The Emperor has no Clothes:
Limits to Risk Modelling

Jón Daníelsson[*]
Financial Markets Group
London School of Economics
www.RiskResearch.org

September 2001

Abstract

This paper considers the properties of risk measures, primarily Value–at–Risk (VaR), from both internal and external (regulatory) points of view. It is argued that since market data is endogenous to market behavior, statistical analysis made in times of stability does not provide much guidance in times of crisis. In an extensive survey across data classes and risk models, the empirical properties of current risk forecasting models are found to be lacking in robustness while being excessively volatile. For regulatory use, the VaR measure may give misleading information about risk, and in some cases may actually increase both idiosyncratic and systemic risk.

1 Introduction

Recent years have witnessed an enormous growth in financial risk modelling both for regulatory and internal risk management purposes. There are many reasons for this; stronger perception of the importance of risk management, deregulation enabling more risk taking, and technical advances encouraging both risk taking and facilitating the estimation and forecasting of risk. The impact has been felt in regulatory design: market risk regulations are now model based, and Basel–II proposals recommend basing credit, liquidity, and operational risk on modelling. The motivations for market risk modelling are obvious. Data is widely available, a large number of high quality models exist, rapid advances in computer technology enable the estimation of the most complicated models, and the ever–increasing supply of human capital, all led to a sense of "can do" within the technical modelling community. Empirical modelling has been of enormous use in applications such as derivatives pricing and risk management, and is being applied successfully to credit risk.

But, has risk management delivered? We don't know for sure, and probably never will. If regulatory risk management fulfills its objectives, systemic failures should not be observed, but only the absence of crisis can prove the system works. There is however an increasing body of evidence that inherent limitations in risk modelling technology, coupled with imperfect regulatory design, acts more like a placebo rather than a scientifically proven preventer of crashes it is sometimes made out to be. Below I survey some of this evidence; the general inaccuracy and limitations of current risk models, the impact of externally imposed uniform risk constraints on firm behavior, the relevance of statistical risk measures, and the feedback between market data, risk models, and the beliefs of market participants. I finally relate this evidence to the current debate on regulatory design where I argue against the notion of model based regulations, be it for market, credit, liquidity, or operational risk.

An explicit assumption in most risk models is that market data follows a stochastic process which only depends on past observations of itself and other market variables. While this assumption is made to facilitate modelling, it relies on the hypothesis that there are so many market participants, and they are so different that in the aggregate their actions are essentially random and can not influence the market. This implies that the role of the risk forecaster is akin to the job of a meteorologist, who can forecast the weather, but not influence it. This approach to modelling has a number of shortcomings. If risk measurements influence people's behavior, it is inappropriate to assume market prices follow an independent stochastic process. This becomes especially relevant in times of crisis when market participants hedge related risks leading to the execution of similar trading strategies. The basic statistical properties of market data are not the same in crisis as they are during stable periods; therefore, most risk models provide very little guidance during crisis periods. In other words, risk properties

of market data change with observation. If, in addition, identical external regulatory risk constraints are imposed, regulatory demands may perversely lead to the amplification of the crisis by reducing liquidity. There is some evidence that this happened during the 1998 crisis. Indeed, the past 4 years have not only been the most volatile in the second half of the 20th century but also the era of intensive risk modelling. The theoretical models of Daníelsson and Zigrand (2001) and Daníelsson et al. (2001c) demonstrate exactly this phenomena.

In order to forecast risk, it is necessary to assume a model which in turn is estimated with market data. This requires a number of assumptions regarding both model design and the statistical properties of the data. It is not possible to create a perfect risk model, and a risk forecaster needs to weigh the pros and cons of various models and data choices to create what inevitably can only be an imperfect model. I present results from an extensive survey of forecasting properties of market risk models, employing a representative cross–section of data and models across various estimation horizons and risk levels. The results are less than encouraging. All the models have serious problems with lack of robustness and high risk volatility, implying that in many cases model outcomes will be as accurate as if a roulette wheel was used to forecast risk.

Current market risk regulations are based on the 99% Value–at–Risk (VaR) measure obtained from a risk model. The VaR number can, under some assumptions discussed below, provide an adequate representation of risk, however such assumptions are often unrealistic, and result in the misrepresentation of risk. There are (at least) four problems with the regulatory VaR measure. First, it does not indicate potential losses, and as a result is flawed, even on its own terms. Second, it is not a coherent measure of risk. Third, its dependence on a single quantile of the profit and loss distribution implies it is easy to manipulate reported VaR with specially crafted trading strategies. Finally, it is only concerned with the 99% loss level, or a loss which happens 2.5 times a year, implying that VaR violations have very little relevance to the probability of bankruptcy, financial crashes, or systemic failures.

The role of risk modelling in regulatory design is hotly debated. I argue that the inherent flaws in risk modelling imply that neither model based risk regulations nor the risk weighing of capital can be recommended. If financial regulations are deemed necessary, alternative market-based approaches such as *state contingent* capital levels and/or *cross–insurance* systems are needed.

2   Risk Modelling and Endogenous Response

The fundamental assumption in most statistical risk modelling is that the *basic* statistical properties of financial data during stable periods remain (almost) the same as during crisis. This of course is not correct. In normal times people act individually, some are selling while others buy. In contrast, during crisis people's

actions become more similar, there is a general flight from risky assets to safer properties. Herding instincts cause people to behave in a similar way. In other words, the statistical properties of financial risk are *endogenous*, implying that traditional risk models do not work as advertised.

Statistical financial models do break down in crisis. This happens because the statistical properties of data during crisis are different from the statistical properties in stable times. Hence, a model created in normal times may not be of much guidance in times of crisis. Morris and Shin (1999), Daníelsson and Zigrand (2001) and Daníelsson et al. (2001c) suggest that most statistical risk modelling is based on a fundamental misunderstanding of the properties of risk. They suggest that (most) risk modelling is based on the incorrect assumption of a single person (the risk manager) solving decision problems with a natural process (risk). The risk manager in essence treats financial risk like the weather, where the risk manager assumes a role akin to a meteorologist. We can forecasts the weather but can not change it, hence risk management is like a "game against nature". Fundamental to this is the assumption that markets are affected by a very large number of heterogeneous market participants, where in the aggregate their actions become a randomized process, and no individual market participant can move the markets. This is a relatively innocuous assumption during stable periods, or in all periods if risk modelling is not in widespread use. However, the statistical process of risk is different from the statistical process of the weather in one important sense: forecasting the weather does not (yet) change the statistical properties of the weather, but forecasting risk does change the nature of risk. This is related to Goodharts Law:

Law 1 (Goodhart, 1974) *Any statistical relationship will break down when used for policy purposes.*

We can state a corollary to this

Corollary 1 *A risk model breaks down when used for regulatory purposes.*

Current risk modelling practices are similar to pre–rational expectations Keynesian economics in that risk is modelled with behavioral equations that are invariant under observation. However, just as the economic crisis of the 1970's illustrated the folly of the old style Keynesian models, so have events in financial history demonstrated the limitations of risk models. Two examples serve to illustrate this. The Russia crisis of 1998 and the stock market crash of 1987.

Consider events during the 1998 Russia crisis. (See e.g. Dunbar, 1999). At the time risk had been modelled with relatively stable financial data. In Figure 2 on page 25 we see that the world had been in a low volatility state for the preceding half a decade, volatility had somewhat picked up during the Asian crisis of 1997, but those volatility shocks were mostly confined to the Far East,

and were levelling off in any case. In mid year 1998 most financial institutions employed similar risk model techniques and often similar risk constrains because of regulatory considerations. When the crisis hit, volatility for some assets went from 16 to 40, causing a breach in many risk limits. The response was decidedly one–sided, with a general flight from volatile to stable assets. This amplified price movements and led to a sharp decrease in liquidity. In other words, the presence of VaR based risk limits led to the execution of similar trading strategies, escalating the crisis.

This is similar to events surrounding a previous crisis, the 1987 crash when a method called portfolio insurance was very much in vogue. (See e.g. Jacobs, 1999). The key feature of portfolio insurance is that complicated hedging strategies with futures contracts are used to dynamically replicate options in order to contain downside risk. These dynamic trading strategies worked well in the stable pre–crisis periods since they depended on the presence of functioning futures markets. However, one characteristic of the '87 crash was that the futures markets ceased to function properly because the institutions who used portfolio insurance were trying to execute identical trading strategies, which only served to escalate the crisis. This is an important reason why the '87 crash was so short–lived, after the crash, program trading was no longer a factor, enabling prices to recover.

If every financial institution has its own trading strategy, no individual technique can lead to liquidity crisis. However, each institution's behavior does move the market, implying that the distribution of profit and loss is endogenous to the banks decision–making process. Risk is not the separate exogenous stochastic variable assumed by most risk models; risk modelling affects the distribution of risk. A risk model is a model of the aggregate actions of market participants, and if many of these market participants need to execute the same trading strategies during crisis, they will change the distributional properties of risk. As a result, the distribution of risk is different during crisis than in other periods, and risk modelling is not only ineffective in lowering systemic risk, but may exasperate the crisis, by leading to large price swings and lack of liquidity.

The role of regulations during these scenarios is complex. It is rational for most banks to reduce exposure in the event of a risk shock, independent of any regulatory requirements, and if banks have similar incentives and employ similar risk models, that alone can lead to a snowball effect in trading strategies during crisis. Indeed this is what happened during the 1987 crisis when regulation played no direct role. However, if regulations restrict the banks scope for pursuing individually optimal strategies, causing banks to act in a more uniform manner during crisis it may lead to an escalation of the crisis. This is exactly the conclusion of Daníelsson and Zigrand (2001) who model risk and regulation in a general equilibrium model. They also note that unregulated financial institutions, such as hedge funds, are essential for the prevention of systemic crisis. The burden of proof is on the regulators to demonstrate that the regulations do not escalate

financial crisis. They have not yet done so.

3 Empirical Properties of Risk Models

Risk forecasting depends on a statistical models and historical market price data. The modeller makes a number of assumptions about the statistical properties of the data, and consequently specifies the actual model. This model will always be based on objective observations and subjective opinions, and therefore the quality of the model depends crucially on the modeller's skill. As a result, it is not possible to create a perfect model. Each model has flaws, where the modeller weighs the pros and cons of each technique and data set, juggling issues like the choice of the actual econometric model, the length of the estimation horizon, the forecast horizon, and the significance level of the forecasts. In fact, due to these limitations, the resulting model is endogenous to its intended use. Two different users, who have different preferences but identical positions and views of what constitutes risk, require different risk forecasts. This happens because risk modelling is conditional on the user's loss functions. The weighing of the pros and cons is different for different users, resulting in different risk models and hence different risk forecasts, even for identical positions. It is therefore not surprising that some doubts have been raised about the reliability of risk models, both in the literature (see e.g. Lee and Saltoğlu, 2001) as well in the popular press: ˜

> "Financial firms employed the best and brightest geeks to quantify and diversify their risks. But they have all – commercial banks, investment banks and hedge funds – been mauled by the financial crisis. Now they and the worlds regulators are trying to find out what went wrong and to stop it happening again. ... The boss of one big firm calls super–sophisticated risk managers 'high–IQ morons' "
> The Economist, November 18, 1998, pp. 140–145.

In order to address some of these issues I refer below to results from a survey I made of the forecast properties of various models and data sets. The key issues I address are:

Robustness How accurate the risk forecasts really are

Risk volatility Risk forecasts fluctuate considerably from one period to the next, this is measured by the volatility of risk forecasts

Measuring horizon Regulators require that risk be forecasted with at least one year of data, and usually no more than one year

Holding period Regulations require that risk be forecasted over a 10 day holding period

Non–linear dependence Correlations typically underestimate the joint risk of two or more assets

3.1   Background

The data is a sample from the major asset classes, equities, bonds, foreign exchange, and commodities. Each dataset consists of daily observations and spans at least fifteen years. The discussion below is based on forecasts of day–by–day Value–at–Risk during the 1990s (2500 forecasts per data set on average) The risk level is mostly the regulatory 99% but I also consider lower and higher risk levels. The survey was done with a single asset return. While a portfolio approach would be more appropriate, this raises a number of issues which I thought best avoided, e.g. the serious issue of non–linear dependence discussed in Section 3.6 on page 12 . As such, my survey only presents a best case scenario, most favorable to risk models.

A large number of risk models exists and it is not possible to examine each and every one. However, most models are closely related to each other, and by using a carefully selected subsample I am confident that I cover the basic properties of most models in use. The models studied are:

- Variance–covariance models (VCV)[1]
- Unconditional models, historical simulation (HS) and extreme value theory (EVT)

More details on the survey can be found in Appendix A.

3.2   Robustness of Risk Forecasts

For a risk model to be considered reliable, it should provide accurate risk forecasts across different assets, time horizons, and risk levels within the same asset class. The robustness of risk models has been extensively documented, and there is not much reason to report detailed analysis here, my results correspond broadly to those from prior studies. I use violation ratios[2] to measure the accuracy of risk forecasts. If the violation ratio is larger than 1, the model is underforecasting risk (it is thin tailed relative to the data), and if the violation ratio is lower than 1, the model is overforecasting risk (the model is thick tailed relative to the data). Violation ratios are the most common method for ranking models, since they directly address the issue of forecast accuracy. The risk level used is the regulatory 99%, see Table 1 on page 22.

---

[1]Normal and student–t GARCH. These includes RiskMetrics$^{\text{TM}}$ as a special case

[2]The realized number of VaR violations over expected number of violations. By violation I mean that realized loss was larger than the VaR forecast

An ideal model has violation ratios close to 1 across asset classes and significance levels. While what constitutes "close to 1" is subjective, the range of 0.8 to 1.2 is a useful compromise. Based on this criteria, the results are depressing. For example, the normal VCV model produces violation ratios ranging from 0.37 to 2.18, and even for the same data set, e.g. S&P–500, the violation ratios range from 0.91 to 1.46. The other estimation methods have similar problems but not on the same scale. Every method overestimates the bond risk, and underestimates the risk in Microsoft stock. The normal VCV model (and by extension RiskMetrics^{TM}) has the overall worst performance, results for other models are mixed. Considering more considered risk levels, the ranking among models changes. At the 95% risk level, the normal VCV model performs generally best, while at 99.9% EVT is best.

These results show that no model is a clear winner. The forecasts, cover a very wide range, and the lack of robustness is disconcerting. Furthermore, the estimation horizon has considerable impact on forecast accuracy. One conclusion is that none of these models can be recommended, but since these models form the basis of almost every other model, this recommendation is too strong. My approach here is to use off-the–shelf models. A skilled risk manager considering specific situations is able to specify much more accurate models. This is indeed the internal situation in many banks. For reporting purposes, where the VaR number is an aggregate of all the banks risky positions, the use of an accurate specially designed model is much harder and off–the–shelf models are more likely to be used. This coupled with ad hoc aggregation methods for risk across operations, and the lack of coherence in VaR, can only have an adverse effect on model accuracy.

3.3   Risk Volatility

Fluctuations in risk forecasts have serious implications for the usefulness of a risk model; however, risk forecast fluctuations have not been well documented. The reason for this is unclear, but the importance of this issue is real. If a financial institution has a choice between two risk models both of which forecast equally well, but one model produces much less volatile forecasts, it will be chosen. And if risk forecasts are judged to be excessively volatile, it may hinder the use of risk forecasting within a bank. If a VaR number routinely changes by a factor of 50% from one day to the next, and factor 2 changes are occasionally realized, it may be hard to sell risk modelling within the firm. Traders are likely to be unhappy with widely fluctuating risk limits, and management does not like to change market risk capital levels too often. One reason for this is phantom price volatility. Furthermore Andersen and Bollerslev (1998) argue that there is an built–in upper limit on the quality of volatility forecasts (around 47%).

In my survey I use two measures of fluctuations in risk forecasts;

- The volatility of the VaR, i.e. the standard deviation of VaR forecasts over the sample period

- The VaR forecast range, i.e. the maximum and minimum VaR forecast over the sample period

Both measures are necessary. The VaR volatility addresses the issue of day–to–day fluctuations in risk limits, while the VaR range demonstrates the worst–case scenarios.

A representative sample of the results using the S&P–500 index is presented below in Table 2 on page 23. Consider e.g. the regulatory 99% level and the 300 day estimation horizon. The return volatility is 0.9%, and the volatility of the VaR estimates is 0.7% It is almost like we need a risk model to access the risk in the risk forecasts! The largest drop in returns is -7.1% (in 2527 observations), while the lowest normal VCV model forecast is -7.5% at the 99% level or an event once every 100 days. With longer estimation horizons both the volatility and the range decrease, suggesting that longer estimation horizons are preferred. The same results are obtained from the other data sets. Another interesting result is that the least volatile methods are historical simulation (HS) and extreme value theory (EVT). The reason is that conditional volatility models are based on a combination of long estimation horizons (more than 250 days) along with very short run VaR updating horizons (perhaps five days). In contrast, the HS and EVT methods are unconditional and as a result produce less volatile risk forecasts. Note that a hybrid conditional volatility and EVT method, such as the methods proposed by McNeil and Frey (2000) produce VaR forecasts which are necessarily more volatile than the VCV methods.

The contrast between VCV and EVT volatility can be seen in Figure 3 on page 26 which shows Hang Seng index returns during the last quarter of 1997, and Figure 4 on page 26 which shows risk forecasts for that data set with two different models, VCV and HS. Set in the middle of the Asian crisis, the Hang Seng index is very volatile, with the largest one day drop of more than 15%. Both models have an excessive amount of violations, but while the HS forecast is relatively stable throughout the quarter, the VCV forecast is very volatile. The lowest VCV VaR is 19%, and the model takes more than a month to stabilize after the main crash. In addition, the main contributor to the VCV VaR volatility is the positive return of 18% following the main crash. Since conditional volatility models like VCV have a symmetric response to market movements, a positive and negative market movement has the same impact on the VaR.

Because Value–at–Risk numbers are quantiles of the profit and loss distribution it is not surprising that they are volatile. However, I find them surprisingly volatile. It is not uncommon for VaR numbers to double from one day to the next, and then revert back. If VaR limits were strictly adhered to, the costs of portfolio rebalancing would be large. This has not gone unnoticed by the financial industry and regulators. Anecdotal evidence indicates that many firms

9

employ ad hoc procedures to smooth risk forecasts. For example, a bank might only update its covariance matrix every three months, or treat risk forecasts from conditional volatility models as an ad hoc upper limit for daily risk limits. Alternatively, covariance matrices are sometimes smoothed over time using non–optimal procedures. If Value–at–Risk is used to set risk limits for a trading desk, strict adherence to a VaR limit which changes by a factor of two from one day to the next is indeed costly. The same applies to portfolio managers who need to follow their mandate, but would rather not rebalance their portfolios too often. In addition, since regulatory VaR is used to determine market risk capital, a volatile VaR leads to costly fluctuations in capital if the financial institution keeps its capital at the minimum level predicted by the model. This may in turn cause lack of confidence in risk models and hinder their adoption within a firm. Anecdotal evidence indicates that (some) regulators consider bank capital as a constant to be allocated to the three categories of risk, market, credit, and operational, and not the widely fluctuating quantity predicted by the models.

3.4   Model Estimation Horizon

The estimation of a risk model depends on sufficiently long historical data series being available. The regulatory suggestion is (at least) 250 days, and most supervisors do not allow longer estimation horizons. This must be based on one of two assumptions;

- Older data is not available, or is irrelevant due to structural breaks
- Long run risk dynamics are so complicated that they can't be modelled

While the first assumption is true in special cases, e.g. immediately after a new instrument is introduced such as the Euro, and in emerging markets, in general, it is not correct. The second assumption is partially correct: long run risk dynamics are complicated and often impossible to model explicitly; however, long run patterns can be incorporated, it just depends on the model.

Long run risk dynamics are not a well understood and documented phenomena, but it is easy to demonstrate the existence of long cycles in volatility. Consider Figure 2 on page 25 which demonstrates changes in average daily volatility for the second half of the 20th century. Daily volatility ranges from 0.5% to almost 2% in a span of few years. The 1990s demonstrate the well–known U–shaped pattern in volatility.

Observing these patterns in volatility is one thing, modelling them is another. Although existing risk models do not yet incorporate this type of volatility dynamics, conceivably it is possible. Most conditional volatility models, e.g. GARCH, incorporate both long run dynamics (through parameter estimates) and very short–term dynamics (perhaps less than one week). Long memory volatility models may provide the answers; however, their risk forecasting properties are

still largely unexplored.

The empirical results presented in Table 1 on page 22 and Table 2 on page 23 show that shorter estimation horizons do not appear to contribute to more accurate forecasting, but longer estimation horizons do lead to lower risk volatility. For that reason alone longer estimation horizons are preferred.

3.5   Holding Periods and Loss Horizons

Regulatory Value–at–Risk requires the reporting of VaR for a 10 day holding period. This is motivated by a fear of a liquidity crisis where a financial institution might not be able to liquidate its holdings for 10 days straight. While this may be theoretically relevant, two practical issues arise;

- The contradiction in requiring the reporting of a 10 day 99% Value–at–Risk, i.e. a two week event which happens 25 times per decade, in order to catch a potential loss due to a liquidity crisis which is unlikely to happen even once a decade. Hence the probability and problem are mismatched.

- There are only two different methods of doing 10 day Value–at–Risk in practice:
  - Use non–overlapping (Overlapping returns cannot be used) 10 day returns to produce the 10 day Value–at–Risk forecast
  - Use a scaling law to convert one day VaRs to 10 day VaRs (recommended by the Basel Committee on Banking Supervision (1996))

  Both of these methods are problematic

If 10 day returns are used to produce the Value–at–Risk number, the data requirements obviously also increase by a factor of 10. For example, if 250 days (one year) are used to produce a daily Value–at–Risk number, ten years of data are required to produce a 10 day Value–at–Risk number with the same statistical accuracy. If however 250 days are used to produce 10 day Value–at–Risk numbers, as sometimes is recommended, only 25 observations are available for the calculation of something which happens in one observation out of every hundred, clearly an impossible task. Indeed, at least 3,000 days are required to directly estimate a 99% 10 day VaR, without using a scaling law.

To bypass this problem most users tend to follow the recommendation in the Basel regulations Basel Committee on Banking Supervision (1996) and use the so–called '*square–root–of–time*' rule, where a one day VaR number is multiplied by the square root of 10 to get a 10 day VaR number. However, this depends on surprisingly strong distribution assumptions, i.e. that returns are normally identically and individually distributed (iid):

- Returns are normally distributed

- Volatility is independent over time
- The volatility is identical across all time periods

All three assumptions are violated. However, creating a scaling law which incorporates violations of these assumptions is not trivial. For example, it is almost impossible to scale a one day VaR produced by a normal GARCH model to a 10 day VaR (see Drost and Nijman, 1993; Christoffersen and Diebold, 2000). Using *square–root–of–time* in conjunction with conditional volatility models implies an almost total lack of understanding of statistical risk modelling. The problem of time scaling for a single security, e.g. for option pricing, is much easier than the scaling of an institution wide VaR number, which currently is impossible. When I ask risk managers why they use the *square–root–of–time* rule, they reply that they do understand the issues (these problems have been widely documented), but they are required to do this anyway because of the demand for 10 day VaRs for regulatory purposes. In other words, regulatory demands require the risk manager to do the impossible! I have encountered risk managers who use proper scaling laws for individual assets, but then usually in the context of option pricing, where the pricing of path dependent options depends crucially on using the correct scaling method and accurate pricing has considerable value.

In fact, one can make a plausible case for the *square–root–of–time* rule to be twice as high, or alternatively half the magnitude of the real scaling factor. In other words, if a daily VaR is one million, a 10 day VaR equal to 1.5 million or 6 million is as plausible. Indeed, given current technology, it is not possible to come up with a reliable scaling rule, except in special cases. The market risk capital regulatory multiplier of 3 is sometimes justified by the uncertainty in the scaling laws, e.g. by Stahl (1997), however as suggested by Daníelsson et al. (1998), it is arbitrary.

Estimating VaR for shorter holding horizons (intra day VaR, e.g. one hour) is also very challenging due to intraday seasonal patterns in trading volume, as frequently documented. Any intraday VaR model needs to incorporate these intraday patterns explicitly for the forecast to be reliable, a non–trivial task.

Considering the difficulty given current technology of creating reliable 10 day VaR forecasts, regulatory market risk capital should not be based on the 10 day horizon. If there is a need to measure liquidity risk, other techniques than VaR need to be employed. If the regulators demand the impossible, it can only lead to a lack of faith in the regulatory process.

3.6   Non–Linear Dependence

One aspect of risk modelling which has been receiving increasing attention is the serious issue of non–linear dependence, sometimes known as changing correlations. It is well known that estimated correlations are much lower when

the markets are increasing in value, compared to market conditions when some assets increase and other decrease, and especially when the markets are falling. Indeed, the worse market conditions are, the higher the correlation is: in a market crash, all assets collapse in value and correlations are close to hundred percent. However, most risk models do not take this into consideration and produce correlation estimates based on normal market conditions. As a result, these models will underestimate portfolio risk. Furthermore, since the dependence increases with higher risk levels, a VCV model which performs well at the 95% level, will not perform as well at the 99% or beyond. In contrast, this problem is bypassed in some unconditional methods which preserve the dependence structure, e.g. historical simulation and extreme value theory.

One study which demonstrates non–linear dependence is by Erb et al. (1994) who consider monthly correlations in a wide cross–section of assets in three different market conditions (bull markets , bear markets, mixed). They rank data according to the market conditions, and report correlations for each subsample. A small selection of their results is reported below:

Asset Pair		Up–Up	Down–Down	Mixed	Total
US	Germany	8.6	52	-61	35
	Japan	21	41	-54	26
	UK	32	58	-60	50
Germany	Japan	4.6	24	-47	40
	UK	22	40	-62	42
Japan	UK	12	21	-54	37

We see that correlations are low when both markets are increasing in value, for example for the U.S. and Germany the correlation is only 8.6%. When both of these markets are dropping in value, the correlation increases to 52%. Similar results have been obtained by many other authors using a variety of data samples.

These problems occur because the distribution of risk is non–elliptical, and as a result, correlations provide misleading information about the actual dependence. In most cases, correlations will underestimate the downside risk. To properly measure the dependence structure, one needs to use the related concepts of bi–variate extreme value theory and copulas. See e.g. Longin (1998), Hartmann et al. (2000), Frey and McNeil (2002), or Embrechts et al. (2000) for more information.

Another problem in correlation analysis relates to international linkages. Since not all markets are open exactly the same time, volatility spillovers may be spread over many days. This happened during the 1987 crisis where the main crash day was spread over two days in Europe and the Far East. A naïve analysis would indicate that the US markets experienced larger shocks than other markets, however this is only an artifact of the data. This also implies that it is very

hard to measure correlations across timezones. Any cross–country analysis is complicated by market opening hours, and is an additional layer of complexity.

4 The Concept of (Regulatory) Risk

In order to manage and regulate risk, we need a useful definition of risk. It is therefore unfortunate that no uniform definition of financial risk exists, it depends on factors such as intended applications and amount of available data. We can identify three main risk concepts, in increasing order of complexity:

Volatility Traditional risk measure, but mostly unsuitable for financial risk

Value–at–Risk The foundation of market risk regulations. Flawed but unavoidable

Coherent risk measures The preferred way to measure risk, but hard to work with

4.1 Volatility

A textbook definition of risk is volatility, i.e. the standard deviation of returns, however, volatility is a highly misleading concept of risk. It depends on the notion of returns being normally distributed, but as they are not, volatility only gives a partial picture.

Consider Figure 5 on page 27 which shows 500 realizations of two different return processes, A which is normally distributed and B which is not normal. For the purpose of risk, returns B are clearly more risky, for example the 99% VaR for asset A is 2, while the VaR for B is 7. However, the volatility of A is 1, while the volatility of B is 0.7. If volatility was used as a risk measure, we would incorrectly select B as the less risky asset. Furthermore, most risk models are based on volatility and hence would also select B.

4.2 Value–at–Risk

The Basel Committee on Banking Supervision (1996) chose VaR as the foundation of market risk regulations, and the 2001 Basel Committee proposals leave the 1996 amendment unchanged. Furthermore, in the 2001 proposals a VaR type methodology underpins the credit and operational risk recommendations.

VaR is in many aspects an attractive measure of risk, it is not only distribution independent and relatively easy to implement, it is also easy to explain to non–expert audiences. Therefore is is surprising how many authors confuse VaR with other risk measures, c.f. expected shortfall. There are also disadvantages in

using VaR, e.g. it is only one point on the distribution of profit and loss (P&L), and is easy to manipulate. In addition, VaR has the surprising property that the VaR of a sum may be higher than the sum of the VaRs implying that VaR is not *subadditive*. Consider two portfolios, A and B. It is possible that:

$$VaR^{A+B} > VaR^A + VaR^B$$

e.g. diversification will lead to more risk being reported.

Marco Avellaneda suggests an example, illustrates why VaR is not subadditive, and how that leads to interesting results.

Example 1 (Avellaneda) *Let $X(i), i = 1, \ldots, 100$ be 100 iid defaultable bonds with the following P&L: $X(i)$ defaults with probability 1% at the end of the period, and has a period end coupon of 2%, when is does not default. Suppose each bond has face value $100. The P&L of each bond is then:*

$$X(i) = \begin{cases} -100 & \text{with probability } 0.01 \\ +2 & \text{with probability } 0.99 \end{cases} \quad i = 1, \ldots, 100.$$

Consider the following two portfolios:

$$\begin{array}{lll} P1: & X(1) + \ldots + X(100) & \text{(diversified)} \\ P2: & 100 \times X(1) & \text{(non–diversified)} \end{array}$$

It is easy to check that at 95%,

$$VaR(P1) > VaR(P2) = 100 \times VaR(X(1)),$$

hence VaR is non–subadditive.

In the example, the non–diversified portfolio, which intuitively is more risky, is nevertheless chosen by VaR as the less risky portfolio. This demonstrates how overreliance on VaR can lead to perverse outcomes.

4.3   Coherent Risk Measures

Value–at–Risk does not take into account the entire lower tail of the profit and loss (P&L) distribution, only one point, i.e. one quantile. However this may not be a relevant measure in many cases. What matters is how much money a bank loses when a disaster strikes, not the minimum amount of money it loses on a bad day. If the VaR is 1 million, one has no way of knowing whether the maximum possible loss is 1.1 million or 100 million. While users may implicitly map the VaR number into a more useful measure, perhaps relying on dubious distributional assumptions, this can not be recommended. If a different measure is needed, it

should be modelled explicitly. Furthermore, as noted by Artzner et al. (1999) VaR is not a coherent measure of risk because it fails to be *subaddditive*.[3]

As a result, alternative risk measures have been proposed which properly account for the lower tail of the P&L distribution while remaining subadditive. Artzner et al. (1999) discuss several such measures, most notably expected shortfall (ES) which measures the *expected loss* conditional on the VaR being violated. The ES risk measure is intuitively more appealing than VaR and indeed, many people confuse the two. Unfortunately, in spite of its theoretic advantages, ES and related risk measures may be hard to implement in practice, primarily, as demonstrated by Yamai and Yoshiba (2001), ES requires much more data for backtesting than VaR. Alternatively, conditional VaR (Rockafellar and Uryasev, 2002) is a coherent risk measure which also captures some of the downside risk.

The fact that VaR is not subadditive may not be a serious flaw anyway. Cumperayot et al. (2000) and Daníelsson et al. (2001b) show that these ES and VaR provide the same ranking of risky projects under second order stochastic dominance, implying that VaR is a sufficient measure of risk. This however only happens sufficiently far the tails. Whether the regulatory 99% is sufficiently far out, remains an open question.

4.4   Moral Hazard — Massaging VaR Numbers

The reliance on a single quantile of the P&L distribution by VaR is conducive to the manipulation of reported risk. Suppose a bank considers its VaR unacceptably high. Ahn et al. (1999) discuss the optimal response by the bank to this by the use of options. Consider a richer example where the bank sets up trading strategies to manipulate the VaR.

Example 2 (VaR Manipulation) *Assume that the VaR before any response is VaR_0 and that the bank really would like the VaR to be VaR^D where the desired VaR is $VaR^D > VaR_0$. One way to achieve this is to write a call with a strike price right below VaR_0 and buy a put with a strike right above VaR^D, i.e. $X^c = VaR_0 - \epsilon$ and $X^p = VaR^D + \epsilon$. The effect of this will be to lower expected profit and increase downside risk, see Figure 1 on page 17.*

This is possible because the regulatory control is only on a single quantile, and the bank is perfectly within its rights to execute such a trading strategy. A risk measure like expected shortfall renders this impossible. The regulatory focus on a simple measure like VaR may thus perversely increase risk and lower profit, while the intention is probably the opposite.

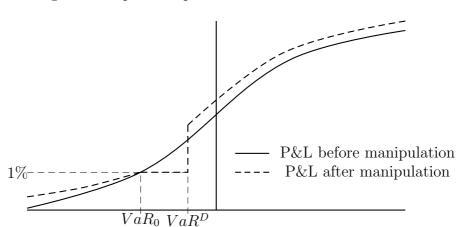[3]A function f is subadditive if $f(x_1 + ... + x_N) \leq f(x_1) + ... + f(x_N)$.

Figure 1: Impact of options on the C.D.F. of Profit and Loss

4.5 The Regulatory 99% Risk Level

The regulatory risk level is 99%. In other words, we expect to realize a violation of the Value–at–Risk model once every hundred days, or 2.5 times a year on average. Some banks report even lower risk levels, JP Morgan (the creator of VaR and RiskMetrics) states in its annual report that in 1996 its average daily 95% VaR was $36 million. Two questions immediately spring to mind: why was the 99% level chosen, and how relevant is it?

The first question may have an easy answer. Most risk models only have desirable properties in a certain probability range and the risk level and risk horizon govern the choice of model. For example, at the 95% risk level, variance–covariance models such as GARCH or RiskMetrics are the best choice. However, the accuracy of these models diminishes rapidly with lower risk levels, and at the 99% risk level they cannot be recommended, and other, harder to use techniques must be employed. In general, the higher the risk level, the harder it is to forecast risk. So perhaps the 99% level was chosen because it was felt that more extreme risk levels were too difficult to model?

The question of the relevance of a 99% VaR is harder to answer because it depends on the underlying motivations of the regulators. The general perception seems to be that market risk capital is required in order to prevent systemic failures. However not only is systemic risk hard to define (see De Bandt and Hartmann, 2000, for a survey), it is unclear if VaR regulations decrease or increase the probability of systemic crisis as demonstrated by Daníelsson and Zigrand (2001). In any case, systemic failures are very rare events, indeed so rare that one has never been observed in modern economies. We have observed near–systemic collapses, e.g. in the Scandinavian banking crisis, but in that case even a meticulously measured VaR would not have been of much help.

The regulatory risk level is clearly mismatched with the event it is supposed to be relevant for, i.e. systemic collapse. In other words, the fact that a bank violates its VaR says nothing about the probability of the banks probability of bankruptcy, indeed, one expects the violation to happen 2.5 times a year. There is no obvious mapping from the regulatory risk level to systemic risk, however defined. Whether there is no link between regulatory risk and systemic risk is still an open question.

5 Implications for Regulatory Design

The analysis above focuses on market risk models. Market risk modelling is much easier than the modelling of other risk factors due to the relative abundance of accurate market risk data and simpler and more established methodology. The fundamental criticism of risk modelling remains however, and the discussion consequently applies equally to credit, liquidity, and operational risk. Therefore, considering the challenges in market risk modelling, it is disappointing to read the Basel–II proposals which basically extend the VaR regulatory methodology to credit and operational risk.

The arguments voiced above suggest that using risk modelling as a regulatory tool is not as straightforward as usually claimed. Assuming that regulations are here to stay, the important question must be whether it is possible to create a regulatory mechanism that is successful in reducing systemic risk, but not too costly.

For most of the 1990s the answer to this question seemed to be the VaR methodology. It is only after the Asian and Russian crisis that modelling as a regulatory tool has come under serious criticism. Given current technology, risk modelling is simply too unreliable, it is too hard to define what constitutes a risk, and the moral hazard issues are too complicated for risk modelling to be an effective part of regulatory design, whether for market, credit, liquidity, or operational risk. This is a reflection of the current state of technology. Empirical risk modelling is a rapidly developing field, and there is every reason to believe that modelling can become an effective regulatory tool, just as it is now for internal risk management.

At the moment, if the authorities want banks to hold minimum capital, crude capital adequacy ratios may be the only feasible way since model based risk weighing of capital does not work effectively. Furthermore, basing capital on risk weighed assets leaves open a serious question. If capital is risk sensitive, in times of crisis, when risk is increasing, financial institutions will have to raise capital in order to remain compliant. This can only serve to exasperate the crisis. As a result, such capital is probably *de facto state contingent* since the regulators will have to relax the standards in times of crisis. Countries like the UK where the supervisor ultimately decides on implementation have a considerable advantage over countries, e.g. like the USA, where the supervisors are required by an act of Congress to implement the capital standards. Supervisors have not addressed this issue publicly, not only would this directly contradict the idea behind the present capital adequacy standards, it also leaves open the question of what constitutes a crisis.

The authorities could follow the lead of New Zealand and do away with minimum capital, but require banks instead to purchase insurance, in effect require financial institutions to cross insure each other. This market solution has the advantage

that much more flexibility is built into the system while at the same time sifting the burden of risk modelling back to the private sector. While such a system may work for small country like New Zealand which can insure in larger markets, it is still an open question whether this would work for larger economies.

Ultimately, we do not yet know enough about systemic crisis to design an effective supervisory system. Unfortunately, the Basel–II proposals only address the more tractable aspects of financial risk, i.e. the measuring and management of financial risk in non–crisis periods. I fear that an implementation of the 2001 Basel–II proposals will not help in understanding the likelihood or the nature of financial crisis. These issues have been pointed out by several comments on the Basel–II proposals, e.g. by Daníelsson et al. (2001a).

Much research is being done in this area, both theoretic and applied, with work which combines empirical market microstructure and financial theory especially promising. The hope is that in the near future we will know enough to properly regulate, or perhaps realize that regulations are not needed. It is therefore important that policy makers carefully consider their options and do not rush into the implementation of ineffective or even harmful regulations.

6   Conclusion

Empirical risk modelling forms the basis of the market risk regulatory environment as well as internal risk control. This paper identifies a number of shortcomings with regulatory Value–at–Risk, where both theoretic and empirical aspects of VaR are analyzed.

Most existing risk models break down in times of crisis because the stochastic process of market prices is endogenous to the actions of market participants. If the risk process becomes the target of risk control, it changes its dynamics, and hence risk forecasting becomes unreliable. This is especially prevalent in times of crisis, such as events surrounding the Russia default of 1998 demonstrated.

In practice, Value–at–Risk is forecasted using an empirical model in conjunction with historical market data. However, current risk modelling technology still in the early stages of development, is lacking in the robustness of risk forecasts, and produces excessively volatile risk forecasts. If risk modelling is not done with great skill and care, the risk forecast will be unreliable to the point of being useless. Or even worse, it may impose significant but unnecessary costs on the financial institution, due to the misallocation of capital and excessive portfolio rebalancing.

This, however, is only a reflection on the current state of technology. A risk model which incorporates insights from economic and financial theory, in conjunction with financial data during crisis, has the potential to provide much more accurate answers by directly addressing issues such as liquidity dynamics. There is a need

for a joint market and liquidity risk model, covering both stable and crisis periods.

The theoretic properties of the VaR measure, conceptually result in VaR providing misleading information about a financial institution's risk level. The very simplicity of the VaR measure, so attractive when risk is reported, leaves the VaR measure wide open to manipulation. This in turn implies that founding market risk regulations on VaR, not only can impose considerable costs on the financial institution, it may act as a barrier to entry, and perversely increase both bank and systemic risk.

The problems with risk modelling have not gone unnoticed. Risk modelling does, however, serve a function when implemented correctly internally within a firm, but its usefulness for regulatory purposes is very much in doubt for the present.

A Empirical Study

The empirical results are a subset of a survey I made. I used there 4 common estimation methods

- Normal GARCH
- Student t GARCH
- Historical simulation
- Extreme value theory

as well as representative foreign exchange, commodity, and equity datasets containing daily observations obtained from DATASTREAM, from the first recorded observation until the end of 1999.

- S&P 500 index
- Hang Seng index
- Microsoft stock prices
- Amazon stock prices
- Ringgit pound exchange rates
- Pound dollar exchange rates
- Clean U.S. government bond price index
- Gold prices
- Oil prices

I estimated each model and dataset with a moving 300, 1,000, and 2,000 day estimation window, and forecast risk one day ahead. Then I record the actual return, move the window and reestimate. This is repeated until the end of the sample.

Table 1: 99% VaR Violation Ratios 1990–1999

Data	Estimation horizon	GARCH Normal	GARCH Student-t	Historical simulation	Extreme value theory
S&P–500	300	1.46	1.07	0.79	0.79
S&P–500	1,000	1.27	0.83	0.95	0.99
S&P–500	2,000	0.91	0.67	1.07	1.07
US bond	300	0.94	0.66	0.49	0.49
US bond	1,000	0.53	0.57	0.66	0.41
US bond	2,000	0.37	0.49	0.67	0.37
Oil	300	1.76	1.38	1.17	1.17
Oil	1,000	1.67	1.30	0.92	1.00
Oil	2,000	1.64	1.04	0.35	0.35
Hang Seng	300	2.18	1.41	0.69	0.69
Hang Seng	1,000	1.90	1.29	1.17	1.21
Hang Seng	2,000	2.02	1.21	1.09	1.09
Microsoft	300	2.24	1.78	1.58	1.58
Microsoft	,1000	1.98	1.60	1.84	1.74
Microsoft	2,000	2.25	1.69	2.06	1.87
GBP/USD	300	2.13	1.42	0.79	0.79
GBP/USD	1,000	1.85	1.18	0.63	0.63
GBP/USD	2,000	1.62	1.14	0.47	0.47

Notes: Each model was estimated with three different estimation horizons, 300, 1000, and 2000 days. The expected value for the violation ratio is one. A value larger than one indicates underestimation of risk, and a value less than one indicates overestimation.
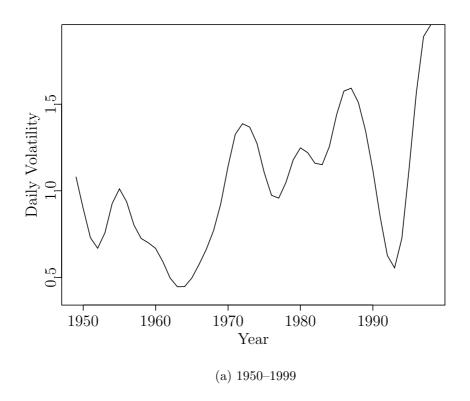
B tables

Table 2: S&P–500 Index 1990–1999. VaR Volatility

Risk Level	Statistic	Estimation Horizon	Returns	GARCH Normal	GARCH Student-t	Historical simulation	Extreme value theory
5%	SE	300	0.89	0.47	0.44	0.41	0.41
5%	Min	300	−7.11	−5.32	−4.17	−2.19	−2.19
5%	Max	300	4.99	−0.74	−0.66	−0.73	−0.73
5%	SE	2,000	0.89	0.42	0.41	0.12	0.12
5%	Min	2,000	−7.11	−3.70	−3.31	−1.55	−1.55
5%	Max	2,000	4.99	−0.81	−0.74	−1.15	−1.15
1%	SE	300	0.89	0.66	0.68	0.71	0.71
1%	Min	300	−7.11	−7.52	−6.68	−3.91	−3.91
1%	Max	300	4.99	−1.05	−1.09	−1.26	−1.26
1%	SE	2,000	0.89	0.60	0.64	0.29	0.33
1%	Min	2,000	−7.11	−5.23	−5.45	−2.72	−2.84
1%	Max	2,000	4.99	−1.14	−1.26	−1.90	−1.90
0.4%	SE	300	0.89	0.76	0.82	1.94	1.94
0.4%	Min	300	−7.11	−8.58	−8.25	−7.11	−7.11
0.4%	Max	300	4.99	−1.19	−1.36	−1.81	−1.81
0.4%	SE	2,000	0.89	0.68	0.80	0.74	0.63
0.4%	Min	2,000	−7.11	−5.96	−6.84	−4.27	−4.18
0.4%	Max	2,000	4.99	−1.30	−1.61	−2.50	−2.49
0.1%	SE	300	0.89	0.88	1.06	1.94	1.94
0.1%	Min	300	−7.11	−9.99	−10.92	−7.11	−7.11
0.1%	Max	300	4.99	−1.39	−1.82	−1.81	−1.81
0.1%	SE	2,000	0.89	0.80	1.06	2.10	1.42
0.1%	Min	2,000	−7.11	−6.94	−9.31	−8.64	−7.49
0.1%	Max	2,000	4.99	−1.52	−2.26	−3.13	−3.15

Notes: For each model and the risk level, the table presents the standard error (SE) of the VaR forecasts, and the maximum and minimum forecast throughout the sample period.

Figure 2: Estimated Daily Unconditional Volatility (Smoothed) S&P–500
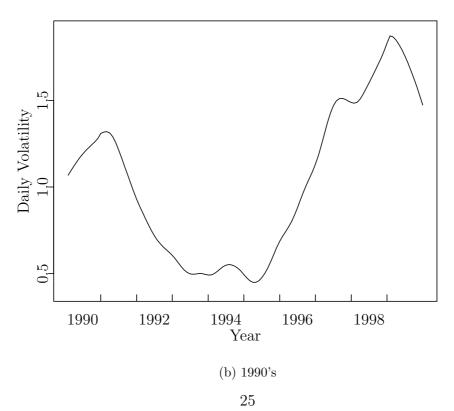

(a) 1950–1999

(b) 1990's

25

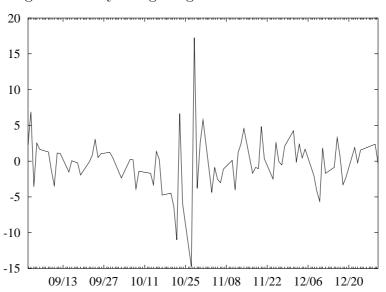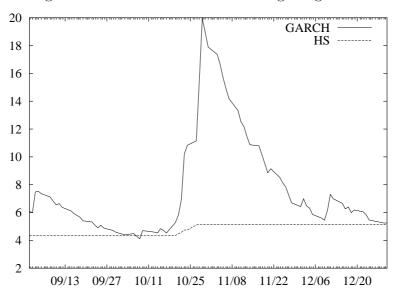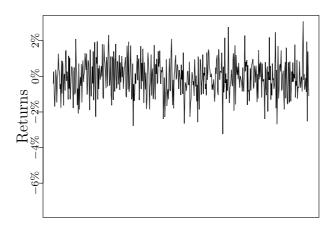Figure 3: Daily Hang Seng Index 1997 and 99% VaR

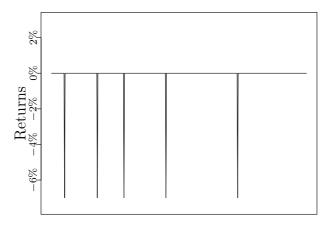Figure 4: Risk forecasts for the Hang Seng Index

Impact on the sample economy of adjusting the risk constraint, $log(\bar{v})$. At values $log(\bar{v}) < 12$.

26

Figure 5: Which is More Volatile and which is more Risky?

(a) Returns A

(b) Returns B

References

Ahn, D. H., Boudukh, J., Richardson, M., and Whitelaw, R. F. (1999). Optimal risk management using options. *Journal of Finance*, 54 (1):359–375.

Andersen, T. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 38:885–905.

Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999). Coherent measure of risk. *Mathematical Finance*, 9 (3):203–228.

Basel Committee on Banking Supervision (1996). *Overview of the amandment to the capital accord to incorporate market risk.*

Christoffersen, P. F. and Diebold, F. X. (2000). How relevant is volatility forecasting for financial risk management? *Review of Economics and Statistics*, 82:1–11.

Cumperayot, P. J., Daníelsson, J., Jorgensen, B. N., and de Vries, C. G. (2000). On the (ir)relevancy of value–at–risk regulation. In Franke, J., Härdle, W., and Stahl, G., editors, *Measuring Risk in Complex Stochastic Systems*, volume 147 of *Lecture Notes in Statistics*, chapter 6. Springer Verlag.

Daníelsson, J., Embrechts, P., Goodhart, C., Keating, C., Muennich, F., Renault, O., and Shin, H. S. (2001a). An academic response to Basel II. The New Basel Capital Accord: Comments received on the Second Consultative Package `http://www.bis.org/bcbs/cacomments.htm` or `www.RiskResearch.org`.

Daníelsson, J., Hartmann, P., and de Vries, C. G. (1998). The cost of conservatism: Extreme returns, value-at-risk, and the basle 'multiplication factor'. *Risk*, January 1998. *www.RiskResearch.org*.

Daníelsson, J., Jorgensen, B. N., and de Vries, C. G. (2001b). Risk measures and investment choice. Unfinished working paper, `www.RiskResearch.org`.

Daníelsson, J., Shin, H. S., and Zigrand, J.-P. (2001c). Asset price dynamics with value–at–risk constrained traders. Working paper, LSE, `www.RiskResearch.org`.

Daníelsson, J. and Zigrand, J.-P. (2001). What happens when you regulate risk? evidence from a simple equilibrium model. Working paper, LSE, `www.RiskResearch.org`.

De Bandt, O. and Hartmann, P. (2000). Systemic risk: a survey. Discussion paper series, no. 2634, CEPR.

Drost, F. C. and Nijman, T. E. (1993). Temporal aggregation of GARCH processes. *Econometrica*, 61:909–927.

Dunbar, N. (1999). *Inventing Money: The Story of Long–Term Capital Management and the Legends Behind It.* John Wiley.

Embrechts, P., McNeil, A., and Straumann, D. (2000). Correlation and dependency in risk management: properties and pitfalls. In Dempster, M. and HK, H. M., editors, *Risk management: value at risk and beyond*, volume forthcomming. Cambridge University Press.

Erb, Harvey, and Viskanta (1994). Forecasting international correlations. *Financial Analysts Journal*.

Frey, R. and McNeil, A. J. (2002). Modelling dependent defaults. *Journal of Banking and Finance*, forthcomming.

Goodhart, C. A. E. (1974). Public lecture at the reserve bank of Australia.

Hartmann, P., Straetmans, S., and de Vries, C. G. (2000). Asset market linkages in crisis periods. working paper.

Jacobs, B. (1999). *Capital Ideas and Market Realities.* Blackwell.

Lee, T.-H. and Saltoğlu, B. (2001). Evaluating predictive performance of value-at-risk models in emerging markets: A reality check. `http://mimoza.marmara.edu.tr/~saltoglu/jef.pdf`.

Longin, F. (1998). Correlation structure of international equity markets during extremely volatile periods. *babel.essec.fr:8008/domsite/cv.nsf/WebCv/Francois+Longin.*

McNeil, A. J. and Frey, R. (2000). Estimation of tail–related risk measures for heteroskedastic financial time series: An extreme value approach. *Journal of Emprical Finance*, 7:271–300.

Morris, S. and Shin, H. S. (1999). Risk management with interdependent choice. *Oxford Review of Economic Policy*, Autumn 1999, 52 - 62, reprinted in the Bank of England *Financial Stability Review*, 7, 141 - 150, `http://www.bankofengland.co.uk/fsr/fsr07art5.pdf`.

Rockafellar, R. T. and Uryasev, S. (2002). Conditional Value–at–Risk for general loss distributions. *Journal of Banking and Finance*, forthcomming.

Stahl, G. (1997). Three cheers. *Risk*, 10:67–69.

Yamai, Y. and Yoshiba, T. (2001). Comparative analyses of expected shortfall and var: their estimation error, decomposition, and optimization. IMES Discussion Paper Series 2001-E-12, http://www.imes.boj.or.jp/english/publication/edps/2001/01-E-12.pdf.