses while privatising profits. These circumstances made it clear to the global political community that changes needed to be made to the financial system. It should be noted that there has always been an exposure of operational risk to financial institutions, ‘however, there is strong reason to believe that the exposure to operational risks in the future will increase. The reason is that systems, financial products and IT solutions tend to become increasingly complex and interconnected, especially if financial institutions decide to outsource vital parts of their services’ (Rose, 2009:30).

Some recommendations on how operational risk management may be improved are provided below, given the lessons learnt in Section 4.

5.1 Principal-agent risk

One of the most significant risks of a major corporation is principal-agent risk. Managers representing the corporation should ensure that compensation structures of agents should be well-structured in the interest of the corporation so that agents who take on too much risk on behalf of the organisation are penalised (Andersen et al., 2011). In other words the self-centred agent mentality of taking profit but shifting losses to somebody else should change in order to mitigate this risk. This means personnel engaged in value-destroying activities should lose, but they must lose in proportion to the amount of damage caused. Those in violation of fiduciary responsibilities knowingly must experience negative consequences (Kirkpatrick, 2009).

5.2 Risk management practices

The way in which risk is managed and measured is an important issue. Organisations struggle to incorporate the impact of rare events accurately so they typically underestimate their level of risk. When a company’s risk-adjusted performance measures are based
on flawed models, some ‘risk-reward arbitrage’ opportunities will exist and some decision-makers could take excessive risks to maximise their personal rewards (Andersen et al., 2011). To prevent a repeat of the 2008 financial crisis, a shift in risk management practices is required. Risk models need to be validated by independent and objective experts whilst it can provide valuable insight into complex problems. It must also incorporate expert opinion and empirical data in a transparent, credible and theoretically valid manner. In the light of the above, regulators should develop enhanced guidance to strengthen institutions’ risk management practices, and encourage financial firms to re-examine their internal controls and implement strengthened policies for sound risk management (Kirkpatrick, 2009; Jobst, 2010).

5.3 Risk mitigating strategies
Andersen et al. (2011:15) suggest that mitigating undesired events in general can be pursued along two complementary paths. Mitigating strategies can be designed to reduce the probability of an event occurring, or reduce the magnitude of associated consequences, or both. If faced with a choice between probability reducing and consequence reducing measures, it stands to reason that working towards avoiding an undesired event altogether is preferable to being good at handling the consequences. For example, it is better to prevent a fire from starting rather than mitigating the consequences of the fire after the conflagration. It could be a good idea to enforce regulations that requires companies to give priority to probability-reducing measures above consequence-reducing measures.

5.4 Incentive and performance management
Remuneration practices were at an early stage identified as one of the prime suspects causing the observed reckless behaviour by actors within the financial industry. Considering the size of bonuses that were paid in the years leading up to the crisis (Crotty, 2009) and a bonus regime providing a seemingly infinite upside and a downside limited to zero (i.e. no bonus), the observed behaviour can possibly be considered ‘rational’. In 2009, the Financial Stability Forum (FSF) issued nine principles for sound compensation practices (FSF, 2009) where it was emphasised that governing bodies of financial firms have to acknowledge the effect of remuneration practices on risk taking. Among the principles suggested are risk adjusted bonuses and compensation schedules sensitive to the time horizon of the risk to which the employee has subjected the firm. Other initiatives to reform the remuneration practices of financial firms include guidelines issued by the Committee of European Banking Supervision (CEBS, 2010) which have been implemented in the legislation of several European countries. Curbing excessive risk taking and avoiding a focus on short term profits are central motivators for the bonus regulations introduced. When senior management and the board of directors place their own interest above the interest of shareholders, they must be held more accountable. The improvement of the understanding of risk and risk management (especially to board members and senior executives) must improve considerably. The ability of management to manage risk must keep pace with all the other business innovations.

5.5 Complexity of financial products and skill levels of risk managers
Since early 2003, non-transparent and complicated financial products were developed. The widespread use of these products was one of the most significant reasons for the occurrence of the 2008 financial crisis in that risk assessment procedures failed. It is important that risk managers possess the necessary skills to assess the inherent risks in these complex instruments (US Senate, 2011).

5.6 Rogue trading
Operational risk events (particularly trading events) are often driven by market volatility. Cagan (2009) warns that ‘when volatility rises, there should be no tolerance for traders who breach their limits.’ Any area in a financial institution that can result in large unauthorised trades or fraud must be supervised and volatility should translate into oversight and attention to controls. Controls on traders and the supervisory functions such as market risk
5.7 Capital level
Andersen et al. (2011) suggest that reform of the general level of capitalisation of organisations within the financial system and within banks in particular is needed. During the years leading up to the financial crisis investment banks systematically moved assets off their balance sheets in order to reduce capital requirements, and in so doing, the regulatory demands for capital (as stipulated under the Basel II Accord) also decreased. Several studies emphasised the role of thinly capitalised firms as a major cause to the systemic weakness revealed by the crisis (see e.g. Hellwig, 2009; Bielecki et al., 2011). As a result the Basel III regulations were finalised in record time and unveiled in 2010.

The revised Basel Accord has been updated specifically to strengthen the level of capitalisation within the financial industry (BCBS, 2011c). Key changes included in the Basel III Accord is an increase in lower limit of asset to capital ratio and market sensitive capital requirements providing increased capital levels in ‘boom’ times to dampen the effects of (often) subsequent ‘busts’.

5.8 Capital adequacy for operational risk
The BCBS have made considerable improvements to the Basel II accord's recommendations to operational risk capital calculations (BCBS, 2011b). The new rules argue in favour of separating the body and tail distributions, but caution than banks should consider the choice of the body-tail modelling threshold that distinguishes the two regions carefully. Because the threshold can have profound implications for capital requirements, banks seeking approval for the advanced measurement approach for operational risk are required to document statistical support, and provide supplemental information of all qualitative elements, for the selected threshold. The estimate of the body-tail modelling threshold should be made conjunctly with the parameters of the distribution. The BCBSA identifies the Hill plot and the Mean Excess Function plot as useful identifiers of the threshold.

Esterhuysen et al., (2010) and Hess (2011) analysed operational risk in the context of the 2008 crisis and the loss distribution approach. Hess found that the shape parameters of the GPD model for the business lines trading and sales and retail brokerage increased significantly for different thresholds due to losses caused by the financial crisis. The market failure of auction rate securities (ARSs), and the corresponding large losses from some internationally operating investment banks that marketed and distributed these securities, explains 85% of the shape parameter rise for the business lines trading and sales and the complete parameter rise for the BL retail brokerage.

Esterhuysen et al., (2010) researched low-frequency, high-severity operational risk events (events that occur in the upper tail of loss distributions) since these are of particular interest to operational risk managers. Peak over threshold (POT) models focus on loss events above high thresholds and then fit distributions to data above these thresholds (see e.g. McNeil et al., 2005). For a sufficiently large threshold, the conditional excess distribution of such extreme observations converges to the generalised Pareto distribution (GPD). The cumulative distribution function of GPD is:

$$F(x) = 1 - \left( 1 + \frac{x - \xi}{\beta} \right)^{-\frac{1}{\xi}}$$

where $\beta > 0$ the scale parameter; $\xi$ the distribution shape parameter and $\mu$ the location parameter. Note that $x \approx 0$ when $\xi \approx 0$ and $0 < x < -\beta/\xi$ if $\xi < 0$. When $\xi = 0$ the GPD becomes the light-tailed exponential distribution, when $\xi < 0$ a short tail Pareto Type II distribution is obtained and when $\xi > 0$ heavy tailed distributions are obtained. Of course, the larger $\xi$, the heavier the tails of the GPD. The latter distribution is then fitted to the excess losses over some threshold.

It is important to obtain accurate estimates of the shape parameter $\xi$ and the Hill estimator (Hill, 1975) has proved to be a
reasonable estimator (Cruz, 2002:221; Perry & de Fontnouvelle, 2005:332). In Figure 2 below, using the data of Esterhuysen et al., (2010), a Hill plot is constructed which show the estimated shape parameters as a function of the order statistic. From the graph it is clear that the pre-crisis estimated shape parameters stabilise around 1 while the estimated shape parameters during the crisis stabilises around 1.5. Both these values are indicative of a heavy-tailed distribution, but the ‘during crisis’ estimate is 50% more than the ‘pre-crisis’ shape parameter.

Figure 2
Comparison of estimated shape parameters, using Hill’s method, for the period January 2003 to June 2007 (pre-crisis) and July 2007 to July 2009 (during the crisis) against the order statistic number

Conducting a similar study, Hess (2011) computed a 157% higher VaR for trading and sales BLs and a 52% higher VaR for the retail brokerage BL due to the financial crisis. Since the ARS market failure is mainly responsible for these results, he suggests that financial institutions intensify their risk management regarding the handling of market failures. This can be achieved by scenario analyses that simulate the consequences of a collapse of the markets in which the institution operates. Afterward, it is possible to decide whether the liquidity and capital situation is sufficient to bear the risks that arise from financial intermediation.

5.9 Reserve bank stability fund
To reduce the effect of crises in the financial system, the introduction of a government financial stability fund (or crisis fund), financed through a financial stability tax (or bank tax) has been suggested (Andersen et al., 2011:9). Following the crisis, several governments argued that financial firms should finance their own bailouts in future, in advance. For example, in 2009, the Swedish government introduced a permanent stability fee for banks and other credit institutions, and Germany, France and the UK are planning to, or have already followed suit. While it has been argued that government bailouts could result in moral hazard, the requirement that the industry fund future bailouts itself may create incentives for improved risk management and governance. Conditioning access to the stability fund on compliance with sound risk management standards as well as robust capitalisation could create incentives towards a sustainable and stable financial system.

6 Conclusion
After the 2007/8 financial crisis the general public is more aware of the complexity of the global environment. There is a greater need to
focus on the risks that matter, rather than on the not so important ones that consume considerable resources and energy. In the present (May 2012) global economy businesses depend in some capacity on agent performance. Effective risk management protocols must be in place if day-to-day control over agents is not feasible. It is important that management focuses on the management of operational risks, particularly when financial engineering is prevalent and complex. If this task is neglected, fatal consequences could await financial institutions. Management of financial institutions should thoroughly understand all the different risks associated with their products. Risk cannot be outsourced or ignored as it returns in other more odious and influential forms (outsourcing credit risk through the mortgage securitisation process increased liquidity and operational risk substantially).

In summary, the following guidelines are proposed for improving operational risk management:

• Institutions should ensure that compensation structures of agents are well-structured in the interest of the corporation.
• A shift in risk management practices is required, entailing amongst others, the redesign of compensation and performance management policies and procedures to incorporate risk-adjusted performance measurement of executive and non-executive directors as well as top and middle management.
• The ability of management to manage risk must keep pace with other business innovations. This would include setting up risk training programmes for the assessment of risk inherent in new complex financial instruments.
• Institutions should ensure that managers possess the necessary skills to identify, assess, measure and mitigate inherent risks in complex instruments.
• Supervisory functions should exercise more stringent control over traders.
• Financial institutions should improve their risk management procedures and governance structures in order to guard against market failures.
• Financial institutions should identify vulnerabilities and pro-actively map risk scenarios so that corrective policies can be effectively implemented. Although operational risk management enjoys some success in the identification of underlying vulnerabilities, i.e., the predisposition to shocks, it has a shoddy record of large loss timing predictions.
• Financial institutions should identify and link multiple vulnerabilities to determine the severity of potential threats.
• Operational risk management processes should be honed to warn of imminent tail event risks.
• Operational risk management needs to assist in the prioritisation of policy recommendations and the formulation of contingency plans based on both impact and event probabilities.
• The introduction of a government financial stability fund, financed thorough a financial stability tax.
• The ‘best practice’ Advanced Measurement Approach (AMA) procedures for economic capital allocation should be revisited and improved. Currently the assumptions and guidelines underlying these approaches result in widely varying capital estimates. Although significant improvement has been made (BCBS, 2011 a&b), more research is needed.

Endnotes:

1 The credit quality of particular borrowers, who have weaker credit histories and a greater risk of loan default than prime borrowers.
2 A CDO is created by bundling a pool of similar loans, e.g. mortgages into a single investment (securitisation)

Acknowledgements

We would like to thank the referees for insightful remarks that led to the substantial improvement of this manuscript. This work is based on the research supported in part by the National Research Foundation (NRF) of South Africa reference number (UID: TP1207243988 ). The authors acknowledge that opinions, findings
and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the author(s), and that the NRF accepts no liability whatsoever in this regard.

References


ORX. 2012. Available at: http://www.orx.org/ [accessed 2012-09-17].


