Model Risk Analysis for Risk Management and Option Pricing

(Modelrisicoanalyses voor risicomanagement en derivatenwaardering)

PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Tilburg,

op gezag van de rector magnificus,

prof. dr. F.A. van der Duyn Schouten,

in het openbaar te verdedigen ten overstaan van

een door het college voor promoties aangewezen commissie

in de aula van de Universiteit op

vrijdag 7 november 2003, om 14.15 uur

door

Franciscus Lambertus Johannes Kerkhof,

geboren op 5 februari 1976 te Waspik.

Promotor: Prof. dr. J.M. Schumacher

Copromotor: Dr. B. Melenberg

Voor diegenen die het van kaft tot kaft lezen...

Acknowledgements

Four years of work resulted in the thesis that lies in front of you. During these four years I had the opportunity to (unevenly) divide my time between Tilburg University and ABN-AMRO Bank. I am grateful to both institutions for offering me the possibility to combine academic research and finance practice.

During my quest to finish this thesis, I have been guided and supported by numerous people, some directly others indirectly, without whom this thesis could not have been written in its current form. Among these people, I owe great gratitude to my supervisors Hans Schumacher and Bertrand Melenberg. They provided me with the freedom needed to develop my own ideas and turned them in more realistic ones along the way. They taught me academic thinking and writing, being precise, and much more. I would also like to thank Bas Werker for carefully reading and commenting most of the thesis during its realization.

At ABN-AMRO Bank I would like to thank all my colleagues and, in particular, Mark de Vries and Marcel van Regenmortel. Mark gave me the opportunity to get a feel for finance practice in both the product analysis as the product development group and explaining the different focuses of academia and practice. By discussing his work with me Marcel gave me insights in many different topics and aspects of being a finance "quant".

Along these four years I have been lucky to visit various interesting places such as the department of Finance at the University of Maryland at College Park. I would like to thank Dilip Madan for his hospitable hosting. Furthermore, I would like to thank Jenke ter Horst for sharing his office and apartment during this period. I also had the pleasure of visiting the department of Mathematics at the Scuola Normale Superiore in Pisa for which I would like to thank Paolo Guasoni and Maurizio

Pratelli.

I would like to thank Antoon Pelsser for persuading me to publish our results on string models and sharing his insights on the interest rate derivatives literature. I would also like to thank Theo Nijman and Theo Dijkstra for reading my thesis and being part of the PhD committee.

During my stay at Tilburg University I had the pleasure of sharing the excitements and troubles of the life of a PhD student with my friend and officemate Laurens Swinkels. Especially, the conference "holidays" are to be remembered. Furthermore, I would like to thank my friends, the secretaries, and fellow PhD students at Tilburg University for the enjoyable times I have had.

Special thanks goes to my mother who has always given me freedom in choosing my own way and supporting me along it. Finally, I would like to thank Astrid for making everything I do more enjoyable.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and motivation

During the last decennia we witnessed an increasing complexity of products in the financial markets. This led financial institutions to rely more and more on the use of quantitative models. Besides the growth in complexity we also saw a spectacular growth in trading, especially in derivative instruments, that continues in the new millennium. For example, the turnover of exchange-traded financial derivatives at the end of 2002 is estimated about $ 170 trillion and the notional amount of outstanding OTC contracts is estimated about $ 128 trillion (see Jeanneau (2002)).

The growing complexity and trading size of the financial markets makes the task for financial regulators more difficult and important. An important part of banks' portfolios these days consists of derivative securities whose value can depend on some traditional underlyings such as stocks, bonds, and currencies or some more exotic underlyings such as volatility and credit. One of the most apparent differences between trading derivatives and fundamental securities, like stocks and bonds, is the importance of theoretical valuation and hedging models in derivatives markets. For example, these models are used for predicting the behavior of the term structure of interest rates, modeling the dynamics of assets underlying derivative contracts, the volatility surface of option prices, determining the most appropriate hedge instruments, etc. This induces a somewhat new type of risk, *model risk*. Model risk manifests itself, for instance, when a delta hedge for a put or call option on a stock index is based on a Black-Scholes (lognormal) model with a certain volatility, whereas the index in fact follows a lognormal process with a different volatility. When the assumed value of the volatility is less than the actual value, the party that is writing the option will not only suffer a loss from quoting a too low price but will also be subject to a variance resulting from mishedging. When the volatility is overestimated the writer may in a competitive market not be able to sell the option at the high price that correspond to the overestimated volatility; even when the option is sold, there is still a mishedging risk. Potentially more severe cases of model risk may be envisaged as well, such as when the true process does not follow a lognormal process but rather one that has heavier tails or is time-dependent.

Derivatives practitioners are well aware of the fact that their models are not entirely correct and try to adjust the models to market conditions. For example, having to deal with the constant volatility assumption in the Black-Scholes model, while knowing very well that the true value of volatility is uncertain, traders often use implied volatilities as model input. Usually there are numerous options trading on the same underlying with different strikes resulting in multiple implied volatilities contradicting the Black-Scholes assumption. Despite the apparent inconsistency of the Black-Scholes model with market prices, traders continue to use it trying to adjust it using rules of thumb based on market knowledge.

Risk managers, who are further removed from the market than traders, are usually less capable of making these subjective corrections necessary to get the models to work. Therefore, model risk is harder to grasp for risk managers as more traditional sources of risk such as market risk, credit risk, legal risk, etc. A solution is to set model reserves for trading desks. Ideally, these model reserves depend on the market and the product traded, as some markets and/or products are more easily reliably modeled than others.

Though most prominent in derivatives markets, model risk is certainly not restricted to derivatives. Risk managers prefer to use (downside) risk measures to estimate the risk of a portfolio (possibly, but not necessarily containing derivatives). They employ models for describing the dynamics of the portfolios in order to compute these risk measures. However, if the models they employ do not accurately describe the underlying dynamics this results in an opaque estimate of the risk.

The need for an accurate estimate of the risk profile of a bank is first of all important to the banks itself, but second to central banks regulating the financial system. The Bank for International settlements (BIS) puts great emphasis in its Basel Accord (and its upcoming sequel, Basel II) on the accurate risk representations. For example, capital requirements are based on the results of a backtest of the model of a bank used to predict its value-at-risk. It is the purpose of this thesis to investigate model risk in general and related issues such as accuracy testing of risk management models and option pricing models.

1.2 Overview and contribution of thesis

The first part of the thesis investigates model risk issues in risk management. In *Chapter 2* we present a framework for backtesting all currently popular risk measurement methods for quantifying market risk (including value-at-risk and expected shortfall) using the functional delta method. Estimation risk can be taken explicitly into account. Based on a simulation study we provide evidence that tests for expected shortfall with acceptable low levels have a better performance than tests for value-at-risk in realistic financial sample sizes. We propose a way to determine multiplication factors, and find that the resulting regulatory capital scheme using expected shortfall compares favorably to the current Basel Accord backtesting scheme.

We test several risk management models for computing expected shortfall for one-period hedge errors of hedged derivatives positions in *Chapter 3*. Contrary to value-at-risk, expected shortfall cannot be tested using a standard binomial test, since we need information about the distribution in the tail. As derivatives positions change characteristics and thereby the size of risk exposures over time changes as well, one cannot apply the standard tests based on stationarity. To overcome this problem, we present a transformation procedure. For comparison purposes the tests are also performed for value-at-risk.

Chapter 4 proposes a general framework for quantification of model risk. This framework allows one to allocate regulatory capital to positions in a given market depending on the extent to which this market can be reliably modelled. Our approach is based on computing worst-case risk measures over sets of models that are in some appropriate sense close to a nominal model. The method is general in the sense that it can be applied with any of the usual risk measures such as value-at-risk and expected shortfall. In as far as risk measures can also be used as pricing tools or as determinants of margin requirements, the chapter provides a quantification of model risk in these settings as well. We present an application to stock portfolios and find that for usually applied specifications misspecification risk is much more important than estimation risk.

In the early days of option pricing plain vanilla options were not traded very frequently and, therefore, their pricing posed a modelling challenge. Currently, the market for (short-term) plain vanilla options is so liquid that the pricing of them does not require much modelling. A simple interpolation of the implied volatility surface would already give a reasonable price. Therefore, the pricing model risk is negligible. However, modelling remains crucial for hedging. Depending on the hedge strategy used the risk profile of a derivative can take (very) different forms (see Green and Figlewski (1999). The issue of hedging model risk for derivatives portfolios is addressed in *Chapter 5*. We empirically investigate the S&P 500 market and the \$/£, \$/¥, and £/¥ foreign exchange markets. Furthermore, we advocate the bootstrap as an alternative to the historical simulation method to determine estimation and misspecification risk. We find that in our samples estimation risk and misspecification risk are considerable and often significant. Furthermore, we find that in the S&P 500 market a risk premium seems to be demanded for bearing these risks while this is not the case in the FX markets investigated.

The second part of the thesis deals with pricing interest rate derivatives. In *Chapter 6* we show that discrete string models are observationally equivalent to market models. We also derive the parsimony of the models. As a consequence of the observational equivalence discrete string models are a special case of the HJM framework. The discrete string models can be estimated/calibrated using principal components analysis in the same manner as the market models.

In *Chapter 7* we investigate factor dependence and estimation risk for commonly traded exotic interest rate derivatives. We employ the popular Libor market model, and suggest the (stationary) bootstrap method to compute the estimation risk for exotic products. We find that autocaps, sticky caps, and ratchet caps are sensitive to the number of factors used and have considerable estimation risk.

Chapter 8 presents the main conclusions and directions for further research.

Part I

Model risk issues in risk management

Chapter 2

Backtesting for Risk-Based Regulatory Capital

2.1 Introduction

Regulators face the important but difficult task of determining appropriate capital requirements for regulated banks. Such capital requirements should protect the banks against adverse market conditions and prevent them from taking extraordinary risks (where, in this paper, we focus on market risk). At the same time, regulators should not prevent banks from practicing one of their core businesses, namely trading risk. The crucial ingredients in the process of risk based capital requirement determination are the use of a risk measurement method (to quantify market risk), a backtesting procedure, and multiplication factors, based on the outcomes of the backtesting procedure. Regulators apply multiplication factors to the risk measurement method they use in order to determine the capital requirements. The multiplication factors depend on the backtesting results, where a bad performance of the risk measurement method results in a higher multiplication factor. Consequently, to guarantee an appropriate process of capital requirement determination, regulators need an accurate backtesting procedure, combined with a suitable way of determining multiplication factors. Based on these requirements the regulators will assign the risk measurement method.

Since its introduction in the 1996 amendment to the Basel Accord (see Basel

Committee on Banking Supervision (1996a) and Basel Committee on Banking Supervision (1996b)) the value-at-risk has become the standard risk measurement method. However, although the value-at-risk may be interesting from a practical point of view, it has a serious drawback: it does not necessarily satisfy the property of subadditivity, which means that one can find examples where the value-at-risk of a portfolio as a whole is higher than that of the sum of the value-at-risks of its mutually exclusive sub-portfolios. An alternative, practically viable risk measurement method that satisfies the subadditivity property (and other desirable properties [1]) is the expected shortfall. Currently, a debate is going on whether the use of expected shortfall should be recommended in Basel II. So far, it is not in Basel II due to the expected difficulties concerning backtesting (see Yamai and Yoshiba (2002b)). Thus, although the value-at-risk does not necessarily satisfy the subadditivity property, it is still prescribed by regulators, because of its perceived superior performance in case of backtesting.

Both the value-at-risk and the expected shortfall (as well as many other risk measurement methods) are level-based methods, meaning that one first has to choose a level; given this level, the risk depends on the corresponding left-hand tail of the profit and loss distribution. For the value-at-risk the Basel Committee chooses a level of 0.01, meaning that the value-at-risk is based on the 1% quantile of the profit and loss distribution. For the sake of comparison, one might be tempted to choose the same level for alternative risk measurement methods, like the expected shortfall, so that they are calculated based on the same left-hand tail of the profit and loss distribution. When the level in both cases equals 0.01 it seems obvious to expect that backtesting expected shortfall will be much harder than backtesting the value-at-risk, even without trying it out. However, comparing alternative risk measurement methods by equating their levels does not seem to be appropriate from the viewpoint of capital reserve determination. From that perspective it seems much better to choose the levels such that the risk measurement methods result in (more or less) the same quantiles of the profit and loss distribution. The 0.01-level of value-at-risk will then correspond to a higher level in case of the expected shortfall. But then it is no longer clear which method will perform better in backtesting. It

[1] Namely, translation invariance, monotonicity, and positive homogeneity. These three properties are also satisfied by value-at-risk.

is the aim of this paper to make this comparison.

The contribution of the paper is threefold. First, we provide a general backtesting procedure for a large class of risk measurement methods, which contains all major risk measurement methods used nowadays. In particular, as a result a test for expected shortfall is derived which appears to be new in the literature. Using the functional delta method we provide a framework that requires the regulator only to determine the influence function of the risk measurement method in order to determine the critical levels of the capital requirements table. We show that the present backtesting methodology in the Basel Accord is a special case. Furthermore, a simple method to incorporate estimation risk is presented. The fact that banks have time-varying portfolio sizes and risk exposures complicates the use of standard statistical techniques. We deal with this issue using a standardization procedure based on the probability integral transform also used by Diebold, Gunther and Tay (1998) and Berkowitz (2001). The key idea of the standardization procedure is that banks should not only report whether or not the realized profit/loss is beyond the value-at-risk, but also which quantile of the predicted profit and loss distribution is realized. Second, we establish, via simulation experiments, that backtests for expected shortfall have a more promising performance than for value-at-risk, when the comparison is based on (more or less) equal quantiles instead of equal levels. In this way we provide evidence for a viable risk based regulatory capital scheme using expected shortfall with good backtesting properties. Finally, we suggest a general method to determine multiplication factors for the risk measurement methods using the backtest procedure developed.

The setup of the paper is as follows. In Section 2.2 we review the most popular risk measurement methods in current quantitative risk management. In Section 2.3 we present the standardization procedure in order to take account of the time-varying portfolio sizes and risk exposures. Section 2.4 treats the backtesting of the Basel Accord, its generalization using the functional delta method, and the incorporation of estimation risk. Simulation experiments are presented in Section 2.5. In Section 2.6 a suggestion for determination of multiplication factors is given. Finally, Section 2.7 concludes.

2.2 Risk measurement methods

2.2.1 Definitions and notation

Though risk profiles contain much relevant information for risk managers, they become unmanageable for large firms with many divisions and portfolios. Therefore, for risk management purposes, risk managers prefer low dimensional characteristics of the risk profiles. In order to compute these low dimensional characteristics they use a financial model $m = (\Omega, \mathbb{P})$, where Ω denotes the states of the world, and $\mathbb{P}$ the postulated probability distribution.[2] A risk is defined as follows.[3]

Definition 2.1 Let a financial model m be given. A *risk* defined on m is an element of $\mathcal{R}(m)$, the set of random variables defined on Ω.

This definition, in which a "risk" is a random variable, follows the terminology of Artzner, Delbaen, Eber and Heath (1999) and Delbaen (2000). Artzner et al. (1999) defined a risk measure for a particular financial model.

Definition 2.2 Let a financial model m be given. A *risk measure*, ρ, defined on m is a map from $\mathcal{R}(m)$ to $I\!R \cup \{\infty\}$.[4]

In order to allow for several financial models, we use a class of financial models denoted by $\mathcal{M}$. Each of these models defines a set of risks $\mathcal{R}(m)$. Following Kerkhof, Melenberg and Schumacher (2002) we denote a mapping defined on $\mathcal{M}$ that assigns a risk measure defined on m for each $m \in \mathcal{M}$ by a *risk measurement method defined on* $\mathcal{M}$, RMM. The most well-known risk measurement method nowadays is the value-at-risk method which was supported by the Basel Committee in the 1996 amendment to the Basel Accord (see Basel Committee on Banking Supervision (1996a)).

Before coming to the formal definitions of the popular risk measurement methods we present the quantile definitions.

Definition 2.3 (Quantiles) Let $X \in \mathcal{R}(m)$ be a risk for model $m = (\Omega, \mathbb{P})$.

[2]Formally, a model is defined by $m = (\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ is the information available.

[3]Formally, $\mathcal{R}(m)$ is defined as the space of all equivalence classes of real-valued measurable functions on $(\Omega, \mathcal{F})$.

[4]Including ∞ allows risks to be defined on more general probability spaces, see Delbaen (2000).

1. $Q_p(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq p\}$ is the lower p -quantile of X.

2. $Q^p(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leq x) > p\}$ is the upper p -quantile of X.

The definition of the value-at-risk method can then be given by

Definition 2.4 The *value-at-risk* method with *reference asset* N and *level* $p \in (0,1)$ assigns to a model $m = (\Omega, \mathbb{P})$ the risk measure VaR_m^p given by

$$\mathrm{VaR}_m^p : \mathcal{R}(m) \ni X \mapsto -Q^p(X/N_m) = Q_{1-p}(-X/N_m) \in I\!R \cup \{\infty\}, \qquad (2.1)$$

where N_m denotes the reference asset in model m.

We use a reference asset N (for example, the money market account) to measure the losses in terms of money lost relative to the reference asset. This allows comparison of risk measures for different time horizons.

Since the introduction of value-at-risk by RiskMetrics (1996), the literature on value-at-risk has surged (see, for example, Risk Magazine (1996), Duffie and Pan (1997), and Jorion (2000) for overviews). Though value-at-risk is an intuitive risk measure, the reasoning behind it was more practical than theoretically grounded. Recently, Artzner, Delbaen, Eber and Heath (1997) introduced the notion of coherent risk measures having the properties of translation invariance, monotonicity, positive homogeneity, and subadditivity. Their ideas were formalized in Artzner et al. (1999) and Delbaen (2000), amongst others. The value-at-risk method does not necessarily satisfy the relevant subadditivity property. This means that we can find examples where the value-at-risk of a portfolio is higher than that of the sum of the value-at-risks of a set of mutually exclusive sub-portfolios (see, for example, Artzner et al. (1999), Acerbi and Tasche (2002), and Tasche (2002)). A practically usable coherent risk measure is the expected shortfall as given in Acerbi and Tasche (2002).

Definition 2.5 The *expected shortfall method* with *reference asset* N and *level*

$p \in (0,1)$ assigns to a model $m = (\Omega, \mathbb{P})$ the risk measure ES_m given by

$$\mathrm{ES}_m : \mathcal{R}(m) \ni X \mapsto -\frac{1}{p}\left(\mathbb{E}X\mathbf{I}_{(-\infty, Q_p(X/N_m)]}\right.$$
$$\left. + Q_p\left(X/N_m\right)\left(p - \mathbb{P}\left(X/N_m \leq Q_p\left(X/N_m\right)\right)\right)\right) \in I\!R \cup \{\infty\}. \quad (2.2)$$

In case that $p = \mathbb{P}\left(X/N_m \leq Q_p\left(X/N_m\right)\right)$, the expected shortfall equals[5]

$$\mathrm{ES}_m(X) = -\frac{1}{p}\mathbb{E}\left[X\mathbf{I}_{(-\infty, Q_p(X/N_m)]}\right] = \mathbb{E}\left[X \mid X \leq Q_p\left(X/N_m\right)\right]. \quad (2.3)$$

Thus, informally, value-at-risk gives "the *minimum potential loss* for the worst $100p$ % cases"[6] while expected shortfall gives the "*expected potential loss* for the worst $100p$ % cases". Therefore, the expected shortfall takes the magnitude of the exceedance of the value-at-risk into account, while for value-at-risk the magnitude of exceedance is irrelevant.

2.2.2 Which levels?

Both the value-at-risk and the expected shortfall risk measurement method are defined for arbitrary levels $p \in (0,1)$. This leaves the issue of the choice of p open. Since we are interested in protecting against adverse market conditions it is clear that p should be chosen small. But how small? For value-at-risk the most common choices are $p = 0.05$ or $p = 0.01$ (the level chosen by the Basel Committee). In combination with the current multiplication factors used by the Basel Committee, the 1% value-at-risk results in more or less satisfactory capital reserves. In order to get a risk based capital reserve scheme based on expected shortfall, we need to determine a level p for the expected shortfall. In most comparisons between value-at-risk and expected shortfall their levels are taken to be equal. This seems to lead to the general opinion that, although expected shortfall has nice theoretical properties, it is much harder to backtest than value-at-risk (see Yamai and Yoshiba (2002b)), the

[5]The additional term $Q_p\left(X/N_m\right)\left(p - \mathbb{P}\left(X/N_m \leq Q_p\left(X/N_m\right)\right)\right)$ is needed in order to make the expected shortfall coherent, see Acerbi and Tasche (2002).

[6]Most value-at-risk devotees prefer the alternative formulation of "the maximum loss in the 100(1-p)% best cases."

main reason why expected shortfall is still absent in Basel II.[7] However, for capital reserve determination it seems to make sense to look at comparable quantiles instead of levels. For example, take the median shortfall, that is, take the median in the tail instead of the expectation. The median shortfall with level $2p$ corresponds to value-at-risk with level p. If we would compare the backtest results of the median shortfall and the value-at-risk with the same level, we probably find that value-at-risk has a better performance than median shortfall. But for a valid comparison, we should use the median shortfall with twice the level of value-at-risk, in which case we find equal performance. A similar reasoning applies to expected shortfall. In order to have a valid comparison of the backtest results we should look at the quantiles and not the levels. Doing this for the Gaussian distribution (as a reference distribution), we find $p = 0.025$ for the expected shortfall when $p = 0.01$ for value-at-risk.[8] In case of excess kurtosis we need to take a higher level for the expected shortfall for it to equal the 1% value-at-risk. Since, in practice, we usually encounter distributions with heavier tails than the Gaussian distribution, the level of 2.5% can be seen as a lower bound on the level for equal capital requirement.

2.3 Standardization procedure

Let $(h_t)_{t\in\mathcal{T}_T}$ with $\mathcal{T}_T = \{1, ..., T\}$ (the test period) be a time-series of (in our case daily) returns on a profit and loss account (P&L) of a bank. Usually, the sequence $(h_t)_{t\in\mathcal{T}_T}$ cannot be modelled appropriately as a sample from one single distribution, say F, due to the fact that banks change the composition of their portfolio frequently. In general, the risk profile (the distribution of the P&L) of the bank changes over time. Therefore, we allow $(h_t)_{t\in\mathcal{T}_T}$ to be drawn from a different (marginal) distribution each period, that is,

$$h_t \sim F_t \quad t \in \mathcal{T}_T. \tag{2.4}$$

A bank is required to report the riskiness of its portfolio every day by means of

[7] We thank Jon Danielsson for pointing this out to us.

[8] Notice that for the value-at-risk at level $p = 0.01$ we have $-\Phi^{-1}(0.01) = 2.33$, while for the expected shortfall at level $p = 0.025$ we have $\Phi^{-1}(0.025) = -1.96$ and $-\mathbb{E}[X|X < -1.96] = \phi(-1.96)/\Phi(-1.96) = 2.34$ (see (2.3)), when X follows a standard normal distribution (where ϕ and Φ denote the density and distribution function of the standard normal distribution, respectively).

a risk measure $\rho(h_t)$, where $\rho(h_t)$ denotes the risk measure for period t using the information up to time $t-1$.[9] In order to compute these risk measures the bank uses a sequence of forecast distributions $(P_t)_{t\in\mathcal{T}_T}$, with corresponding densities $(p_t)_{t\in\mathcal{T}_T}$.

Often F_t is assumed to belong to a location-scale family; that is, it is assumed that the sequence $\{(h_t-\mu_t)/\sigma_t\}_{t\in\mathcal{T}_T}$ is identically distributed (see, for example, McNeil and Frey (2000) and Christoffersen, Hahn and Inoue (2001)). However, this restricts the way in which the procedure takes portfolio changes of banks into account. In this set-up moments higher than two are only allowed to vary over time through the first two moments. More generally, we can use the probability integral transform (see, for example, Van der Vaart (1998)) to go from a non-identically distributed sequence $(h_t)_{t\in\mathcal{T}_T}$ to an identically distributed sequence $(y_t)_{t\in\mathcal{T}_T}$. This transform is defined as

$$y_t = G^{-1}\left(\int_{-\infty}^{h_t} p_t(u)\,du\right) = G^{-1}(P_t(h_t)), \qquad t\in\mathcal{T}_T, \tag{2.5}$$

In case $P_t = F_t$ for each $t\in\mathcal{T}_T$, the distribution of y_t equals G, otherwise, the distribution of y_t is equal to, say, Q_t, unequal to G (for at least one time period t). The following lemma (see special cases in Diebold et al. (1998) and Berkowitz (2001)) gives the density q_t of y_t.

Lemma 2.1 Let $f_t(\cdot)$ denote the density of h_t, $p_t(\cdot)$ the density corresponding to $P_t(\cdot)$, g the density associated with G, and $y_t = G^{-1}(P_t(h_t))$. If $\frac{dP_t^{-1}(G(y_t))}{dy_t}$ is continuous and nonzero over the support of h_t, y_t has the following density:

$$\begin{aligned} q_t(y_t) &= \left|\frac{dG^{-1}(P_t(h_t))}{dh_t}\right|^{-1} f_t(h_t) \\ &= g(y_t)\frac{p_t(h_t)}{f_t(h_t)}. \end{aligned} \tag{2.6}$$

Proof. Just apply the change of variables transformation to $y_t = G^{-1}(P_t(h_t))$ and the result follows. ∎

In case the forecast distributions of the bank are correct, i.e., $P_t = F_t$, $t\in\mathcal{T}_T$, we have that $q_t(y_t) = g(y_t)$. Thus, under the hypothesis that $P_t = F_t$, $t\in\mathcal{T}_T$ we can

[9] It would be more appropriate to write $\rho_{t-1}(h_t)$, but we suppress the subscripts for notational convenience.

go from a non-identically distributed sequence $(h_t)_{t\in\mathcal{T}_T}$ to an identically distributed sequence $(y_t)_{t\in\mathcal{T}_T}$ with distribution G. We denote this procedure as *standardization to G*. For example, Berkowitz (2001), uses $G = \Phi$, the standard normal distribution, in order to use the Gaussian likelihood for his Likelihood Ratio tests.[10]

2.4 Backtest procedure

After assigning a risk measurement method the regulator faces the important task of determining the quality of the models that the regulated banks use in order to compute the risk measure. One of the reasons that the value-at-risk approach is often preferred to the coherent risk measures is the fact that the quality of value-at-risk models seems more easily verifiable. Therefore, the choice of risk measurement method by the regulator is based on the tools available to the regulator to verify model quality. In order to motivate the regulated institutions to improve their models, regulators often impose model reserves or multiplication factors (see, for example, the multiplication factors by the Basel Committee). In Section 2.4.1 we review the backtest procedure of the Basel Committee. Then we provide an alternative and more general procedure, in Section 2.4.2 ignoring estimation risk, and in Section 2.4.3 taking estimation risk into account.

2.4.1 Backtest procedure of Basel Committee

In this section we briefly describe the backtest procedure used by the BIS for determining the multiplication factors for capital requirements. A full exposition can be found in the Basel Committee on Banking Supervision (1996b).

Banks need to produce T ($T = 250$ in the current BIS implementation) value-at-risk forecasts (1% value-at-risk in the current BIS implementation) $(\mathrm{VaR}_t)_{t\in\mathcal{T}_T}$, where VaR_t denotes the value-at-risk forecast for day t using the information up to time $t-1$. It is assumed that these value-at-risk forecasts $(\mathrm{VaR}_t)_{t\in\mathcal{T}_T}$ are such that the exceedances sequence $(e_t)_{t\in\mathcal{T}_T}$ consists of independent elements with a Bernoulli distribution with probability p, that is, $\mathrm{Bern}(p)$, where p denotes the quantile rel-

[10] Notice, however, that when $P_t \neq F_t$, for at least one $t \in \mathcal{T}_T$, the standardization procedure will result in distributions Q_t, not necessarily equal for different $t \in \mathcal{T}_T$.

Table 2.1: **BIS multiplication factors**
The table shows the plus factors (multiplication factor = 3 + plus factor) used by the BIS for capital requirements based on a sample of 250. Tables for other sample sizes can be constructed by letting the yellow zone start when the cumulative probability exceeds 95% and the red zone when it exceeds 99.99%.

zone	Number of exceedances	Plus factor	Cumulative probability
green zone	0	0,00	8,11
	1	0,00	28,58
	2	0,00	54,32
	3	0,00	75,81
	4	0,00	89,22
yellow zone	5	0,40	95,88
	6	0,50	98,63
	7	0,65	99,60
	8	0,75	99,89
	9	0,85	99,97
red zone	≥ 10	1,00	99,99

evant to the value-at-risk method employed. The exceedances $(e_t)_{t\in\mathcal{T}_T}$ are defined by

$$e_t = \mathbf{I}_{(-\infty,-\mathrm{VaR}_t)}\left(h_t\right), \quad t \in \mathcal{T}_T. \tag{2.7}$$

By definition we have that

$$\mathbb{P}\left(e_t = 1\right) = \mathbb{P}\left(h_t < -\mathrm{VaR}_t\right), \quad t \in \mathcal{T}_T. \tag{2.8}$$

If $-\mathrm{VaR}_t = F_t^{-1}(p)$, with F the cumulative distribution function of h_t, we have that $\mathbb{P}(e_t = 1) = p$ and, consequently, the distribution of e_t indeed follows a Bernoulli-distribution. Using the cumulative distribution of the binomial distribution one may then compute multiplication factors based on the number of exceedances. For completeness, we present Table 2 from Basel Committee on Banking Supervision (1996b) in Table 2.1.

The capital requirement can then be computed as the product of the value-at-risk at time t, $\mathrm{VaR}_t^{0.01}$, multiplied by a multiplication factor, mf_t, that is determined by the results of a backtest of model m on the previous T ($T = 250$ in Basel Accord) days,[11]

$$\mathrm{CR}_t = \mathrm{mf}_t \cdot \mathrm{VaR}_t^{0.01}. \tag{2.9}$$

The backtest procedure given by the Basel Committee described above has some serious shortcomings. It assumes that under the null hypothesis the exceedances $(e_t)_{t=1}^T$ are *i.i.d.* while empirical evidence shows a clustering phenomenon in the exceedances (see, for example, Berkowitz and O'Brien (2002)). However, in case of dependence, one could adapt the test procedure by applying, for instance, the Newey-West (1987) approach which allows for quite general forms of dependence over time. Another drawback is that the above procedure does not take estimation risk into account which manifests itself in the fact that $\mathrm{VaR}_t = \widehat{F}_t^{-1}(p)$ which is not necessarily equal to $F_t^{-1}(p)$. Due to the limited amount of data there is likely some inaccuracy in the estimate for the value-at-risk which in effect causes an estimation error in the exceedances (compare West (1996)). This issue is treated in Section 2.4.3. A final drawback is that by transforming the information of the distribution into one characteristic (exceeding of value-at-risk or not) we lose relevant information of the return distribution (see also Berkowitz (2001)). In Section 2.5 we see that the power of the test is affected by removing this information.

2.4.2 General backtest procedure

We assume given a sample of transformed data $(y_t)_{t\in\mathcal{T}_T}$ to which the standardization procedure, described in Section 2.3 has been applied; this yields observations drawn from actual distributions Q_t, some or all possibly unequal to the postulated standardized distribution G. In this subsection we abstract from possible estimation risk in estimating the distribution function. This will be discussed in the next subsection.

The null hypothesis $H_0 : Q_t = G$ can be tested against numerous alternatives. We shall formulate these alternatives under the additional assumption of station-

[11] Actually, the used value-at-risk is $\max\{\mathrm{VaR}_t^{0.01}, \frac{1}{60}\sum_{i=1}^{60}\mathrm{VaR}_{t-i}^{0.01}\}$ instead of $\mathrm{VaR}_t^{0.01}$ (see Basel Committee on Banking Supervision (1996b)). Furthermore, the multiplication factors are set every 3 months.

arity, i.e., $Q_t = Q$.[12] For example, Berkowitz (2001) tests this hypothesis using a likelihood ratio (LR) test using the Gaussian likelihood ($H_1 : Q \neq G = \Phi$) and a censored Gaussian likelihood ($H_1 : Q_{(-\infty, Q^{-1}(p)]} \neq G_{(-\infty, G^{-1}(p)]}$).[13] Using the censored Gaussian likelihood has the advantage that it ignores model failures in the interior of the distribution: only the tail behavior matters.

Following this line of reasoning, we use risk measurement methods which focus by construction on the tail behavior to evaluate the null hypothesis. Our main concern is not conservatism, that is, the true risk $\varrho(Q)$ is smaller than or equal to $\varrho(G)$, the risk expected by our model. Since we do not want that the model underestimates the risk, the alternative is taken to be $H_1 : \varrho(Q) > \varrho(G)$.

In Section 2.2, we defined risk measurement methods as functions of random variables (defined on a financial model $m = (\Omega, \mathbb{P})$) following the quantitative risk measurement literature. For the purpose of testing it is more convenient to define the risk measurement method as a functional, $\varrho : D_F \to \mathbb{R}$, of a distribution function to $\mathbb{R} \cup \infty$.[14] Thus, $\text{RMM}_m(X) = \varrho(F)$ for risk X if F is the distribution function of X associated with model m.

If $\varrho : D_F \to \mathbb{R}$ is Hadamard differentiable on D_F, we can apply the functional delta method (see, for example, Van der Vaart (1998) Thm. 20.8)

$$\sqrt{T}\left(\varrho(Q_T) - \varrho(Q)\right) = \sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\psi_t(Q) + o_p(1), \qquad \mathbb{E}\psi_t(Q) = 0,\ \mathbb{E}\psi_t^2(Q) < \infty, \tag{2.10}$$

where Q_T denotes the empirical distribution of the random sample $(y_t)_{t \in \mathcal{T}_T}$ and $\psi_t(Q)$ denotes the influence function of the risk measurement method ϱ at observation t. As can easily be shown, the common risk measures such as value-at-risk and expected shortfall are Hadamard differentiable.[15] We can then use the following test

[12] When presenting the test statistics, we maintain this assumption and implicitly assume that this stationarity is transferred in the risk measures $\varrho(Q_t)$. Notice, however, the testing procedure is more generally applicable than just for the case of stationarity.

[13] For distribution function F, $F_{(-\infty, F^{-1}(p)]}$ denotes the left tail of the distribution up to the p^{th} quantile.

[14] D_F denotes the space of all distribution functions, that is, all non-decreasing cadlag functions F on $[-\infty, \infty]$ with $F(-\infty) \equiv \lim_{x \to -\infty} F(x) = 0$ and $F(\infty) \equiv \lim_{x \to \infty} F(x) = 1$. D_F is equipped with the metric induced by the supremum norm.

[15] For the value-at-risk, see, for example, Van der Vaart and Wellner (1996) Lemma 3.9.20. In case of the expected shortfall, the influence function is easily obtained by applying the chain rule

statistic:

$$S_T = \sqrt{T}\frac{(\varrho(Q_T) - \varrho(Q))}{\sqrt{V}} \xrightarrow[H_0]{d} \mathcal{N}(0,1), \tag{2.11}$$

with $V = \mathbb{E}\psi_t^2(Q)$ and $\varrho(Q)$ evaluated under the null hypothesis, $Q = G$.[16] Some important examples are:

Example 2.1 (Value-at-risk) In the case of value-at-risk written as a function of the distribution function

$$\varrho(Q) = -Q^{-1}(p), \tag{2.12}$$

the influence function $\psi(Q)$ is given by

$$\psi_{\text{VaR}}(Q) = -\frac{p - \mathbf{I}_{(-\infty, Q^{-1}(p)]}(x)}{q(Q^{-1}(p))}, \tag{2.13}$$

and

$$\mathbb{E}\psi_{\text{VaR}}^2(Q) = \frac{p(1-p)}{q^2(Q^{-1}(p))}. \tag{2.14}$$

This leads to the following test statistic

$$S_{\text{VaR}} = \sqrt{T} q\left(Q^{-1}(p)\right)\frac{(\varrho(Q_T) - \varrho(Q))}{\sqrt{p(1-p)}} \tag{2.15}$$

The critical value-at-risk levels for the yellow and red zones are given by

$$\begin{aligned} \text{VaR}_{yellow} &= \sqrt{\frac{z_{0.95}}{T}\frac{p(1-p)}{q^2(Q^{-1}(p))}} + \text{VaR}(Q) \\ \text{VaR}_{red} &= \sqrt{\frac{z_{0.9999}}{T}\frac{p(1-p)}{q^2(Q^{-1}(p))}} + \text{VaR}(Q), \end{aligned} \tag{2.16}$$

where z_p denotes the p^{th} quantile of the standard Gaussian distribution.

Example 2.2 (Exceedances) In the case of the number of exceedances written as

for Hadamard differentiable functions to the quantile function and the mean, see, for example, Van der Vaart and Wellner (1996) Lemma 3.9.3.

[16] Under the assumption of stationarity, i.e., $Q_t = Q$, we could also evaluate V under the alternative as $V = \frac{1}{T}\sum_{t=1}^{T}\left(\psi_t(Q_T) - \frac{1}{T}\sum_{t=1}^{T}\psi_t(Q_T)\right)^2$. However, our simulation study indicates a much worse performance of the test statistics using this estimate than when evaluating V under the null.

a function of the distribution function

$$\varrho(Q) = \mathbf{I}_{(-\infty, Q^{-1}(p)]}, \tag{2.17}$$

the influence function $\psi(Q)$ is given by

$$\psi_{\text{exc}}(Q) = p - \mathbf{I}_{(-\infty, Q^{-1}(p)]}(x), \tag{2.18}$$

and

$$\mathbb{E}\psi^2_{\text{exc}}(Q) = p(1-p). \tag{2.19}$$

This gives the following test

$$S_{\text{exc}} = \sqrt{T}\frac{(\varrho(Q_T) - \varrho(Q))}{\sqrt{p(1-p)}} \tag{2.20}$$

The critical numbers of exceedances for the yellow and red zones are given by

$$\begin{aligned} \text{Exc}_{yellow} &= \sqrt{z_{0.95}Tp(1-p)} + pT \\ \text{Exc}_{red} &= \sqrt{z_{0.9999}Tp(1-p)} + pT \end{aligned} \tag{2.21}$$

For the regular backtest size of 250, these critical values are equal to the exact setting of the binomial distribution used by the BIS.

Example 2.3 (Expected shortfall) In the case of ES written as a function of the distribution function

$$\varrho(Q) = -\int_{-\infty}^{Q^{-1}(p)} x dQ(x) + Q^{-1}(p)\left(p - \int_{-\infty}^{Q^{-1}(p)} dQ(x)\right), \tag{2.22}$$

the influence function $\psi(Q)$ is given by

$$\begin{aligned} \psi_{\text{ES}}(Q) &= -\frac{1}{p}\Big[\big(x - Q^{-1}(p)\big)\mathbf{I}_{(-\infty, Q^{-1}(p)]}(x) \\ &\quad + \psi_{\text{VaR}}(Q)\left(p - \int_{-\infty}^{Q^{-1}(p)} dQ(x)\right)\Big] - \text{ES}(Q) + \text{VaR}(Q) \end{aligned} \tag{2.23}$$

and

$$\begin{aligned} \mathbb{E}\psi_{\mathrm{ES}}^{2}(Q) &= \frac{1}{p}\mathbb{E}\left[X^{2}|X \le Q^{-1}(p)\right] - \mathrm{ES}(Q)^{2} \\ &+2\left(1-\frac{1}{p}\right)\mathrm{ES}(Q)\,\mathrm{VaR}(Q) - \left(1-\frac{1}{p}\right)\mathrm{VaR}(Q)^{2}. \end{aligned} \tag{2.24}$$

This leads to the following test statistic

$$S_{\mathrm{ES}} = \sqrt{T}\frac{\left(\varrho\left(Q_{T}\right)-\varrho\left(Q\right)\right)}{\sqrt{\mathbb{E}\psi_{\mathrm{ES}}^{2}(Q)}} \tag{2.25}$$

The critical ES levels for the yellow and red zones are given by

$$\begin{aligned} \mathrm{ES}_{yellow} &= \sqrt{\frac{z_{0.95}}{T}\mathbb{E}\psi_{\mathrm{ES}}^{2}(Q)^{2}} + \mathrm{ES}(Q) \\ \mathrm{ES}_{red} &= \sqrt{\frac{z_{0.9999}}{T}\mathbb{E}\psi_{\mathrm{ES}}^{2}(Q)^{2}} + \mathrm{ES}(Q) \end{aligned} \tag{2.26}$$

We conclude this subsection by illustrating that the test statistics can easily be implemented for the Gaussian case $G = \Phi$, by presenting the outcomes of $\mathbb{E}\psi_{t}^{2}(G)$ in case of value-at-risk and expected shortfall. For this, let $\phi(x)$ denote the density function of the standard Gaussian $\mathcal{N}(0,1)$ distribution and z_p the p^{th} quantile of the standard normal distribution. The value-at-risk in case of a normal distribution $\mathcal{N}(0,1)$ is given by

$$\mathrm{VaR}_{p}(X) = z_{p}, \tag{2.27}$$

and the expected shortfall is given by

$$\mathrm{ES}_{p}(X) = -\phi(z_{p})/p. \tag{2.28}$$

$\mathbb{E}\psi_{t}^{2}(\Phi)$ for value-at-risk and expected shortfall are then given by,

$$\mathbb{E}\psi_{t}^{2}(\Phi) = \frac{p(1-p)}{\phi(z_{p})},$$

for value-at-risk and

$$\mathbb{E}\psi_t^2(\Phi) = 1 - z_p\frac{\phi(z_p)}{p} - \left(\frac{\phi(z_p)}{p}\right)^2 - 2\left(1-\frac{1}{p}\right)\frac{\phi(z_p)}{p}z_p - \left(1-\frac{1}{p}\right)z_p^2,$$

for expected shortfall.

2.4.3 Estimation risk

The backtesting procedures described in this section assume that the forecasted distributions $(P_t)_{t\in\mathcal{T}_T}$ of the profit/loss are given. It seems natural to penalize banks with a plus factor for using inappropriate model families, but not for just having to estimate a correctly specified model (assuming that they use their data efficiently). In order to do so, we derive in this section backtest procedures that take estimation risk into account.

Again, we use the standardization procedure described in Section 2.3. We assume given a random estimation sample $(y_t)_{t\in\mathcal{T}_e}$, $\mathcal{T}_e = \{-N+1,...,0\}$, and a random testing sample $(y_t)_{t\in\mathcal{T}_T}$ $\mathcal{T}_T = \{1,...,T\}$ with $y_t \sim Q$ ($Q = G$ under the null). We then have

$$\sqrt{n}\left(\varrho(Q_n) - \varrho(Q)\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\psi^2(Q)\right), \qquad n = T, N$$

where $\psi(\cdot)$ is the influence function of $\varrho(\cdot)$. This yields (still under the null)

$$\begin{aligned}\sqrt{T}\left(\varrho(Q_T) - \varrho(Q_N)\right) &= \sqrt{T}\left(\varrho(Q_T) - \varrho(Q)\right) - \sqrt{\frac{T}{N}}\sqrt{N}\left(\varrho(Q_N) - \varrho(Q)\right) \\ &\xrightarrow{d} \mathcal{N}\left(0, (1+c)\,\mathbb{E}\psi^2(G)\right), \end{aligned} \tag{2.29}$$

when $\frac{T}{N} \to c$ as $N \to \infty$ and $T \to \infty$.

If the estimation period would grow with time, c would tend to zero. In practice, one usually specifies a finite fixed estimation period (for example, 2 years) and computes the risk measure based on this estimation period. This is a so-called rolling window estimation procedure, which can be approximated in our setting by taking $c = \frac{T}{N}$ in (2.29).

For the examples in 2.4.2 we can derive the critical values for the yellow and red zones in the same way by replacing V by $(1+c)V$. With the incorporation of

Table 2.2: **Simulation results for size of tests**
This table presents the Type I errors (in percentages) if $F_t = P_t = \mathcal{N}(0,1)$ for $t \in T_T$ for $T = 125$, 250, 500, and 1000. The argument H_0 indicates that the variance used is $\mathbb{E}\psi_t^2(G)$ and H_1 indicates that the variance used is $V = \frac{1}{T}\sum_{t=1}^{T}\left(\psi_t(Q) - \frac{1}{T}\sum_{t=1}^{T}\psi_t(Q)\right)^2$. $\text{Tail}_{0.025}$ signifies Berkowitz tail test. The number of simulations equals 10,000.

T	Exceedances	$\text{VaR}_{0.01}(H_0)$	$\text{VaR}_{0.01}(H_1)$	$\text{ES}_{0.025}(H_0)$	$\text{ES}_{0.025}(H_1)$	$\text{Tail}_{0.025}$
125	3.75	2.75	1.81	2.64	3.24	3.05
250	4.17	4.81	2.87	5.14	4.64	5.42
500	6.63	2.91	2.27	9.38	8.10	5.16
1000	4.51	3.87	2.98	4.34	2.63	5.33

estimation risk in the backtesting procedure we introduce an additional degree of freedom for the regulator, namely the choice of c (or N, since T could already be chosen by the regulator).

2.5 Simulation results

In this section we compare the finite sample behaviors of the backtest procedures. First, we determine the actual size of the tests for the exceedances ratio, value-at-risk, and expected shortfall. For simplicity, we take $F_t = \mathcal{N}(0,1)$, the standard normal distribution, for $t \in \mathcal{T}_T$. To check the performance of the tests for size, we take $P_t = F_t$, $t \in \mathcal{T}_T$, and set the significance level $\alpha = 0.05$. We verify the performance of the tests given in the examples in Section 2.4.2 using $G = \Phi$, the standard normal distribution function.[17] The tests are compared to the censored LR test of Berkowitz (2001), which we refer to as the Berkowitz tail test. Table 2.2 shows the results of the performance of the size of the tests. We see that the size for the three tests (Exceedances, value-at-risk, and expected shortfall) seem reasonable for the common sample size of 250. The Berkowitz tail test seems to converge a bit faster.

[17] Using $G = U[0,1]$ results in very poor results for smaller sample sizes. The reason is that by transforming the data to uniform random numbers the symmetry in the test is lost due to the non-linear shape of F.

Next, we investigate the power of the different tests. In practice, financial time series often exhibit excess kurtosis with respect to the normal distribution and have longer left tails. We consider three alternatives that replicate (parts of) this behavior. First, we use the student $t-$distribution with 5 degrees of freedom, that is, $F_t = t_5$. This distribution has heavier tails than the normal distribution, but is still symmetric. Second, we use two alternatives from the Normal Inverse Gaussian (NIG) family.[18] The NIG distribution allows one to control both the level of excess kurtosis and the skewness. We consider two cases: a symmetric case with a moderately high kurtosis, $\beta = 0,\ \alpha = \sqrt{\beta^2+1},\ \delta = 1/\left(1+\beta^2\right),\ \mu = 0$ and a case where the distribution is very skewed to the left and has a large kurtosis, $\beta = -0.25,\ \alpha = \sqrt{\beta^2+1},\ \delta = 1/\left(1+\beta^2\right),\ \mu = 0$. Third, we take a GARCH(1,1)-process,[19] with parameter values $\omega = 0.05$, $\gamma_1 = 0.25$, and $\gamma_2 = 0.7$ to allow for a time-dependent distribution under the alternative hypothesis. For the time-independent cases we present the results for VaR and ES with the test statistic estimated under the null as well as under the alternative (see footnote 16). Table 2.3 contains the results. We see that in case of a time-independent alternative for both the value-at-risk and the expected shortfall, the tests with variance evaluated under the null hypothesis have (far) more power. The difference with the test using the estimated variance under the alternative narrows when the sample size increases. The test for expected shortfall performs best in detecting the misspecification, also when the alternative is GARCH$(1,1)$ for $T \geq 250$; the number-of-exceedances test has less power than the value-at-risk test and the expected shortfall test. The Berkowitz tail test also performs well and, therefore, seems a worthwhile auxiliary test, but, in

[18]The density of the $NIG\left(\alpha,\beta,\mu,\delta\right)$ is given by

$$f_{NIG}\left(x\right) = \frac{\alpha \exp\left(\delta\sqrt{\alpha^2-\beta^2}-\beta\mu\right)}{\pi} q\left(\frac{x-\mu}{\delta}\right)^{-1} K_1\left\{\delta\alpha q\left(\frac{x-\mu}{\delta}\right)\right\}\exp\left\{\beta\left(x-\mu\right)\right\},$$

with $q\left(x\right) = \sqrt{1+x^2}$ and $K_1\left(x\right)$ the modified Bessel function of the third kind. See, for example, Barndorff-Nielsen (1996).

[19]The GARCH(1,1) model (see Bollerslev (1986)) is given by the following return and volatility equations:

$$\begin{aligned} r_t &= \sqrt{h_t}\epsilon_t \\ h_t &= \omega + \gamma_1 r_{t-1}^2 + \gamma_2 h_{t-1} \end{aligned}$$

Table 2.3: **Simulation results for power of tests**

This table presents the Type II errors (in percentages) if $F_t = t_5$, $F_t = NIG(\alpha, 0, \delta, \mu)$, $F_t = NIG(\alpha, -0.25, \delta, \mu)$, and $F_t = N(0, \sigma_t^2)$ (GARCH(1,1)) ; $\alpha = \sqrt{\beta^2 + 1}$, $\delta = 1/(1+\beta^2)$, $\mu = 0$. σ_t^2 follows the volatility equation of a GARCH(1,1) model with $\omega = 0.05$, $\gamma_1 = 0.25$, and $\gamma_2 = 0.7$. $P_t = N(0,1)$ for $t \in \mathcal{T}_T$ for $T = 125$, 250, 500, and 1000. The number of simulations equals 10,000.

T	Exceedances	$\text{VaR}_{0.01}(H_0)$	$\text{VaR}_{0.01}(H_1)$	$\text{ES}_{0.025}(H_0)$	$\text{ES}_{0.025}(H_1)$	$\text{Tail}_{0.025}$
t_5						
125	11.72	22.44	10.41	26.77	6.73	20.51
250	17.64	35.98	14.98	45.65	14.22	42.43
500	32.86	38.57	17.54	69.86	35.93	63.13
1000	42.89	57.60	32.68	82.39	52.12	87.91
$NIG(\alpha, 0, \delta, \mu)$						
125	16.08	25.08	14.22	30.27	0.00	22.84
250	25.53	44.73	22.93	52.51	22.72	45.29
500	47.06	51.17	29.25	78.51	51.11	69.90
1000	63.32	74.38	53.43	90.13	71.44	91.41
$NIG(\alpha, -0.25, \delta, \mu)$						
125	33.94	45.81	31.03	54.26	21.41	41.52
250	52.97	71.94	47.48	81.00	48.41	72.54
500	83.40	85.53	67.25	97.15	87.42	92.96
1000	95.97	97.93	91.87	99.76	98.39	99.71
GARCH(1,1)						
125	11.08	11.60		13.66		17.63
250	14.45	20.49		24.02		19.23
500	24.17	20.10		40.66		25.78
1000	27.34	29.63		43.37		39.93

general, trails the test for expected shortfall. Especially for the shorter sample sizes the test for expected shortfall performs better with only GARCH$(1,1)$ for $T = 125$ as an exception.

Finally, we take estimation risk into account. In Table 2.4 the results are shown for an equal estimation and testing period. It gives the expected result that the longer the samples the better the power of the tests. However, the performance of the test for value-at-risk with the variance evaluated under the alternative (in the time-independent cases) is quite bad. In Table 2.5 we fixed the testing period at 1 year (250 days) and varied the estimation period. As expected the results improve for longer estimation periods. Again, the performance of the test for value-at-risk with the variance evaluated under the (time-independent) alternative is quite bad.

Concluding, we find that the performances of the tests with the variance evaluated under (a time-independent) H_0 have far more power than the tests with the variance evaluated under H_1 for sample sizes realistic for financial data. Furthermore, we find that the performance for the size of the tests of the 2.5% expected shortfall is about equal to the 1% value-at-risk. However, the power of the 2.5% expected shortfall test is much better than that of the 1% value-at-risk.

2.6 Multiplication factors

In this section we propose a method to compute multiplication factors for capital requirements determination. Our starting point is the test statistic (2.11). If the test statistic results in rejection of the null hypothesis, then we might conclude that $\varrho(G)$ is taken too low. The question then is by which multiplication factor $\varrho(G)$ at least should be increased, such that the test statistic does no longer result in rejection of the null. Let $\varrho^*(Q_T)$ the realized value of $\varrho(Q)$. Then the minimum multiplication factor, mf, for which the null hypothesis would not be rejected follows from setting (2.11) equal to k_α, the critical value of the test at the significance level α

$$\sqrt{T}\frac{(\varrho(Q_T) - \mathrm{mf}(s_T^*)\varrho(G))}{\sqrt{V}} = z_\alpha, \tag{2.30}$$

Table 2.4: **Simulation results for power of tests in case of estimation risk (testing period equals estimation period))**

This table presents the Type II errors (in percentages) if $F_t = t_5$, $F_t = NIG(\alpha, 0, \delta, \mu)$, $F_t = NIG(\alpha, -0.25, \delta, \mu)$, and $F_t = N(0, \sigma_t^2)$ (GARCH(1,1)) ; $\alpha = \sqrt{\beta^2 + 1}$, $\delta = 1/(1+\beta^2)$, $\mu = 0$. σ_t^2 follows the volatility equation of a GARCH(1,1) model with $\omega = 0.05$, $\gamma_1 = 0.25$, and $\gamma_2 = 0.7$. $P_t = N(0,1)$ for $t \in \mathcal{T}_T$ and $\mathcal{T}_T$ for $T = 125$, 250, 500, and 1000. The number of simulations equals 10,000.

$N = T$	Exceedances	$\text{VaR}_{0.01}(H_0)$	$\text{VaR}_{0.01}(H_1)$	$\text{ES}_{0.025}(H_0)$	$\text{ES}_{0.025}(H_1)$	$\text{Tail}_{0.025}$
t_5						
125	18.40	15.49	0.34	22.91	4.87	15.93
250	13.51	22.84	0.38	37.81	6.69	27.49
500	19.25	21.30	0.27	59.23	15.85	47.79
1000	28.50	30.91	1.40	72.24	23.92	74.94
$NIG(\alpha, 0, \delta, \mu)$						
125	21.85	17.07	0.23	24.79	6.66	15.84
250	18.11	25.00	0.38	41.62	10.68	26.15
500	27.89	26.99	0.63	66.16	25.62	48.37
1000	45.44	41.88	3.66	80.08	40.71	76.80
$NIG(\alpha, -0.25, \delta, \mu)$						
125	38.36	31.03	0.85	45.10	12.81	33.55
250	41.01	47.08	1.95	69.98	24.90	54.97
500	61.90	57.61	4.67	91.94	58.48	81.22
1000	86.61	81.47	20.17	98.71	85.74	97.86
GARCH(1,1)						
125	18.79	12.10		13.13		7.83
250	13.31	13.22		19.78		10.28
500	16.28	11.37		31.81		14.46
1000	20.30	13.81		32.28		21.09

Table 2.5: **Simulation results for power of tests in case of estimation risk (one year testing period)** This table presents the Type II errors (in percentages) if $F_t = t_5$, $F_t = NIG(\alpha, 0, \delta, \mu)$, $F_t = NIG(\alpha, -0.25, \delta, \mu)$, and $F_t = N(0, \sigma_t^2)$ (GARCH(1,1)) ; $\alpha = \sqrt{\beta^2 + 1}$, $\delta = 1/(1+\beta^2)$, $\mu = 0$. σ_t^2 follows the volatility equation of a GARCH(1,1) model with $\omega = 0.05$, $\gamma_1 = 0.25$, and $\gamma_2 = 0.7$. $P_t = N(0,1)$ for $t \in \mathcal{T}_T$ and $\mathcal{T}_T$ for $T = 125$, 250, 500, and 1000. The number of simulations equals 10,000.

(N, T)	Exceedances	$\text{VaR}_{0.01}(H_0)$	$\text{VaR}_{0.01}(H_1)$	$\text{ES}_{0.025}(H_0)$	$\text{ES}_{0.025}(H_1)$	$\text{Tail}_{0.025}$
t_5						
$(125, 250)$	17.58	16.71	0.02	33.06	4.46	43.28
$(250, 250)$	13.56	22.91	0.33	37.53	6.51	55.80
$(500, 250)$	21.46	28.14	0.92	42.29	9.18	63.21
$(1000, 250)$	20.17	31.37	1.34	44.09	12.26	68.02
$NIG(\alpha, 0, \delta, \mu)$						
$(125, 250)$	18.31	13.33	0.10	33.01	5.03	18.60
$(250, 250)$	18.00	25.00	0.36	42.75	10.19	27.00
$(500, 250)$	29.45	34.51	1.30	47.71	16.56	34.48
$(1000, 250)$	29.61	40.17	2.35	50.37	20.21	39.54
$NIG(\alpha, -0.25, \delta, \mu)$						
$(125, 250)$	41.32	30.37	0.52	62.26	13.57	9.52
$(250, 250)$	41.17	47.10	1.83	70.38	24.84	27.49
$(500, 250)$	55.13	57.11	5.31	74.50	35.27	57.55
$(1000, 250)$	54.40	62.11	7.87	76.06	41.90	85.70
GARCH(1,1)						
125	18.98	12.82		19.07		8.72
250	13.36	13.60		19.82		9.98
500	17.98	15.84		22.37		13.38
1000	16.27	17.47		23.04		15.23

where s_T^* denotes the realized value of the test statistic. More generally, we may want to use a basis multiplication factor (bmf) and we may want to cap the multiplication factor at some upper value (limit). Using the fact that $\rho(Q_T) = \rho(G) + \sqrt{\frac{Vs_T^*}{T}}$ our proposal for the multiplication factor becomes

$$\mathrm{mf}\left(s_T^*\right) = \min\left\{\left(\mathrm{bmf} \cdot \max\left\{1, 1 + \frac{\sqrt{\frac{Vs_T^*}{T}} - \sqrt{\frac{Vk_\alpha}{T}}}{\varrho\left(G\right)}\right\}\right), \mathrm{limit}\right\}, \tag{2.31}$$

We show the results for our proposed multiplication factor applied to value-at-risk, and expected shortfall in Figure 2.1, where we use $G = \Phi$, $\alpha = 0.05$, bmf $= 3$, and limit $= 4$. As the variance in (2.29) is larger than without estimation risk, the basis multiplication factor should be taken higher if one takes estimation risk into account. This is probably also one of the reasons why the multiplication factor of the BIS is rather high. For reasons of comparison with the BIS scheme, we use here a bmf of 3 and a limit of 4. See Kerkhof et al. (2002) for suggestions on setting the bmf for markets depending on the reliability with which the market can be modeled. On the horizontal axis we plot the quantiles of the distribution of the test statistic in (2.11) under the null hypothesis and on the vertical axis the resulting multiplication factors. As a benchmark we also plot the multiplication factors when using the current Basel procedure (now as a function of the quantiles of the corresponding test under the null). We see that the multiplication factors according to our proposal seem to compare favorably with those according to the Basel procedure. Moreover, the multiplication factors for expected shortfall are slightly lower than for value-at-risk. This has to do with the result that expected shortfall is more accurately estimated under the null than value-at-risk, i.e., the variance V in case of expected shortfall is smaller than in case of value-at-risk.

In Figure 2.2 we report the results of applying the multiplication factors from (2.31) to value-at-risk and expected shortfall, using again the outcomes of the Basel procedure as a benchmark. We consider two cases: first, we look at the case where the model is correct, $P_t = F_t = \mathcal{N}\left(\mu, \sigma^2\right)$; second, the case of a seriously misspecified model, $P_t = \mathcal{N}\left(\mu, \sigma^2\right)$ and $F_t = NIG(\alpha, -0.25, \delta, \mu)$ with α, δ, μ as before, being the

Figure 2.1: **Multiplication factors**

This figure shows the multiplication factors on the vertical axis against the quantiles of the test statistic on the horizontal axis. We used $G = \Phi$, $\alpha = 0.05$, and a basic multiplication factor bmf= 3.

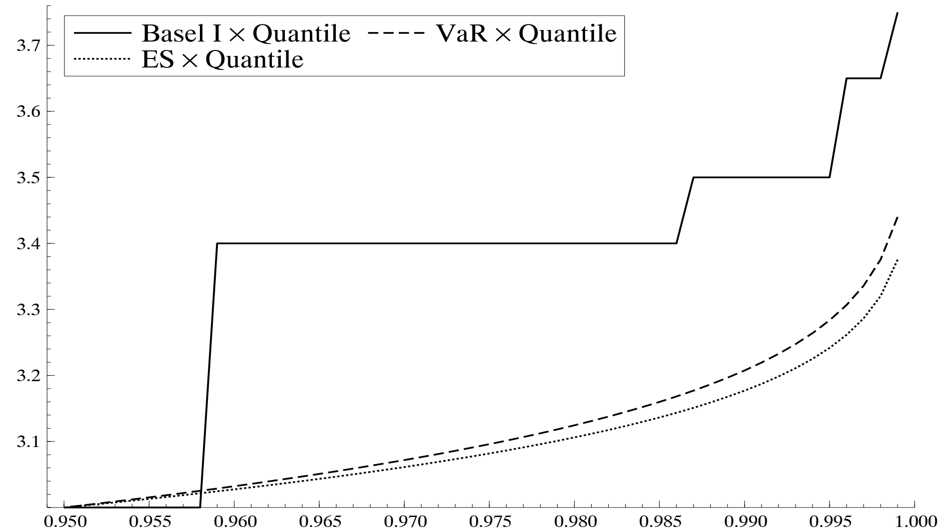
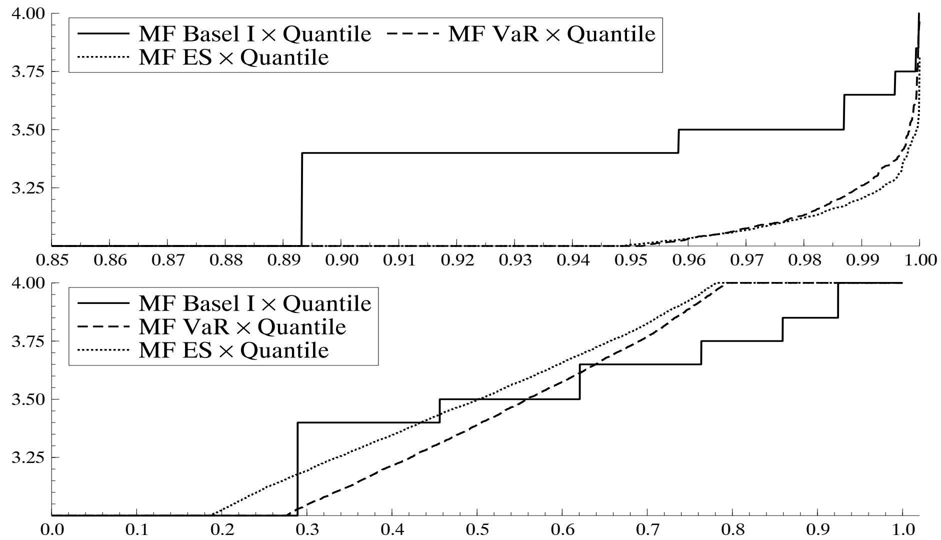


Figure 2.2: **Multiplication factors (size, power)**

This figure shows the simulated cdf of the multiplication factors. In the upper panel the case of $F_t = \mathcal{N}(\mu, \sigma^2)$ is shown. In the lower panel we have the case where $F_t = NIG(\alpha, -0.25, \delta, \mu)$. In both panels $P_t = \mathcal{N}(\mu, \sigma^2)$. The number of days equals 250 and the number of simulations equals 10,000.



case where the distribution is very skewed to the left and has a large kurtosis.

The results of the correctly specified case reflect the outcomes presented in the previous figure: expected shortfall, having the lowest multiplication factors, performs best. Notice that the multiplication factor scheme from the current Basel Accord results in (too) large multiplication factors. In the second case of a misspecified model we see that the test using expected shortfall results in higher factors in more cases (due to the higher power) than the test using value-at-risk. For both expected shortfall and value-at-risk the penalty depends smoothly on the outcome of the test. The multiplication factors according to the current Basel Accord more or less correspond to those of value-at-risk and expected shortfall, but in a heavily non-smooth way.

Concluding, in the case that the bank uses a correctly specified model, we find that the capital requirement scheme using expected shortfall leads to the least severe penalties. On the basis of the current Basel Accord banks would be punished more often and then also severely. Furthermore, in case of a misspecified model, we find that the capital requirement scheme using expected shortfall rejects the misspecified models most often, the multiplication factor depends smoothly on the size of the misspecification found and the variance in the multiplication factors is low.

2.7 Conclusions

In this paper we suggested a backtest framework for a large and relevant group of risk measurement methods using the functional delta method. We showed that, for a large group of risk measurement methods containing all currently used risk measurement methods, the backtest procedure can readily be found after computing the appropriate influence function of the risk measurement method. The influence functions for value-at-risk and expected shortfall are provided. Since this general framework is based on asymptotic results, we investigated whether the procedure is appropriate for realistic finite sample sizes. The results indicate that this is indeed the case, and that, contrary to common belief, expected shortfall is not harder to backtest than value-at-risk if we adjust the level of expected shortfall. Furthermore, the power of the test for expected shortfall is considerably higher than that of value-at-risk. Since the probability of detecting a misspecified model is higher for a given value of the test statistic, this allows the regulator to set lower multiplication factors. We suggested a scheme for determining multiplication factors. This scheme results in less severe penalties for the backtest based on expected shortfall compared to backtests based on value-at-risk and to the current Basel Accord backtesting scheme in case the test incorrectly rejects the model. In case of a misspecified model the multiplication factors are on average about the same for all tests. However, the multiplication factors based on the expected shortfall test are smooth and have low variance.

Thus, the prospects for setting up viable capital determination schemes based on expected shortfall seem promising.

Chapter 3

Testing Expected Shortfall Models for Derivatives Positions

3.1 Introduction

Managing the risks of derivative assets has always been one of the major challenges in risk management. With the strong increase in derivative positions in the portfolios of financial institutions the task of managing these risks has become more daunting than ever. An equally daunting task is testing the quality of models used to quantify the risk of derivatives positions.

Since the Basel Committee advised the use of value-at-risk (VaR) in the 1996 amendment to the Basel Accord for determination of regulatory capital, many studies have investigated VaR (see, for example, the overviews in Jorion (2000) and Dowd (1998) and the references therein). Recently, a literature emerged advocating alternative risk measures, namely, coherent risk measures and, in particular, expected shortfall (see, for example, Artzner et al. (1999), Delbaen (2000), Acerbi and Tasche (2002), and Tasche (2002)). The advantages of expected shortfall over VaR are that it satisfies the property of subadditivity and the fact that portfolio optimization under expected shortfall constraints yields reasonable portfolios, contrary to VaR (see, for example, Yamai and Yoshiba (2002a) for the constrained portfolio optimization). Though most people agree that from a theoretical point of view expected shortfall is to be preferred to VaR, it is still less widely used due to the lack of a solid backtesting procedure. Recently, Kerkhof and Melenberg (2003) (see also Chapter 2 of this thesis) introduced a test for expected shortfall and found that for appropriately adjusted levels, expected shortfall has more desirable backtesting properties than VaR.

Though quite a number of studies have tested the performance of several VaR models, derivatives positions were rarely explicitly taken into account (see, for example, McNeil and Frey (2000), Christoffersen et al. (2001), and Berkowitz and O'Brien (2002)). In cases where derivative positions were explicitly taken into account, the literature usually focused on the computation of VaR rather than on the testing of the VaR models, since the standard binomial test can be applied (see, for example, Kupiec (1995) and El-Jahel, Perraudin and Sellin (1999)). However, the standard binomial test cannot be applied to expected shortfall. In order to test expected shortfall we need information of the distribution of bank's profit and losses (P&L) account, or more specifically its tail behavior.

One of the problems that one faces in determining the P&L distribution of (non-linear) derivatives is that their risk characteristics change over time. For example, an option can change from a 1 year at-the-money option into a 3 month far out-of-the-money option, resulting in completely different risk characteristics. In this chapter, we propose a method to take into account the differences in risk exposures between options with different characteristics.

We consider several methods to estimate the risk measure for the one-day hedge error. The first method we consider is a simple Black-Scholes based model which assumes normal asset returns and constant implied volatilities. Method 2 relaxes the assumption of normal asset return and uses a nonparametric asset return distribution based on historical simulation. Method 3 is a full historical simulation method that assumes a nonparametric asset returns distribution and a nonparametric implied volatility distribution. The fourth method is a Vector AutoRegessive (VAR) model for asset returns and implied volatilities returns with Gaussian errors, while method 5 considers nonparametric errors instead.

We test the models on the FX market and, in particular, the mutual exchange rates of the US, the UK, and Japan. Furthermore, we test the models on S&P 500 options. We find that the historical simulation method and the VAR models perform reasonably well.

The remainder of the chapter is structured as follows. Section 3.2 describes daily market risk for derivative positions. Section 3.3 discusses the aging, moneyness, and level effects of derivative positions and a possible transformation to standardize the risk exposures. The models used are described in Section 3.4. Section 3.5 describes the test used and Section 3.6 presents the empirical results. Finally, Section 3.7 concludes.

3.2 Quantifying daily market risk

Consider the situation where a financial institution manages a portfolio which is short in options. Due to this position the financial institution is subject to a risk exposure with respect to the value of the options. To decrease this risk exposure the financial institution hedges the derivative using a particular hedge strategy. To illustrate, consider a derivative whose price at day t equals f_t. The financial institution hedges the derivative using some underlying instruments with prices $S_t = \left(S_t^1, ..., S_t^k\right)$. Let the money market account at time t be given by N_t. Let the financial institution hedge the derivative by buying amounts $\gamma_t = \left(\gamma_t^1, ..., \gamma_t^k\right)$ of the underlying instruments. Define

$$\alpha_t = \frac{f_t - \gamma_t \cdot S_t}{N_t}. \tag{3.1}$$

Then we will have as accounting identity

$$f_t = \gamma_t \cdot S_t + \alpha_t N_t. \tag{3.2}$$

The next day the price of the derivative will be f_{t+1}, while the hedging position (if there are no intermediate adaptations) will be valued $\gamma_t \cdot S_{t+1} + \alpha_t N_{t+1}$. The difference between the next period's derivative's price and the hedge position induces the *daily market risk*. A financial institution can quantify this daily market risk by assuming some method to estimate or calibrate the next day's probability distribution of $(f_{t+1}, S_{t+1}, N_{t+1})$. Taking a numeraire whose future value at $t + 1$ is known (for example, a one-period discount bond) reduces the problem to estimating or calibrating the next day's probability distribution of (f_{t+1}, S_{t+1}), but now with respect to the numeraire instead of cash. This allows for estimation or calibration of the daily market risk measures (for instance, value-at-risk or expected shortfall). Our interest in this chapter is in risk measures of the daily market risk profile. Specifically, we are interested in the distribution of

$$E_t^1 \equiv \Delta f_t - \gamma_t \Delta S_t, \tag{3.3}$$

where $\Delta x_t \equiv x_{t+1} - x_t$ for $x = f, S$. E_t^1 denotes the *one-period hedge error* and its distribution which is termed the *daily market risk profile* is denoted by $\mathcal{L}\left(E_t^1\right)$. Examples of the daily market risk profile are given in the upper panel of Figure 3.1, which presents daily market risk profiles of a delta hedged 3 month at-the-money, 1 year at-the-money, and 3 year at-the-money (ATM) call option in a Black-Scholes world with annual instantaneous drift $\mu = 0.1$, instantaneous volatility $\sigma = 0.2$, and instantaneous riskless interest rate equal to $r = 0.05$. Time t is measured in days (1 year equals 250 days). In line with Boyle and Emanuel (1980) a shifted non-central χ^2-distribution is found as an approximation for the market risk profile.

3.3 Aging, moneyness, and level effect

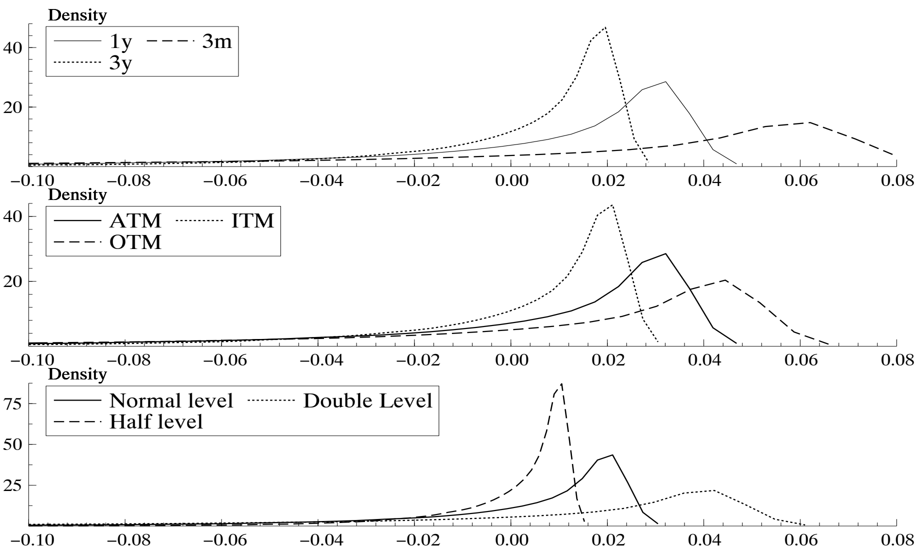
Figure 3.1 clearly shows that the distribution of hedge errors of options depends on the time to maturity, $\tau = T - t$. We refer to the fact that the daily market risk profile changes with the time to maturity as the *aging effect*. For shorter maturities, the daily risk profile is more spread out. The middle panel of Figure 3.1 shows the dependence of the daily risk profile on moneyness which is termed the *moneyness effect*.[1] Out-of-the-money options have more variability than in-the-money options. Finally, in the lower panel we see the influence of the level on the risk profiles, the so-called *level effect*. It is easy to show that this effect is linearly dependent on the level.

The three effects shown in Figure 3.1 indicate the problems one encounters when using time series data of a particular option to extract information of the daily

[1] Moneyness is defined as $m = \log\left(e^{rt} S_t / k\right)$. A call option is called in-the-money (ITM) if $m > 0$, at-the-money if (ATM) $m = 0$, and out-of-the-money (OTM) if $m < 0$.

Figure 3.1: **Aging, moneyness, and level effects**

On the horizontal axis the return on the hedged portfolio is given in percentages. Total number of simulations = 100,000. The upper panel shows the daily risk profiles of delta hedged ATM call option with a maturities of 3 months, 1 year, and 3 years and level 100. The middle panel shows the daily risk profiles of delta hedged OTM ($m = -0.1$) , ATM ($m = 0$), and ITM ($m = 0.1$) call option with a maturity 1 year and level of 100. The lower panel shows the daily risk profiles of a delta hedged ATM call option with a maturity of 1 year and levels of 50, 100, and 200.



market risk profile of that option. The observations of hedge errors of the option are taken with different times to maturity and potentially different moneyness and levels. Since the distribution differs for these situations, these hedge errors are hard to compare. In order to suppress the level effect we first determine a level-independent distribution of relative hedge errors in the following way:

$$\tilde{E}_t^1 = \frac{E_t^1}{f_t - \gamma_t \cdot S_t}. \tag{3.4}$$

The dependence on the daily market risk profile on the aging and moneyness effect is more complicated to resolve. To get rid of the aging and moneyness effect, it is natural to use data on derivatives with the same moneyness and time to maturity, if possible. For FX derivatives and interest-rate derivatives these data are available, since these are quoted in the market with a fixed time to maturity. For equity derivatives, however, this is more complicated due to the fact that these derivatives have fixed maturity dates. Therefore, we have to transform our data.

3.3.1 Transformation of the data

A possible way to correct the daily market risk profile for the aging and moneyness effect is to assume a parametric option pricing model so that one can use the characteristics of such a model to find the appropriate corrections. In this section we correct for the aging and moneyness effect using the Black-Scholes model.[2] Denoting the model price by $f(\xi_t)$ with $\xi_t = (S_t, t)$ a Taylor series expansion gives

$$f(\xi_{t+\Delta t}) = f(\xi_t) + \Delta(\xi_t)\Delta S_t + \frac{1}{2}\Gamma(\xi_t)\,(\Delta S_t)^2 + \Theta(\xi_t)\Delta t + O(\Delta t^{3/2}), \tag{3.5}$$

where

$$\Delta S_t = S_{t+\Delta t} - S_t. \tag{3.6}$$

$\Delta(\xi_t) \equiv \frac{\partial f}{\partial S}(\xi_t)$ denotes the first-order partial derivative of f with respect to the underlying, $\Gamma(\xi_t) \equiv \frac{\partial^2 f}{\partial S^2}(\xi_t)$ denotes the second order partial derivative with respect to the underlying, and $\Theta(\xi_t) \equiv \frac{\partial f}{\partial t}(\xi_t)$ denotes the first partial derivative with respect to the current time.

[2]Other models with sufficiently smooth pricing formulas can also be used.

We take $\Delta t = 1$. Let E_t^1 denote the one-period hedge error from time t to $t+1$ and let $\{\gamma_t\}_{t=1}^T$ denote the hedging strategy. Neglecting the remainder term from now on, we get

$$\begin{aligned} E_t^1 &= \Delta f_t - \gamma_t \Delta S_t \\ &= (\gamma_t - \Delta(\xi_t)\Delta S_t) + \frac{1}{2}\Gamma(\xi_t)\left(\Delta S_t\right)^2 + \Theta(\xi_t). \end{aligned}$$

In general, the hedge errors $E_1^1, ..., E_T^1$ resulting from the hedge strategy $\{\gamma_t\}_{t=1}^T$ do not have the same distribution. To evaluate the performance of a hedge strategy, we want to "standardize" the hedge errors such that they have the same distribution. As reference distribution, we use the distribution, $\mathcal{L}\left(E_{t^*}^1\right)$, for some t^* such that $0 \leq t^* < T$.

We assume strict stationarity of the (conditional) differenced underlying process, implying,

$$\mathcal{L}\left(\Delta S_t \mid \mathcal{F}_t\right) = \mathcal{L}\left(\Delta S_{t^*} \mid \mathcal{F}_{t^*}\right) \tag{3.7}$$

for $t = 1, ..., T$. Using the auxiliary process γ_t^*

$$\gamma_t^* = \Delta(\xi_t) + \frac{\Gamma(\xi_t)}{\Gamma(\xi_{t^*})}\left(\gamma_{t^*} - \Delta(\xi_{t^*})\right),$$

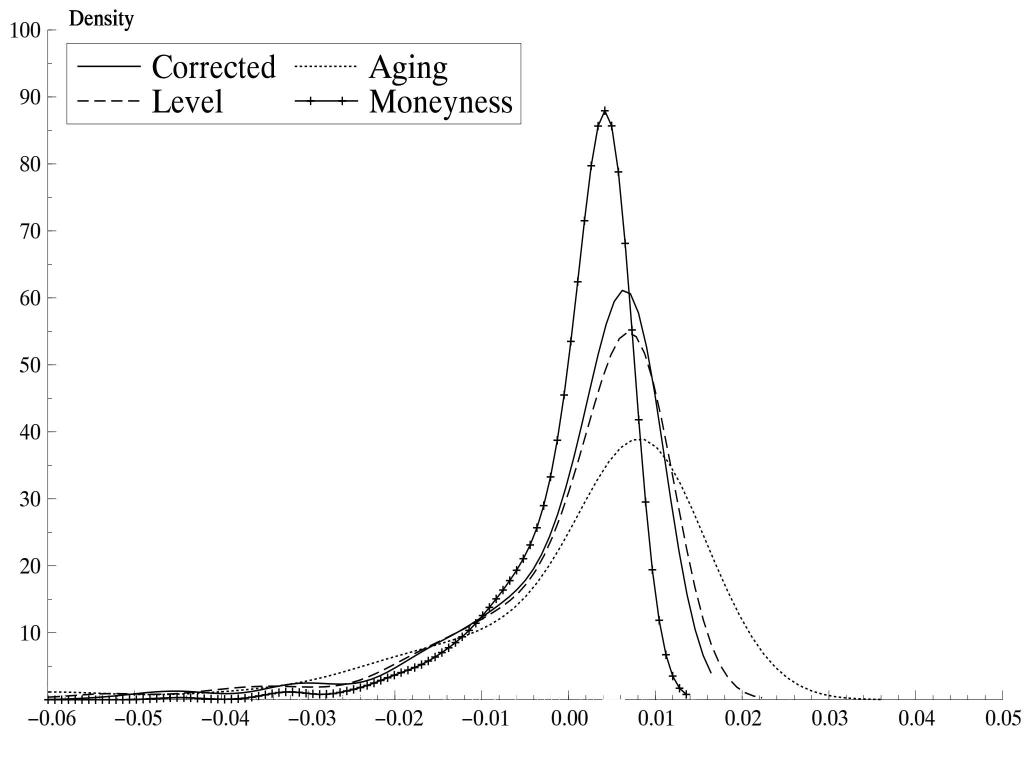
we find the following relationship between the distributions of the hedge errors at different times.

$$\mathcal{L}\left(E_{t^*} \mid \mathcal{F}_t\right) = \mathcal{L}\left(\frac{\Gamma(\xi_{t^*})}{\Gamma(\xi_t)}E_t^1 + \frac{\Gamma(\xi_{t^*})}{\Gamma(\xi_t)}\left(\gamma_t^* - \gamma_t\right)\Delta S_t - \frac{\Gamma(\xi_{t^*})}{\Gamma(\xi_t)}\Theta(\xi_t) + \Theta(\xi_{t^*}) \mid \mathcal{F}_{t^*}\right). \tag{3.8}$$

In (3.8) we found a relation between the one-period hedge error from t^* to $t^* + 1$ with characteristics $(S_{t^*}, m_{t^*}, \tau_{t^*})$ and the one-period hedge error from t to $t+1$ with characteristics (S_t, m_t, τ_t). Therefore, we can transform the data set of realizations drawn from not identically distributed distributions to one of realizations drawn from approximately identically distributed distributions. To obtain (3.8) we neglected the remainder term and used a parametric model in (3.5), so that this can only be seen as a good practical approximation and not as a strict identity.

Suppose we have a time series of hedge errors from a one year ATM call option.

Figure 3.2: **Transformation**
This figure shows the effect of the transformation described in 3.3.1. The graph shows the risk profile when the hedge errors of a delta hedged ATM option position are corrected for all effects, and the risk profiles corrected for all effects but the aging effect, the moneyness effect, and the level effect, respectively. The reference distribution is a one year ATM call option.



In Figure 3.2 we see the result of correcting the time series hedge errors for aging, moneyness, and level effect. This is in accordance with the "true" distribution determined by cross-sectional simulation. Furthermore, the distribution is given in case one of the corrections is left out. We see that the distribution is more spread out, if we leave out the aging effect correction. This follows from the fact that gamma is higher for short term options. Not correcting for the moneyness results (for an originally ATM option) in a less spread out distribution due to the fact that the gamma is lower for ITM and OTM options. Finally, we see that the distribution which is not corrected for level is rather similar to the corrected one. The level effect, however, becomes more important in case the sample is longer and the underlying moves further away from its starting position.

3.4 Daily market risk forecasting methods

In this section, we discuss several methods that can be used to compute risk measures, such as value-at-risk and expected shortfall, of the daily market risk. In doing so, the models need to estimate $F \equiv \mathcal{L}\left(\tilde{E}_t^1\right)$ or more specifically the (joint) distribution of $\Delta f_t^{m_i}$ and $\Delta S_t^{m_i}$ for models $i = 1, ..., 5$. After applying the standardizing procedure in (3.8) we assume a stationary time series of[3]

$$\tilde{E}_t^{1,m_i} = \frac{\Delta f_t^{m_i} - \gamma_t \Delta S_t^{m_i}}{f_t - \gamma_t S_t}, \quad t = 1, ..., T,$$

with $\Delta f_t^{m_i} \equiv f_{t+1}^{m_i} - f_t$ and $\Delta S_t^{m_i} \equiv S_{t+1}^{m_i} - S_t$ where f_t and S_t denote observed prices.

We have returns data available of the underlying (S), the implied volatility (σ), the domestic interest rate (r^d), and the foreign interest rate (r^f), $h_{-N+1} = \left(h_{-N+1}^s, h_{-N+1}^\sigma, h_{-N+1}^{r^d}, h_{-N+1}^{r^f}\right), ..., h_T = \left(h_T^s, h_T^\sigma, h_T^{r^d}, h_T^{r^f}\right)$, with $h_t^x = \log\left(x_t/x_{t-1}\right)$. From these data we use $\{h_1, ..., h_T\}$ for testing and we refer to this set as the testing sample. The testing sample is used in the backtest to determine the quality of the method. All the models discussed below are used to estimate the distribution of the relative one-period hedge error $\tilde{E}^1, m_i$, denoted by F^{m_i}. For notational conve-

[3]The time series is also assumed to be ergodic and to satisfy the necessary regularity conditions needed for Central limit theorems used later on.

nience, we neglect the dependence of $\tilde{E}^{1,m_i}$, $\Delta f_t^{m_i}$, and $\Delta S_t^{m_i}$ on m_i in the following enumeration of models.

1. Method 1 is a naive method which more or less follows the Black-Scholes world assumptions, but with potentially changing mean and volatility. It assumes that $\mathcal{L}\left(h_{t+1}^s\right) = \mathcal{N}\left(\mu_t, \sigma_t^2\right)$ and that σ, r^d, and r^f are constant. To estimate μ_t and σ_t^2 we use the returns data of the underlying, $h_t^s, ..., h_{t-N}^s$, to get μ_t and σ_t^2, the so-called rolling window estimators for μ_t and σ_t^2. For $t = 1, ..., T$ we draw h_t^s from $\mathcal{N}\left(\mu_t, \sigma_t^2\right)$ to construct $\left(\Delta S_t\right)_{t=1}^T$ and $\left(\Delta f_t\right)_{t=1}^T$.[4] Given the hedge strategy γ we construct $\left(\tilde{E}_t^1\right)_{t=1}^T$ from which we produce an estimate $\hat{\varrho}^{m_1} = \varrho\left(\hat{F^{m_1}}\right)$.

2. Method 2 is a historical simulation method for the underlying asset. The implied volatilities, domestic and foreign interest rates are as in method 1. Method 2 allows a distribution for the underlying that differs from the normal distribution. It assumes $\left(h_t^s\right)_{t=-N+1}^T$ is an *i.i.d.* sample. We estimate the distribution, $\mathcal{L}\left(h_t^s\right)$, by the empirical distribution of $\left(h_{t^*}^s\right)_{t^*=t-N+1}^t$.[5] Drawing (with replacement) from $\left(h_{t^*}^s\right)_{t^*=t-N+1}^t$ for $t = 1, ..., T$ allows us to construct $\left(\Delta S_t\right)_{t=1}^T$ and $\left(\Delta f_t\right)_{t=1}^T$. Given the hedge strategy γ we construct $\left(\tilde{E}_t^1\right)_{t=1}^T$ from which we produce an estimate $\hat{\varrho}^{m_2} = \varrho\left(\hat{F^{m_2}}\right)$.

3. Method 3 is a full historical simulation method. This type of method is often used in practice and assumes that $\left(h_t\right)_{t=-N+1}^T$ is an *i.i.d.* sample. We estimate the distribution, $\mathcal{L}\left(h_t\right)$ by the empirical distribution of $\left(h_{t^*}\right)_{t^*=t-N+1}^t$. Drawing h_t (with replacement) from $\left(h_{t^*}\right)_{t^*=t-N+1}^t$ for $t = 1, ..., T$ allows us to construct to construct $\left(\Delta S_t\right)_{t=1}^T$ and $\left(\Delta f_t\right)_{t=1}^T$. Given the hedge strategy γ we construct $\left(\tilde{E}_t^1\right)_{t=1}^T$ from which we produce an estimate $\hat{\varrho}^{m_3} = \varrho\left(\hat{F^{m_3}}\right)$.

4. In method 4 a first-order Vector AutoRegressive (VAR) model for estimation of the distribution of $\left(S_{t+1}, \sigma_{t+1}\right)_{t=1}^T$ is estimated using $\left(h_{t^*}\right)_{t^*=t-N+1}^t$ for $t =$

[4]Note that the sequence $\left(\Delta S_t\right)_{t=1}^T$ is not used to produce a price path $\left(S_t\right)_{t=1}^T$ of the underlying. It only serves to compute a series of hedge errors. The price path of the underlying is given by the data.

[5]Considering the stationarity assumption, it would be more efficient to use all available data, but we use this nonparametric rolling window estimator because it is often used in practice.

$1, ..., T$

$$\begin{bmatrix} h^s_{t+1} \\ h^\sigma_{t+1} \end{bmatrix} = \Phi_0 + \Phi_1 \begin{bmatrix} h^s_t \\ h^\sigma_t \end{bmatrix} + \begin{bmatrix} u^s_{t+1} \\ u^\sigma_{t+1} \end{bmatrix} \tag{3.9}$$

$$h_{t+1} = \Phi_0 + \Phi_1 h_t + u_{t+1}, \quad t = 1, ..., T \tag{3.10}$$

with

$$\mathcal{L}\left(u_{t+1}|\mathcal{F}_t\right) = \mathcal{N}\left(0, \Sigma_t\right),$$

where $\mathcal{F}_t$ denotes the information set at point t. This gives $(\Phi_{0,t}, \Phi_{1,t})_{t=1}^T$, and $(\Sigma_t)_{t=1}^T$ to generate $(\Delta S_{t+1}, \Delta\sigma_{t+1})$ for $t = 1, ..., T$ and $(\Delta f_t)_{t=1}^T$. Given the hedge strategy γ we construct $\left(\tilde{E}^1_t\right)_{t=1}^T$ from which we produce an estimate $\hat{\varrho}^{m_4} = \varrho\left(\hat{F^{m_4}}\right)$.

5. In method 5 a first-order Vector AutoRegressive (VAR) model for estimation of the distribution of $(S_{t+1}, \sigma_{t+1})_{t=1}^T$ is estimated using $(h^s_{t^*}, h^\sigma_{t^*})_{t^*=t-N+1}^t$ for $t = 1, ..., T$

$$\begin{bmatrix} h^s_{t+1} \\ h^\sigma_{t+1} \end{bmatrix} = \Phi_0 + \Phi_1 \begin{bmatrix} h^s_t \\ h^\sigma_t \end{bmatrix} + \begin{bmatrix} u^s_{t+1} \\ u^\sigma_{t+1} \end{bmatrix} \tag{3.11}$$

$$h_{t+1} = \Phi_0 + \Phi_1 h_t + u_{t+1}, \quad t = 1, ..., T \tag{3.12}$$

with

$$\mathcal{L}\left(u_{t+1}|\mathcal{F}_t\right) = F^N_t,$$

where F^N_t denotes the empirical distribution function of u at time t estimated from $u_{t-N+1}, ..., u_{t-1}$. This gives $(\Phi_{0,t}, \Phi_{1,t})_{t=1}^T$, and $(\Sigma_t)_{t=1}^T$ to generate $(\Delta S_{t+1}, \Delta\sigma_{t+1})$ for $t = 1, ..., T$ and $(\Delta f_t)_{t=1}^T$. Given the hedge strategy γ we construct $\left(\tilde{E}^1_t\right)_{t=1}^T$ from which we produce an estimate $\hat{\varrho}^{m_5} = \varrho\left(\hat{F^{m_5}}\right)$.

3.5 Test procedure

In this section, we present a test to evaluate daily market risk evaluation models described in Section 3.4. Time t runs from $-N+1$ to T. The last T observations are used for testing. At each point in time the method is estimated from the previous N observations, that is, we use the so-called rolling window estimator.

The predicted daily market risk will be obtained from the distribution F^{m_i} of[6]

$$\tilde{E}_t^{1,m_i} = \frac{\Delta f_t^{m_i} - \gamma_t \Delta S_t^{m_i}}{f_t - \gamma_t S_t}, \tag{3.13}$$

while the actual daily market risk is induced by the distribution F of

$$\tilde{E}_t^{1} = \frac{\Delta f_t - \gamma_t \Delta S_t}{f_t - \gamma_t S_t} \tag{3.14}$$

We would like to test whether the predicted risk measures are the same for the method hedge errors as for the empirical hedge errors. Let $\varrho(F^{m_i})$ represent the characteristic of interest of F^{m_i} and let $\varrho(F)$ represent the corresponding characteristic of interest of F.

Denote by $\varrho\left(\hat{F}^{m_i}\right)$ an appropriate estimator for $\varrho(F^{m_i})$ such that

$$\sqrt{T}\left(\varrho\left(\hat{F}^{m_i}\right) - \varrho(F^{m_i})\right) = \sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\Psi_t^{m_i} + o_p(1), \qquad \mathbb{E}\Psi_t^{m_i} = 0, \mathbb{E}\left(\Psi_t^{m_i}\right)^2 < \infty, \tag{3.15}$$

and, similarly, let $\varrho\left(\hat{F}\right)$ be an appropriate corresponding estimator for $\varrho(F)$ such that

$$\sqrt{T}\left(\varrho\left(\hat{F}\right) - \varrho(F)\right) = \sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\Psi_t + o_p(1), \qquad \mathbb{E}\Psi_t = 0, \mathbb{E}\left(\Psi_t\right)^2 < \infty, \tag{3.16}$$

where $\Psi_t^{m_i}$ and Ψ_t are called the influence functions. In Appendix A the influence functions for VaR and expected shortfall are given. Then, under the null hypothesis $H_0 : \varrho(F^{m_i}) = \varrho(F)$, we have

$$\sqrt{T}\left(\varrho\left(\hat{F}^{m_i}\right) - \varrho(F^{m_i})\right) - \sqrt{T}\left(\varrho\left(\hat{F}\right) - \varrho(F)\right) =$$
$$\begin{bmatrix} 1 & -1 \end{bmatrix}\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \Psi_t^{m_i} \\ \Psi_t \end{bmatrix} + o_p(1) \xrightarrow[H_0]{d} N(0, V) \tag{3.17}$$

[6]Notice the difference in notation: $\Delta x_t^{m_i} \equiv x_{t+1}^{m_i} - x_t$ and $\Delta x_t \equiv x_{t+1} - x_t$ for $x = S, f$. In both cases we use the observed prices as starting point.

with

$$V = \begin{bmatrix} 1 & -1 \end{bmatrix} \left[\lim_{T\to\infty} \mathbb{E} \left[T^{-1} \left(\sum_{t=1}^{T} \begin{bmatrix} \Psi_t^{m_i} \\ \Psi_t \end{bmatrix} \begin{bmatrix} \Psi_t^{m_i} \\ \Psi_t \end{bmatrix}' \right) \right] \right] \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{3.18}$$

So, with $\hat{V}$ (using, for example, the estimator of Newey and West (1987)) satisfying $\hat{V} \xrightarrow{p} V$, we can take as a test statistic

$$T \frac{\left(\varrho\left(\hat{F}^{m_i}\right) - \varrho\left(\hat{F}\right)\right)^2}{\hat{V}} \xrightarrow[H_0]{d} \chi_1^2 \tag{3.19}$$

Since we can simulate from F^{m_i} as often as we would like, we can strengthen the test above by using $\bar{\varrho} \equiv \frac{1}{K} \sum_{k=1}^{K} \varrho_k(\hat{F}_t^{m_i})$, with K equal to the number of trials, instead of $\hat{\varrho}(F_t^{m_i})$. This gives for fixed K

$$\sqrt{T}\left(\bar{\varrho} - \varrho(F^{m_i})\right) = \sqrt{T}\frac{1}{TK}\sum_{k=1}^{K}\sum_{t=1}^{T} \Psi_{t,k}^{m_i} + o_p(1),$$

$$\mathbb{E}\Psi_{t,k}^{m_i} = 0, \mathbb{E}\left(\Psi_{t,k}^{m_i}\right)^2 < \infty. \tag{3.20}$$

The expression in (3.20) converges in probability to zero as $K \to \infty$ and so we can take as a test statistic

$$T \frac{\left(\bar{\varrho} - \varrho\left(\hat{F}\right)\right)^2}{\hat{v}} \xrightarrow[H_0]{d} \chi_1^2, \tag{3.21}$$

where $\hat{v}$ denotes a consistent estimator for

$$v = \lim_{T\to\infty} \mathbb{E}\left[T^{-1} \sum_{t=1}^{T} \Psi_t^2 \right]. \tag{3.22}$$

3.6 Empirical Results

3.6.1 FX market

The FX market is by far the most liquid market in the world with a daily turnover of about 1.5 trillion US dollars (for comparison, the NYSE has a daily turnover of about 30 billion US dollar). In this section, we apply the test outlined above to call options on the dollar-yen, dollar-pound, and pound-dollar exchange rates. Quotes are in implied volatilities in the FX market and prices can be computed using the Garman-Kohlhagen model (see Garman and Kohlhagen (1983)). This is a version of the Black-Scholes model applicable to currency options. Call option prices are given by

$$c^{GK}\left(S,k,r^d,r^f,\sigma,\tau\right)=S_t e^{-r^f\tau}\Phi\left(d_+\right)-Xe^{-r^d\tau}\Phi\left(d_-\right), \tag{3.23}$$

where

$$d_{\pm}=\frac{\log\left(S_t/k\right)+\left(r^d-r^f\right)\tau}{\sigma\sqrt{\tau}}\pm\frac{1}{2}\sigma\sqrt{\tau}. \tag{3.24}$$

r^d is the domestic instantaneous riskless interest rate, r^f is the foreign instantaneous riskless interest rate, σ denotes the instantaneous volatility of the exchange rate and $\Phi\left(\cdot\right)$ denotes the Gaussian cumulative distribution function.

The daily data available consist of implied volatilities of 3 month ATM call options on dollar-yen, dollar-pound, and pound-dollar exchange rates, the corresponding exchange rates, and the US, UK, and Japanese interest rates.[7] The data run from August 9, 1995 until December 13, 2002 and are shown in Figure 3.3.

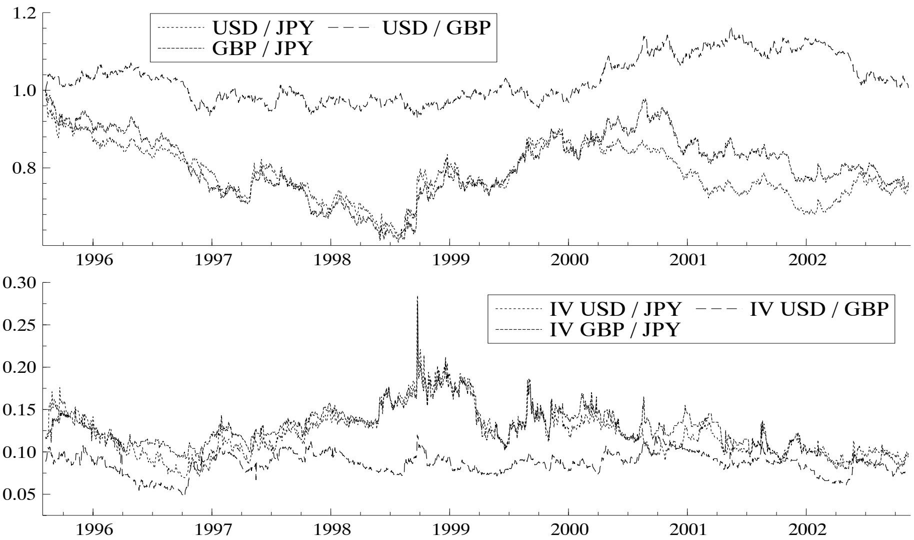
This results in 1918 data points. We use a two year rolling window estimation period for all the models. Taking the number of trading days per year equal to 250 gives us estimation periods of 500 observations and 1418 observations for testing. In Kerkhof and Melenberg (2003) (see also Chapter 2 of this thesis) it is argued that for fair comparison with a 1% value-at-risk the level of expected shortfall should be about 2.5%.[8] The quality of the models is tested by tests whether the variances, the 1%-value-at-risk, and 2.5% expected shortfall of the hedge error as predicted by the models and empirical hedge errors are equal.[9] The level for value-at-risk is chosen at 1% such that it equals the current level in the 1996 amendment to the Basel Accord (see Basel Committee on Banking Supervision (1996b)). Table 3.1 reports the variance, 1% value-at-risk, and 2.5% expected shortfall for an investment of $100 in a portfolio of ATM call options and the underlying exchange rate with as ratio

[7] The data have been kindly shared by ABN-AMRO Bank.

[8] This argument is based on the normal distribution, but seems to be approximately correct in our sample.

[9] In the absence of data on ITM and OTM options, we have assumed a flat volatility smile for the FX options. Since we are looking at one-day hedge errors and the FX volatility smile is rather flat near the money, this should not lead to severe biases.

Figure 3.3: **FX data**
In the upper panel the normalized price paths of the USD/JPY, USD/GBP, and GBP/JPY are given. In the lower panel the implied volatilities for the 3m ATM call options are given.



the hedge strategy.

For all exchange rates we find that the methods 1 and 2 are rejected for all risk measures. The full historical simulation method (method 3) performs well for all exchange rates and all risk measures. The parametric VAR method, method 4, is rejected for the USD/JPY exchange rate for being too conservative, while it is rejected in the GBP/JPY exchange rate for underestimating the risk. The nonparametric VAR method, method 5, is conservative in all markets and is rejected for the USD/JPY and USD/GBP exchange rates.

3.6.2 S&P 500 options

We have available option data on the S&P 500 ranging from January 2, 1992 till August 29, 1997. Quotes on the options are the end-of-day quotes with synchronous observations of the underlying index. For the FX options we have data on fixed time to maturity and moneyness options available. For the S&P 500 we have fixed maturity and varying moneyness option data. Therefore, we apply the transformation method of Section 3.3.1. We analyze the models for calculating the risk measures

Table 3.1: **Tests of Risk measures for the 3 month exchange rates**
This table shows the empirical standard deviations, $VaR_{0.01}$, and $ES_{0.025}$ and those obtained from methods 1,...,5 for the USD/JPY, GBP/JPY, and USD/GBP exchange rate. We test whether the method predictions correspond to the empirical quantities. The p-values of these tests are given in parentheses. In order to reduce sampling error we used $K = 10,000$.

	emp	method 1	method 2	method 3	method 4	method 5
USD/JPY						
std. dev.	0.23	0.05	0.09	0.20	0.22	0.21
		(0.00)	(0.00)	(0.36)	(0.85)	(0.63)
$VaR_{0.01}$	-0.58	−0.18	−0.25	−0.62	−0.80	−0.72
		(0.00)	(0.00)	(0.39)	(0.00)	(0.00)
$ES_{0.025}$	-0.74	−0.19	−0.33	−0.70	−0.90	−0.85
		(0.00)	(0.00)	(0.74)	(0.43)	(0.81)
USD/GBP						
std. dev.	0.11	0.04	0.04	0.11	0.12	0.13
		(0.00)	(0.00)	(0.19)	(0.09)	(0.00)
$VaR_{0.01}$	-0.30	−0.13	−0.17	−0.35	−0.35	−0.44
		(0.00)	(0.00)	(0.11)	(0.12)	(0.00)
$ES_{0.025}$	-0.38	−0.14	−0.19	−0.40	−0.36	−0.47
		(0.00)	(0.00)	(0.19)	(0.75)	(0.02)
GBP/JPY						
std. dev.	0.19	0.06	0.08	0.17	0.16	0.17
		(0.00)	(0.00)	(0.27)	(0.07)	(0.30)
$VaR_{0.01}$	-0.52	−0.19	−0.30	−0.54	−0.40	−0.52
		(0.00)	(0.00)	(0.78)	(0.03)	(0.91)
$ES_{0.025}$	-0.61	−0.20	−0.34	−0.60	−0.41	−0.61
		(0.00)	(0.00)	(0.83)	(0.02)	(0.92)

Table 3.2: **Tests of risk measures for delta hedged 3 month S&P 500 options** This table shows the empirical standard deviations, $VaR_{0.01}$, and $ES_{0.025}$ and those obtained from methods 1,...,5 for a delta-hedged positions in 3 month ATM S&P 500 options. We test whether the method predictions correspond to the empirical quantities. The p-values of these tests are given in parentheses. In order to reduce sampling error we used $K = 10,000$.

	emp	method 1	method 2	method 3	method 4	method 5
ATM 3 months						
std. dev.	0.24	0.04 (0.00)	0.05 (0.00)	0.23 (0.34)	0.24 (0.83)	0.23 (0.55)
$VaR_{0.01}$	−0.74	−0.13 (0.00)	−0.19 (0.00)	−0.67 (0.27)	−0.80 (0.42)	−0.71 (0.55)
$ES_{0.025}$	−0.85	−0.13 (0.00)	−0.19 (0.00)	−0.76 (0.35)	−0.83 (0.73)	−0.79 (0.48)
ATM 1 year						
std. dev.	0.41	0.06 (0.00)	0.06 (0.00)	0.43 (0.50)	0.43 (0.49)	0.43 (0.28)
$VaR_{0.01}$	−1.26	−0.13 (0.00)	−0.17 (0.00)	−1.39 (0.18)	−1.12 (0.15)	−1.42 (0.10)
$ES_{0.025}$	−1.27	−0.17 (0.00)	−0.19 (0.00)	−1.41 (0.22)	−1.16 (0.12)	−1.44 (0.12)

for 3 month ATM options. For this we use the options with time to maturity closest to 3 months and closest to the ATM level. Again we investigate a portfolio of \$100 invested in options and the underlying asset. As hedge ratio we apply the standard Black-Scholes delta with continuous dividend yield.

We find that the empirical risks for S&P500 options are higher than for the FX options. We find that the positions in the 1 year options are more risky than the positions in the 3 months options. For the tests of the S&P500 options we get more or less the same results as for the FX options. Only models 3, 4, and 5 have a acceptable prediction behavior.

Overall, we see that models 1 and 2 do not perform well and underestimate the risk of delta hedged derivatives positions in almost all cases. This can be explained by the fact that they do not take fluctuations in the levels of implied volatilities into account. The historical simulation method and both VAR models perform about the same, although the VAR models for changes in the underlying and implied volatilities are sometimes a bit too conservative. Since the historical simulation method and the VAR model with historical simulation take more time to compute than the Gaussian VAR model where VaR and ES can be computed analytically, it seems easiest to compute both VaR and ES based on the Gaussian VAR model.

3.7 Conclusions

In this chapter we tested several risk management models for computing expected shortfall and value-at-risk for one-period hedge errors of hedged derivatives positions. Though value-at-risk can be tested using a binomial test, this is not the case for expected shortfall and we need information of the distribution in the tail. By nature, the characteristics of derivatives positions are changeable and as a consequence the size of risk exposures varies over time. To overcome this problem, we present a transformation procedure.

We empirically test the performance of several models, based on tests for standard deviation, value-at-risk, and expected shortfall. We find that in order to get good indication of the risk of a hedged derivative in both the FX and the equity market it is of crucial importance to take the variation in the implied volatilities into account. We find that a historical simulation method, which is commonly used in practice, produces the best results. A parametric and non-parametric VAR model perform reasonably well, but their performance trails that of the historical simulation method.

A Influence functions for value-at-risk and expected shortfall

Let F_t denote the distribution of the one-day hedge error E_t^1. The influence functions of value-at-risk and expected shortfall are then given by:

1. Value-at-risk: In the case of VaR_p the influence function $\Psi(F_t)$ is given by

$$\Psi_{\mathrm{VaR}}(F_t) = \frac{p - \mathbf{I}_{\left[-\infty, F_t^{-1}(p)\right]}(x)}{q\left(F_t^{-1}(p)\right)}, \tag{3.25}$$

   and

$$\mathbb{E}\Psi_{\mathrm{VaR}}^2(F_t) = \frac{p(1-p)}{q^2\left(F_t^{-1}(p)\right)}. \tag{3.26}$$

2. Expected shortfall: In the case of ES_p the influence function $\Psi(F_t)$ is given by

$$\begin{aligned}\Psi_{\mathrm{ES}}(F_t) &= -\frac{1}{p}\Big[\left(x - F_t^{-1}(p)\right)\mathbf{I}_{\left[-\infty, F_t^{-1}(p)\right]}(x) \\ &\quad + \Psi_{\mathrm{VaR}}(F_t)\left(p - \int_{-\infty}^{F_t^{-1}(p)} dF(x)\right)\Big] \\ &\quad -\mathrm{ES}(F_t) + \mathrm{VaR}(F_t)\end{aligned} \tag{3.27}$$

   and

$$\begin{aligned}\mathbb{E}\Psi_{\mathrm{ES}}^2(F_t) &= \frac{1}{p}\mathbb{E}\left[X^2 | X \leq F_t^{-1}(p)\right] - \mathrm{ES}(F_t)^2 \\ &\quad +2\left(1 - \frac{1}{p}\right)\mathrm{ES}(F_t)\,\mathrm{VaR}(F_t) \\ &\quad -\left(1 - \frac{1}{p}\right)\mathrm{VaR}(F_t)^2.\end{aligned} \tag{3.28}$$

Chapter 4

Model Risk and Regulatory Capital

4.1 Introduction

Due to the growing complexity of financial markets, financial institutions rely more and more on the use of models to assess the risks to which they are exposed. The accuracy of these risk assessments depends crucially on the extent to which a market can be reliably modeled. Choosing an appropriate model to compute market risk measures is an important and difficult task. It is a widespread feeling among both academics and practitioners that, although some models do a better job than others, the search for one ultimate model is futile. An approach that takes the limitations of our knowledge into account is to develop models—depending on the application (pricing, hedging, ...) — that capture the most important aspects of a particular market, and to somehow control for the fact that the assessment of risk is based on a possibly misspecified model (see Derman (1996)).

The hazard of working with a potentially misspecified model is called *model risk*. Currently no explicit capital requirements are set by regulators in connection with model risk. This is done indirectly using the so called multiplication factors. However, the Basel Committee has indicated that it plans to expand the current capital adequacy framework to improve the charting of risks to which financial institutions are exposed (see Basel Committee on Banking Supervision (1999)). In particular, the Committee intends to set capital requirements for operational risk, which is often taken to consist for an important part of model risk (see Basel Committee on Banking Supervision (2003)). Just as the 1996 Amendment of the Basel Committee stimulated financial institutions to refine their market risk models, banks are likely to make more detailed assessments of model risk after incorporation of model risk regulation in the Basel Accord. As part of their internal risk management systems, most large financial institutions already set aside reserves for model risk (the so-called *model reserves*). This means that booking of certain profits on trades is postponed if it is felt that these profits are sensitive to the model used.

The aim of this paper is to provide a quantitative basis for the incorporation of model risk in regulatory capital requirements. The same framework may also be used for the computation of model reserves in the context of internal risk management procedures within financial institutions; in addition, the method may be used in margin setting by clearing house exchanges, or as a pricing tool. To extend the current practice of computing market risk measures on the basis of some given ("nominal") model, we determine a set of plausible alternative models. In recognition of the fact that each of these models is a (reasonable) candidate for representing reality, we propose to compute a worst-case market risk measure over some set of alternative models. Model risk is then defined as the difference between this measure and the market risk measure computed from the nominal model. Using sets of alternative models restricted and unrestricted to a model class, we distinguish between model risk due to estimation error and model risk due to misspecification. Previous studies on model risk have focused on the risk of using incorrect parameter values in a parametric setting, i.e., estimation error (see, for example, Gibson, Lhabitant, Pistre and Talay (1999), Talay and Zheng (2002), and Bossy, Gibson, Lhabitant, Pistre, Talay and Zheng (2000)). However, our study suggests that the leading factor in model risk is often misspecification rather than estimation error.

One major area where financial models play an important role is the risk management of the portfolios of financial institutions. We discuss value-at-risk and expected shortfall when using a simple Gaussian model and a GARCH$(1, 1)$ model for portfolio returns to illustrate the model risk measurement tools. We consider the S&P 500 and USD/GBP exchange rate as investments. The results can be interpreted in terms of a multiplication factor that should be applied to account for model risk in a given market. Our results for these models indicate that about half of the regulatory capital set by the Basel Committee can be explained in this way when computing the 1% value-at-risk at a 95% confidence level. We find that the model risk due to misspecification is much larger than the model risk due to estimation error.

Another area which relies heavily on financial models is constituted by derivatives trading. For instance, Hull and Suo (2002) investigate the model risk associated with the calculation of prices and deltas for illiquid exotic options based on an implied-volatility model that is calibrated using current prices of liquid products. Their paper clearly demonstrates the presence of model risk in a number of situations. In a companion paper (see Chapter 5), we assess hedging model risk in derivative products on the basis of the total hedging error rather than the error in computing Greeks, and we propose a quantitative measure of model risk that could be used, for instance, in the determination of model reserves.[1] We illustrate the approach for the Black-Scholes family of option pricing models. The results indicate that also in this setting model risk due to misspecification is much larger than the model risk due to estimation error.

The remainder of the paper is structured as follows. In the next section, we give an overview of market risk measurement. We discuss some of the popular risk measures with some emphasis on coherent market risk measurement which fits neatly with the model risk measurement method proposed in Section 4.2. In Section 4.3 we propose a general framework for incorporation of model risk. This is based on a worst-case analysis. A decomposition of model risk in a parametric and a nonparametric part is proposed. Section 4.4 provides an application to portfolio risk management. We discuss the value-at-risk and the expected shortfall approach. Finally, Section 4.5 concludes.

[1] Steps towards the quantification of model risk for derivative contracts have been taken by Green and Figlewski (1999), who show that the risk of trading derivative securities can be decreased substantially by delta hedging. We follow this line of thought by considering the risk of derivative products in combination with a given hedging strategy. The proposed methodology encompasses the methodology proposed by Hull and Suo (2002). Furthermore, robustness issues as treated by El Karoui, Jeanblanc-Picqué and Shreve (1998) fit into the proposed setup.

4.2 Market Risk Measurement

By *market risk* we understand the risk caused by random fluctuations in future asset prices. For each given position, the most basic question that a risk manager must be able to answer is whether or not the risk associated to this position is acceptable. This qualitative decision is often based on the computation of a risk measure which in some way represents the "distance to (un)acceptability". Such a risk measure may, for instance, be arrived at as follows. Since, in the context of finance, risk is usually measured in terms of a univariate distribution (profit/loss), an unacceptable position can be made acceptable if enough of a suitable "sweetener" is added.[2] The amount of sweetener that has to be added to make a given position just acceptable is a natural measure of the distance to acceptability.

4.2.1 Use of market risk measures

Market risk measures may be used for a number of different purposes.

1. Regulatory capital requirements for securities firms and banks are computed on the basis of risk measures. Specifically, the value-at-risk method has been adopted by the Basel Committee on Banking Supervision (1996a).

2. Some banks set reserves for trading desks as part of their internal risk management procedures. The size of the reserve is coupled to some measure of the riskiness of the positions taken by the desk.

3. Exchanges need to guarantee the promises to all parties involved in a contract. To guarantee these promises they use clearing margins for their members. For example, the Chicago Mercantile Exchange (CME) and many other exchanges use SPAN to determine the clearing margins. For more detailed information, see Artzner et al. (1999) and SPAN (1995).

4. Market risk measures may also be used as pricing tools, since they can be used to compare different risks so that good deals can be identified. This point of

[2]For the "sweetener" one can think of 1) a premium in pricing applications, 2) capital reserve in case of regulatory applications, and 3) margin in case of clearing houses.

view is elaborated, for instance, in Cochrane and Saá-Requejo (2000), Jaschke and Küchler (2001) and Carr, Geman and Madan (2001).

4.2.2 Notation and definitions

Since in this paper we are interested in model risk, we will be working with classes of models rather than with a single model. It is not always convenient to use the same probability space for each of these models. Therefore, we start by a formal description of a setting that allows the use of multiple probability spaces.

Definition 4.1 A *model* is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

One could imagine more elaborate probabilistic settings; in particular, a filtration might be assumed given. However, the above notion will be sufficient for the purposes of this paper. For any model m, let $\mathcal{R}(m)$ denote the space of equivalence classes of measurable real-valued functions on $(\Omega, \mathcal{F})$.

Definition 4.2 Let a model m be given. A *risk* defined on m is an element of $\mathcal{R}(m)$.

This definition, in which a "risk" is a random variable defined on a given probability space, follows the terminology of Artzner et al. (1999) and Delbaen (2000). We introduce a similar concept for model classes rather than for individual models.

Definition 4.3 Let $\mathcal{M}$ be a class of models. A *product* defined on $\mathcal{M}$ is a mapping that assigns to each model $m \in \mathcal{M}$ a risk defined on m. The set of all products defined on $\mathcal{M}$ is denoted by $\mathcal{X}(\mathcal{M})$.

The risk induced by a product Π on a model m will be denoted by Π_m. Since $\mathcal{R}(m)$ is a vector space, the set of products $\mathcal{X}(\mathcal{M})$ has the structure of a vector space as well. For instance, if Π_1 and Π_2 are products defined on the same class of models $\mathcal{M}$, then $\Pi_1 + \Pi_2$ is the product that associates to a model m in $\mathcal{M}$ the risk $(\Pi_1)_m + (\Pi_2)_m$. Similarly, we can also define products relative to a reference product (if the reference product is nonzero), and we have a partial ordering on products.

We now proceed to risk measures, starting with the definition for an individual model.

Definition 4.4 Let a model m be given. A *risk measure* defined on m is a map from $\mathcal{R}(m)$ to $\mathbb{R} \cup \{\infty\}$.[3]

Definition 4.5 Let a class of models $\mathcal{M}$ be given. A *risk measurement method* defined on $\mathcal{M}$ is a mapping that assigns to each model $m \in \mathcal{M}$ a risk measure defined on m.

Risk measures can be used to separate acceptable from unacceptable risks in the following way.

Definition 4.6 Let a model m be given, and let ρ be a risk measure defined on m. The acceptance set associated with ρ is the set

$$\mathcal{A}_\rho = \{X \in \mathcal{R}(m) \mid \rho(X) \leq 0\}. \tag{4.1}$$

So far we did not discuss specific properties for risk measures and related notions that would justify the nomenclature. We come to this in the next section.

4.2.3 Popular risk measurement methods and their properties

Most risk measures used in practice can be viewed as risk measurement methods in the formal sense of the previous section. Due to its prominent role in the amendment of 1996 by the Basel Committee, the value-at-risk approach is currently the most popular method used in risk measurement (see, for example, Duffie and Pan (1997), Basel Committee on Banking Supervision (1996a), Dowd (1998), and Risk Magazine (1996)). A formal description of VaR may be given as follows.

Definition 4.7 (Value at Risk (VaR)) Let a model class $\mathcal{M}$ be given. The *value-at-risk* method with *reference asset* $N \in \mathcal{X}(\mathcal{M})$ and *level* $p \in (0,1)$ assigns to a model $m = (\Omega, \mathcal{F}, \mathbb{P}) \in \mathcal{M}$ the risk measure VaR_m given by

$$\text{VaR}_m : \mathcal{R}(m) \ni X \mapsto -\inf\{q \in I\!R : \mathbb{P}(X/N_m \leq q) \geq p\} \in I\!R \cup \{\infty\}. \tag{4.2}$$

[3]Including ∞ allows risks to be defined on more general probability spaces, see Delbaen (2000).

We now list a number of properties that risk measures and risk measurement methods may satisfy. We start with individual models. So, let a model m be given, and let ρ be a risk measure defined on m. Since m will be fixed for the moment, we write $\mathcal{R}(m)$ simply as $\mathcal{R}$. Some properties of interest will be stated as axioms. In the first axiom we also assume that a reference risk $N \in \mathcal{R}$ has been given.

Axiom 4.1 (Translation invariance) For all $X \in \mathcal{R}$ and $\tau \in \mathbb{R}$, we have $\rho(X + \tau N) = \rho(X) - \tau$.

Adding (subtracting) an initial investment of size τ in the reference asset N decreases (increases) the risk measure ρ by τ. Therefore, τ can be interpreted as the amount of the sweetener added to the risk X to make it more acceptable (or less, in case τ is negative).

Axiom 4.2 (Monotonicity) For all X and $Y \in \mathcal{R}$ with $X \leq Y$, we have $\rho(X) \geq \rho(Y)$.

It seems natural to assign a higher value to risks that always have a lower payoff. Note that the axiom of monotonicity rules out the commonly used mean-variance measure $\rho(X) = -\mathbb{E}_{\mathbb{P}}(X) + \gamma \mathbb{V}\mathrm{ar}_{\mathbb{P}}(X)$, where γ is a risk aversion parameter. The VaR measure, on the other hand, is monotonic.

Axiom 4.3 (Positive homogeneity) For all $X \in \mathcal{R}$ and $\lambda \geq 0$, $\rho(\lambda X) = \lambda \rho(X)$.

Again, this axiom is satisfied by VaR. The homogeneity axiom may be considered reasonable as a local approximation, or when size effects (due, for instance, to liquidity risk or to regulatory constraints) are taken into account in the future net worth of a position.

Axiom 4.4 (Subadditivity) For all X and $Y \in \mathcal{R}$, we have $\rho(X+Y) \leq \rho(X)+\rho(Y)$.

If the risk measure ρ satisfies the subadditivity property, the risk manager/supervisor is sure that the sum of two separate risks X and Y can be estimated conservatively by the sum of the risk measures of the separate risks. If a risk measure does not satisfy the subadditivity property, a risk might be disguised by splitting it up. The

VaR measure does not satisfy the subadditivity property (see Artzner et al. (1999) for a counterexample).

In some situations the risk of a portfolio might be increasing in a nonlinear way with the position size (for example due to increasing liquidity risk). This led Föllmer and Schied (2002) to introduce the axiom of convexity.

Axiom 4.5 (Convexity) For all X and $Y \in \mathcal{R}$, and $\lambda \in [0,1]$, we have $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$.

Convexity means that diversification does not increase risk.

The above axioms can be transferred to risk measurement methods in a straightforward way. We shall say that a risk measurement method RMM defined on a model class $\mathcal{M}$ satisfies Axiom i ($i = 1, \ldots, 5$) if for each $m \in \mathcal{M}$ the risk measure RMM_m on m satisfies Axiom i with $\mathcal{R} = \mathcal{R}(m)$. In the case of the translation invariance axiom, it is assumed that a reference product in $\mathcal{X}(\mathcal{M})$ is given.

The fact that VaR does not satisfy the subadditivity or convexity property is often seen as a disadvantage of this risk measurement method; see Artzner et al. (1999) and Acerbi and Tasche (2002) for a more extensive discussion. Alternative risk measures have been proposed that do satisfy the desirable subadditivity property. Artzner et al. (1997) introduced the notion of coherent risk measures. Their ideas were formalized in Artzner et al. (1999), Artzner (1999), and Delbaen (2000).

Definition 4.8 A *coherent risk measure* is a risk measure that satisfies the axioms of translation invariance, monotonicity, subadditivity, and positive homogeneity.

Definition 4.9 A *convex risk measure* is a risk measure that satisfies the axioms of translation invariance, monotonicity, and convexity.

The above definitions can immediately be extended to produce the notion of a *coherent risk measurement method* and a *convex risk measurement method*, respectively.

The five axioms still allow many measurement methods, so even when one decides to use a coherent or convex measure one needs further considerations to arrive at a specific method. An example of a coherent risk measurement method is the worst conditional expectation (WCE). Contrary to VaR, this measure takes the size of losses under the VaR limit into account. Therefore, it is not possible to increase the expected return of a portfolio under WCE restrictions by taking extremely risky bets with a very low probability of a very high loss.

Definition 4.10 (Worst Conditional Expectation (WCE)) Let a model class $\mathcal{M}$ be given. The *worst conditional expectation* method with *reference product* $N \in \mathcal{X}(\mathcal{M})$ and *level* $p \in (0,1)$ assigns to a model $m = (\Omega, \mathcal{F}, \mathbb{P}) \in \mathcal{M}$ the risk measure WCE_m given by

$$\mathrm{WCE}_m : \mathcal{R}(m) \ni X \mapsto -\inf_{A\in\mathcal{F},\, \mathbb{P}(A)>p} \mathbb{E}_{\mathbb{P}}\left[X/N_m \mid A\right] \in I\!R \cup \{\infty\}. \tag{4.3}$$

It is a straightforward exercise to show that WCE satisfies the axioms of translation invariance, monotonicity, positive homogeneity, and subadditivity. Though WCE has nice theoretical implications, it is difficult to compute in practice. A practically usable coherent risk measure is the expected shortfall as given in Acerbi and Tasche (2002).

Definition 4.11 (expected shortfall (ES)) The *expected shortfall method* with *reference asset* N and *level* $p \in (0,1)$ assigns to a model $m = (\Omega, \mathcal{F}, \mathbb{P})$ the risk measure ES_m given by

$$\begin{aligned}\mathrm{ES}_m : \mathcal{R}(m) \ni X \mapsto -\frac{1}{p}\big(&\mathbb{E} X \mathbf{I}_{(-\infty, Q_p(X/N_m)]} \\ &+ Q_p\left(X/N_m\right)\left(p - \mathbb{P}\left(X/N_m \le Q_p\left(X/N_m\right)\right)\right)\big) \in I\!R \cup \{\infty\}. \end{aligned} \tag{4.4}$$

4.3 Model risk

Market risk measures are typically based on a class of scenarios together with a base probability measure; both items are provided by a model m. At a higher level, however, there is uncertainty about which model to use. A financial institution's perception of market risk can deviate substantially from the actual market risk due to the fact that the actual dynamics are insufficiently represented by the model dynamics. Due to the use of an incorrect model, the financial institution may accept risks that it would find unacceptable in case it would know the actual dynamics. The risk associated to the mismatch between model dynamics and actual dynamics is called *model risk*.

In the sequel we propose a framework to quantify this model risk. Since the true dynamics are unknown, it makes sense to form a set of alternative dynamics $\mathcal{K}$ (containing a nominal model m) which is likely to contain the true dynamics. A natural candidate for a model risk measure is the difference between the worst-case risk measure among all models in the neighborhood $\mathcal{K}$ and the risk measure under the dynamics of the nominal model m. If the market risk measurement method is translation invariant, the difference between these two quantities gives the extra position in the reference product which has to be added to the market risk measure of the nominal model to make the risk acceptable, even under the worst case dynamics. In the next section, this intuition is formalized.

4.3.1 Measuring model risk

Suppose that the financial institution uses a risk measurement method RMM to assess the acceptability of a product (portfolio) Π. In model m, the risk of the product Π is computed as $\text{RMM}_m(\Pi_m)$. To take into account model uncertainty, we take a set of alternative dynamics $\mathcal{K}$ around m and compute the *worst-case market risk measure* (with respect to $\mathcal{K}$), which is given by $\sup_{k\in\mathcal{K}} \text{RMM}_k(\Pi_k)$. Model risk may now be quantified as follows.

Definition 4.12 (Model risk measure)[4] Let $\mathcal{M}$ be a class of models, let m be a model in $\mathcal{M}$, and let $\mathcal{K}$ be a subset of $\mathcal{M}$ containing m. Furthermore, let Π be a product defined on $\mathcal{M}$ and let RMM be a risk measurement method satisfying the axiom of translation invariance for $\mathcal{M}$. The *model risk* associated to the method RMM of product Π, with respect to the *nominal model* m and the *tolerance set* $\mathcal{K}$, is given by

$$\phi_{\text{RMM}}(\Pi, m, \mathcal{K}) = \sup_{k\in\mathcal{K}} \text{RMM}_k(\Pi_k) - \text{RMM}_m(\Pi_m). \tag{4.5}$$

[4]The case where $\text{RMM}_m(\Pi_m) = \infty$ is uninteresting since the financial institution will never accept the product Π in its portfolio.

Artzner et al. (1999), Delbaen (2000), Föllmer and Schied (2002), and others use one particular model m to compute the market risk measure $\mathrm{RMM}_m(\Pi_m)$. With the definition above we extend their risk measurement framework by including model risk. The amount $\phi_{\mathrm{RMM}}(\Pi, m, \mathcal{K})$ can be thought of as a model reserve that should be held to cover the worst-case dynamics of $\mathcal{K}$. This interpretation depends on the translation invariance axiom of the risk measurement method which is therefore crucial in the definition. Consider, for example, value-at-risk. From empirical data we can determine whether the VaR limit given by a nominal model is exceeded as often as predicted or more often. If the model is accurate in predicting the VaR limit we would like to set a small model reserve. On the other hand, we want to set a large model reserve in case the model does a poor job predicting the VaR limit. Adding the model reserve $\phi_{\mathrm{RMM}}(\Pi, m, \mathcal{K})$ to the nominal market risk measure $\mathrm{RMM}_m(\Pi_m)$ gives a total risk measure equal to $\sup_{k \in \mathcal{K}} \mathrm{RMM}_k(\Pi_k)$. In appendix 4.3.2 we illustrate the procedure for coherent risk measures, in particular, the WCE and SPAN. The size of the model reserve (and thereby the total risk measure) is controlled by the size of $\mathcal{K}$. In the next section, we discuss the determination of $\mathcal{K}$ and the dependence of the model reserve on model accuracy in more detail.

The model risk measure that we have defined may have some desirable properties depending on the market risk measurement method from which it has been derived. Of the properties of the risk measurement method chosen translation invariance is of special importance. It allows for the intuitive invariance property of the model risk measure.

Theorem 4.1 (Invariance) Let RMM be a risk measurement method that is translation invariant with respect to a reference product N. Then the model risk measure associated to RMM is invariant in the sense that

$$\phi_{\mathrm{RMM}}(\Pi + \tau N, m, \mathcal{K}) = \phi_{\mathrm{RMM}}(\Pi, m, \mathcal{K}) \tag{4.6}$$

for all $\tau \in \mathbb{R}$.

Proof. Take $\tau \in \mathbb{R}$. We have

$$\begin{aligned}
\phi_{\mathrm{RMM}}(\Pi + \tau N, m, \mathcal{K}) &= \sup_{k\in\mathcal{K}} \mathrm{RMM}_k\left(\Pi_k + \tau N_k\right) - \mathrm{RMM}_m\left(\Pi_m + \tau N_m\right) \\
&= \sup_{k\in\mathcal{K}} \mathrm{RMM}_k\left(\Pi_k\right) - \tau - \mathrm{RMM}_m\left(\Pi_m\right) + \tau \\
&= \phi_{\mathrm{RMM}}(\Pi, m, \mathcal{K}).
\end{aligned}$$

■

The addition of a constant payoff should not alter the model risk, since it is model independent. Another way to look at it is that the constant payoff can be fully hedged by a position in the reference product. In a similar way, one can easily prove that for positive homogeneous RMM the model risk measure is positively homogeneous.

The model risk measure does not, in general, satisfy the monotonicity, subadditivity, and convexity property. However, if the underlying market risk measurement method satisfies any of these properties, these properties hold for what might be called *total market risk*, viz. the sum of nominal market risk and model risk. For example, in case of subadditivity, this can be seen from the fact that total market risk is given by the formula $\sup_{k\in\mathcal{K}} \mathrm{RMM}_k(\Pi_k)$, and from the general fact that $\sup_i(a_i+b_i) \le \sup_i(a_i)+\sup_i(b_i)$. As noted above, the reason why market risk measures are often required to be subadditive is to prevent companies, trading desks, etc. from covering up large risks by splitting them into separate positions that do satisfy the risk criteria. If total market risk is reported, then subadditivity of this risk measure is sufficient for this.

We choose a worst-case approach to quantify model risk. An alternative would be a Bayesian approach, in which the model risk measure is a weighted average of risk measures according to some prior. Depending on its risk attitude, the financial institution can give more weight to unfavorable dynamics. However, the choice of a prior is difficult and arbitrary. In a worst-case approach, one only needs to specify the tolerance set $\mathcal{K}$; this may be seen as an acknowledgment of the restrictions of statistical modeling in the face of limited data and limited understanding of the true dynamics.

4.3.2 Model risk for popular risk measures

In this section we illustrate the model risk measure for coherent risk measures and, in particular, the worst conditional expectation and SPAN.

A coherent risk measure method ρ_m for model $m = (\Omega, \mathcal{F}, \mathbb{P})$ can be written in the form[5]

$$\rho_m(\Pi) = \sup_{\mathbb{Q}\in\mathcal{P}(m)} \mathbb{E}_{\mathbb{Q}}[\Pi].$$

Different choices of $\mathcal{P}(m)$ produce different risk measures. We specify $\mathcal{P}(m)$ for WCE and SPAN.

Example 4.1 (WCE) Given is a model m with a base probability $\mathbb{P}$, $m = (\Omega, \mathcal{F}, \mathbb{P})$. The class of models $\mathcal{P}(m)$ is given by

$$\mathcal{P}_{\text{WCE}}(m) = \{\mathbb{P}(.\mid A) \mid \mathbb{P}(A) > \alpha\}.$$

Example 4.2 (SPAN) Given is a model m with a base probability $\mathbb{P}$, $m = (\Omega, \mathcal{F}, \mathbb{P})$, with $\Omega = \omega_1, ..., \omega_k$ and $\mathcal{F} = 2^\Omega$, where ω_i denotes a scenario. Let

$$\mathbb{P}(\omega_1) = ... = \mathbb{P}(\omega_k) = \frac{1}{k}.$$

The SPAN method is such that[6]

$$\mathcal{P}_{\text{SPAN}}(m) \subset \{\mathbb{Q} \mid \mathbb{Q} << \mathbb{P}\}.$$

Note the difference between $\mathcal{P}(m)$ and $\mathcal{K}$ in Def. 4.12. $\mathcal{P}(m)$ is a set of probability measures based on *one* base probability measure $\mathbb{P}$ to compute a coherent market risk measure. However, $\mathcal{K}$ denotes a set of models. The models in this set can have different measurable spaces and *different* base probability measures. The model risk measure for a general coherent risk measure is then given by

$$\phi_{\text{RMM}}(\Pi, m, \mathcal{K}) = \sup_{k\in\mathcal{K}} \sup_{\mathbb{Q}\in\mathcal{P}(k)} \mathbb{E}_{\mathbb{Q}}[\Pi] - \sup_{\mathbb{Q}\in\mathcal{P}(m)} \mathbb{E}_{\mathbb{Q}}[\Pi].$$

[5]For simplicity we use the definition given by Artzner et al. (1999). The definition for general probability spaces is given in Delbaen (2000).

[6]Of course, any probability measure $\mathbb{P}^*$ equivalent to $\mathbb{P}$ could serve as a base probability measure for $\mathcal{P}_{\text{SPAN}}$ (see SPAN (1995) for details or Artzner et al. (1999) for a summary).

4.3.3 Decomposition of Model Risk

The exposition given in section 4.3.1 was rather general. We did not specify a model m or a set of alternative models $\mathcal{K}$. In this section we discuss some possible choices for the set of alternative dynamics $\mathcal{K}$. In practice, one starts with a (usually parametric) model class, say $\mathcal{M}(\Theta) \equiv \{(\Omega, \mathcal{F}, \mathbb{P}_\theta) : \theta \in \Theta\} \subset \mathcal{M}$, where Θ denotes the parameter space.[7] Using an estimation or calibration procedure, a particular element $m(\hat{\theta})$ is chosen from $\mathcal{M}(\Theta)$. Even if the actual dynamics, say m_0, belong to the parametric model class $\mathcal{M}(\Theta)$, that is $m_0 = m(\theta_0)$ for some $\theta_0 \in \Theta$, the financial institution faces the risk of selecting the wrong element $m(\hat{\theta})$. This risk is called *model risk due to estimation error*. To define a neighborhood of plausible values around $m(\hat{\theta})$, one typically uses confidence regions. Specifically, we can place a confidence region around the estimator $\hat{\theta}$ for θ_0 to define some neighborhood around $m(\hat{\theta})$. Depending on a chosen level α we take a $(1-\alpha)$ confidence region around $\hat{\theta}$.[8] In this way we arrive at a set of alternative models of the following form:

$$\mathcal{K}(\alpha) = \left\{ m(\theta) \in \mathcal{M}(\Theta) : \theta \in CI_{1-\alpha}(\hat{\theta}) \right\} \tag{4.7}$$

In situations where one is interested in a specific market risk measurement method RMM and a specific product Π, an alternative approach which focuses more directly on the given situation is to use the set $\mathcal{K}$ defined by

$$\mathcal{K}(\alpha) = \left\{ m(\theta) \in \mathcal{M}(\Theta) : \mathrm{RMM}_{m(\theta)}\left(\Pi_{m(\theta)}\right) \in CI_{1-\alpha}\left(\mathrm{RMM}_{m(\hat{\theta})}\left(\Pi_{m(\hat{\theta})}\right)\right) \right\}. \tag{4.8}$$

We define *model risk due to estimation error*, or simply *estimation risk*, as the model risk that is obtained from a tolerance set derived from confidence regions in within the model class.

Now let us consider the situation where the actual dynamics may not belong to $\mathcal{M}(\Theta)$. The risks that we are considering are real-valued random variables and so a natural idea is to work on the basis of the associated distribution functions. Suppose that a cumulative distribution function $\hat{F}(x)$ has been obtained by some nonparametric estimation method. This allows us to define a tolerance set $\mathcal{K}$ depending on

[7] In the usual parametric case the parameter space is a subset of $\mathbb{R}^k$.

[8] $CI_{1-\alpha}(\hat{\theta})$ denotes the $(1-\alpha)$-confidence interval for θ_0.

confidence level α in the following way:

$$\mathcal{K}(\alpha) = CI_{1-\alpha}(\hat{m}) := \left\{ m = (\Omega, \mathcal{F}, \mathbb{P}) : F(x) = \mathbb{P}((-\infty, x]) \in \left[\hat{F}(x) \pm \frac{k_{\alpha/2}}{\sqrt{n}} \right] \quad \forall x \in \mathbb{R} \right\},$$

where $k_{\alpha/2}$ is the critical value of the Kolmogorov-Smirnov statistic.[9] As above, one may also define tolerance sets that are more specifically tied to a given risk measurement method and a given product. Along this line, one may estimate $\mathrm{RMM}_m(\Pi_m)$ first and define a tolerance set based on a confidence region $CI_{1-\alpha}$ for the estimate

$$\mathcal{K}(\alpha) = \{ m : \mathrm{RMM}_m(\Pi_m) \in CI_{1-\alpha} \}.$$

In general, we can determine tolerance sets that are restricted to a (parametric) subclass $\mathcal{M}(\Theta)$ or that are not restricted to such a model class (which we will refer to as unrestricted, in the sequel). As above, one may define model risk due to estimation error as the model risk restricted to the model class $\mathcal{M}(\Theta)$. The amount that has to be added to arrive at the model risk determined from the unrestricted method may be termed *model risk due to misspecification* or simply *misspecification risk*. In other words, if $\mathcal{K}_r$ is the restricted tolerance set and $\mathcal{K}_u$ is the unrestricted one, then we define the misspecification risk for a given product Π as

$$\phi_{\mathrm{RMM}}(\Pi, \mathcal{K}_r, \mathcal{K}_u) = \sup_{k \in \mathcal{K}_u} \mathrm{RMM}_k(\Pi_k) - \sup_{k \in \mathcal{K}_r} \mathrm{RMM}_k(\Pi_k). \tag{4.9}$$

However, the quantity defined above may in some cases be less than zero, whereas we would prefer to define misspecification risk in such a way that it is always non-negative. To achieve this with the above definition, we have to make sure that the set $\mathcal{K}_r$ is nested in $\mathcal{K}_u$. In case the misspecification risk in (4.9) turns out to be negative, one could argue that that the unrestricted set $\mathcal{K}_u$ is too small, i.e., based upon too low a confidence level. Therefore, we use a family $\{\mathcal{K}_u(\gamma)\}$ of tolerance sets parameterized by the confidence level γ. For a given confidence level α and a given tolerance set $\mathcal{K}_r$, which may have been selected on the basis of the same

[9] Formally, the use of the Kolmogorov-Smirnov statistic requires $(\Omega, \mathcal{F})$ to be model independent. For the empirical applications we have in mind, this is not a restriction. Alternative uniform confidence bounds around a nonparametric distribution may be obtained from the Cramér-von Mises statistic or the Kuiper statistic (see, for example, Shorack and Wellner (1986)).

confidence level, we then take $\mathcal{K}_u = \mathcal{K}_u(\beta)$ where β is defined by[10]

$$\beta = \min(\alpha, \sup\{\gamma \in (0,1) : \mathcal{K}_r \in \mathcal{K}_u(\gamma)\}). \tag{4.10}$$

The analogs of Theorem 4.1 (invariance) and positive homogeneity can easily be shown to hold for both estimation risk and misspecification risk separately. Proofs follow the lines of the proofs of the cited results.

4.4 Regulatory Capital

One of the most important tasks of a risk management department is to compute the risk of the portfolio of the financial institution. In this section we illustrate how the methodology can be used in portfolio risk management, using two financial time series, the Standard and Poor's 500 and the £/ $ exchange rate. The data were obtained from Thomson Datastream (definitions and sources of the data can be found in Appendix B).

The Bank for International Settlements (BIS) has suggested to set risk-based capital requirements which are closely related to the value-at-risk methodology. Here, we show first how the model risk measurement approach can be taken into account for a (simple) value-at-risk methodology, followed by an example using the expected shortfall. The derivation of the formulas and the asymptotic distributions used in these examples are presented in Appendix A.

4.4.1 Examples

Example 4.3 Value-at-Risk

We have available a data set of returns $\left(h_1^T, ..., h_n^T\right)$ of the portfolio under consideration for a period of length nT years ($nT = 20$, $T = 1/252$ (one day)). An elementary VaR model assumes that the data is a realization of a random sample

[10] An alternative way to ensure nesting is to form convex combinations. Note that, in a context in which we are concerned with a specific product, it is reasonable to identify models with the cumulative distribution functions induced by the given product, and in this way it is indeed possible to consider convex combinations of models. The nesting property can then be guaranteed by replacing the set $\mathcal{K}_u$ by the convex hull of $\mathcal{K}_r$ and the original $\mathcal{K}_u$. However, our proposal seems to be more transparent.

$\left(H_1^T, ..., H_n^T\right)$ where $H_j^T \sim \mathcal{N}\left(\mu T, \sigma^2 T\right)$ for $j = 1, .., n$ where μ and σ^2 denote annualized mean and variance, respectively. We estimate μ by $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n h_i^T$ and σ by $\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^n \left(h_i^T - \hat{\mu}\right)}$. Let $\theta = (\mu, \sigma)$. The model class $\mathcal{M}$, with typical element $m(\theta) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_\theta)$ where $\mathbb{P}_\theta\left((-\infty, x]\right) = \Phi\left(\frac{\log x - \mu}{\sigma}\right)$, belongs to the class of lognormal distributions.

Let $X_0 \in \mathbb{R}$ denote the (model independent) initial capital, and let $\Pi \in \mathcal{X}(\mathcal{M})$ denote the portfolio at time T. To compute the worst cases, we follow the approach of focusing directly on the given risk measurement method (VaR in this case) and the given product, as discussed in 4.3.3 above.

First, assume that asset returns are normally distributed and let $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$ be the estimate of the parameter $\theta = (\mu, \sigma)$. We shall call $m(\hat{\theta})$ the *nominal parametric model*. The corresponding value-at-risk of portfolio Π at level p is given by

$$\mathrm{VaR}_{m(\hat{\theta})}\left(\Pi_{m(\hat{\theta})}\right) = -X_0 \exp\left(z_p \hat{\sigma}\sqrt{T} + \hat{\mu} T\right), \tag{4.11}$$

where z_p denotes the p^{th} quantile of the (standard) normal distribution. Still assuming normality of asset returns, the parametric worst-case value-at-risk is the lower bound of the $(1-\alpha)$ confidence interval around $\mathrm{VaR}_{m(\hat{\theta})}\left(\Pi_{m(\hat{\theta})}\right)$. This lower bound is given (based on an asymptotic approximation) by

$$\mathrm{VaR}_{m(\hat{\theta})}\left(\Pi_{m(\hat{\theta})}\right) - z_{\alpha/2}\sqrt{\Sigma_{\mathrm{VaR}}/n}, \tag{4.12}$$

where

$$\Sigma_{\mathrm{VaR}} = T\,\mathrm{VaR}_{m(\hat{\theta})}\left(\Pi_{m(\hat{\theta})}\right)^2 \left[\sigma^2\left(1 + T/4\right) + z_p \sigma^3 \sqrt{T} + z_p^2 \sigma^2/2\right].$$

Nonparametric versions of VaR may be computed on the basis of the empirical distribution function, F_n. We denote by m_n the model $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_n)$ where $\mathbb{P}_n$ is given by $\mathbb{P}_n((-\infty, x]) = F_n(x)$. The nominal empirical value-at-risk is given by

$$\mathrm{VaR}_{m_n}\left(\Pi_{m_n}\right) = -X_0 \exp\left(h^T_{n(\lfloor pn \rfloor + 1)}\right), \tag{4.13}$$

where $n\,(i)$ denotes the i^{th} order statistic of $\left(h_1^T, ..., h_n^T\right)$ and $\lfloor a \rfloor$ is the largest integer that is less than or equal to a. Finally, the worst-case empirical VaR is the lower bound of the (nonparametric) confidence interval around $\mathrm{VaR}_{m_n}(\Pi_{m_n})$, which may be computed as

$$\mathrm{VaR}_{m_n}(\Pi_{m_n}) - z_{\beta/2}\sqrt{\frac{p(1-p)}{nf^2(F_n^{-1}(p))}} \tag{4.14}$$

where β is as defined in (4.10) and where $f(x)$ can be estimated using, for instance, the Rosenblatt-Parzen kernel estimator.[11]

Example 4.4 Expected Shortfall

For the empirical part we use expected shortfall as well as value-at-risk, so we repeat the exercise of the previous example for ES. Assume the same setting as before. The nominal parametric ES (at level p) under normality of the asset returns is given by

$$\mathrm{ES}_{m(\hat{\theta})}\left(\Pi_{m(\hat{\theta})}\right) = -\frac{1}{p}\exp\left(\hat{\mu}T + \tfrac{1}{2}\hat{\sigma}^2T\right)\Phi\left(z_p - \hat{\sigma}\sqrt{T}\right). \tag{4.15}$$

The worst-case parametric ES may be computed as

$$\mathrm{ES}_{m(\hat{\theta})}\left(\Pi_{m(\hat{\theta})}\right) - z_{\alpha/2}\sqrt{\Sigma_{ES}/n} \tag{4.16}$$

where (writing $\phi = \Phi'$)

$$\Sigma_{\mathrm{ES}_{m(\theta)}} = T\Psi^2_{\mathrm{ES}}\left[\sigma^2 + \left(T\sigma - \sqrt{T}\frac{\phi}{\Phi}\left(z_p - \sigma\sqrt{T}\right)\right)\sigma^3 + \left(\sqrt{T}\sigma - \frac{\phi}{\Phi}\left(z_p - \sigma\sqrt{T}\right)\right)^2\sigma^2/2\right], \tag{4.17}$$

where $\Psi_{(ES)} = \mathrm{ES}_{m(\hat{\theta})}\left(\Pi_{m(\hat{\theta})}\right)$ and $\frac{\phi}{\Phi}(x) = \frac{\phi(x)}{\Phi(x)}$.

The empirical ES can be computed by

$$\mathrm{ES}_{m_n}(\Pi_{m_n}) = -\frac{1}{\lfloor np\rfloor + 1}\sum_{i=1}^{n} X_0\exp\left(h_j^T\right)\mathbf{I}_{\{(-\infty,\mathrm{VaR}_{m_n}(\Pi_{m_n})]\}}\left(h_j^T\right). \tag{4.18}$$

[11] In our applications below we have approximately normal data and so we do bandwidth selection by taking $h = 1.06sn^{-1/5}$, which is the optimal bandwidth in case of a normal $\mathcal{N}(\mu,\sigma^2)$ distribution, where s denotes the usual estimate for σ.

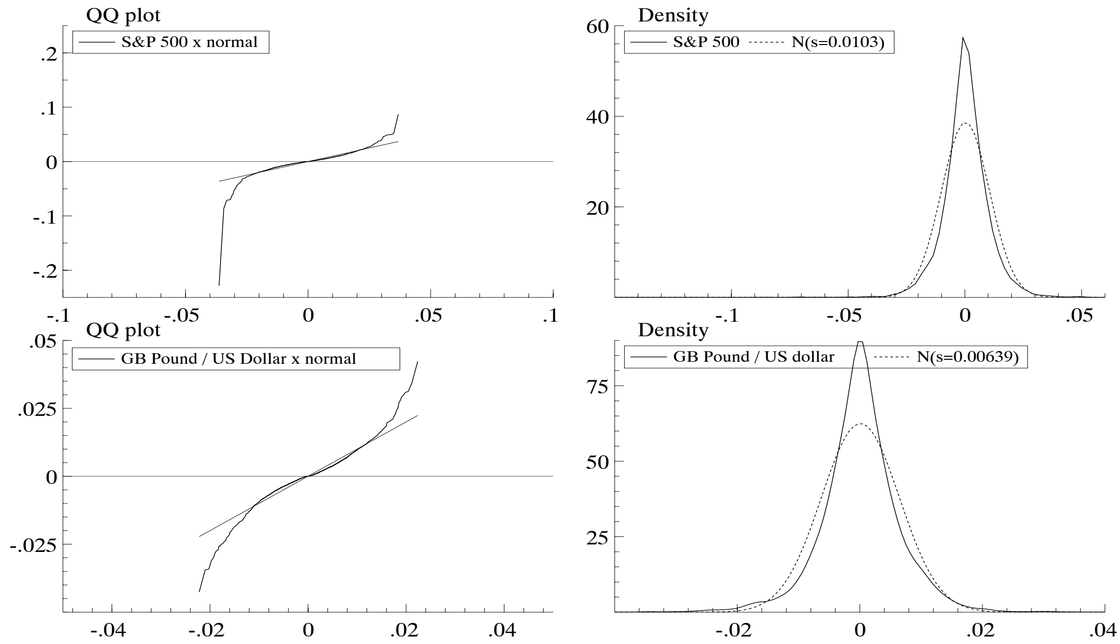


Figure 4.1: **Data descriptions**
QQ-plot and density comparison of the normal density with nonparametric density estimate (using a the Rosenblatt-Parzen kernel estimator with Gaussian kernel and bandwidth $h = 1.06sn^{-1/5}$) of the daily (total) returns of the S&P 500 and £/ \$ exchange rate. The data periods are 26-10-'81 – 29-04-'03 for the S&P 500 and 03-01-'86 – 29-04-'03 for the £ / \$ exchange rate.

The empirical worst-case ES can be computed as

$$\mathrm{ES}_{m_n}\left(\Pi_{m_n}\right) - z_{\alpha/2}\sqrt{\Sigma_{\mathrm{ES}_{m_n}}/n} \tag{4.19}$$

where

$$\begin{aligned} \Sigma_{\mathrm{ES}_{m_n}} &= \frac{1}{p}\mathbb{E}\left[Y^2 \mid Y \leq \mathrm{VaR}_{m_n}\left(\Pi_{m_n}\right)\right] - \mathrm{ES}^2_{m_n}\left(\Pi_{m_n}\right) \\ &- \left(1 - \frac{1}{p}\right)\mathrm{VaR}^2_{m_n}\left(\Pi_{m_n}\right) + \left(2 - \frac{2}{p}\right)\mathrm{ES}_{m_n}\left(\Pi_{m_n}\right)\mathrm{VaR}_{m_n}\left(\Pi_{m_n}\right). \end{aligned}$$

4.4.2 Empirical results

Based on these examples, we now illustrate the methodology. In addition to the Gaussian models described in the examples we also investigate a GARCH$(1,1)$ model with Gaussian innovations that should be more capable of capturing time varying risk.[12] Figure 4.1 shows the normal density with variance equal to the sample variances of the S&P 500 data and the British pound / US dollar (£/$) exchange rate data and compares this with a nonparametric density[13] estimate of the densities of the S&P 500 and the £/$ exchange rate. We see that the returns from the S&P 500 and an investment in British money market (for a US investor) account exhibit more kurtosis than could be expected on the basis of normally distributed returns. This could be the result of time-varying volatility and, therefore, we use rolling window versions of our models with a window of 2 years. Based on the graphical analysis of Figure 4.1, we expect some misspecification error when calculating the value-at-risk and expected shortfall on the basis of a nominal model assuming normally distributed returns.

Figure 4.2 shows the estimators for the (annualized) mean and volatility of both models. We see that both models predict more or less the same means. The difference between the models is in the predictions of the volatilities. The predictions of the GARCH$(1,1)$ model are as expected much more erratic than those of the Gaussian model.

Next, we investigate the performance of the various models for VaR and ES from a statistical point of view. We start by investigating VaR. Ideally, the frequency of excessive losses (FOEL), i.e., the number of days at which the loss exceeds the predicted VaR, should be close to the VaR levels. As a benchmark we choose the 1% level for VaR, since this is the quantile required by BIS (see Basel Committee on Banking Supervision (1996a)). In Tables 4.1 and 4.2 we present the results of a one-sided FOEL test with 95% confidence intervals. We denote by n the number of

[12] Berkowitz and O'Brien (2002) find that an ARMA$(1,1)$-GARCH$(1,1)$ model with Gaussian innovations does a good job in forecasting value-at-risk for their portfolios of actual investment banks. Since we did not find any statistically significant ARMA structure in our data, we restricted the model to a GARCH$(1,1)$. For more advanced volatility estimation methods see, for example, Eberlein, Kallsen and Kristen (2003).

[13] In view of the approximate normality of the data, the bandwidth h has been set equal to $h = 1.06\hat{\sigma}n^{-1/5}$ which is the optimal bandwidth selection for normally distributed data.

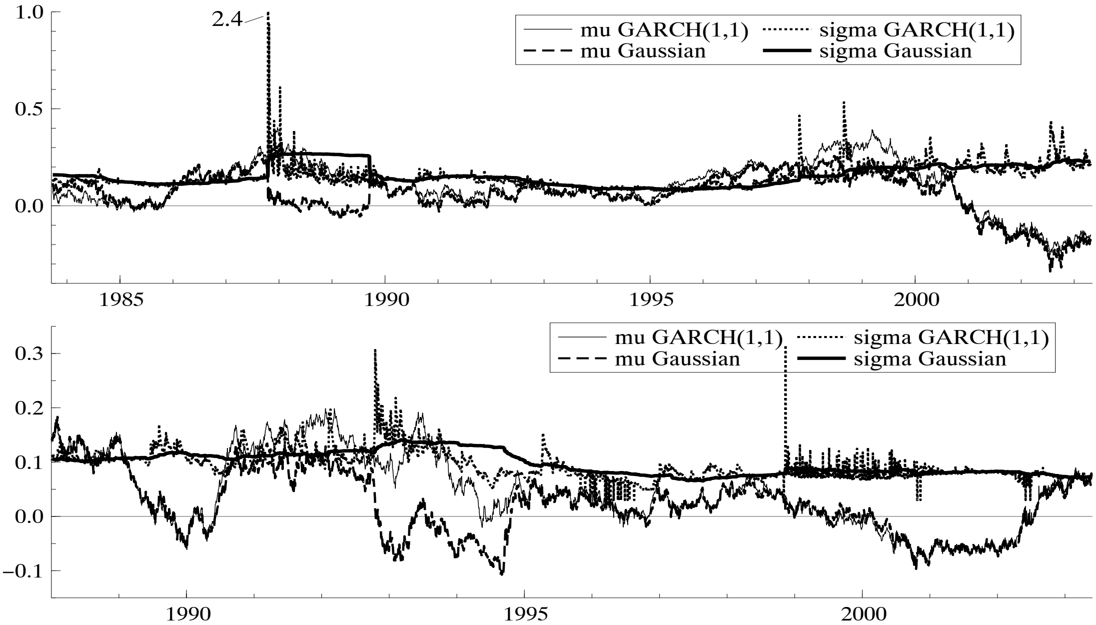


Figure 4.2: **Parameter estimates**
The upper panel displays the parameter estimates of the mean and volatility in the S&P 500 market for both the Gaussian and the GARCH(1, 1) model. The lower panel displays the parameter estimates of the mean and volatility in the £/$ FX rate market for both the Gaussian and the GARCH(1, 1) model. For both markets the estimates are using two-year rolling window models. The data runs from 26-10-'81−29-04-'03 for the S&P 500 market and from 03-01-'86−29-04-'03.

Table 4.1: **FOEL test VaR for S&P 500**
FOEL test for Gaussian VaR, GARCH(1, 1) VaR, and non-parametric VaR models and their worst-case equivalents (for definitions, see main text). Daily data on S&P 500 (total return) index from 26-10-'81 to 29-04-'03.

model	VaR level	FOEL	1-sided 95% CI	T	p-value	VaR level hypothesis rejected
Gauss.	2.5%	3.2%	(2.7%; –)	2.9	0.00	yes
Gauss. wc	2.5%	2.6%	(2.1%; –)	0.37	0.35	no
GARCH(1, 1)	2.5%	3.3%	(2.8%; –)	3.1	0.00	yes
GARCH(1, 1) wc	2.5%	3.3%	(2.8%; –)	3.1	0.00	yes
NP	2.5%	3.1%	(2.6%; –)	2.3	0.01	yes
NP wc	2.5%	2.1%	(1.7%; –)	−1.9	0.97	no
Gauss.	1%	2.0%	(1.6%; –)	5.2	0.00	yes
Gauss. wc	1%	1.7%	(1.4%; –)	3.6	0.00	yes
GARCH(1, 1)	1%	2.0%	(1.6%; –)	5.2	0.00	yes
GARCH(1, 1) wc	1%	1.7%	(1.6%; –)	5.1	0.00	yes
NP	1%	1.6%	(1.2%; –)	3.3	0.00	yes
NP wc	1%	1.0%	(0.7%; –)	−0.45	0.67	no

days in the backtesting period, by f the number of times the VaR level has been exceeded, and by $1-p$ the predicted level of VaR (2.5% or 1% in our case). The test statistic of the FOEL test (see, for example, Kupiec (1995)) is given by

$$T = \sqrt{n}\frac{f/n - p}{p(1-p)} \tag{4.20}$$

The results indicate that the Gaussian rolling window and GARCH(1, 1) VaR models are strongly rejected both in case of the S&P 500 data and in case of the £/$ data. For both the S&P 500 and the £/$ exchange rate, taking estimation risk into account seems sufficient in case of the 2.5% level for the Gaussian model. In case of the 1% level, however, taking estimation risk into account does not prevent the VaR limit from being exceeded too often. If we take misspecification risk into account by looking at the non parametric worst-case model, the number of times the VaR limit is crossed does not exceed the level predicted by the model in a statistically significant way.

Table 4.2: **FOEL test VaR for £/ $ FX rate**
FOEL test for Gaussian VaR, GARCH(1,1) VAR, and non-parametric VaR models and their worst-case equivalents (for definitions, see main text). Daily data on £/ $ from 03-01-'86 to 29-04-'03.

model	VaR level	FOEL	1-sided 95% CI	T	p-value	VaR level hypothesis rejected
Gauss.	2.5%	3.3%	(2.7%; –)	2.7	0.00	yes
Gauss. wc	2.5%	2.8%	(1.2%; –)	1.0	0.15	no
GARCH(1,1)	2.5%	3.3%	(2.7%; –)	2.6	0.01	yes
GARCH(1,1) wc	2.5%	3.2%	(2.6%; –)	2.5	0.01	yes
NP	2.5%	2.6%	(2.1%; –)	0.3	0.40	no
NP wc	2.5%	1.6%	(1.2%; –)	−4.9	1.00	no
Gauss.	1%	1.8%	(1.4%; –)	3.8	0.00	yes
Gauss. wc	1%	3.3%	(0.9%; –)	1.5	0.06	no
GARCH(1,1)	1%	2.0%	(1.6%; –)	4.6	0.00	yes
GARCH(1,1) wc	1%	1.9%	(1.5%; –)	4.4	0.00	yes
NP	1%	1.1%	(0.8%; –)	0.4	0.33	no
NP wc	1%	0.5%	(0.3%; –)	−3.9	1.00	no

In order to test the performance of the models for predicting expected shortfall, we use the recently proposed test for expected shortfall in Chapter 2. For ES we adopt a higher level, namely, the 2.5% level, following arguments given in Chapter 2, which motivate that for an appropriate comparison between VaR and ES, the latter should have a higher level. In order to apply the ES test, the return series $\{h_t\}_{t=1}^n$ is transformed using a probability integral transform to a standardized return series $\{y_t\}_{t=1}^n$ which have distributions, $\{Q_t\}_{t=1}^n$, which under the null hypothesis that the model correctly predicts the ES equal a standard Gaussian distribution, $Q_t = \Phi$, for every t:

$$y_t = \Phi^{-1}\left(\int_{-\infty}^{h_t} p_t(u)\,du\right) = \Phi^{-1}\left(P_t(h_t)\right), \tag{4.21}$$

where $\{P_t\}_{t=1}^n$ denotes the cdf of the model forecast distributions ($\mathcal{N}(\mu_t, \sigma_t)$ in the case of the Gaussian and GARCH(1,1) model, where μ_t and σ_t are model specific).

Table 4.3: **ES tests**
Test of expected shortfall for the nominal Gaussian and GARCH(1, 1) and the nominal non-parametric ES model (for definitions, see main text). The four upper rows present the results of the S&P 500 and the four bottom rows represent the results of the £/ $ FX rate. Daily data on the S&P 500 (total return) index from 26-10-'81 to 29-04-'03. Daily data on the £/ $ FX rate index from 03-01-'86 to 29-04-'03.

market	model	ES level	T	p-value	ES model rejected
S&P 500	Gauss.	5%	−13.5	0.00	yes
	GARCH(1, 1)	5%	−12.6	0.00	yes
	NP	5%	−0.8	0.22	no
	Gauss.	2.5%	−19.5	0.00	yes
	GARCH(1, 1)	2.5%	−17.6	0.00	yes
	NP	2.5%	−0.6	0.28	no
£/ $	Gauss.	5%	−7.3	0.00	yes
	GARCH(1, 1)	5%	−9.8	0.00	yes
	NP	5%	1.0	0.84	no
	Gauss.	2.5%	−9.5	0.00	yes
	GARCH(1, 1)	2.5%	−13.0	0.00	yes
	NP	2.5%	1.58	0.94	no

The test statistic is then given by[14]

$$T = \sqrt{n}\frac{ES_m(Q_t) - ES_m(\Phi)}{V}, \tag{4.22}$$

where V, calculated under the null hypothesis, is given by

$$V = 1/p - (1 - 1/p^2)z_p^2 + (1 + 1/p)\phi(z_p)z_p/p - \phi^2(z_p)/p^2. \tag{4.23}$$

The results indicate that both the rolling window Gaussian model and the GARCH(1, 1) model are strongly rejected. The rolling-window nonparametric model cannot be rejected for both series. Since the worst-case non-parametric expected-shortfall is below the nominal non-parametric expected shortfall, it serves as a lower bound.[15]

[14]For convenience, contrary to definition 4.4, we write ES as a function of the distribution function.

[15]We did not report any results of tests for the worst-case variants of the models for expected shortfall. In order to perform these tests one needs to make assumptions about the tail behavior (for example, a shift in all tail observations.)

One way to investigate the relation between the worst-case risk measure and the risk measure based on the nominal models is in terms of a multiplication factor. We define the multiplication factor for VaR or ES as the ratio between the non-parametric worst-case VaR (ES) based on a 95% confidence interval and the nominal parametric VaR (ES). Plots of the multiplication factors for the Gaussian and GARCH$(1,1)$ models are shown in Figure 4.3. Our definition of the model risk multiplication factors implies that the capital requirements for a bank are the same irrespective of the nominal model used. This seems reasonable, since the amount of regulatory capital should depend on the position that the bank takes and not on the model it uses.

However, the regulator does not know the position of the bank. The information that the regulator gathers is based on the results reported by the banks. Thus, if banks use more accurate models, the regulator has more insight in the risks for the bank and the financial system. Therefore, the regulator wants to provide incentives for the banks to use accurate models. One way to do this is to vary the non-parametric worst case VaR (ES) depending on the backtest. A scheme providing these incentives would be: in case of a rejected model based on backtesting banks should use a model risk multiplication factor based on a (95+penalty)% confidence interval non-parametric VaR (ES), where the penalty increasing with the degree of rejection (higher penalties for lower p-values).

We see in the upper panels of Figure 4.3 that for the Gaussian model in case of 1% VaR multiplication factors of 2 for the S&P 500 and 1.6 for the £/$ exchange rate comfortably cover model risk at the 95% confidence level during the full sample period. In case of the 2.5% ES we find that multiplication factors of 1.7 for the S&P 500 and 1.5 for the £/$ exchange rate are sufficient for the Gaussian model. The lowest BIS multiplication factor for both VaR and ES, a multiplication factor of three, would correspond to a confidence level of about 99.99%. In the right panels of Figure 4.3 we see that for the GARCH$(1,1)$ model the model risk multiplication factors are much higher than for the Gaussian model. This can be explained by the fact that the GARCH$(1,1)$ model responds more quickly to periods of low volatility and then forecasts low values of VaR and ES contrary to the non-parametric worst case.

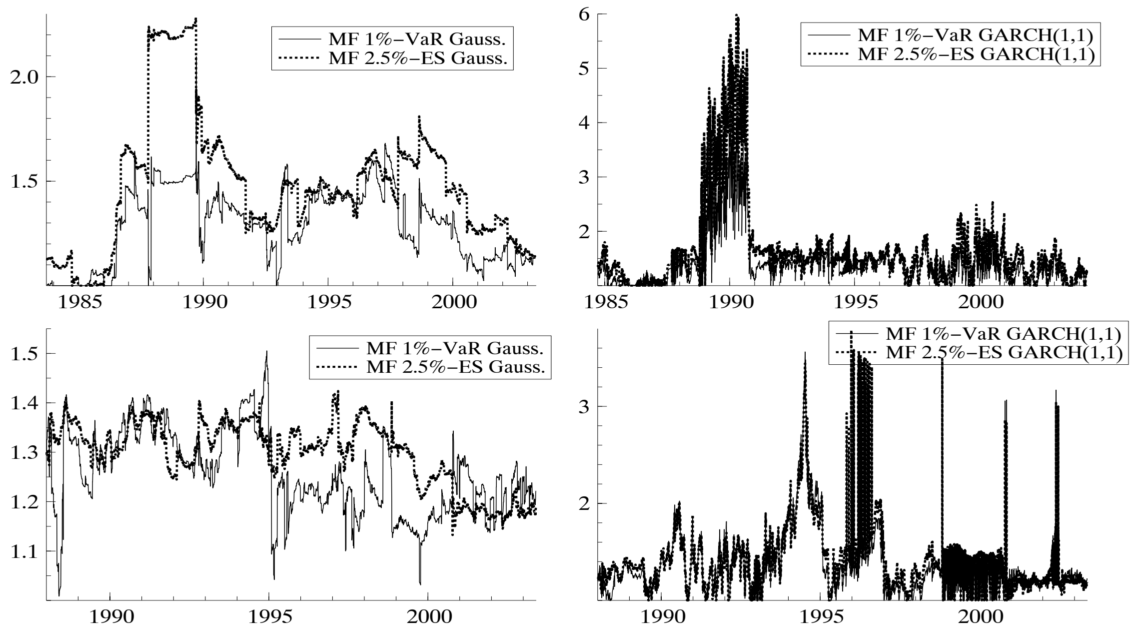


Figure 4.3: **Model risk multiplication factors**
The upper panels display model risk multiplication factors (on the vertical axis) of the 1%-VaR and 2.5%-ES for the S&P 500 during the period 26-10-'83 – 29-04-'03 (left is Gaussian and right GARCH$(1,1)$). The lower panels display model risk multiplication factors of the 1%-VaR and 2.5%-ES for the £/ $ FX rate during the period 03-01-'88 – 29-04-'03 (left is Gaussian and right GARCH$(1,1)$).

In Figures 4.4 and 4.5 we give the capital requirements based on the BIS capital requirements. All requirements are based on investments of $100 in the market. Results can therefore be interpreted as percentages. We have used the BIS backtest procedure (see Basel Committee on Banking Supervision (1996b)) to backtest the Gaussian and the GARCH$(1,1)$ models and to determine the multiplication factors.[16] The capital requirement can then be determined by multiplying the daily value-at-risks by the multiplication factor and $\sqrt{10}$.[17] The capital requirements are compared to the two-week returns. In addition to the BIS capital requirements we plot the capital requirements based on the model risk multiplication factors shown in Figure 4.3. In Figures 4.4 and 4.5 we see that the capital requirements for the GARCH$(1,1)$ model are much more variable than those of the Gaussian model. Furthermore, we see that in normal market conditions the model reserves based on the model risk measures cover the losses safely. The performance in terms of number of exceedances per daily returns, two week returns, and average regulatory capital, is more or less the same for both models as can be seen from Table 4.4. In Table 4.4 we see that the number of exceedances of the two-week VaR and ES's is very small for all capital requirement schemes. Of course, the capital requirements set by the BIS are exceeded least, but they are also very large compared to the model risk multiplication factors. Eventually, the regulator needs to make a trade-off between the cost of exceedance of the capital requirements and the cost of impeding banks in their operations by charging high capital requirements.

4.5 Conclusions

In this paper we have presented a framework to set capital requirements for trading activities in a market, based on the extent to which this market can be reliably modeled. The framework extends the (market) risk framework set out by Artzner et al. (1999) and Delbaen (2000) by considering risk measurement methods for a class of models instead of a risk measure for one particular model. This allows for

[16]Banks only need to do this every three months. However, in this application we did it on a daily basis in order to mitigate the effect of the timing of these three month periods. The BIS capital requirements are therefore not precisely those that would result in practice.

[17]Though the models are backtested using daily VaR, banks should report two-week VaR. The BIS allows the scaling by $\sqrt{10}$. Under the Gaussian model assumptions this would be correct.

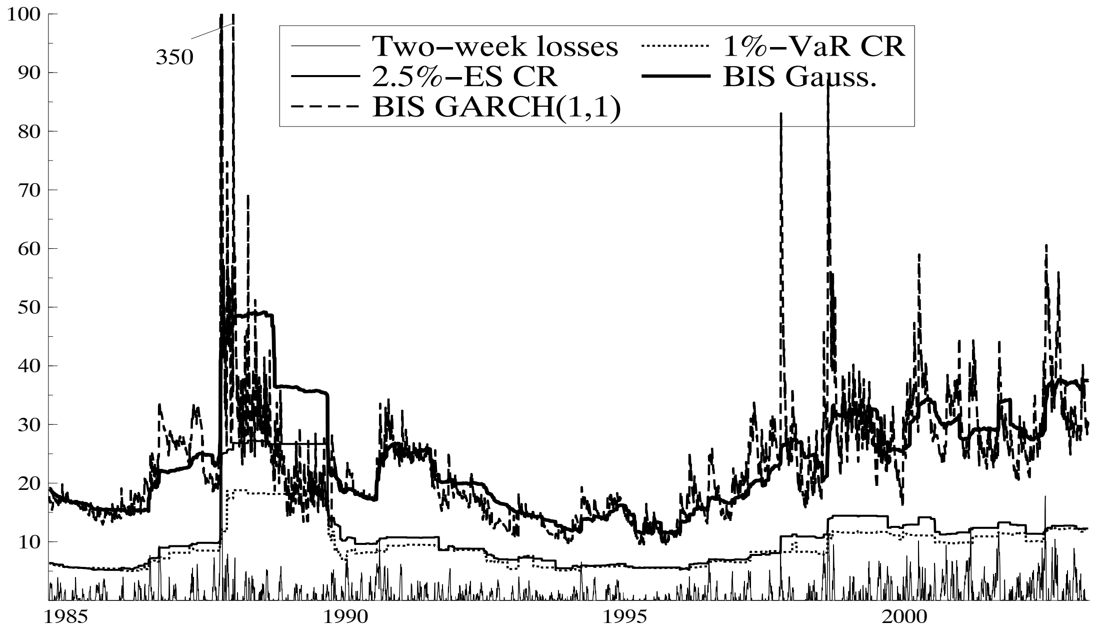


Figure 4.4: **Capital requirements S&P 500**
This figure compares two losses on the S&P 500 to the capital requirements (on the vertical axis) are given for a firm trading in the S&P 500. Given are the capital requirements using the BIS regulation and the capital requirements based on a 1%-VaR and 2.5%-ES model risk multiplication factor. The graph is truncated as in the GARCH$(1,1)$ the BIS CR go up to 350 at the time of 1987 stock market crash.

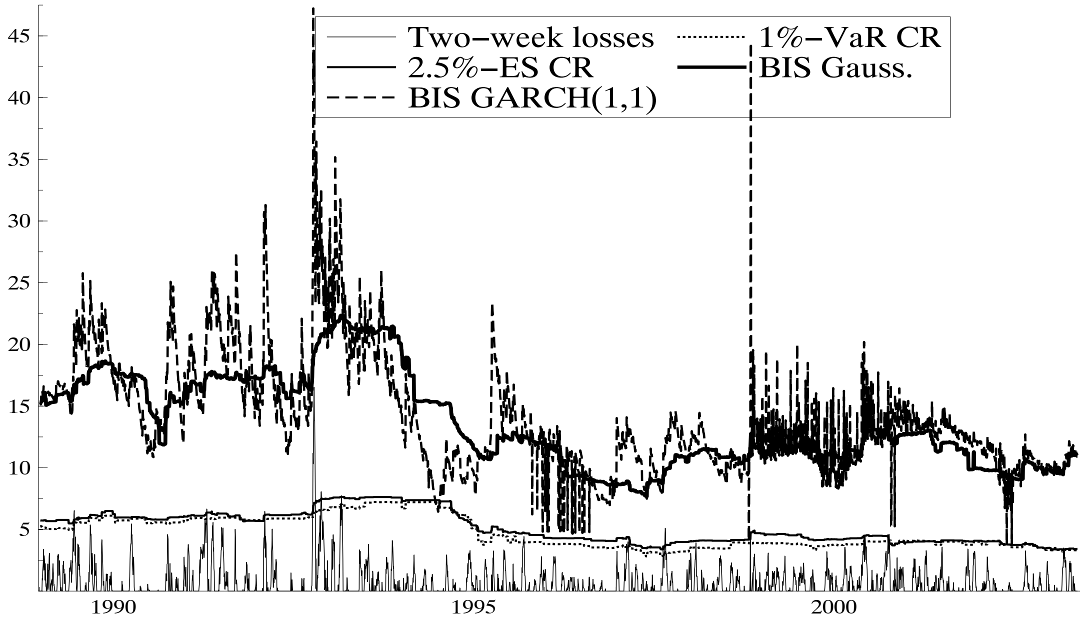


Figure 4.5: **Capital requirements £/ $ FX rate**
This figure compares two losses on the £/ $ FX rate to the capital requirements (on the vertical axis) are given for a firm trading in the £/ $ FX rate. Given are the capital requirements using the BIS regulation and the capital requirements based on a 1%-VaR and 2.5%-ES model risk multiplication factor.

Table 4.4: **Capital requirement schemes**
This table reports the 1-day average exceedance rate, two-week average exceedance rate, and the average capital requirements (CR) the capital requirement schemes. The CR schemes investigated are the BIS CR for the Gaussian model, the BIS CR for the GARCH(1, 1) model, the VaR model risk multiplication factor based CR, and the ES model risk multiplication factor based CR. The S&P 500 (for 26-10-'83 – 29-04-'03) and the £/ $ FX rate (for 03-01-'88 – 29-04-'03) are investigated.

market	scheme	avg. 1-day exceedance per year	avg. two-week exceedance per year	avg. CR
S&P 500	BIS Gauss.	0.05	0.05	24.0
	BIS GARCH	0.10	0.00	23.0
	MRMF VaR	2.52	1.70	9.2
	MRMF ES	1.49	1.08	11.0
£/ $	BIS Gauss.	0.00	0.00	14.7
	BIS GARCH	0.07	0.07	14.9
	MRMF VaR	1.33	0.90	5.9
	MRMF ES	1.00	0.70	6.3

a quantification of model risk on top of market risk measurement.

The general framework presented is elaborated in such a manner that it fits well into the capital adequacy framework set out by the Basel Committee and that of many internal risk management divisions. The use of risk measurement methods extends the currently used value-at-risk and the recently proposed coherent risk measures in a natural way.

We decompose the total model risk into a component due to estimation error and a component due to misspecification. This is established using a tolerance set restricted to a model class in order to quantify estimation risk and an unrestricted tolerance set to quantify misspecification risk. This allows a division of capital requirements currently used (for example, the multiplication factor of the BIS) in market risk, model risk (estimation risk and misspecification risk), and residual risks.

Our results suggest that, for commonly used models, a Gaussian and a GARCH(1, 1) model, misspecification risk dominates estimation risk. The analysis indicates that the multiplication factor set by the BIS is conservative if it would only be intended to cover model risk. In general, the confidence levels chosen by the BIS or any other regulator need to address the trade-off between limiting the probability of excessive losses on the one hand and leaving room for operation in the market on the other hand. Furthermore, besides model risk the multiplication factor set by the BIS should also cover hard-to-measure risks such as operational risk, legal risk, etc.

Concluding, the framework presented allows regulators to differentiate their capital requirements on the basis of the extent to which a market can be reliably modeled on the basis of state-of-the-art technology. Depending on the performance of the model used for market risk assessment by the individual bank, model risk reserves can be determined. A further comparison between markets on the basis of the extent to which they can be reliably modeled and the determination of the size of model risk reserves for different models is left for future empirical research.

A Risk measure derivations

A.1 Computation of ES

To compute the ES under normality we use some well-known properties of the normal and lognormal distribution. We can compute the expected shortfall of X when X is lognormally distributed, that is,[18]

$$\mathcal{L}(\log(X)) = \mathcal{N}\left(\mu, \sigma^2\right).$$

$$\begin{aligned}
\mathrm{ES}_p(X) &= -\frac{1}{p}\int_{-\infty}^{z_p(\mu,\sigma)} \exp(x)\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)dx \\
&= -\frac{1}{p}\exp\left(\mu+\frac{1}{2}\sigma^2\right)\int_{-\infty}^{z_p(\mu,\sigma)} \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma^2}\left(x-\left(\sigma^2+\mu\right)\right)^2\right)dx \\
&= -\frac{1}{p}\exp\left(\mu+\frac{1}{2}\sigma^2\right)\int_{-\infty}^{\frac{z_p(\mu,\sigma)-\mu-\sigma^2}{\sigma}} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y^2\right)dy \\
&= -\frac{1}{p}\exp\left(\mu+\frac{1}{2}\sigma^2\right)\Phi\left(z_p-\sigma\right) \qquad (4.24)
\end{aligned}$$

where $z_p(\mu,\sigma)$ denotes the p-quantile of the $\mathcal{N}(\mu,\sigma^2)$ distribution and is given by $z_p(\mu,\sigma) = z_p\sigma + \mu$, where z_p denotes the p-quantile of the standard normal distribution.

A.2 Asymptotic distribution of VaR and ES

We derive the asymptotic distribution of the VaR and the ES starting with the parametric case.

Parametric case

We have available a data set of n (for convenience, equally spaced) returns $\left(h_1^T, ..., h_n^T\right)$ on the time interval $[0,\tau]$ which is a realization of a random sample $\left(H_1^T, ..., H_n^T\right)$, where $H_j^T \sim \mathcal{N}\left(\mu T, \sigma^2 T\right)$ for $j = 1, .., n$. μ denotes the yearly mean, σ^2 the yearly

[18] $\mathcal{L}(X)$ denotes the law of X and $\mathcal{N}$ refers to the normal distribution.

variance, and $\tau = nT$. It is well-known that the Central limit theorem gives

$$\sqrt{n}\left(\begin{bmatrix}\hat{\mu}\\ \hat{\sigma}^2\end{bmatrix} - \begin{bmatrix}\mu\\ \sigma^2\end{bmatrix}\right) \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix}0\\ 0\end{bmatrix}, \begin{bmatrix}\sigma^2/T & 0\\ 0 & 2\sigma^4\end{bmatrix}\right) \tag{4.25}$$

with

$$\begin{aligned}\hat{\mu} &= \frac{1}{\tau}\sum_{i=1}^{n} h_i^T\\ \hat{\sigma}^2 &= \frac{1}{\tau}\sum_{i=1}^{n}\left(h_i^T - \hat{\mu}T\right)^2,\end{aligned}$$

the maximum likelihood estimators for μ and σ^2, respectively.

Since the VaR and ES are functions of μ and σ, their asymptotic distribution can be computed by applying the delta method (see, for example, Van der Vaart (1998)) to (4.25). We start with VaR. Let $\theta = (\mu, \sigma)$ and X_0 is the initial capital. All computations are made with horizon T. For notational convenience, we write[19]

$$\begin{aligned}\Psi_{\text{VaR}} &\equiv \text{VaR}_{m(\theta)}\left(\Pi_{m(\theta)}\right)\\ &= -X_0 \exp\left(z_p\sigma\sqrt{T} + \mu T\right).\end{aligned} \tag{4.26}$$

Using the delta method we get the asymptotic distribution of $\widehat{\text{VaR}}_{m(\theta)}\left(\Pi_{m(\theta)}\right)$

$$\sqrt{n}\left(\widehat{\text{VaR}}_{m(\theta)}\left(\Pi_{m(\theta)}\right) - \text{VaR}_{m(\theta)}\left(\Pi_{m(\theta)}\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_{\text{VaR}_{m(\theta)}}\right), \tag{4.27}$$

where

$$\Sigma_{\text{VaR}_{m(\theta)}} = T\Psi_{\text{VaR}}^2\left[\sigma^2 + z_p\sigma^3\sqrt{T} + z_p^2\sigma^2/2\right]. \tag{4.28}$$

The worst-case VaR is computed by

$$\begin{aligned}\Psi_{\text{VaR}}^{wc} &\equiv \text{VaR}_{m(\theta)}^{wc}\left(\Pi_{m(\theta)}\right)\\ &= \text{VaR}_{m(\theta)}\left(\Pi_{m(\theta)}\right) - z_{\alpha/2}\sqrt{\Sigma_{\text{VaR}}/n}\end{aligned} \tag{4.29}$$

[19] In the interest of readability the dependence on parameters is suppressed in the notation.

Again using the delta method we get the asymptotic distribution of $\widehat{\mathrm{VaR}}^{wc}_{m(\theta)}\left(\Pi_{m(\theta)}\right)$

$$\sqrt{n}\left(\widehat{\mathrm{VaR}}^{wc}_{m(\theta)}\left(\Pi_{m(\theta)}\right) - \mathrm{VaR}^{wc}_{m(\theta)}\left(\Pi_{m(\theta)}\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma^{wc}_{\mathrm{VaR}}\right), \tag{4.30}$$

where

$$\Sigma^{wc}_{\mathrm{VaR}_{m(\theta)}} = \left(\frac{\partial \Psi^{wc}_{\mathrm{VaR}}}{\partial \mu}\right)^2 \frac{\sigma^2}{T} + \frac{\partial \Psi^{wc}_{\mathrm{VaR}}}{\partial \mu}\frac{\partial \Psi^{wc}_{\mathrm{VaR}}}{\partial \sigma}\sigma^3 + \frac{\sigma^4}{4}\left(\frac{\partial \Psi^{wc}_{\mathrm{VaR}}}{\partial \sigma}\right)^2. \tag{4.31}$$

For the ES we have

$$\begin{aligned}\Psi_{\mathrm{ES}} &\equiv \mathrm{ES}_{m(\theta)}\left(\Pi_{m(\theta)}\right)\\ &= -\frac{X_0}{p}\exp\left(\mu T + \frac{1}{2}\sigma^2 T\right)\Phi\left(z_p - \sigma\sqrt{T}\right).\end{aligned} \tag{4.32}$$

The asymptotic distribution of $\widehat{\mathrm{ES}}_{m(\theta)}\left(\Pi_{m(\theta)}\right)$ is then given by

$$\sqrt{n}\left(\widehat{\mathrm{ES}}_{m(\theta)}\left(\Pi_{m(\theta)}\right) - \mathrm{ES}_{m(\theta)}\left(\Pi_{m(\theta)}\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_{\mathrm{ES}_{m(\theta)}}\right), \tag{4.33}$$

where

$$\begin{aligned}\Sigma_{\mathrm{ES}_{m(\theta)}} = T\Psi^2_{\mathrm{ES}}\Bigg[\sigma^2 &+ \left(T\sigma - \sqrt{T}\frac{\phi}{\Phi}\left(z_p - \sigma\sqrt{T}\right)\right)\sigma^3\\ &+ \left(\sqrt{T}\sigma - \frac{\phi}{\Phi}\left(z_p - \sigma\sqrt{T}\right)\right)^2\sigma^2/2\Bigg].\end{aligned} \tag{4.34}$$

The worst-case ES is given by

$$\begin{aligned}\Psi^{wc}_{\mathrm{ES}} &\equiv \mathrm{ES}^{wc}_{m(\theta)}\left(\Pi_{m(\theta)}\right)\\ &= -\frac{X_0}{p}\exp\left(\mu T + \frac{1}{2}\sigma^2 T\right)\Phi\left(z_p - \sigma\sqrt{T}\right) - z_{\alpha/2}\sqrt{\Sigma_{\mathrm{ES}_{m(\theta)}}/n}.\end{aligned} \tag{4.35}$$

The asymptotic distribution can computed using the delta method to be

$$\sqrt{n}\left(\left(\widehat{\mathrm{ES}}^{wc}_{m(\theta)}\left(\Pi_{m(\theta)}\right) - \mathrm{ES}^{wc}_{m(\theta)}\left(\Pi_{m(\theta)}\right)\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma^{wc}_{\mathrm{ES}_{m(\theta)}}\right), \tag{4.36}$$

where

$$\Sigma^{wc}_{\mathrm{ES}_{m(\theta)}} = T\psi^2_{\mathrm{ES}}(\mu,\sigma)\Bigg[\sigma^2\left(1+\sigma^2T/2\right)+\left(T\sigma-\sqrt{T}\frac{\phi}{\Phi}\left(z_p-\sigma\sqrt{T}\right)\right)\sigma^3 + \left(\sqrt{T}\sigma-\frac{\phi}{\Phi}\left(z_p-\sigma\sqrt{T}\right)\right)^2\sigma^2/2\Bigg]. \quad (4.37)$$

Nonparametric case

We have available a data set of n (equally spaced) returns $\left(h_1^T,...,h_n^T\right)$ on the interval $[0,\tau]$ which is a realization of a random sample $\left(H_1^T,...,H_n^T\right)$, $\tau=nT$. The empirical distribution function (EDF) is given by

$$F_n(y)\equiv\frac{1}{n}\sum_{i=1}^{n}I_{(-\infty,y]}\left(H_i^T\right) \quad (4.38)$$

We have

$$\sqrt{n}\left(F_n(y)-F(y)\right)\xrightarrow{d}\mathcal{N}\left(0,F(y)\left(1-F(y)\right)\right) \quad (4.39)$$

To compute the asymptotic distributions of the VaR and the ES we need to compute the influence functions[20] of the VaR and the ES. The value-at-risk[21] is given by

$$\begin{aligned}\Psi_{\mathrm{VaR}}(F) &\equiv \mathrm{VaR}_{m_n}\left(\Pi_{m_n}\right)\\ &= F^{-1}(p), \end{aligned}\quad (4.40)$$

[20]The influence function of Ψ can be computed as the ordinary derivative

$$\psi(F)=\frac{d}{dt}|_{t=0}\psi\left((1-t)F+t\delta_x\right),$$

where δ_x denotes the Dirac measure.

[21]The quantile function of CDF F is the generalized inverse $F^{-1}:(0,1)\to\mathbb{R}$ given by

$$F^{-1}(\alpha)=\inf\{x:F(x)\geq\alpha\}$$

and its influence function by

$$\psi_{\text{VaR}}(F) = \frac{p - \mathbf{I}_{[y,\infty)}\left(F^{-1}(p)\right)}{f\left(F^{-1}(p)\right)}. \tag{4.41}$$

We can now compute the asymptotic distribution of $\widehat{\text{VaR}}_{m_n}(\Pi_{m_n}) = F_n^{-1}(p)$ as

$$\sqrt{n}\left(\widehat{\text{VaR}}_{m_n}(\Pi_{m_n}) - \text{VaR}_{m_n}(\Pi_{m_n})\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{p(1-p)}{f^2\left(F^{-1}(p)\right)}\right) \tag{4.42}$$

Based on the asymptotic distribution given in (4.42) we can construct a confidence interval for $F^{-1}(p)$, namely

$$CI_{1-\alpha}\left(\text{VaR}_{m_n}(\Pi_{m_n})\right) = \left[\widehat{\text{VaR}}_{m_n}(\Pi_{m_n}) \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{nf^2\left(F^{-1}(p)\right)}}\right], \tag{4.43}$$

where z_α denotes the α-quantile of the standard normal distribution. The density f in (4.43) can be estimated by the Rosenblatt-Parzen kernel estimator

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right). \tag{4.44}$$

The worst-case VaR is given by

$$\begin{aligned}\Psi_{\text{VaR}}^{wc}(F) &= \text{VaR}_{m_n}^{wc}(\Pi_{m_n}) \\ &= \text{VaR}_{m_n}(\Pi_{m_n}) - z_{\alpha/2}\sqrt{\frac{p(1-p)}{nf^2\left(F^{-1}(p)\right)}}.\end{aligned} \tag{4.45}$$

The influence function of the worst-case VaR is then given by

$$\psi_{\text{VaR}}^{wc}(F) = -\psi_{\text{VaR}}(F)\left[1 - z_{\alpha/2}\frac{\sqrt{p(1-p)}}{n}\frac{f'}{f}\left(F^{-1}(p)\right)\right], \tag{4.46}$$

where f' can be estimated by

$$\hat{f}'(x) = \frac{-1}{nh^2}\sum_{i=1}^{n} K'\left(\frac{x_i - x}{h}\right).$$

The asymptotic distribution is given by

$$\sqrt{n}\left(\widehat{\mathrm{VaR}}_{m_n}^{wc}\left(\Pi_{m_n}\right)-\mathrm{VaR}_{m_n}^{wc}\left(\Pi_{m_n}\right)\right)\xrightarrow{d}\mathcal{N}\left(0,\Sigma_{\mathrm{VaR}_{m_n}}^{wc}\right), \tag{4.47}$$

with

$$\Sigma_{\mathrm{VaR}_{m_n}}^{wc}=\mathbb{E}\psi_{\mathrm{VaR}}^{wc^2}\left(F\right).$$

The expected shortfall is given by

$$\begin{aligned}\Psi_{\mathrm{ES}}\left(F\right) &= \mathrm{ES}_{m_n}\left(\Pi_{m_n}\right)\\ &= \mathbb{E}_F\left(Y\mid Y\leq F^{-1}\left(p\right)\right),\end{aligned} \tag{4.48}$$

and its influence function is given by

$$\psi_{\mathrm{ES}}\left(F\right) = \frac{1}{p}\left(y-F^{-1}\left(p\right)\right)\mathbf{I}_{(-\infty,F^{-1}(p)]}\left(y\right)-\Psi_{\mathrm{ES}}\left(F\right)+F^{-1}\left(p\right). \tag{4.49}$$

The asymptotic variance of $\widehat{\mathrm{ES}}_{m_n}\left(\Pi_{m_n}\right)=\mathbb{E}_{F_n}\left(Y\mid Y\leq F^{-1}\left(p\right)\right)$ is given by

$$\begin{aligned}\Sigma_{\mathrm{ES}_{m_n}} &= \mathbb{E}\psi_{\mathrm{ES}}^2\left(F\right)\\ &= \frac{1}{p}\mathbb{E}\left[Y^2\mid Y\leq F^{-1}\left(p\right)\right]-\Psi_{\mathrm{ES}}^2\left(F\right)\\ &\quad-\left(1-\frac{1}{p}\right)F^{-2}\left(p\right)+\left(2-\frac{2}{p}\right)\Psi_{\mathrm{ES}}\left(F\right)F^{-1}\left(p\right).\end{aligned}$$

The asymptotic distribution of $\widehat{\mathrm{ES}}_{m_n}\left(\Pi_{m_n}\right)=\mathbb{E}_{F_n}\left(Y\mid Y\leq F^{-1}\left(p\right)\right)$ is then given by

$$\sqrt{n}\left(\widehat{\mathrm{ES}}_{m_n}\left(\Pi_{m_n}\right)-\mathrm{ES}_{m_n}\left(\Pi_{m_n}\right)\right)\xrightarrow{d}\mathcal{N}\left(0,\Sigma_{\mathrm{ES}_{m_n}}\right). \tag{4.50}$$

A confidence interval for $\mathrm{ES}_{m_n}\left(\Pi_{m_n}\right)=\mathbb{E}\left(Y\mid Y\leq F^{-1}\left(p\right)\right)$ can be constructed using (4.50), namely

$$CI_{1-\alpha}\left(\mathrm{ES}_{m_n}\left(\Pi_{m_n}\right)\right)=\left[\widehat{\mathrm{ES}}_{m_n}\left(\Pi_{m_n}\right)\pm z_{\alpha/2}\sqrt{\frac{1}{n}\Sigma_{\mathrm{ES}_{m_n}}}\right]. \tag{4.51}$$

The worst case expected shortfall is given by

$$\begin{aligned}\Psi^{wc}_{\mathrm{ES}}(F) &= \mathrm{ES}^{wc}_{m_n}(\Pi_{m_n}) \\ &= \mathrm{ES}_{m_n}(\Pi_{m_n}) - z_{\alpha/2}\sqrt{\frac{1}{n}\Sigma_{\mathrm{ES}_{m_n}}},\end{aligned} \tag{4.52}$$

and its influence function is given by

$$\begin{aligned}\psi^{wc}_{\mathrm{ES}}(F) &= -\psi_{\mathrm{ES}}(F) - \frac{z_{\alpha/2}}{2\sqrt{n\mathbb{E}\psi^2_{\mathrm{ES}}(F)}} * \\ &\quad \left[\frac{1}{p^2}\left(y^2 - F^{-1}(p)\right)\mathbf{I}_{(-\infty,F^{-1}(p)]}(y)\right. \\ &\quad -\mathbb{E}\left[Y^2 \mid Y \le F^{-1}(p)\right]/p - F^{-1}(p)/p \\ &\quad -2\Psi_{\mathrm{ES}}(F)\,\psi_{\mathrm{ES}}(F) - 2\left(1-\frac{1}{p}\right)F^{-1}(p)\,\psi_{\mathrm{VaR}}(F) \\ &\quad \left.+\left(2-\frac{2}{p}\right)\left(\psi_{\mathrm{ES}}(F)\,F^{-1}(p) + \Psi_{\mathrm{ES}}(F)\,\psi_{\mathrm{VaR}}(F)\right)\right].\end{aligned} \tag{4.53}$$

The asymptotic distribution of $\widehat{\mathrm{ES}}^{wc}_{m_n}(\Pi_{m_n})$ is then given by

$$\sqrt{n}\left(\widehat{\mathrm{ES}}^{wc}_{m_n}(\Pi_{m_n}) - \mathrm{ES}^{wc}_{m_n}(\Pi_{m_n})\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma^{wc}_{\mathrm{ES}_{m_n}}\right), \tag{4.54}$$

where

$$\Sigma^{wc}_{\mathrm{ES}_{m_n}} = \mathbb{E}\psi^{wc^2}_{\mathrm{ES}}(F).$$

B Data

This appendix describes the data used in the study. For the S&P 500 series we use the total return series from Thomson Datastream code: S&PCOMP(RI). The £/$ exchange rates is given by Thomson Datastream code: USBRITP(ER). For the US risk free interest rate we have transformed the Thomson Datastream series ECUSD3M(IR) to continuously compounded interest rates. For the UK risk free interest rates we use the continuously compounded interest rates of the Thomson Datastream series ECUKP3M(IR).

Chapter 5

How Risky are Written Derivatives Positions? A Model Risk Study

5.1 Introduction

During the nineties we saw a spectacular growth of trading derivatives instruments which continues in the new millennium. For example, the turnover of exchange-traded financial derivatives is estimated at about \$ 192 trillion and the notional amount of outstanding OTC contracts is estimated at \$ 128 trillion (see Jeanneau (2002)). Though the larger part consists of (short-term) interest rate derivatives which are almost exclusively traded between financial institutions, a significant part is due to equity products traded by the public. Taking into account that the option buyer has liability limited to the option premium, while the option seller risks losses that can severely exceed his initial premium it comes as no surprise that the public prefers to be on the buy side. As each contract needs a buyer and a writer, the financial institutions need to be short in options in case the public wants to be long in options. Being short options financial institutions are exposed to several risks.

Derivatives risks can be decomposed into several categories, such as market risk, credit risk, legal risk, and operational risk. The derivatives community has become increasingly aware of these risks. This contributed to more formal methods of derivative risk assessment (see, for example, Basel Committee on Banking Supervision (1996a)).

Due to the growing complexity of derivatives markets, financial institutions rely more and more on the use of models to assess the risks to which they are exposed. The accuracy of these risk assessments depends crucially on the extent to which a market can be reliably modelled. Choosing an appropriate model for risk assessments is an important and difficult task. It is a widespread feeling among both academics and practitioners that, although some models do a better job than others, the search for one ultimate model is futile. An approach that takes the limitations of our knowledge into account is to develop models — depending on the application (pricing, hedging, ...) — that capture the most important aspects of a particular market, and to somehow control for the fact that the assessment of risk is based on a possibly misspecified model (see, for example, Derman(1996, 2001) for a practitioner's view). In practice, it has become customary to set aside so-called *model reserves*. This means that booking of certain profits on trades is postponed if it is felt that these profits are sensitive to the model used. The hazard of working with a potentially misspecified model is termed *model risk*.

In the early days of option pricing plain vanilla options were not traded very frequently and, therefore, their pricing posed a modelling challenge. Currently, the market for (short-term) plain vanilla options is so liquid that pricing does not require much modelling. A simple interpolation of the implied volatility surface would already give a reasonable price. Therefore, the pricing model risk is negligible. However, modelling remains crucial for hedging. Depending on the hedge strategy used, the risk profile of a derivative can take (very) different forms (see Green and Figlewski (1999)).[1] In this study we focus on the hedging model risk for plain vanilla options and aim to quantify this.

The main contribution of this chapter is an empirical investigation of the hedging model risk associated with the industry standard Black-Scholes Greeks in the S&P 500 market and some of the most important currency markets. For this we provide a framework to chart model risk. We use the model risk framework proposed in

[1]For exotic options (or illiquid derivatives in general), markets are not as mature as for plain vanillas. Therefore, in these markets besides hedging model risk also considerable pricing model risk can exist (see, for example, Hull and Suo (2002) and Hirsa, Courtadon and Madan (2002)).

Chapter 4 that builds on the axiomatic market/credit risk framework proposed by Artzner et al. (1999). Furthermore, we provide simulation evidence showing the hazards of relying on historical simulation to quantify the risk associated with the writing of options. We propose and use the bootstrap as an alternative.

As pointed out in Green and Figlewski (1999) three important sources of model risk arise when trading derivatives. First, the model can be misspecified. In case of pricing (liquid) plain vanilla options, this is not of major concern, since prices can be readily found in the market. However, the hedge parameters used are based on modelling assumptions. Second, the option value is derived based on no-arbitrage assumptions. This means that the model prescribes a hedge strategy which is usually specified in continuous time. This is not be feasible in practice, since this entails an infinite number of transactions. Therefore, approximate (delta) hedging in discrete time is used and we focus in this paper on the risk profile of the cost of hedging in case of a discrete (daily) hedge strategy. A third source of model risk is the problem of unobserved model parameters such as, for example, volatility (see the literature on stochastic volatility models such as Hull and White (1987) and Heston (1993) amongst others). Though the method presented allows more advanced option pricing models, we restrict our focus to the Black-Scholes model since this remains the industry standard for hedging plain vanilla options.

To extend the current practice of computing risk assessments on the basis of some given ("nominal") model for the cost of hedging, we also determine a set of plausible alternative models. In recognition of the fact that each of these models is a (reasonable) candidate for representing reality, we propose to compute a worst-case market risk measure of the cost of hedging over the set of alternative models. Model risk is then defined as the difference between this measure and the market risk measure computed from the nominal model. Using sets of alternative models that are or are not restricted to a model class, we distinguish between model risk due to estimation error and model risk due to misspecification.

We investigate the hedging model risk associated with the Black-Scholes delta hedge for plain vanilla options. The markets that we investigate are the S&P 500, and the \$/£, £/¥, and \$/¥ exchange rates. We find that in our sample the misspecification risk is considerable. Further, we find in the S&P 500 market that the market sets a premium which could serve as reward for facing the model risk.

The remainder of the paper is structured as follows. The next section considers the risks associated with derivatives. Section 5.3 sets out the model risk framework. Section 5.4 describes the methodology used and provides supporting simulation evidence and empirical analysis. In Section 5.5 the results are presented. Finally, Section 5.6 concludes.

5.2 Derivatives risks

Theoretically, derivative assets can be exactly replicated (in case of a complete market) by a (dynamic) position in the underlying asset and some numeraire asset. In practice, these replicating strategies are not feasible due to transaction costs and the inability to trade continuously. Therefore, financial institutions rely on hedge strategies in discrete time. However, by hedging in discrete time the position consisting of the derivative and the hedging portfolio is no longer risk free and subjected to market risk. We use this market risk to get to a definition of the risk associated with a derivative.

5.2.1 Cost of Hedging

For a portfolio of basis assets it is natural to determine the market risk (by a risk measure such a value-at-risk or expected shortfall) on the distribution of the portfolio at the relevant time horizon, say T. In principle, the same can be done for derivatives using the maturity date as the relevant time horizon T. However, this would ignore the fact that financial institutions have the possibility (and extensively use it) to hedge their derivatives portfolios. Indeed, Green and Figlewski (1999) found that derivative risks, as expected, can be reduced considerably by delta hedging. Therefore, the hedge strategy of the financial institution should be taken into account when calculating the market risk of a derivative. We suggest to define the market risk of a derivative as the market risk of the cost of hedging of this derivative. The cost of hedging $C(X; \gamma)$ of claim X, say a call option, using trading strategy γ on a discrete set of time points $\{t_0, ..., t_n = T\}$ is given by[2]

$$\frac{C(X;\gamma)}{N}(0) \equiv \frac{X}{N}(T) - \sum_{i=1}^{n} \gamma(t_{i-1}) \cdot \Delta\left(\frac{S}{N}\right)(t_i) - \sum_{i=1}^{n} \gamma(t_{i-1}) \cdot \frac{\delta \odot S}{N}(t_i)(t_i - t_{i-1}), \tag{5.1}$$

where S denotes the prices of the underlying asset(s), N denotes the price of a numeraire asset, δ denotes the dividend process, and $\Delta\left(\frac{S}{N}\right)(t_i) \equiv \left(\frac{S}{N}\right)(t_i) - \left(\frac{S}{N}\right)(t_{i-1})$. The first term of (5.1) denotes the discounted payoff at $T = t_n$, the second gives the discounted gains/losses over time and the final term gives the dividend payouts.

The trading strategy γ used does not have to be related to the pricing model. For example, in the following sections we mostly use market prices (model independent) and hedge using the Black-Scholes delta hedge with historical (implied) volatilities.

5.3 Model risk

Market risk measures such as value-at-risk and expected shortfall are typically based on a class of scenarios together with a base probability measure; both items are provided by a model m. Therefore, the market risk measure can be computed once the model is selected. However, there is uncertainty about which model to use. A financial institution's perception of market risk can deviate substantially from the actual market risk due to the fact that the actual dynamics are insufficiently represented by the model dynamics. Due to the use of an incorrect model, the financial institution may accept risks that it would find unacceptable in case it would know the actual dynamics. The risk associated to the mismatch between model dynamics and actual dynamics is called *model risk*.

5.3.1 Notation and definitions

Since in this paper we are interested in model risk for written derivatives positions, we are working with classes of models rather than with a single model. It is not always convenient to use the same probability space for each of these models. Therefore, we start by a formal description of a setting that allows the use of multiple

[2]The symbol $\odot$ denotes the Hadamard product, that is, $x \odot y = (x_1 y_1, ..., x_n y_n)$ (see, for example, Magnus and Neudecker (1999)).

probability spaces (see also Chapter 4).We define a *model* as a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with Ω the sample space, $\mathcal{F}$ the set of events, and $\mathbb{P}$ the probability measure. This setting could be extended to more elaborate probabilistic settings; in particular, a filtration might be assumed given.[3] For simplicity, we focus on final payoffs of the derivatives in this chapter so that the above notion will be sufficient for the purposes of this chapter. For any model m, let $\mathcal{R}(m)$ denote the space of equivalence classes of measurable real-valued functions on $(\Omega, \mathcal{F})$. If a model m is given, a *risk* is defined on m as an element of $\mathcal{R}(m)$. This definition, in which a "risk" is a random variable defined on a given probability space, follows the terminology of Artzner et al. (1999) and Delbaen (2000). We introduce a similar concept for model classes rather than for individual models. $\mathcal{M}$ denotes the class of models. A *product* can then be defined on $\mathcal{M}$ as a mapping that assigns to each model $m \in \mathcal{M}$ a risk defined on m. The set of all products defined on $\mathcal{M}$ is denoted by $\mathcal{X}(\mathcal{M})$. The risk induced by a product Π on a model m will be denoted by Π_m. The product that we use in this paper is the delta-hedged derivative given in (5.1).

We now proceed to risk measures, again starting with the definition for an individual model. A *risk measure* defined on m is a map from $\mathcal{R}(m)$ to $\mathbb{R} \cup \{\infty\}$.[4] The notion of a risk measure can be generalized to a class of models using so-called risk measurement methods. A *risk measurement method* defined on a class of models $\mathcal{M}$ is a mapping that assigns to each model $m \in \mathcal{M}$ a risk measure defined on m. Suppose that the financial institution uses a risk measurement method RMM to assess the acceptability of a product (portfolio) Π. In model m, the risk of the product Π is computed as $\mathrm{RMM}_m(\Pi_m)$. To take into account model uncertainty, we take a set of alternative dynamics $\mathcal{K}$ around m and compute the *worst-case market risk measure* (with respect to $\mathcal{K}$), which is given by $\sup_{k \in \mathcal{K}} \mathrm{RMM}_k(\Pi_k)$. Model risk may now be quantified as follows.

[3]In case of (5.1) we should introduce a filtration for a formal definition such that the underlying process and the trading strategy are well defined w.r.t. this filtration. However, since we are only interested in (discounted) final payoffs of the derivative we do not need to go beyond the static model. If we would also be interested in payoffs during the life of the option, we would need to extend the framework by using stochastic processes instead of random variables (see Artzner, Delbaen, Eber, Heath and Ku (2002)).

[4]Including ∞ allows risks to be defined on more general probability spaces, see Delbaen (2000).

Definition 5.1 (Model risk measure)[5] Let $\mathcal{M}$ be a class of models, let m be a model in $\mathcal{M}$, and let $\mathcal{K}$ be a subset of $\mathcal{M}$ containing m. Furthermore, let Π be a product defined on $\mathcal{M}$ and let RMM be a risk measurement method for $\mathcal{M}$. The *model risk* associated to the method RMM of product Π, with respect to the *nominal model* m and the *tolerance set* $\mathcal{K}$, is given by

$$\phi_{\mathrm{RMM}}(\Pi, m, \mathcal{K}) = \sup_{k \in \mathcal{K}} \mathrm{RMM}_k(\Pi_k) - \mathrm{RMM}_m(\Pi_m). \tag{5.2}$$

The model risk measure that we have defined may have some desirable properties depending on the market risk measurement method from which it has been derived. In Chapter 4 invariance and positive homogeneity are proved. In general, the model risk measure does not satisfy subadditivity and convexity, but this presents no problems (see Chapter 4 for a discussion).

5.3.2 Decomposition of Model Risk

In the previous section we did not specify a model m or a set of alternative models $\mathcal{K}$. In this section we discuss some possible choices for the set of alternative dynamics $\mathcal{K}$. In practice, one starts with a (usually parametric) model class, say $\mathcal{M}(\Theta) \equiv \{(\Omega, \mathcal{F}, \mathbb{P}_\theta) : \theta \in \Theta\} \subset \mathcal{M}$. Using an estimation or calibration procedure, a particular element $m(\hat{\theta})$ is chosen from $\mathcal{M}(\Theta)$. Even if the actual dynamics, say m_0, belong to the parametric model class $\mathcal{M}(\Theta)$, that is $m_0 = m(\theta_0)$ for some $\theta_0 \in \Theta$, the financial institution faces the risk of selecting the wrong element $m(\hat{\theta})$. This risk is called *model risk due to estimation error*. To define a neighborhood of plausible models around $m(\hat{\theta})$, one typically uses confidence regions. Depending on a chosen level α we take a $(1-\alpha)\,\%$ confidence region around $\hat{\theta}$. A general case is treated in Chapter 4, but here we limit our focus. Since we are interested in a specific market risk measurement method RMM and a specific product Π, we use a

[5] The case where $\mathrm{RMM}_m(\Pi_m) = \infty$ is uninteresting since the financial institution will never accept the product Π in its portfolio.

confidence interval around the risk measurement method $\mathcal{K}$ defined by[6]

$$\mathcal{K}(\alpha) = \left\{ m(\theta) \in \mathcal{M}(\Theta) : \mathrm{RMM}_{m(\theta)}\left(\Pi_{m(\theta)}\right) \in CI_{1-\alpha}\left(\mathrm{RMM}_{m(\hat{\theta})}\left(\Pi_{m(\hat{\theta})}\right)\right)\right\}. \tag{5.3}$$

We define *model risk due to estimation error*, or simply *estimation risk*, as the model risk that is obtained from a tolerance set derived from confidence regions within the model class.

Now let us consider the situation where the actual dynamics may not belong to $\mathcal{M}(\Theta)$. As above, we can define tolerance sets that are specifically tied to a given risk measurement method and a given product. Along this line, one may estimate $\mathrm{RMM}_m(\Pi_m)$ first and define a tolerance set based on a confidence region $CI_{1-\alpha}$ for the estimate

$$\mathcal{K}(\alpha) = \{m : \mathrm{RMM}_m(\Pi_m) \in CI_{1-\alpha}\}. \tag{5.4}$$

In general, we can determine tolerance sets that are restricted or are not restricted to a model class (unrestricted, in the sequel). As above, one may define model risk due to estimation error as the model risk restricted to the model class $\mathcal{M}(\Theta)$. The amount that has to be added to arrive at the model risk determined from the unrestricted method may be termed *model risk due to misspecification* or simply *misspecification risk*. In other words, if $\mathcal{K}_r$ is the restricted tolerance set and $\mathcal{K}_u$ is the unrestricted one, then we define the misspecification risk for a given product Π as

$$\phi_{\mathrm{RMM}}(\Pi, \mathcal{K}_r, \mathcal{K}_u) = \sup_{k\in\mathcal{K}_u} \mathrm{RMM}_k(\Pi_k) - \sup_{k\in\mathcal{K}_r} \mathrm{RMM}_k(\Pi_k). \tag{5.5}$$

However, the quantity defined above may in some cases be less than zero, whereas we would prefer to define misspecification risk in such a way that it is always non-negative. To achieve this with the above definition, we have to make sure that the set $\mathcal{K}_r$ is nested in $\mathcal{K}_u$. We do this by using a family $\{\mathcal{K}_u(\alpha)\}$ of tolerance sets parameterized by confidence level α. For a given confidence level α and a given tolerance set $\mathcal{K}_r$, which may have been selected on the basis of the same confidence

[6] $CI_{1-\alpha}\left(\mathrm{RMM}_{m(\hat{\theta})}\left(\Pi_{m(\hat{\theta})}\right)\right)$ denotes the $(1-\alpha)\,\%$ -confidence interval for $CI_{1-\alpha}\left(\mathrm{RMM}_{m(\theta_0)}\left(\Pi_{m(\theta_0)}\right)\right)$.

level, we then take $\mathcal{K}_u = \mathcal{K}_u(\beta)$ where β is defined by

$$\beta = \min(\alpha, \sup\{\gamma \in (0,1) : \mathcal{K}_r \in \mathcal{K}_u(\gamma)\}). \tag{5.6}$$

Chapter 4 provides some alternatives for nesting $\mathcal{K}_r$ in $\mathcal{K}_u$.

5.4 Methodology

5.4.1 Set-up of experiment

We investigate several major markets on which financial options are actively traded: The Standard and Poor's 500 (SPX) for equity options, the U.S. dollar / British pound (\$/£), British pound / Japanese yen (£/¥), and the U.S. dollar / Japanese yen (\$/¥) foreign exchange (FX) options. Sample periods vary depending on data availability. The data were obtained from Thomson Datastream and ABN-AMRO Bank (definitions and sources of the data can be found in Appendix B). The models that we investigate all come from the Black-Scholes (BS) framework. More advanced option pricing models have been used in the literature (see, for example, Carr, Geman, Madan and Yor (2002)), but these are mainly used for calibration to plain vanilla instruments and pricing of exotics. For hedging of plain vanillas options the Black-Scholes type models remain the most widely used derivatives models.

The original BS model (see Black and Scholes (1973)) was designed for European options on non-dividend paying stock. Since the SPX consists of dividend paying stocks, we use the adjusted BS model of Merton (1973) which allows for a continuous proportional dividend yield δ. Since future dividends are unknown, this represents another source of risk in trading derivatives, namely dividend risk. We neglect this type of risk and compute the option prices using the realized dividend yield. Since dividends are usually quite stable over time this seems to be of minor influence. For FX options we use the Garman-Kohlhagen model which adjusts the original BS model for options on foreign currencies. For the pricing formulas we refer to the original papers (see Garman and Kohlhagen (1983) and the citations above or a standard textbook such as Hull (2002)).

To compare risks of derivatives with different characteristics (call/put flag, moneyness, and time to maturity) we always write enough contracts to generate a premium (initial value) of \$100. For example, if an option is valued at \$5, we write 20 contracts. The results can then be interpreted as a dollar return on an investment of \$100 in the specified contract or percentage returns per dollar of option premium. To limit derivatives risk one can essentially distinguish three strategies. The first strategy is to diversify using derivatives with different characteristics and other risky assets. The second approach is using cash flow matching which consists of creating offsetting positions with different counterparties such that the derivatives contract is replicated. Though cash flow matching is the most precise method of hedging and, furthermore, model independent, it is rarely possible for a financial institution to construct a cash flow matching hedge. In general, the public wants to be long in options which brings about the short options position of the financial industry and, thereby, making it impossible for the financial institution to match all of its cash flows. Finally, the financial institution can hedge using delta hedging to hedge the derivatives risk.[7] Since cash flow matching is impractical and Green and Figlewski (1999) showed that delta hedging is far superior to hedging by diversification, we restrict ourselves to delta hedging which is also the most often used hedge strategy in the financial industry.

Since option traders are usually restricted in taking directional bets on the market, we investigate straddles instead of individual calls and puts. ITM options end in the money much more than OTM options. This could result in differences in the effectiveness of the hedge strategy. To investigate the influence of moneyness we compare the results for ITM, ATM, and OTM options where we take OTM options with a moneyness of 5% out of the money and ITM options with a moneyness of 5% in the money.[8] For a straddle it is not clear what OTM or ITM is, since when the call is ITM the put is OTM and vice versa. In our terminology the moneyness of the call gives the moneyness of straddle.

[7]Traders also often hedge other greeks (gamma, vega, etc.). As argued by Green and Figlewski (1999) this requires, however, other options which need to be bought from other financial institutions. The overall financial industry is therefore restricted to delta hedging.

[8]Moneyness is defined as $m = \log(F/\kappa)$ for calls and as $m = \log(\kappa/F)$ for puts, where F denotes the futures price, e.g. $F = Se^{rT}$ with T the time to maturity and r the riskless interest rate.

5.4.2 Estimation risk and misspecification risk

In order to decompose the model risk into estimation risk and misspecification risk we need to determine the sets (5.3) and (5.4).

Let G denote the distribution function of the cost of hedging. This distribution depends on the distribution of the underlying, F, and the hedging strategy, γ, but for notational convenience we suppress the dependence on the hedging strategy and write $G = G(F)$. In case we can parameterize F, we write F_θ and $\hat{F}$ denotes $F_{\hat{\theta}}$. We can use the standard expansion to take estimation risk into account (see, for example, Van der Vaart (1998))

$$\sqrt{T}(\varrho(G(\hat{F})) - \varrho(G(F)) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\phi^2_{\rho\circ G}(h_t)\right), \tag{5.7}$$

where $\phi_{\rho\circ G}(h_t) = \phi_{G(F)} \circ \psi_F(h_t)$ (see Van der Vaart and Wellner (1996) Lemma 3.9.3) denotes the composite influence function of observation t, $\phi_{G(F)}$ denotes the influence function of the risk measure, $\psi_F(h_t)$ the influence function of the underlying, and $\hat{F} = F_T$, the empirical distribution function, if F cannot be parameterized.

Unfortunately, G is not known analytically. However, it can be retrieved by simulation. Let $\varrho_k(G(\hat{F}))$ denote the estimator for $\varrho(G(\hat{F}))$ using k simulations. The asymptotic distribution can be written as

$$\sqrt{k}\left(\varrho_k(G(\hat{F})) - \varrho(G(\hat{F}))\right) \xrightarrow{d} \mathcal{N}\left(0, S^2_{\varrho(G(\hat{F}))}\right) \tag{5.8}$$

for all risk measures ϱ considered in this paper. Thus, by taking k large enough $\varrho(G(\hat{F}))$ can be approximated as closely as desired.

Though we cannot analytically compute the asymptotic variance of (5.7), we can determine it by applying the bootstrap. We use the fact that $\sqrt{T}(\varrho(G(\hat{F})) - \varrho(G(F))$ has the same asymptotic variance as $\sqrt{T}(\varrho(G(\hat{F^*})) - \varrho(G(\hat{F}))$, where F^* denotes the bootstrap empirical distribution (see Van der Vaart and Wellner (1996) Theorem 3.9.11).[9,10]

[9] Let F_T denote the empirical measure of an *i.i.d.* sample $X_1, ..., X_T$ from a distribution F. Given the sample values, let $X_1^*, ..., X_T^*$ be a *i.i.d.* sample from F_T. The bootstrap empirical distribution is then given by $1/T \sum_{t=1}^T \delta_{X_t^*}$, where $\delta_{X_t^*}$ denotes the Dirac measure in X_t^*.

[10] Due to the simulation error, we actually compute $\sqrt{T}(\varrho_k(G(\hat{F^*})) - \varrho_k(G(\hat{F})))$. However, by using the same random seed for all bootstrap samples, the influence of simulation error is

The bootstrap procedure works as follows: At time t we have n returns $h^s_{t-n+1;t} \equiv \left\{h^s_{t-n+1}, ..., h^s_t\right\}$ for the underlying and returns for a numeraire $h^n_{t-n+1;t} \equiv \left\{h^n_{t-n+1}, ..., h^n_t\right\}$. The period $\{t-n+1, ..., t\}$ is termed the estimation period. We compute for these estimation periods $\varrho_k(G(\hat{F}))$ for $k = 5,000$ as estimator for $\varrho(G(\hat{F}))$. For the determination of the asymptotic variance in (5.7) we create B bootstrap estimation samples $\left(h^{s,b}_{t-n+1;t}\right)^B_{b=1}$ and $\left(h^{n,b}_{t-n+1;t}\right)^B_{b=1}$.[11] Computing $\varrho(G(\hat{F}^{*,b}))$ for $b = 1, ..., B$ we can estimate the asymptotic variance of (5.7) with the sample variance of $\left\{\varrho(G(\hat{F}^{*,b}))\right\}^B_{b=1}$ and construct confidence intervals for $\varrho(G(\hat{F}))$. In the restricted case, $F = F_\theta$, this confidence interval gives the set (5.3) and for the unrestricted case it gives the set (5.4).

5.4.3 Hedging: historical versus implied volatility

For pricing of liquid derivatives, implied volatility is by definition superior to historical volatility, since by definition implied volatilities recover the market prices. Although in practice usually implied volatilities are used for hedging as well, it is by no means obvious that using implied volatilities for hedging outperforms using historical volatilities. The usual argument in favor of using implied volatilities is that the implied volatility should be a better predictor for future volatility than historical volatility. However, the evidence for this statement is rather scant (see, for example, Canina and Figlewski (1993)). Furthermore, implied volatilities are subject to severe measurement errors (see, for example, Christensen and Prabhala (1998)). This comes as no surprise, if one considers the fact that implied volatility can be seen as the garbage bin of the Black-Scholes formula and therefore captures all misspecifications of the model. On the other hand, historical volatility has also not proven to be a very reliable estimator of future volatility, but has the advantage that it is equal for all the derivatives on the same underlying. In order to investigate the hedging performance of historical and implied volatilities we compute one-day relative hedge errors using both the implied and historical volatility estimators for

suppressed.

[11] The bootstrap draws time points from the estimation period and uses both the return of the underlying and the numeraire at these time points.

the \$/¥, \$/£, and £/¥ exchange rates.[12]

We find that in most cases using implied volatilities lead to slightly better hedging performance measured in terms of mean, variance, 1% VaR, and 2.5% ES. This performance is often statistically significant, but comparing the performance measures one may conclude that the economic significance is small. Since using historical volatilities leads to (slightly) worse values for value-at-risk and expected shortfall this implies that the misspecification risk is somewhat overestimated in our analysis.

5.4.4 Historical simulation versus bootstrap

In order to determine an estimate of the empirical distribution of the cost of hedging, we have two alternatives. First, we can rely on historical simulation, following, among others, Galai (1977), Merton, Scholes and Gladstein(1978, 1982), and Green and Figlewski (1999). In case of historical simulation the amount of data available for estimation is rather limited unless one uses overlapping samples as is common in the literature. Though historical simulation produces consistent estimates in case of overlapping samples, standard errors are hard to estimate in finite samples.

To perform the historical simulation we need data of the underlying $\{s_0, ..., s_n\}$ and from this we can get the (daily) log returns $\{h_1, ..., h_n\}$. We start by writing an option f at time $t = 0$ and continue this until $t = n - k$, where $k = T * no$, and T denotes the maturity of the option in years and no is the number of days in a year. In addition to the option prices $\{f(0), ..., f(n-k)\}$ we compute the actual cost of hedging the option, using the specified hedge strategy, to get $\{C(0), ..., C(n-k)\}$ (see eq. (5.1)). Using this data we compute an estimator for the risk profile (cumulative distribution function) of the (discounted) final profit and loss ($P\&L$) account $P\&L \equiv f - C$. The risk profile of the $P\&L$ can be seen as the return distribution of pricing derivatives with the specified pricing model (in our case the BS model with historical or implied volatility) and the specified hedge strategy (daily hedge based on BS deltas using historical volatility). For this risk profile we can then compute the value-at-risk and expected shortfall. In doing so we need to take into account the fact that the set $f(0) - C(0), ..., f(n-k) - C(n-k)$

[12]The one period relative hedge error is defined as $\frac{f(t+1)-\gamma(t)S(t+1)}{f(t)-\gamma(t)S(t)} - 1$. Using relative hedge errors has the advantage that these are invariant to position size.

Table 5.1: **Tests for hedge errors using historical and implied volatilities**
This table presents the sample characteristics of a hedged portfolio. In each row the upper left entry is for the case of using implied volatilities in the delta hedge and the upper right entry is for the case of using (2 year) historical volatilities. The bottom entry between brackets gives the p-value for a test of equality between the implied volatility and historical volatility approach.

market	SPX ITM	SPX ATM	SPX OTM	£/¥	\$/£	\$/¥
Mean	0.021; 0.023	0.015; 0.016	0.006; 0.007	0.026; 0.027	0.005; 0.005	0.024; 0.025
	(0.16)	(0.05)	(0.03)	(0.12)	(0.86)	(0.20)
Variance	0.32; 0.34	0.24; 0.25	0.22; 0.23	0.19; 0.19	0.11; 0.11	0.23; 0.22
	(0.00)	(0.00)	(0.00)	(0.19)	(0.01)	(0.10)
$VaR_{0.01}$	−0.85; −0.95	−0.73; −0.79	−0.65; −0.65	−0.52; −0.52	−0.30; −0.33	−0.58; −0.60
	(0.03)	(0.21)	(0.87)	(0.99)	(0.22)	(0.01)
$ES_{0.025}$	−0.98; −1.05	−0.81; −0.85	−0.76; −0.78	−0.61; −0.61	−0.37; −0.38	−0.74; −0.72
	(0.00)	(0.00)	(0.00)	(0.61)	(0.02)	(0.01)

is subject to the overlapping samples problem. We handle this problem by using the method of Newey and West (1987).

We have performed a simulation study to investigate the performance of confidence intervals resulting from the historical simulation method using both the standard confidence intervals as those computed using Newey-West standard errors. We simulated $K = 5,000$ daily return time series of 20 years following the Black-Scholes world assumptions with mean and variance estimated on daily returns data on the S&P500 from October 26, 1981 to April 26, 2003. For each of these time series we estimated the mean, 1% VaR, and 2.5% ES for a 3 months ATM call option and their asymptotic normal distributions. We compared these with the "true" mean, 1% VaR, and 2.5% ES of the distribution of the cost of hedging of a 3 months ATM call option computed based on 100,000 cross-sectional simulations.[13] We compared these "true" values to the asymptotic distribution by computing for each time series to which the quantile the realization corresponds. In case the asymptotic distribution is correct this should result in a uniform distribution. In Figure 5.1 we see that this is far from the case. In case we correct for the overlapping samples problem using the Newey-West (see Newey and West (1987)) asymptotic distribution the situation is much better than using the standard asymptotic distribution, but still far from good.[14] Only in 70% (22%), 70% (43%), and 74% (38%) of the ideally 95% of the cases the true mean, 1% VaR, and 2.5% ES, respectively are within the computed Newey-West (usual) confidence intervals.[15]

A second approach is to use the bootstrap (see Efron (1979) for the original work and Horowitz (1999) for an overview). A problem arising with the bootstrap methodology is that it is more problematic to use implied volatility for hedging, because for bootstrap time series of the underlying no implied volatilities are available.[16] However, in Section 5.4.4, we saw that, although for hedging purposes his-

[13]The standard errors on the risk measures based on $100,000$ simulations are so small that these can be taken as the "true" values.

[14]In their analysis Green and Figlewski (1999) do not correct for the overlapping samples problem and therefore their standard errors are an underestimation of the true standard errors.

[15]We did the same analysis for different option characteristics such as time to maturity, moneyness and option type, but the conclusions remain the same. We also varied the number of lags to be used in the Newey-West procedure, but this only led to minor changes in the results.

[16]One can, of course, use a simultaneous bootstrap of the returns and the implied volatilities, but this requires assumptions on the dependence between the implied volatilities and the underlying

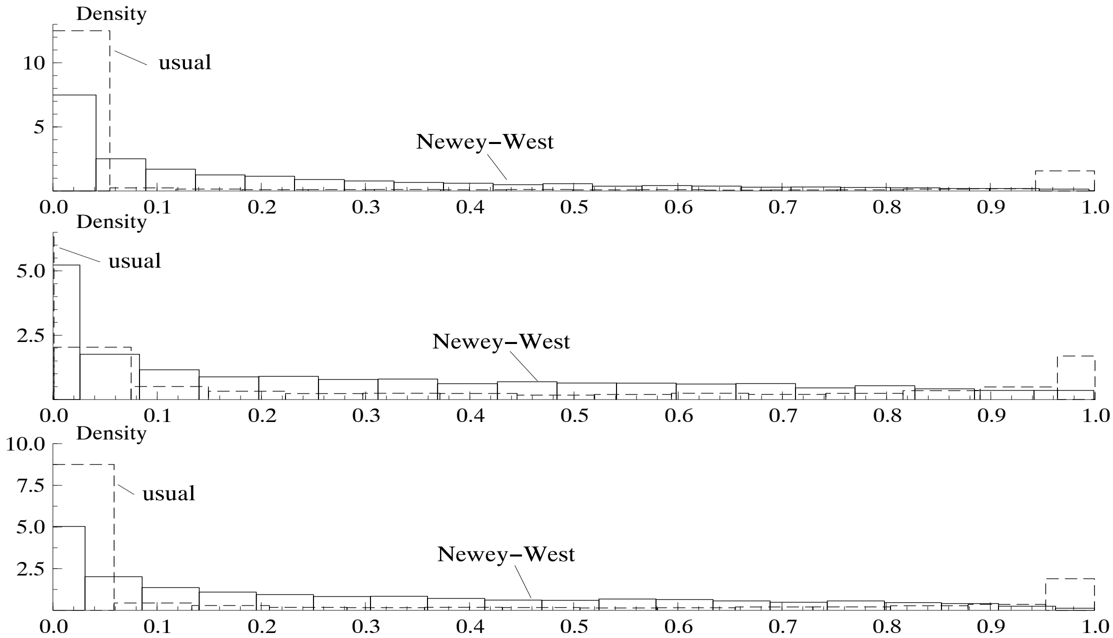


Figure 5.1: **Performance historical simulation**

This figure shows the distribution of the quantiles of the asymptotic distributions (both standard and Newey-West (with the number of lags equal to the overlapping period)) of the mean, 1% VaR, and 2.5% ES corresponding to the "true" mean, 1% VaR, and 2.5% ES. The derivative used in this analysis is a 3 month ATM call option on the S&P 500.

Table 5.2: **Test results bootstrap**
In this table we present the standard errors for market risk estimates. The column of simulation gives the results based on monte-carlo simulation. The column bootstrap mr gives the results for the bootstrap using MC simulation in order to estimate the risk measures. Finally, the column bootstrap ms gives the results for the bootstrap using drawing with replacement to estimate the risk measures.

	simulation		bootstrap mr		bootstrap ms	
	VaR	ES	VaR	ES	VaR	ES
ITM	2.80	2.75	2.88	2.89	3.50	3.45
ATM	3.22	3.18	3.09	3.06	3.15	3.12
OTM	3.26	3.21	2.82	2.82	3.40	3.39

torical volatility estimators slightly underperform implied volatility, the differences are small. Therefore, the misspecification risk is only slightly overestimated. We have also performed a simulation experiment to check the accuracy of the bootstrap methodology. We have simulated $1,000$ times a 3 years history according to the Black-Scholes world assumptions again using the mean and variance estimated on the S&P 500 data from October 26, 1981 to April 26, 2003. For each simulated history, we compute 1% value-at-risk and 2.5% expected shortfall using $K = 5,000$ simulations. This allows us to compute confidence intervals for the risk measures. For one of the simulated histories, we use the bootstrap to generate 199 bootstrap samples. For each of the bootstrap samples the risk measures are determined using 1) simulation from the normal distribution with mean and variance equal to those of the data in order to get the market risk, 2) drawing with replacement from the data in order to get the misspecification risk (which equals zero in this experiment). In Table 5.2 we see that the variance estimated using the bootstrap for market risk is very similar to that of the "true" variance computed using simulation. Furthermore, we see that the variance estimated using bootstrap for misspecification risk is somewhat higher, but not drastically so. Therefore, we may conclude that the bootstrap gives reliable estimators for the variance of our estimates of market risk.

asset.

5.5 Results

We have conducted an empirical analysis of the one-month and three-month S&P 500 5% ITM, ATM, and 5% OTM straddles. Furthermore, we conducted an empirical analysis of the one-month and three-months ATM straddles for the \$/£, \$/¥, £/¥ FX markets. For the FX straddles contracts with maturities equal to one-month and three-months were readily available. For the S&P 500 straddles maturity dates are fixed and therefore option time to maturities vary in our data. We have selected the straddles closest to the desired characteristics (time to maturity and moneyness) and used linear interpolation of the volatility term structure to get the volatilities for the one and three month straddles. In order to get estimates for the market, estimation, and misspecification risk, we used historical volatilities to price and hedge the options. We also conducted the analysis using market prices (still hedging using historical volatilities) to investigate whether the markets want to be compensated for bearing the misspecification risk.

Figure 5.2 presents the historical and ATM implied volatilities for the markets investigated. We see that in most markets the implied volatilities are on average somewhat higher than the historical ones, but in the FX markets periods of higher implied than historical and lower implied than historical volatilities alternate. In case of the S&P 500 we clearly see that the implied volatilities exceed the historical volatilities and, furthermore, that this difference increases under more volatile conditions.

In Figure 5.3 we present the results for the market risk, estimation risk, and misspecification risk based on the 1% value-at-risk measure for the 3 month ITM straddles on the S&P 500 market. We employed a three year rolling window estimation period to estimate the historical volatility. The frequency of the results in two months (other values are interpolated). We find that the estimation risk is rather consistently around 4, while the misspecification risk (on top of the estimation risk) varies between 4 and 30. In percentage terms we find that about 50 to 60% of the total risk can be explained by the market risk estimate. Between 5 and 10% is due to estimation risk and misspecification risk ranges from 20% to around 40% in some periods. In the lower panel of Figure 5.3 we present the results where option values are calculated using implied volatilities (that is, market prices). We see that

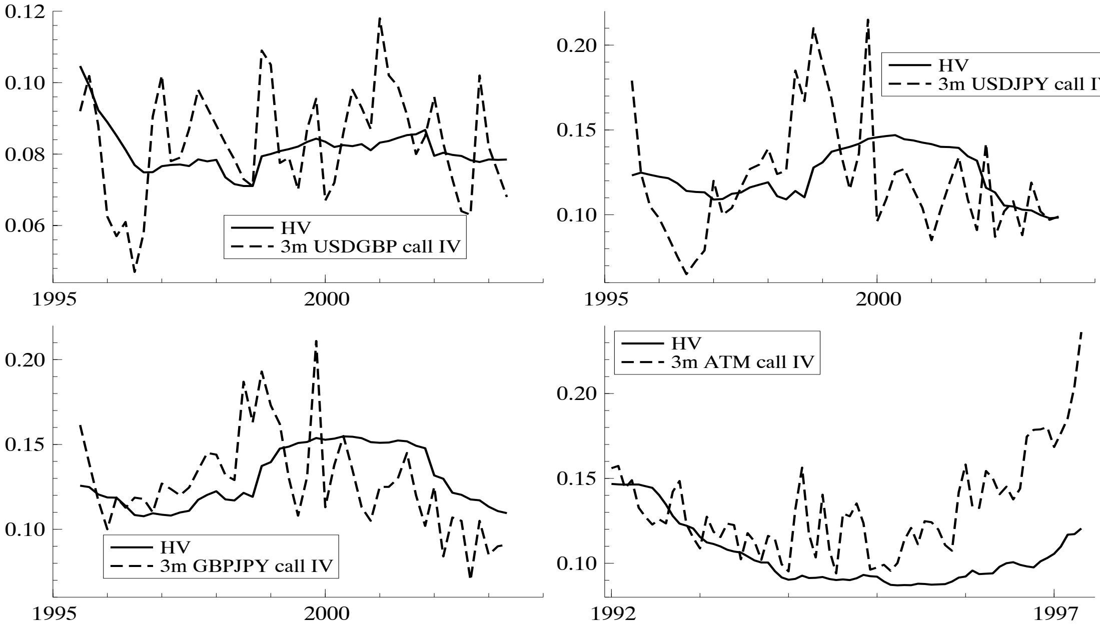


Figure 5.2: **Comparison historical vs ATM implied volatilities**
In this figure we plotted the three-year rolling window historical volatilities against the one-month ATM implied volatilities for the £/$, £/¥, and the $/¥ exchange rate. In case of the S&P 500 the 3 month ATM implied volatilities are used. Data is from 02-01-92−29-08-97 for the S&P 500 and from 09-08-1995−01-04-2003 for the $/£, £/¥, and the $/¥ exchange rates.

after 1994 the gap between historical and implied volatilities is so large that the misspecification risk is reduced to zero.

Figure 5.4 presents the results for the market risk, estimation risk, and misspecification risk based on the 2.5% expected shortfall measure for the 1 month ATM straddles on the £/¥ market. We find that the estimation risk is rather consistently around 8.5 (see Table 5.4), while the misspecification risk (on top of the estimation risk) varies between 30 to over 100. The large values for misspecification risk are mainly due to the highly volatile markets in 1999 around the LTCM crisis. Since we use a three year rolling window this has an aftereffect on the estimates of misspecification risk. Ignoring the "tub" after 1999, market risk accounts for about 50 to 60% of the total risk. Around 5% is due to estimation risk and misspecification risk ranges around 40%. In the lower panels of Figure 5.4 we give the results where option values are calculated using market values. We see that under "normal" mar-

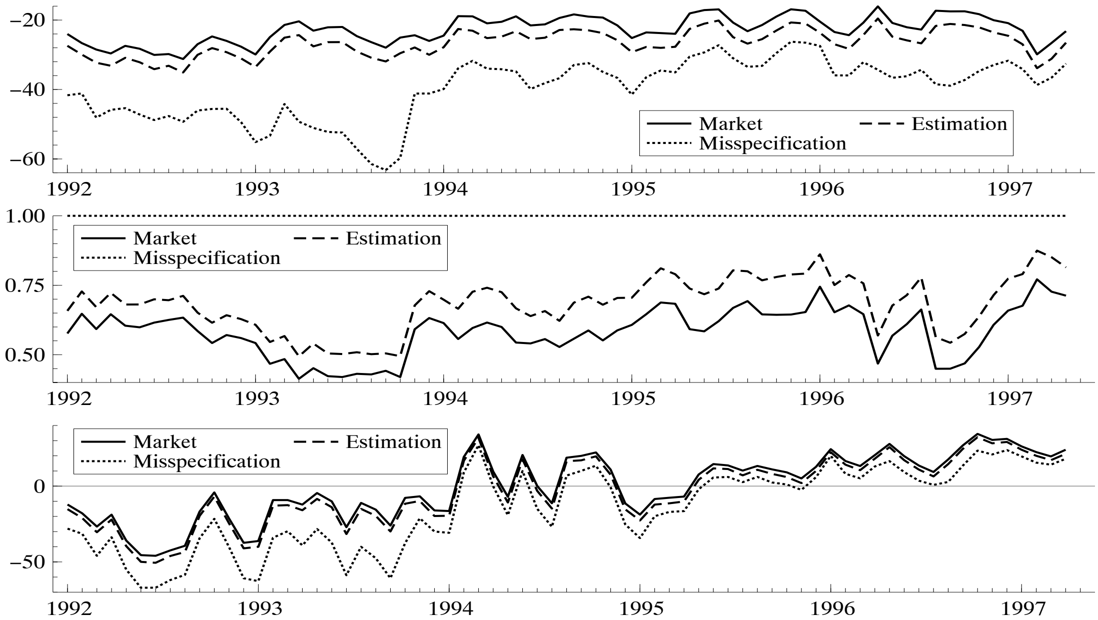


Figure 5.3: **Model Risk VaR S&P 500**
This figure presents the market, estimation, and misspecification risk based on the 1% value-at-risk measure for 3 month ITM straddles in the S&P 500 market. The upper panel gives the 1% value-at-risk for a position of $100 with straddles sold at Black-Scholes prices using historical volatilities. The middle panel gives the percentages due to market, estimation, and misspecification risk. The lower panel again presents the 1% value-at-risk for a position of $100, but the straddles sold at market prices.

ket conditions (that is, excluding the tub) misspecification risk is reduced, while in the tub misspecification risk is higher.

Table 5.3 presents the results of the market, estimation, and misspecification risk for the S&P 500 market. We see that the market risk and misspecification risk is far larger for the ATM straddles than for the ITM and OTM straddles for both maturities. This can be explained by the fact that ATM straddles have a higher gamma position than ITM and OTM straddles, which makes delta hedging more variable. As can be expected, we find that all market risks estimates are statistically significantly different from zero at the 5% level. Furthermore, we find that estimation risk is fairly large for most cases and highly statistically significant. Misspecification estimates for the one-month straddles are economically significant,

Table 5.3: **Results S&P 500 market**

In this table we present the results for the 1 and 3 month S&P 500 straddles based on 1% value-at-risk and 2.5% expected shortfall. We present averages of the market risk, estimation risk, and misspecification risk (standard errors in brackets). The column, TR IV, gives the total risk (that is, cumulative market, estimation, and misspecification risk) if the market prices were used. Asterisks denotes significance at 5% level.

RM	Ttm	moneyness	market risk	est. risk	miss. risk	TR IV	miss. high	miss. low
VaR	1m	OTM	*22.0 (10.9)	*9.5 (1.4)	8.9 (4.9)	15.0 (37.8)	29.5	7.1
VaR	1m	ATM	*55.8 (4.5)	*7.0 (0.4)	51.5 (30.7)	100.0 (68.9)	148.0	26.7
VaR	1m	ITM	*20.3 (9.0)	*6.7 (1.7)	20.3 (22.0)	39.3 (56.1)	89.5	7.4
ES	1m	OTM	*22.1 (10.7)	*9.3 (1.4)	8.9 (5.6)	16.2 (38.4)	29.6	5.4
ES	1m	ATM	*57.0 (4.7)	*6.9 (0.4)	51.4 (30.0)	100.8 (68.1)	178.9	15.8
ES	1m	ITM	*21.1 (8.8)	*6.6 (1.7)	20.0 (20.8)	39.5 (55.1)	108.2	4.4
VaR	3m	OTM	*20.3 (7.0)	*7.1 (0.7)	*11.5 (3.9)	10.5 (25.6)	28.0	12.9
VaR	3m	ATM	*34.1 (2.2)	*3.2 (0.3)	*20.0 (8.3)	28.4 (27.2)	46.8	10.0
VaR	3m	ITM	*22.8 (3.9)	*3.9 (0.4)	*13.2 (6.7)	14.6 (27.0)	35.3	7.0
ES	3m	OTM	*20.6 (6.9)	*7.1 (0.71)	*11.9 (4.1)	11.0 (25.7)	29.0	13.4
ES	3m	ATM	*34.8 (2.1)	*3.1 (0.3)	*20.5 (8.6)	29.3 (27.5)	49.5	10.2
ES	3m	ITM	*23.1 (3.8)	*3.8 (0.4)	*13.6 (6.9)	15.2 (27.2)	36.1	7.1

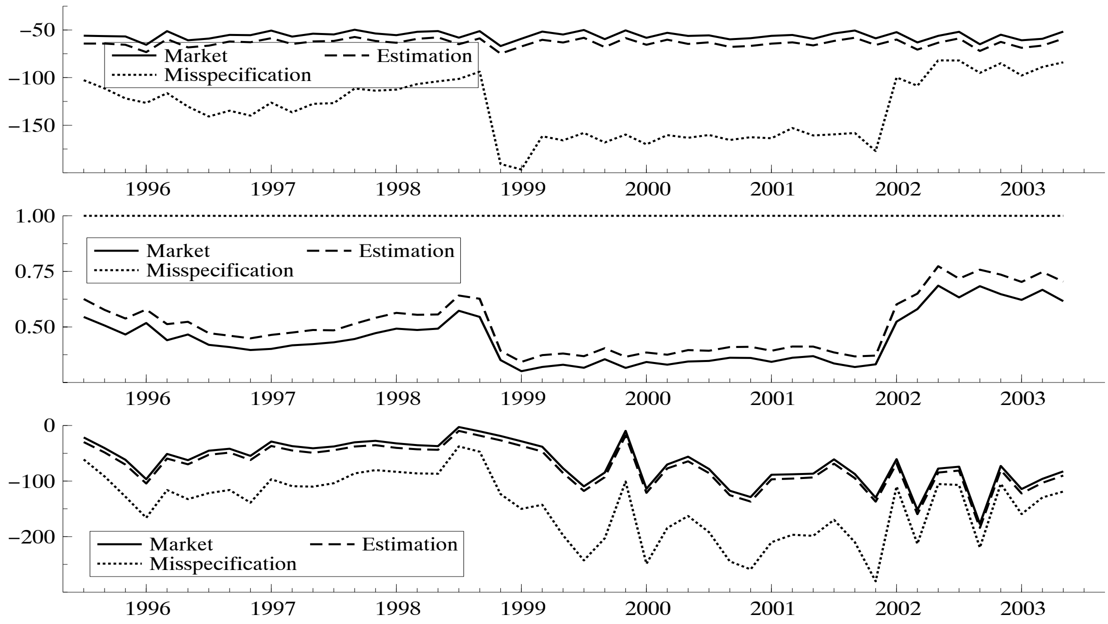


Figure 5.4: **Model Risk ES £/¥ FX rate**
This figure presents the market, estimation, and misspecification risk based on the 2.5% expected shortfall measure for 1 month ATM straddles in the £/¥ market. The upper panel gives the 2.5% value-at-risk for a position of $100 with straddles sold at Black-Scholes prices using historical volatilities. The middle panel gives the percentages due to market, estimation, and misspecification risk. The lower panel presents the 2.5% expected shortfall for a position of $100, but the straddles sold at market prices.

but statistically not significant. For the three-months straddles, we find that the misspecification risks estimates are also high (though below those for the one month), and statistically significant. Interestingly, the average estimates for total risk (that is, market, estimation, and misspecification risk) is below the market risk estimate using prices based on historical volatilities in case of the 3 months straddles, if we compare the hedging cost of the derivative to the market price. The market seems to demand a risk premium for running misspecification risk.

Table 5.4 presents the results for the FX markets. Again we find that market risk and estimation risk are economically and statistically significant for all markets. Comparing the results to the ATM straddles in the S&P 500 market, we find that market and estimation risk estimates are about the same. Misspecification risk

Table 5.4: **Results FX markets**

In this table we present the results for the 1 and 3 month ATM straddles for the \$/£, \$/¥, and £/¥ markets based on 1% value-at-risk and 2.5% expected shortfall. We present averages of the market risk, estimation risk, and misspecification risk (standard errors in brackets). The column, TR IV, gives the total risk (that is, cumulative market, estimation, and misspecification risk) if the market prices were used. Furthermore, the maximum and minimum misspecification risk is given. Asterisks indicate significant at a 5% level.

RM	Ttm	market	market risk	est. risk	miss. risk	TR IV	miss. high	miss. low
VaR	1m	\$/£	*57.7 (6.2)	*6.9 (0.7)	100.5 (78.4)	*165.1 (78.0)	221.9	18.5
VaR	1m	\$/¥	*54.5 (5.2)	*8.5 (0.8)	*152.1 (74.7)	*215.1 (76.8)	286.5	31.1
VaR	1m	£/¥	*55.6 (4.2)	*7.9 (0.5)	*68.5 (30.4)	*132.0 (30.9)	139.3	25.5
ES	1m	\$/£	*58.3 (5.9)	*6.7 (0.7)	98.0 (78.4)	*163.1 (73.5)	208.6	17.9
ES	1m	\$/¥	*55.7 (4.9)	*8.3 (0.8)	*146.0 (68.8)	*210.0 (71.0)	260.0	31.9
ES	1m	£/¥	*56.0 (4.2)	*7.7 (0.5)	*68.9 (30.5)	*132.6 (31.1)	137.3	25.7
VaR	3m	\$/£	*44.0 (7.2)	*6.4 (1.9)	44.6 (36.4)	*88.8 (40.4)	106.4	10.5
VaR	3m	\$/¥	*32.7 (5.1)	*4.9 (0.6)	*72.8 (34.9)	121.2 (62.3)	137.3	19.2
VaR	3m	£/¥	*32.3 (4.4)	*4.2 (0.3)	*33.7 (14.8)	*78.6 (33.8)	69.0	13.6
ES	3m	\$/£	*44.5 (7.1)	*6.2 (1.8)	45.7 (36.7)	*90.1 (40.7)	104.9	10.5
ES	3m	\$/¥	*33.2 (5.1)	*4.8 (0.6)	*74.4 (35.8)	123.3 (63.0)	143.1	19.7
ES	3m	£/¥	*32.8 (4.4)	*4.1 (0.3)	*34.4 (15.2)	*79.8 (34.2)	71.7	13.3

estimates are much higher, but also have higher standard errors. We find that for both ¥markets the misspecification risk is statistically significant. An interesting difference with the S&P market is that the total risk based on market prices is much higher than for the S&P 500 market. Furthermore, contrary to the results for the S&P 500 market almost all total risk estimates are statistically different from zero. It appears that in the FX markets hardly any risk premium is demanded for the misspecification risk.

5.6 Conclusions

In this paper we have empirically investigated the model risk associated with writing plain vanilla straddles in the S&P 500 equity derivatives market and the $/£, £/¥, and the $/¥ FX derivatives markets.

We apply the bootstrap method to take estimation risk and misspecification risk into account when estimating market risk of written derivative positions. To support this method in favor of the more often used method of historical simulation we have provided simulation evidence that historical simulation does a poor job in estimating estimation risk in samples with sample sizes realistic for financial applications. Furthermore, we find that hedging using historical volatilities does not economically significantly underperform hedging using implied volatilities.

We find in our samples that for the S&P 500 market considerable estimation and misspecification risk is present. Estimation risk is found to be significant for all products, while misspecification risk is significant for all three months options. Furthermore, we find that the market demands a risk premium for bearing the misspecification risk and this premium increases towards the end of our sample. For the FX markets we also find substantial misspecification risk, which is found to be statistically significant for the $/¥ and £/¥ markets. Interestingly, in our sample there does not appear to be a risk premium for bearing the misspecification risk.

A Data

We have available option data on the S&P 500 ranging from January 2, 1992 till August 29, 1997. Quotes on the straddles are the end-of-day quotes with synchronous observations of the underlying index. For the rolling volatility estimators before January 2, 1992 we use the (total) return index of the S&P 500 from Thomson Datastream. For the exchange rates we have ATM volatilities, exchange rates, interbank rates matching the option maturities, for the \$/£, \$/¥, and the £/¥. The data runs from $09-08-1995$ to $01-04-2003$.[17]

[17] All option data were kindly shared by ABN-AMRO Bank.

Part II

Pricing interest rate derivatives

Chapter 6

Observational Equivalence of Discrete String Models and Market Models

6.1 Introduction

In this chapter we discuss the discrete string model as used by Longstaff, Santa-Clara and Schwartz (2001a) and Longstaff, Santa-Clara and Schwartz (2001b) (LSS papers) and its relation to familiar models, namely, the LIBOR market model (LMM) as introduced by Miltersen, Sandmann and Sondermann (1997), Brace, Gatarek and Musiela (1997), and Jamshidian (1997) and the HJM framework (see Heath, Jarrow and Morton (1992)). We show that the discrete version of the string model for LIBOR rates is observationally equivalent to the LMM and thereby a special case of the HJM framework. Since there has been some mysticism surrounding string models and, in particular, the estimation/calibration of the correlation matrix we provide some guidelines and references for estimation/calibration.

The structure of this chapter is as follows. In Section 6.2, the discrete version of the string model as used in the LSS papers and the LMM are described. Section 6.3 shows the observational equivalence of the two models and relates them to the HJM framework. Furthermore, the parsimony of the models is determined. For illustrative purposes an example of the estimation/calibration of the models and a numerical example are provided in Section 6.4. Section 6.5 concludes.

6.2 Description of the models

First, the discrete version of the string model as used in the LSS papers is described. Second, a description of the LIBOR market model is given.

6.2.1 Discrete string model

Kennedy (1994) introduced the idea to model the evolution of the term structure of forward rates as a stochastic string. His analysis has been generalized in Kennedy (1997), Goldstein (2000), and Santa-Clara and Sornette (2001). By construction, the string model is high-dimensional (infinite dimensional if we model a continuum of forward rates), since each rate has its own perturbation. Here, we describe the string model based on a finite number of forward LIBOR rates. First, we define a finite set of dates, the so-called *tenor structure*

$$T_0 < T_1 < T_2 < \ldots < T_{N+1}. \tag{6.1}$$

We indicate the current time by T_0 and $T_1, ..., T_{N+1}$ denote the forward tenor dates. This gives a spot LIBOR rate (for $[T_0, T]$) and N forward LIBOR rates from (for $[T_i, T_{i+1}]$, $i = 1, ..., N$). We define $\delta_i = \delta(T_i, T_{i+1})$ as the so-called daycount fractions (for an extensive treatment on day-count fractions, see Miron and Swannell (1992)), which are determined by the maturity of the LIBOR rate and are most often equal to approximately 3 or 6 months. Let the forward LIBOR rate from T_i to T_{i+1} at time T_0 be denoted by $F(T_0, T_i, T_{i+1})$ which is defined as

$$F(T_0, T_i, T_{i+1}) \equiv \frac{1}{\delta_i} \left(\frac{D(T_0, T_i) - D(T_0, T_{i+1})}{D(T_0, T_{i+1})} \right), \tag{6.2}$$

where $D(T_0, T)$ denotes the value of a discount bond at time T_0 with maturity T. For notational convenience, we define $F_i(T_0) \equiv F(T_0, T_i, T_{i+1})$. The string model specifies the following dynamics for the N individual forward rates

$$\frac{dF_i(t)}{F_i(t)} = \alpha_i^M(t)\,dt + \sigma_i dZ_i^M(t), \quad i = 1, ..., N, \tag{6.3}$$

where $\left\{Z_i^M(t)\right\}_{i=1}^N$ are (correlated) Wiener processes under probability measure M with

$$d\left[Z_i^M, Z_j^M\right](t) = \rho_{ij}dt, \quad i, j = 1, ..., N.$$

If $\mathbb{Q}^{i+1}$ denotes the probability measure (equivalent to M) associated with the numeraire $D(\bullet, T_{i+1})$ [1] we know from the first fundamental theorem of asset pricing (see Delbaen and Schachermayer (1994)) that in order to exclude arbitrage possibilities $\alpha_i^{\mathbb{Q}^{i+1}}(t)$ equals 0. Due to market completeness and absence of arbitrage possibilities the drift term $\alpha_i^M(t)$ is uniquely determined for the equivalent probability measure M used in (6.3). The volatility functions $\{\sigma_i\}_{i=1}^N$ and the correlation parameters $\{\rho_{ij}\}_{i,j=1}^N$ do not change under a change of measure and are taken to be constant for ease of exposition, but can easily be extended to be deterministic functions of time.

We can stack the individual Wiener processes in a vector $Z^M = \begin{bmatrix} Z_1^M & \cdots & Z_N^M \end{bmatrix}'$. The correlation matrix Ψ (with rank $K \leq N$) of Z is given by

$$\Psi = \begin{bmatrix} 1 & \cdots & \rho_{1N} \\ \vdots & \ddots & \vdots \\ \rho_{N1} & \cdots & 1 \end{bmatrix}. \tag{6.4}$$

The volatility functions $\{\sigma_i\}_{i=1}^N$ together with the correlations of the Wiener processes determine the covariance matrix of the forward rate changes. In case Ψ is of full rank, we have to estimate $N(N+1)/2$ parameters (N volatility functions σ_i and $N(N-1)/2$ correlation parameters ρ_{ij} ($\rho_{ij} = \rho_{ji}$)). For Ψ of rank $K < N$,

[1] Under $\mathbb{Q}^{i+1}$ the value of every tradable asset, say X, satisfies

$$\frac{X(t)}{D(t, T_{i+1})} = \mathbb{E}^{\mathbb{Q}^{i+1}}\left[\frac{X(T)}{D(T, T_{i+1})}|\mathcal{F}_t\right]$$

for $T \leq T_{i+1}$.

see Section 6.3.2 . We can write the model in matrix notation as follows

$$\begin{bmatrix} \frac{dF_1(t)}{F_1(t)} \\ \vdots \\ \frac{dF_N(t)}{F_N(t)} \end{bmatrix} = \begin{bmatrix} \alpha_1^M(t) \\ \vdots \\ \alpha_N^M(t) \end{bmatrix} dt + \begin{bmatrix} \sigma_1 & & \emptyset \\ & \ddots & \\ \emptyset & & \sigma_N \end{bmatrix} dZ^M(t) . \tag{6.5}$$

The covariance matrix of the log forward rate changes is then given by

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \cdots & \rho_{1N}\sigma_1\sigma_N \\ \vdots & \ddots & \vdots \\ \rho_{1N}\sigma_1\sigma_N & \cdots & \sigma_N^2 \end{bmatrix} . \tag{6.6}$$

6.2.2 LIBOR market model

The LMM as introduced by Miltersen et al. (1997), Brace et al. (1997), and Jamshidian (1997) is usually specified in the following form:

$$\frac{dF_i(t)}{F_i(t)} = \mu_i^M(t)\,dt + \Gamma_i^{'} dW^M(t) , \quad i = 1, ..., N, \tag{6.7}$$

where W^M denotes an K-dimensional standard Wiener process ($K \leq N$) under probability measure M, $W^M = \begin{bmatrix} W_1^M & \cdots & W_K^M \end{bmatrix}'$ and Γ_i is a constant K-dimensional volatility function $\Gamma_i = \begin{bmatrix} \Gamma_{i1} & \cdots & \Gamma_{iK} \end{bmatrix}$. Just as the volatility functions $\{\sigma_i\}_{i=1}^N$ in the discrete string model, the volatility functions Γ_i can easily be extended to be deterministic functions of time. In some papers the LMM is specified with a correlated Wiener process W^{*M}, but by suitable rotation of Γ_i and W^{*M} we can always rewrite it in the form of (6.7) (see Section 6.4.1).

As in the discrete string model the drift term $\mu_i^M(t)$ is uniquely determined by the probability measure M used in (6.7) due to market completeness and absence of arbitrage possibilities. Again the first fundamental theorem of asset pricing gives that in order to exclude arbitrage possibilities $\mu_i^{\mathbb{Q}^{i+1}}(t)$ equals 0. Putting the LMM in matrix notation gives

$$\begin{bmatrix} \frac{dF_1(t)}{F_1(t)} \\ \vdots \\ \frac{dF_N(t)}{F_N(t)} \end{bmatrix} = \begin{bmatrix} \mu_1^M(t) \\ \vdots \\ \mu_N^M(t) \end{bmatrix} dt + \begin{bmatrix} \Gamma_{11} & \cdots & \Gamma_{1K} \\ \vdots & \ddots & \vdots \\ \Gamma_{N1} & \cdots & \Gamma_{NK} \end{bmatrix} dW^M(t)$$
$$= \mu^M(t)\,dt + \Gamma dW^M(t)\,. \tag{6.8}$$

The covariance matrix of the log forward rate changes is given by

$$\Gamma\Gamma' = \begin{bmatrix} \|\Gamma_1\|^2 & \cdots & \Gamma_1'\Gamma_N \\ \vdots & \ddots & \vdots \\ \Gamma_N'\Gamma_1 & \cdots & \|\Gamma_N\|^2 \end{bmatrix}, \tag{6.9}$$

where $\|\cdot\|$ denotes the Euclidean norm, Γ is of dimension $N \times K$, and the N volatility functions $\{\Gamma_i\}_{i=1}^N$ are of dimension $1 \times K$.

6.3 Observational equivalence

In this section we show that the discrete string model and the LMM are observationally equivalent. By observational equivalence we mean that for every specification in the class of discrete string models one can find a specification in the class of market models with the same probabilistic properties and vice versa. By Girsanov's theorem we know that the volatility and correlation structure determines the change of drift in (6.3) and (6.7) in case of a change of measure. We saw already that $\alpha_i^{\mathbb{Q}^{i+1}}(t) = \mu_i^{\mathbb{Q}^{i+1}}(t) = 0$ and therefore the drift terms in both models are equal under each equivalent measure M iff the volatility and correlation structure is the same. Thus, given a discrete string model specification $\sigma = (\sigma_1, ..., \sigma_N)$ and Ψ, we need to show that $\{\Gamma_i\}_{i=1}^N$ exist such that $\Gamma_i' W^M(t) \overset{d}{=} \sigma_i Z_i^M(t)$ for every i, where $\overset{d}{=}$ denotes 'equal in distribution'. Given a LMM specification $\{\Gamma_i\}_{i=1}^N$ we need to find a σ and Ψ such that $\sigma_i Z_i^M(t) \overset{d}{=} \Gamma_i' W^M(t)$ for every i.

The first part of showing the observational equivalence consists of finding a specification for the LMM in case the discrete string model is given with σ and $\Psi^{(K)}$, where superscript (K) denotes the rank of Ψ.

The spectral decomposition of $\Psi^{(K)}$ in (6.4) is given by $\Psi^{(K)} = U\Lambda U'$, where U is a matrix of orthonormal eigenvectors and Λ an ordered diagonal matrix with the eigenvalues on the diagonal.[2] We have

$$U\Lambda^{\frac{1}{2}} = \left[\sqrt{\lambda_1}u_1 \ldots \sqrt{\lambda_N}u_N\right],$$

where u_i denotes the orthonormal eigenvector corresponding to eigenvalue λ_i. In case of a K-factor model, $\lambda_{K+1} = ... = \lambda_N = 0$. We take

$$A = \left[\sqrt{\lambda_1}u_1 \ldots \sqrt{\lambda_K}u_K\right], \tag{6.10}$$

which gives $\Psi^{(K)} = AA'$. Taking

$$\Gamma = \begin{bmatrix} \sigma_1 & & \emptyset \\ & \ddots & \\ \emptyset & & \sigma_N \end{bmatrix} A \tag{6.11}$$

gives $\Sigma^{(K)} = \Gamma\Gamma'$. Therefore,

$$Z(t) \overset{d}{=} AW(t) \tag{6.12}$$

and

$$\begin{bmatrix} \sigma_1 & & \emptyset \\ & \ddots & \\ \emptyset & & \sigma_N \end{bmatrix} Z(t) \overset{d}{=} \Gamma W(t). \tag{6.13}$$

Alternatively, we could have computed $\Sigma^{(K)} = \text{diag}(\sigma)\,\Psi^{(K)}\text{diag}(\sigma)$, where $\text{diag}(\sigma)$ denotes a diagonal matrix with size of σ and its elements on the diagonal. Decomposing $\Sigma^{(K)} = VDV'$ gives $\Gamma^* = VD^{\frac{1}{2}} = \left[\sqrt{\eta_1}v_1 \ldots \sqrt{\eta_K}v_K\right]$, where (η_i, v_i) denotes the i^{th} (eigenvalue, eigenvector) pair of $\Sigma^{(K)}$. Note that Γ^* has orthogonal columns and is in general not equal to Γ, but of course $\Gamma^*\Gamma^{*\prime} = \Gamma\Gamma' = \Sigma^{(K)}$.

The second part consists of finding a specification of the discrete string model when a LMM with Γ is given. We have $\Sigma^{(K)} = \Gamma\Gamma'$. Since $\Sigma^{(K)}$ is as in (6.9), we

[2] Without loss of generality, we can take (λ_i, u_i) to denote the i^{th} largest (eigenvalue, eigenvector) pair.

can take

$$\sigma = \begin{bmatrix} \|\Gamma_1\| \\ \vdots \\ \|\Gamma_N\| \end{bmatrix} \text{ and } \Psi^{(K)} = \begin{bmatrix} 1 & \cdots & \frac{\Gamma_1'\Gamma_N}{\|\Gamma_1\|\|\Gamma_N\|} \\ \vdots & \ddots & \vdots \\ \frac{\Gamma_N'\Gamma_1}{\|\Gamma_N\|\|\Gamma_1\|} & \cdots & 1 \end{bmatrix} \tag{6.14}$$

as the discrete string model specification.

The discrete string model is therefore always just a convenient way to model term structure dynamics when the correlation structure is an input to the model.

6.3.1 Relation with HJM framework

The continuous tenor string model described in Santa-Clara and Sornette (2001) extends the HJM framework (see Heath et al. (1992)). However, in practice one is limited to using discrete string models and it is therefore interesting from a practical point of view to know whether discrete string models belong to the HJM framework.

If one specifies a discrete string model for the instantaneous forward rates one can show analogously to the procedure described in Section 6.3 that it is observationally equivalent to the HJM model. It is somewhat harder to show that the discrete string model for forward LIBOR rates fits into the HJM framework. The discrete string model for forward LIBOR is defined only on a discrete tenor structure.

In order to find a HJM specification that results in the same behavior for the forward LIBOR rates as the discrete string model, we need to use a continuous tenor HJM specification. The resulting HJM specification is derived in Miltersen et al. (1997) for the LMM.

By the observational equivalence of the discrete string model for forward LIBOR rates and the LMM established above, we know that this HJM specification also applies to the discrete string model for forward LIBOR rates.

6.3.2 Parsimony of the models

From casual observation one might be inclined to think that a K-factor LIBOR market models needs NK parameters, while the discrete string model only needs $K(K+1)/2$ parameters (see LSS papers). Note, however, that as a consequence of the observational equivalence of the two models, it necessarily follows that they must need the same number of identifying parameters. Below we demonstrate that in fact both models can be identified by $NK - K(K-1)/2$ parameters. As a simple example demonstrating that a discrete string model needs more than $K(K+1)/2$ parameters, note that the $K = 1$ dimensional discrete string model requires $N\,(> K(K+1)/2 = 1)$ parameters to specify the volatility functions. Further, we demonstrate below that there are some "hidden" restrictions that reduce the number of free parameters in the LMM from NK to $NK - K(K-1)/2$.

We can represent the covariance matrix $\Sigma^{(K)}$ in its spectral decomposition[3]

$$\Sigma^{(K)} = VDV' = \Gamma\Gamma' = \sum_{i=1}^{K} \eta_i v_i v_i' \tag{6.15}$$

where V is a matrix with orthonormal eigenvectors and D is an ordered diagonal matrix with the eigenvalues of $\Sigma^{(K)}$ with $\eta_{K+1} = ... = \eta_N = 0$. At first it seems that NK parameters are necessary. However Γ is not unique. Consider a $K \times K$ orthonormal matrix T. Then using $\Gamma^* = AT$ and $W^* = T'W$ gives the same dynamics as using Γ and W, where W^* is also a standard Wiener process, since $T'W(t) \overset{d}{=} \mathcal{N}(0, T'Tt) = \mathcal{N}(0, It)$.

The number of necessary parameters to be estimated can be found using (6.16). We have K unknown eigenvalues $\{\eta_i\}_{i=1}^{K}$. Furthermore, we have K N-dimensional eigenvectors $\{v_i\}_{i=1}^{K}$ which gives an additional NK unknown parameters. These eigenvectors $\{v_i\}_{i=1}^{K}$ need to be orthonormal which leads to $K(K+1)/2$ restrictions as can be seen from Table 6.1. Using

$$\text{number of parameters} = \text{degrees of freedom} + \text{number of restrictions} \tag{6.16}$$

we find that the degrees of freedom equals $NK - K(K+1)/2$. Adding the K eigenvalues $\{\eta_i\}_{i=1}^{K}$ we have $NK - K(K-1)/2$ parameters to estimate. Therefore, by suitable rotation of Γ we can get a $\widetilde{\Gamma}$ such that the first K rows and columns

[3]The spectral decomposition can also be performed on the correlation matrix $\Psi^{(K)}$. This would lead to different eigenvalues and eigenvectors. The number of parameters that need to be estimated is the same (see Basilevsky (1995)).

Table 6.1: **Restrictions on the eigenvectors**
Restrictions on the eigenvectors $\{v_i\}_{i=1}^{K}$ of the spectral decomposition in (6.15).

	v_1	v_2	$\cdots$	v_K
v_1	$\|v_1\|^2 = 1$			
v_2	$v_1 \cdot v_2 = 0$	$\|v_2\|^2 = 1$		
$\vdots$	$\vdots$	$\vdots$	$\ddots$	
v_K	$v_1 \cdot v_K = 0$	$v_2 \cdot v_K = 0$	$\cdots$	$\|v_K\|^2 = 1$

form a lower triangular matrix, that is,

$$\widetilde{\Gamma} = \begin{bmatrix} \widetilde{\Gamma}_{11} & & \emptyset \\ \vdots & \ddots & \\ \widetilde{\Gamma}_{K1} & \cdots & \widetilde{\Gamma}_{KK} \\ \vdots & \ddots & \vdots \\ \widetilde{\Gamma}_{N1} & \cdots & \widetilde{\Gamma}_{NK} \end{bmatrix}. \tag{6.17}$$

6.4 Estimation of the models and numerical example

Since string models have been introduced only recently, there is some mysticism surrounding them, in particular, the treatment and estimation of the covariance/correlation matrix. We showed in Section 6.3 that the discrete string model is observationally equivalent to the LMM. This implies that we can apply the same calibration techniques for the discrete string model as for the LMM, which we briefly outline below. One starts by determining the covariance matrix Σ or correlation matrix Ψ and volatilities σ of the log forward rates $\{F_i(T_0)\}_{i=1}^{N}$ based on historical rates or calibration to caps and swaptions. A common approach is to use principal components analysis (PCA) (see Basilevsky (1995) for an in-depth discussion of PCA) in order to determine the number of factors and estimation of the eigenvalues and eigenvectors. The PCA can be done both on the covariance or correlation matrix. The LSS papers and de Jong, Driessen and Pelsser (2002) capably illustrate this technique empirically. Therefore, we provide some stylised examples to illustrate advantages and disadvantages of both methods.

6.4.1 PCA on covariance matrix

We start from a two factor LMM specification with correlated Wiener processes for four six month forward LIBOR rates ($K = 2, N = 4$),

$$\frac{dF_i(t)}{F_i(t)} = \sigma_1 \exp(\kappa_1 (T_i - t))\, dW_1^{\mathbb{Q}_{i+1}}(t) + \sigma_2 \exp(\kappa_2 (T_i - t))\, dW_2^{\mathbb{Q}_{i+1}}(t), \quad i = 1, ..., 4, \tag{6.18}$$

with $d[W_1, W_2](t) = \rho dt$. As parameter values we take $\sigma_1 = 0.3$, $\sigma_2 = 1.5$, $\kappa_1 = -0.1$, $\kappa_2 = -4.0$, and $\rho = -0.6$.

We rewrite the model in the set-up of (6.7) and derive the discrete string model specification. We can compute the corresponding covariance matrix $\Sigma^{(2)}$ as

$$\Sigma^{(2)} = \begin{bmatrix} \sigma_1 \exp(0.5\kappa_1) & \sigma_2 \exp(0.5\kappa_2) \\ \sigma_1 \exp(\kappa_1) & \sigma_2 \exp(\kappa_2) \\ \sigma_1 \exp(1.5\kappa_1) & \sigma_2 \exp(1.5\kappa_2) \\ \sigma_1 \exp(2\kappa_1) & \sigma_2 \exp(2\kappa_2) \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} \sigma_1 \exp(0.5\kappa_1) & \sigma_2 \exp(0.5\kappa_2) \\ \sigma_1 \exp(\kappa_1) & \sigma_2 \exp(\kappa_2) \\ \sigma_1 \exp(1.5\kappa_1) & \sigma_2 \exp(1.5\kappa_2) \\ \sigma_1 \exp(2\kappa_1) & \sigma_2 \exp(2\kappa_2) \end{bmatrix}'$$

Decomposing $\Sigma_{\Sigma}^{(2)}$ (subscript denotes that the PCA is performed on Σ) as VDV' gives $\Gamma_{\Sigma}^{(2)} = \left[v_1 \sqrt{\eta_1}\ v_2 \sqrt{\eta_2} \right]$ for the LMM,

$$\Gamma_{\Sigma}^{(2)} = \begin{bmatrix} 0.1883 & -0.1330 \\ 0.2551 & 0.0207 \\ 0.2529 & 0.0396 \\ 0.2420 & 0.0403 \end{bmatrix}. \tag{6.19}$$

For the discrete string model we have $\Sigma_{\Sigma}^{(2)} = \operatorname{diag}(\sigma_{\Sigma})\, \Psi_{\Sigma}^{(2)} \operatorname{diag}(\sigma_{\Sigma})$ with

$$\sigma_{\Sigma} = \begin{bmatrix} 0.2305 \\ 0.2559 \\ 0.2560 \\ 0.2453 \end{bmatrix}, \quad \Psi_{\Sigma}^{(2)} = \begin{bmatrix} 1 & 0.7675 & 0.7178 & 0.7108 \\ 0.7675 & 1 & 0.9972 & 0.9964 \\ 0.7178 & 0.9972 & 1 & 1 \\ 0.7108 & 0.9964 & 1 & 1 \end{bmatrix}. \tag{6.20}$$

If we would have done PCA for a two factor model on the correlation matrix in the above example, we would have found exactly the same results, since we started with an exact two factor specification. In practice, however, one starts with an estimated covariance matrix, which is usually of full rank, and wants to estimate a K factor model on it. Then the results differ in general in finite samples, though they should converge if the data is generated from a true K-factor model.

6.4.2 PCA on correlation matrix

In our second example, we start with an estimated full rank covariance matrix Σ and use PCA in order to determine the two-factor versions of the discrete string model and the LMM. We do this for four six-month forward LIBOR rates ($K = 2, N = 4$). For ease of exposition, we have given an estimated correlation matrix, Ψ with $\Psi_{ij} = \rho^{|i-j|}$, $\rho = 0.8$ a parametric structure, and $\Sigma = \operatorname{diag}(\sigma)\,\Psi\operatorname{diag}(\sigma)$

$$\sigma = \begin{bmatrix} 0.13 \\ 0.14 \\ 0.15 \\ 0.14 \end{bmatrix}, \quad \Psi = \begin{bmatrix} 1 & 0.800 & 0.640 & 0.512 \\ 0.800 & 1 & 0.800 & 0.640 \\ 0.640 & 0.800 & 1 & 0.800 \\ 0.512 & 0.640 & 0.800 & 1 \end{bmatrix}. \tag{6.21}$$

We perform the PCA on $\Psi = U\Lambda U'$ with $K = 2$. This gives $A_\Psi^{(2)} = U\Lambda^{\frac{1}{2}}$ and $\Psi_\Psi^{(2)} = A_\Psi^{(2)} A_\Psi^{(2)\prime}$ (subscripts denote that the PCA is performed on Ψ). However, we should note that $\Psi_\Psi^{(2)}$ is not a proper correlation matrix in case $K < N$, since it does not have ones on the diagonal.[4]

$$\Psi_\Psi^{(2)} = \begin{bmatrix} 0.9319 & 0.8759 & 0.6680 & 0.4653 \\ 0.8759 & 0.8993 & 0.8067 & 0.6680 \\ 0.6680 & 0.8067 & 0.8993 & 0.8759 \\ 0.4653 & 0.6680 & 0.8759 & 0.9319 \end{bmatrix}$$

The covariance matrix $\Sigma_\Psi^{(2)}$ resulting from $\Sigma_\Psi^{(2)} = \operatorname{diag}(\sigma)\,\Psi_\Psi^{(2)}\operatorname{diag}(\sigma)$ is, however, a

[4] In practice, the diagonal elements are usually much closer to one, but we chose this stylised example to emphasize the point of improper correlation matrices.

valid covariance matrix. Therefore, we can take $\Gamma_{\Psi}^{(2)} = \mathrm{diag}\,(\sigma)\, A_{\Psi}^{(2)}$ for the LMM,

$$\Gamma_{\Psi}^{(2)} = \begin{bmatrix} 0.1087 & -0.0628 \\ 0.1293 & -0.0301 \\ 0.1385 & 0.0323 \\ 0.1170 & 0.0676 \end{bmatrix}. \tag{6.22}$$

Decomposing $\Sigma_{\Psi}^{(2)}$ as $\Sigma_{\Psi}^{(2)} = \mathrm{diag}\,(\sigma_{\Psi})\, \Psi_{\Psi}^{(2)*}\mathrm{diag}\,(\sigma_{\Psi})$ with $\Psi_{\Psi}^{(2)*}$ a proper correlation matrix gives the specification for the discrete string model,

$$\sigma_{\Psi} = \begin{bmatrix} 0.1255 \\ 0.1328 \\ 0.1422 \\ 0.1352 \end{bmatrix} \text{ and } \Psi_{\Psi}^{(2)*} = \begin{bmatrix} 1 & 0.9767 & 0.8597 & 0.7300 \\ 0.9797 & 1 & 0.9493 & 0.8597 \\ 0.8597 & 0.9493 & 1 & 0.9767 \\ 0.7300 & 0.8597 & 0.9767 & 1 \end{bmatrix}. \tag{6.23}$$

Note that the two largest eigenvalues and corresponding eigenvectors of Ψ do not match the eigenvalues and eigenvectors of $\Psi_{\Psi}^{(2)*}$. As noted before, if the data on which we estimate our K-factor model would be generated by a true K-factor model they converge asymptotically. However, in finite samples the problem of an non-proper $\Psi_{\Psi}^{(2)}$ still exists. For this reason one might prefer to use the PCA on the covariance matrix. On the other hand the PCA on the correlation matrix is preferred from a numerical point of view, since it does not have scaling problems. In the above example performing the PCA on the covariance matrix only gives marginally different results. For practical purposes we would advise to do both PCA on the covariance and on the correlation matrix and compare the results. For some recent work on optimal calibration of the covariance/correlation matrix of the LMM and discrete string model, see, for example, Zhang and Wu (2001).

6.4.3 Numerical example

Since LIBOR market models and the discrete string model allow for Black-type analytical formulas for the valuation of caplets (and floorlets), we illustrate that both models result in the same prices for caplets. The Black formula for pricing of

Table 6.2: **Prices of caplets and floorlets for the discrete string model and LIBOR market model.**

Maturity	$\sigma_{\Sigma,i}$	$\left\Vert\Gamma_{\Sigma,i}^{(2)}\right\Vert$	caplet_{Σ}	$\sigma_{\Psi,i}$	$\left\Vert\Gamma_{\Psi,i}^{(K)}\right\Vert$	caplet_{Ψ}
$T_1 = 0.5$	0.2305	0.2305	3,247.5	0.1255	0.1255	1,803.8
$T_2 = 1.0$	0.2559	0.2559	5,090.8	0.1328	0.1328	2,719.5
$T_3 = 1.5$	0.2560	0.2560	6,228.6	0.1422	0.1422	3,566.6
$T_4 = 2.0$	0.2453	0.2453	6,885.7	0.1352	0.1352	3,880.2

caplets with strike κ is given by

$$\text{cpl}_i\left(t\right) = \delta_i P\left(t, T_{i+1}\right)\left[F_i\left(t\right)\Phi\left(d_+\right) - \kappa\Phi\left(d_-\right)\right] \tag{6.24}$$

with

$$d_{\pm} = \frac{\log\left(F_i\left(t\right)/\kappa\right) \pm \frac{1}{2}\left(T_i - t\right)\nu_i}{\sqrt{\left(T_i - t\right)\nu_i}}, \tag{6.25}$$

where $\nu_i = \sigma_i^2$ for the discrete string model and $\nu_i = \left\Vert\Gamma_i\right\Vert^2$ for the LMM. In Table 6.2 we price 4 ATM caplets/floorlets (caplet and floorlet prices are the same for ATM options) with for both examples a flat initial term structure with $F_i\left(T_0\right) = 0.05$, $i = 1, ..., 4$ and notional equal to one million.

6.5 Conclusion

In this chapter we have shown that discrete string models are observationally equivalent to market models. We derive that the number of parameters needed for the estimation these models equals $NK - K\left(K-1\right)/2$. As a consequence of the observational equivalence discrete string models are a special case of the HJM framework. The discrete string models can be estimated/calibrated using principal components analysis in the same manner as the market models.

Chapter 7

Factor Dependence and Estimation Risk for Cap-Related Interest Rate Exotics

7.1 Introduction

During the nineties we witnessed a spectacular growth of trading derivative instruments that continues in the new millennium. For example, the turnover of exchange-traded financial derivatives is estimated at about $ 192 trillion (see Jeanneau (2002)). The notional amount of outstanding OTC contracts is estimated at $ 128 trillion (see Jameson and Gadanecz (2003)) consisting for the larger part of (short-term) interest rate products.

The two main liquid interest rate derivatives markets are the caps (floors) and swaptions market. Though the largest part of derivatives traded in these markets are the plain vanilla caps, floors, and swaptions there exists a sizable market in more exotic products. These products can be traded separately or, as is often the case, as part of a more complex structured deal. In either case these deals are over-the-counter and not very liquid. Therefore, in these markets besides hedging model risk also considerable pricing model risk can exist (see, for example, Hull and Suo (2002)). Both for traders and for risk management divisions it is, therefore, important to get an idea about the price range of the value of the exotic. In this chapter we investigate the estimation risk involved in pricing exotic interest rate derivatives. Estimation risk can provide risk management divisions guidelines in setting model reserves for the various products and traders about bid-ask spreads.

We adopt the popular Libor market model (see Brace et al. (1997), Miltersen et al. (1997), and Jamshidian (1997)) for analyzing several cap related interest rate exotics: deferred caps, autocaps, sticky caps, ratchet caps, and discrete barrier caps. Contrary to ordinary caps, these products are sensitive to the joint distribution of the forward rate term structure. In case of the Libor market model, this means the specification of the correlation matrix of the forward rates. In practice, this correlation matrix is either estimated historically or calibrated to the swaption market as swaptions are correlation sensitive liquid products. Unfortunately, the relation of caps and swaptions markets remains somewhat unclear (see, for example, Longstaff et al. (2001a) and de Jong, Driessen and Pelsser (2001)). Calibration of the Libor market model to swaption prices often results in unrealistically unstable correlation matrices and thereby prices and especially sensitivities. Therefore, we adopt historical correlation in this chapter.

The Libor market model can be specified in numerous ways. Among other things, the number of factors needs to be specified. In this chapter we investigate to which extent the cap related exotics are sensitive to the number of factors used in estimating the correlation matrix. We find that the autocap, ratchet cap, and sticky caps are very sensitive to the number of factors used. Furthermore, we investigate the estimation risk in pricing the exotics. This is done using the stationary bootstrap of Politis and Romano (1994) and the window length selection method of Politis and White (2003). As one would expect, the prices of the correlation sensitive products also contain a substantial degree of estimation risk.

The remainder of the chapter is structured as follows. The next section introduces notation and reviews the Libor market model and discrete string model. Section 7.3 describes the exotic derivatives investigated. Section 7.4 discusses the bootstrap technique for determining estimation risk. Results are presented in Section 7.5. Finally, Section 7.6 concludes.

7.2 Notation and Models

7.2.1 Notation

In this section we introduce notation necessary for the remainder and give a description of the Libor market model / Discrete string model. First, we define a finite set of dates, the so-called *tenor structure*

$$t = T_0 < T_1 < T_2 < \ldots < T_{N+1}. \tag{7.1}$$

We indicate the current time by t and $T_1, ..., T_{N+1}$ denote the forward tenor dates. This gives a spot LIBOR rate (for $[T_0, T_1]$) and N forward LIBOR rates from (for $[T_i, T_{i+1}]$, $i = 1, ..., N$). We define $\delta_i = \delta(T_i, T_{i+1})$ as the so-called daycount fractions (for an extensive treatment on day-count fractions, see Miron and Swannell (1992)), which are determined by the maturity of the LIBOR rate and are most often approximately equal to 3 or 6 months. Let the forward LIBOR rate from T_i to T_{i+1} at time $t \leq T_i$ be denoted by $F(t, T_i, T_{i+1})$ which is defined as

$$F(t, T_i, T_{i+1}) \equiv \frac{1}{\delta_i}\left(\frac{D(t, T_i) - D(t, T_{i+1})}{D(t, T_{i+1})}\right), \qquad t \leq T_i \text{ and } i = 1, ..., N, \tag{7.2}$$

where $D(t, T)$ denotes the value of a discount bond at time t with maturity T. For notational convenience, we define $F_i(t) \equiv F(t, T_i, T_{i+1})$.

7.2.2 Libor market model

In this section, we briefly describe the Libor market model (introduced by Brace et al. (1997), Miltersen et al. (1997), and Jamshidian (1997)). In Chapter 6 we have shown that this model is observationally equivalent to the discrete string model (introduced by Longstaff et al. (2001b) and Longstaff et al. (2001a)). The models are characterized by the following dynamics for the N individual forward rates

$$\begin{aligned} \frac{dF_i(t)}{F_i(t)} &= \alpha_i^M(t)dt + \sigma_i(t)dZ_i^M(t) && (7.3) \\ &= \alpha_i^M(t)dt + \Gamma_i(t)' dW^M(t), \quad i = 1, ..., N, && (7.4) \end{aligned}$$

where $\left\{Z_i^M(t)\right\}_{i=1}^N$ are (correlated) Wiener processes under probability measure M with

$$d\left[Z_i^M, Z_j^M\right](t) = \rho_{i,j}dt, \quad i,j = 1,...,N,$$

W^M denotes an K-dimensional ($K \leq N$) standard Wiener process and $\rho_{i,j}$ denotes the instantaneous correlation between Z_i and Z_j. The first line, that is (7.3), corresponds to the discrete string model (DSM) setting, while the second line, that is (7.4), corresponds to the Libor market model (LMM) setting. Due to absence-of-arbitrage restrictions, $\alpha_i^M = 0$ if $M = \mathbb{Q}^{i+1}$, the $T_{i+1}-$ forward martingale measure associated with numeraire $D(\bullet, T_{i+1})$.

The covariance matrix (in DSM notation, and constant for notational convenience) of the log forward rate changes is given by

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \cdots & \rho_{1,N}\sigma_1\sigma_N \\ \vdots & \ddots & \vdots \\ \rho_{1,N}\sigma_1\sigma_N & \cdots & \sigma_N^2 \end{bmatrix}. \tag{7.5}$$

7.2.3 Monte-Carlo Pricing

The forward rate dynamics of F_i given in (7.3) are easily generated using Monte-Carlo methods under the T_i-forward martingale measure using an Euler method, for all i. In the sequel we are interested in pricing derivatives that depend on several forward rate settings, which complicates the pricing. Therefore, we want to model the dynamics of all forward rates under one measure. Though in principle the choice of measure is arbitrary, it is convenient to use the $T_{N+1}-$forward measure, also referred to as the terminal measure. Changing the measure from the T_i-forward martingale measure to the terminal measure gives the following drifts for (7.3) and (7.4)

$$\alpha_i^{N+1}(t) = -\sigma_i(t)\sum_{k=i+1}^{N}\frac{\delta_k F_k(t)}{1+\delta_k F_k(t)}\sigma_k(t)\rho_{i,k}, \quad i = 1,...,N. \tag{7.6}$$

Note that the dynamics now have a stochastic drift term, and therefore in an Euler scheme the dynamics of the forward rates are no longer exact but for the last forward rate. More advanced methods exist to approximate the stochastic drift term in (7.6), such as the predictor-corrector method proposed in Hunter, Jäckel and Joshi (2001) (see Kloeden and Platen (1999) for more general predictor-corrector methods) and are employed in this chapter.

For the exotic derivatives treated in the sequel no analytical formula is available in the Libor market model. In computing the prices several variance reduction techniques can be applied. Since for caps and floors analytical prices are available, these are used as control variates. Furthermore, antithetic variables are used.

7.3 Exotic interest rate derivatives

In this section we describe a number of commonly used cap related exotic interest rate derivatives. For a more extensive list of interest rate exotics, see, for example, Brigo and Mercurio (2001), Rebonato (2002), Pelsser (2000), and Hunt and Kennedy (2000). For all products a unit notional amount is assumed.

7.3.1 Cap / floor

We start with the most basic instruments, caps and floors. A cap (floor) is portfolio of options on LIBOR rates called caplets (floorlets). Caps (floors) provide protection against high (low) interest rates. The payoff of a caplet (floorlet) fixing at T_i and paying at T_{i+1} is given by

$$\mathrm{Cpl}_i(T_{i+1}) = \delta_i \left(F_i(T_i) - \kappa\right)^+ \tag{7.7}$$

$$\mathrm{Frl}_i(T_{i+1}) = \delta_i \left(\kappa - F_i(T_i)\right)^+ . \tag{7.8}$$

Choosing the discount bond maturing at T_{i+1}, $D(T_i, Ti + 1)$, as numeraire and working with the associated martingale measure (usually denoted as the T_{i+1}-*forward measure*) we find convenient pricing equations for caplets and floorlets.

$$\mathrm{Cpl}_i(T_0) = \delta_i D(T_0, T_{i+1}) \mathbb{E}_0^{i+1}\left[\left(F_i(T_i) - \kappa\right)^+\right] \tag{7.9}$$

$$\mathrm{Frl}_i(T_0) = \delta_i D(T_0, T_{i+1}) \mathbb{E}_0^{i+1}\left[\left(\kappa - F_i(T_i)\right)^+\right], \tag{7.10}$$

where $\mathbb{E}_0^{i+1}$ denotes the conditional expectation with respect to the measure $\mathbb{Q}^{i+1}$ at time T_0.

In the LMM, where the forward LIBOR rates follow a lognormal distribution, caplets and floorlets can be priced explicitly using the Black formula (see Black (1976))

$$\begin{aligned}\mathrm{Cpl}_i(T_0) &= \delta_i D(T_0, T_{i+1})\left(F_i(T_i)\Phi(d_+) - \kappa\Phi(d_-)\right) &(7.11)\\ \mathrm{Frl}_i(T_0) &= \delta_i D(T_0, T_{i+1})\left(\kappa\Phi(-d_-) - F_i(T_i)\Phi(-d_+)\right), &(7.12)\end{aligned}$$

where Φ denotes the cumulative distribution of the standard Gaussian distribution and

$$d_{\pm} = \frac{\log(F_i(T_i)/\kappa) \pm \frac{1}{2}\Sigma_i^2}{\Sigma_i}, \qquad \text{and} \qquad \Sigma_i^2 = \int_0^{T_i} \sigma_i^2(s)ds. \tag{7.13}$$

7.3.2 Deferred cap

The standard cap / floor caplet / floorlet payment occur at their "natural" times, that is, one period after their setting. There also exist caps / floors in the market for which all caplet / floorlet payments occur at the final time, T_{N+1}. We term this a *deferred cap / floor*. The value of a deferred cap is given by

$$\mathrm{DC}(T_0) = D(T_0, T_{N+1}) \sum_{i=1}^{N} \delta_i \mathbb{E}_0^{N+1}\left[(F_i(T_i) - \kappa)^+\right], \tag{7.14}$$

Similar valuation formulas can be derived for deferred floors. We expect the deferred cap / floor to be sensitive to the shape of the term structure. It should be slightly sensitive to correlation (see correlation in the drift term in (7.6)), but probably not too much.

7.3.3 Autocap

An *autocap* (*autofloor*) is similar to an ordinary cap (floor), but the holder is only allowed to exercise $l \le N$ instead of the normal N caplets (floorlets). Furthermore, the caplets (floorlets) must be exercised when they are in the money at a settlement date. The value of an autocap can be written as

$$\mathrm{AC}_{1:N}^{l}(T_0) = \sum_{i=1}^{N} D(T_0, T_{i+1})\delta_i \mathbb{E}_0^{i+1}\left[(F_i(T_i) - \kappa)^+\right] \mathbf{I}\left\{E_i\right\}, \tag{7.15}$$

where

$$E_i = \left\{ \sum_{j=1}^{i-1} \mathbf{I}(F_j(T_j) > \kappa) \leq l - 1 \right\}.$$

A similar formula can be derived for autofloors. One should note that the $\{E_i\}_{i=1}^N$ not only depends on the current rate, but also on the previously observed spot Libor rates, $\{F_n(T_n)\}_{n=1}^i$.

7.3.4 Discrete barrier cap

A *discrete barrier cap* (*floor*) is similar to a standard cap (floor) with the difference that the payoff of its underlying caplets (floorlets) are conditional on the event that previous spot Libor rates have or have not (depending on the type of barrier option) hit a certain level, known as the barrier. Though more complex variants exist, we consider discrete barrier caps where the barrier condition is only reviewed at fixing dates. The discrete barrier cap (floor) is characterized by a series of strikes, $\{\kappa_i\}_{i=1}^N$, and barriers, $\{H_i\}_{i=1}^N$. The value of an up-and-out discrete barrier cap is given by

$$\mathrm{DBC}^{uo}(T_0) = \sum_{i=1}^{N} D(T_0, T_i)\delta_i \mathbb{E}_0^{i+1} \left[(F_i(T_i) - \kappa_i)^+ \prod_{n=1}^{i} \mathbf{I}(F_n(T_n) < H_n) \right]. \quad (7.16)$$

From (7.16) we see that caplet i only pays out in case all previous spot Libor rates, $\{F_n(T_n)\}_{n=1}^i$, are below their barriers. The discounted payoff of a down-and-in barrier cap is given by

$$\mathrm{DBC}^{di}(T_0) = \sum_{i=1}^{N} D(T_0, T_i)\delta_i \mathbb{E}_0^{i+1} \left[(F_i(T_i) - \kappa_i)^+ \left(1 - \prod_{n=1}^{i} \mathbf{I}(F_n(T_n) > H_n) \right) \right]. \quad (7.17)$$

Thus, a down-and-in discrete barrier caplet i pays out if at least one of the previous spot Libor rates, $\{F_n(T_n)\}_{n=1}^i$ was set below its barrier. A portfolio of an up and in and an up and out barrier caplet equals an ordinary caplet. Combining the following features: in/out, up/down, standard/digital and cap/floor results in sixteen different types of discrete (digital) barrier caps and floors.

7.3.5 Ratchet cap

A *ratchet option* (also denoted a one-way floater) is similar to a standard cap/floor with the difference that the strikes are variable over time and depend on settings of previous spot Libor rates. The value of a ratchet option is given by

$$\mathrm{Ra}(T_0) = \sum_{i=1}^{N} D(T_0, T_{i+1})\delta_i \mathbb{E}_0^{i+1}\left[(F_i(T_i) - \kappa_i)^+\right], \tag{7.18}$$

where κ_1 is given and for $i > 1$

$$\kappa_i = \kappa\left(\kappa_{i-1}, ..., \kappa_1; F_{i-1}(T_{i-1}), ..., F_1(T_1)\right). \tag{7.19}$$

In the sequel we consider a ratchet option with as strike the spot Libor at the previous setting

$$\kappa_i = F_{i-1}(T_{i-1}). \tag{7.20}$$

Furthermore, we consider a *sticky cap* where the strike equals the minimum of the spot Libor at the previous setting and the previous strike

$$\kappa_i = F_{i-1}(T_{i-1}) \wedge \kappa_{i-1}. \tag{7.21}$$

Both the ratchet and the sticky cap are expected to be very sensitive to the correlation matrix.

7.4 Estimation risk in the LMM

From the dynamics of the forward rates given in (7.3) and (7.4) combined with (7.6) it is clear that the prices of the exotic derivatives treated in the previous section are dependent on the volatility functions, $\{\sigma_i(t)\}_{i=1}^N$ and the correlation matrix, which we denote by ρ. If we have an analytical pricing formula, say f, which is a function of $\theta \equiv (\sigma_1, ..., \sigma_N, \rho_{1,2}, ..., \rho_{N-1,N})$ an application of the delta method (see, for example, Van der Vaart (1998)) gives the standard errors of the option prices

$$\sqrt{T}\left(f(\hat{\theta}) - f(\theta)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial f}{\partial \theta'}\Sigma_\theta \frac{\partial f}{\partial \theta}\right), \tag{7.22}$$

where Σ_θ denotes the covariance matrix of θ. However, for the exotic derivatives discussed in Section 7.3 no analytical formula is available and prices need to be determined using numerical methods. In principle one could try to determine the standard errors by numerically computing the partial derivatives of f, but this does not look very attractive and it contains more information than needed.

A second and more promising method to get standard errors of the option prices is using the bootstrap (see Efron (1979) for the original work). Loosely speaking, the bootstrap theorem (under some regularity conditions) says that

$$\sqrt{T}\left(f(\hat{\theta}^*) - f(\hat{\theta})\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial f}{\partial \theta'}\Sigma_\theta \frac{\partial f}{\partial \theta}\right), \tag{7.23}$$

where $\hat{\theta^*}$ is estimated on a bootstrap sample of the original data. For the ordinary bootstrap, a bootstrap sample is a sample drawn with replacement from the original sample,$\{X_1, ..., X_T\}$, say, $\{X_1^*, ..., X_T^*\}$ with the original sample length.[1]

A problem with the ordinary bootstrap is that it does not take time dependence of the observations into account. A method that can take time-dependence into account is the moving block bootstrap, where one draws blocks of observations instead of single observations. This method can generate time dependence in the data; however the bootstrap samples resulting from a moving block bootstrap are not necessarily stationary even if the original data are. In case of the LMM the (return) data are generated from a stationary process.[2] An alternative method able to take time dependence and stationarity into account is the stationary bootstrap (see Politis and Romano (1994)). This method can be described as follows. First, we randomly select a bootstrap observation X_1^* from the original T observations. Suppose that $X_1^* = X_j$. The second bootstrap observation is then equal to X_{j+1} with probability $1-p$; otherwise, it is picked at random from the original T observations, where p is a given positive constant smaller than or equal to 1.[3] Thus, having the ith bootstrap observation $X_i^* = X_{i^*}$, we take the next bootstrap observation according

[1] One can also draw samples smaller than the original sample, in which case we speak of subsampling (see Politis, Romano and Wolf (1999) for an excellent treatment.)

[2] Strictly speaking this only holds for a forward rate under its own forward martingale measure due to the stochastic drift under a different measure, but this effect is minor.

[3] Of course, $p = 1$ gives the ordinary bootstrap.

to

$$X_{i+1}^* = \begin{cases} X_{i^*+1} & \text{with prob. } 1-p \\ \text{a random draw from } \{X_1, ..., X_T\} & \text{with prob. } p, \end{cases} \tag{7.24}$$

for $i = 1, ..., T$.

By applying the stationary bootstrap to our data of forward rates we can get B bootstrap estimators $\hat{\theta}^*$, where B denotes the number of bootstraps. For each bootstrap estimator $\hat{\theta}^*$ we can compute the prices of the exotic options using the Monte-Carlo simulation method resulting in an estimator for standard errors of the option prices. Using (7.22) confidence intervals are easily constructed.

7.5 Results

We use the US forward curve from March 19, 1997 to May 14, 2003 on weekly basis to estimate correlations. We investigate quarterly compounding cap products, as these are the most liquid for the US market. First, we investigate which products are most sensitive to the number of factors used. In case of a K factor model, we perform principal components analysis (PCA) on the historical covariance matrix and set all but the K largest eigenvalues equal to zero. From the resulting covariance matrix, we compute the correlation matrix and transform this into the LMM covariance matrix using the volatility term structure which can be calibrated to market prices of caps.[4] PCA can also be done on the correlation matrix (see, for example, Chapter 6 for an application to Libor market models and Basilevsky (1995) in general).

In Table 7.1 we present the prices in basis points of the various products using a one factor LMM with a flat term structure at 4% and a flat volatility term structure at 20%. Besides prices also simulation standard errors are given. Cap prices are computed using the analytical Black formula and therefore do not contain simulation error.

In Figure 7.1 we investigate the influence of the number of factors used in fitting the correlation matrix. We see that the deferred cap is as expected hardly sensitive to correlation. The up-and-in barrier cap is somewhat sensitive to the number of

[4] Historical and implied volatility functions can therefore differ.

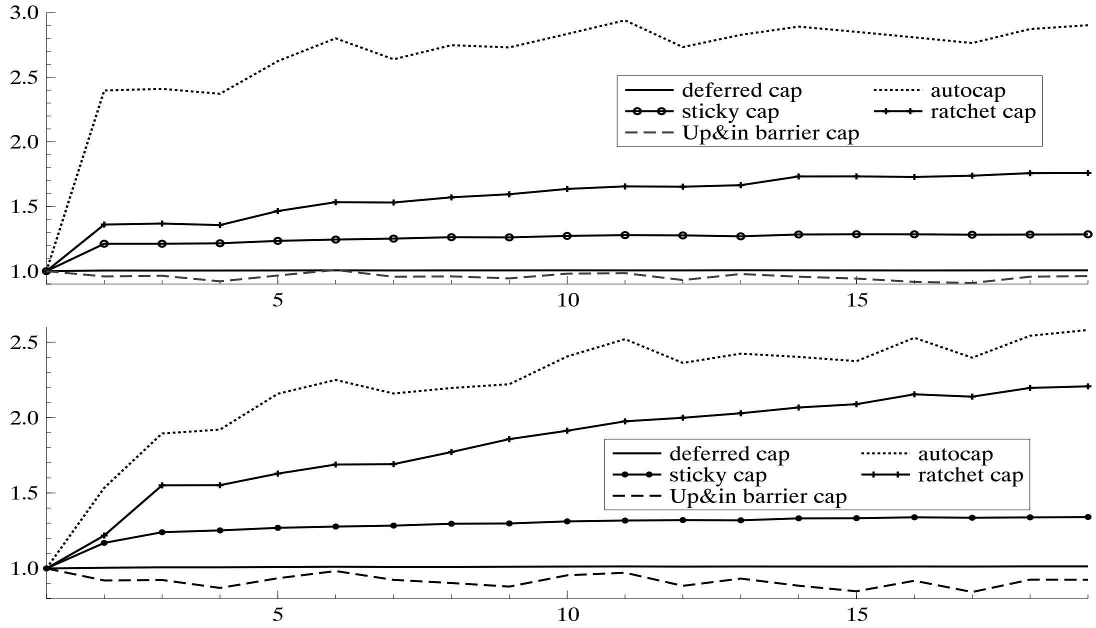


Figure 7.1: **Factor dependence**

This figure gives the prices of the $K-$factor model relative to the one-factor model for the deferred cap, autocap, sticky cap, ratchet cap, and up and in discrete barrier cap. The term structure is assumed to be flat at 4% and the volatility term structure is assumed flat to be at 20%. All products are ATM, the barrier for the up-and-in barrier cap equals 5%, the start strike for the sticky cap and ratchet cap equals 4% and the number of caps for the autocap equals half the number of settings. The upper panel gives a 3 year deal and the lower panel a 5 year deal. The number of simulations equals 10,000.

Table 7.1: **One factor model prices**
This table gives the prices for the cap, deferred cap, autocap, sticky cap, ratchet cap, and up-and-in barrier cap for a one factor Libor market model. The initial term structure is flat at 4% and the volatility term structure is flat at 20%. The strike for the deferred cap, autocap, sticky cap (initial strike), and ratchet cap (initial strike) equals 4%. The barrier for the up and out barrier equals 5%. The number of simulations equals 10,000 and the simulation standard errors are between brackets.

maturity	Cap	Deferred cap	Autocap	Sticky cap	Ratchet cap	UI barrier cap
1Year	16.01	15.87	1.32	19.01	11.71	10.98
	(0.00)	(0.07)	(0.13)	(0.03)	(0.08)	(0.12)
2Years	50.79	49.28	4.29	67.96	26.70	20.67
	(0.00)	(0.31)	(0.42)	(0.12)	(0.35)	(0.44)
3Years	94.82	89.85	7.89	133.45	40.81	26.04
	(0.00)	(0.64)	(0.81)	(0.24)	(0.70)	(0.84)
4Years	145.22	134.15	12.11	210.32	54.52	29.49
	(0.00)	(1.00)	(1.27)	(0.37)	(1.11)	(1.30)
5Years	200.31	180.14	17.23	295.33	68.05	32.21
	(0.00)	(1.37)	(1.76)	(0.51)	(1.57)	(1.79)

factors in the correlation matrix; prices computed using more than one factor are between 90% and 100% of the one factor price. We find that the sticky and ratchet cap are rather sensitive to the number of factors, although in case of the sticky cap the prices flatten out after factor 4 at about 1.3 times the one-factor price. The autocap is very sensitive to correlation matrix. Even for a small number of factors prices are already up twice the one-factor price. However, for all products, one should be careful when looking at high factor models as part of the correlation in the correlation matrix is due to noise. It is quite likely that high factor models are overfitting noise instead of “true” correlation.

In Figure 7.3 we also investigate the factor dependence of the various products, this time with a hump shaped volatility term structure given in Figure 7.2. We find that in this case all products but the ratchet are less sensitive to the number of factors used. Besides the ratchet, the autocap is most sensitive. However, in this situation the autocap is valued higher with the one-factor model than the more than one factor models.

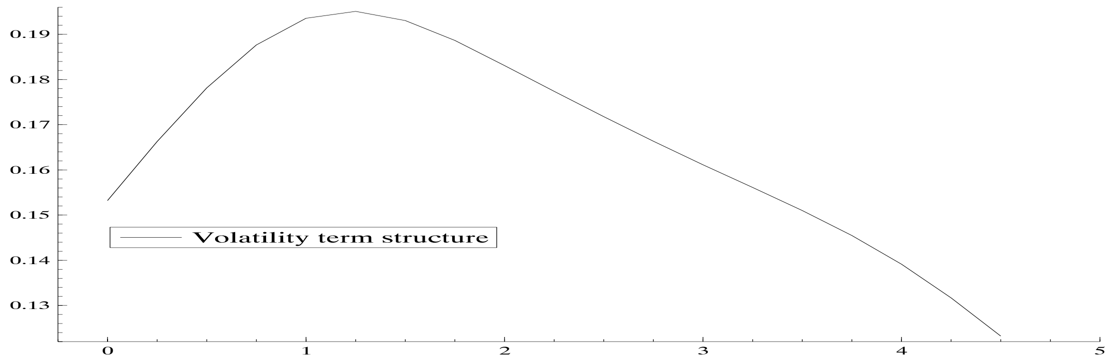


Figure 7.2: **Hump shaped volatility term structure**
This figure presents the hump shaped volatility term structure used for Figures 7.3 and 7.5. On the vertical axis the level of caplet volatility is given and on the horizontal axis the time to maturity.

In order to compute the standard errors of the computed prices, we use the stationary bootstrap with $p = 0.1$ (expected block size equals ten), where p was determined using the automatic block size determination method of Politis and White (2003) on spot Libor rates. For each bootstrap sample we compute the LMM covariance matrix using PCA as described above. In Figures 7.4 and 7.5 we present the relative prices of the autocap, sticky cap, ratchet cap, and up-and-in barrier cap with their 95% confidence intervals. Results of the deferred cap are not shown as it has hardly any estimation risk. We clearly see that, in case one uses a model with more than one factor, estimation risk can be considerable. Furthermore, we see that for all but the up-and-in barrier cap the confidence regions grow with the number of factors, but as expected the growth rate decreases with the number of factors. We see that for the up-and-in barrier option the price range computed using two factors more or less covers all prices computed by higher factor models and one might be inclined to conclude that two factors suffice in pricing the up-and-in barrier. However, risk management divisions are advised to set a model reserve larger than or equal to the worst-case limit of the price range (depending on whether the bank is on the buy or sell side, this is the upper or lower confidence limit). For autocaps and sticky caps, the price range given by the 5 factor model more or less covers the prices computed by the higher factor models. For the ratchet cap things are not so clear and risk management divisions are advised to set considerable model reserves.

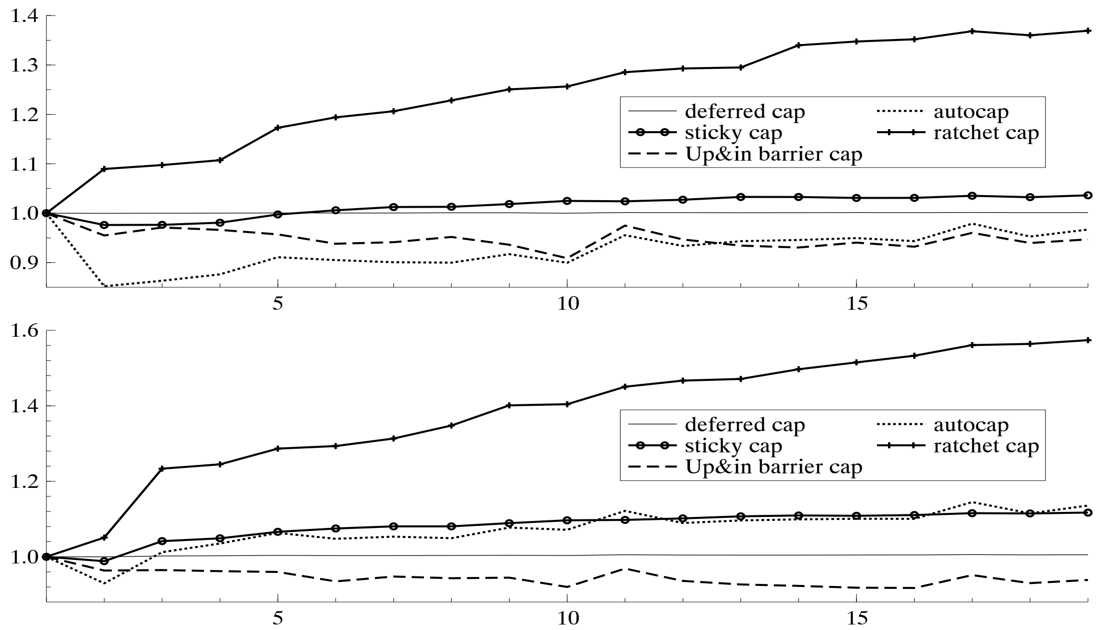


Figure 7.3: **Factor dependence hump shaped volatility**
This figure gives the prices of the $K-$factor model (number of factors on the horizontal axis) relative to the one-factor model for the deferred cap, autocap, sticky cap, ratchet cap, and up and in discrete barrier cap. The term structure is assumed to be flat at 4% and the volatility term structure is assumed hump shaped. All products are ATM, the barrier for the up-and-in barrier cap equals 5%, the start strike for the sticky cap and ratchet cap equals 4% and the number of caps for the autocap equals half the number of settings. The upper panel gives a 3 year deal and the lower panel a 5 year deal. The number of simulations equals 10,000.

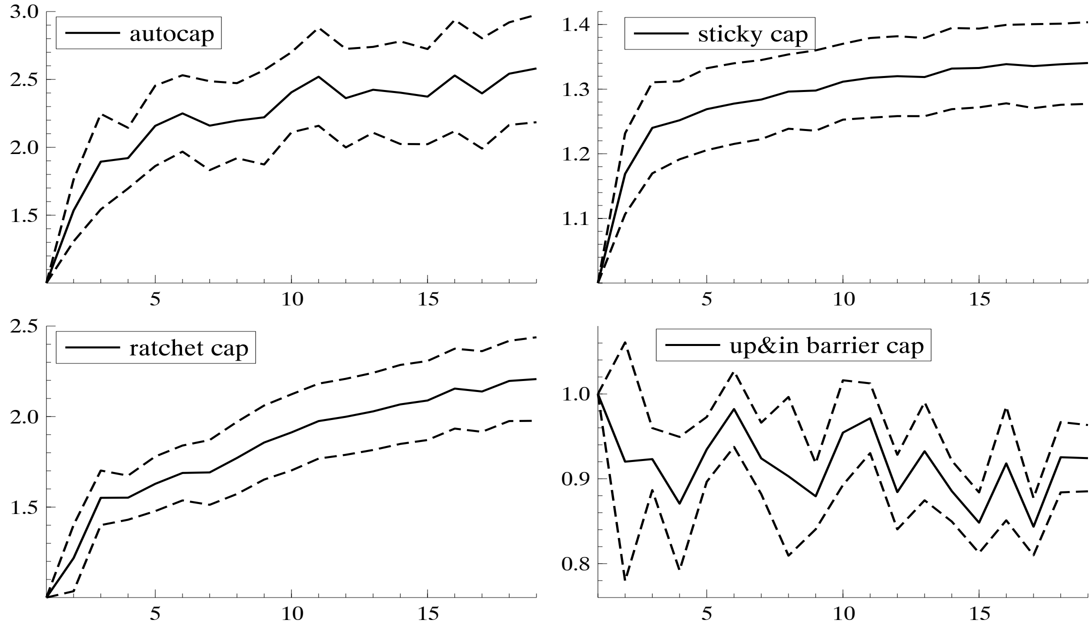


Figure 7.4: **Estimation risk**

This figure gives the prices of the $K-$factor model (number of factors on the horizontal axis) relative to the one-factor model including the 95% confidence regions. The term structure is assumed flat at 4% and the volatility term structure is assumed flat at 20%. All products are ATM, the barrier for the up-and-in barrier cap equals 5%, the start strike for the sticky cap and ratchet cap equals 4% and the number of caps for the autocap equals half the number of settings. The upper panels present the results of a 3 year deal and the lower panels present the results of a 5 year deal. The number of simulations equals 10,000.

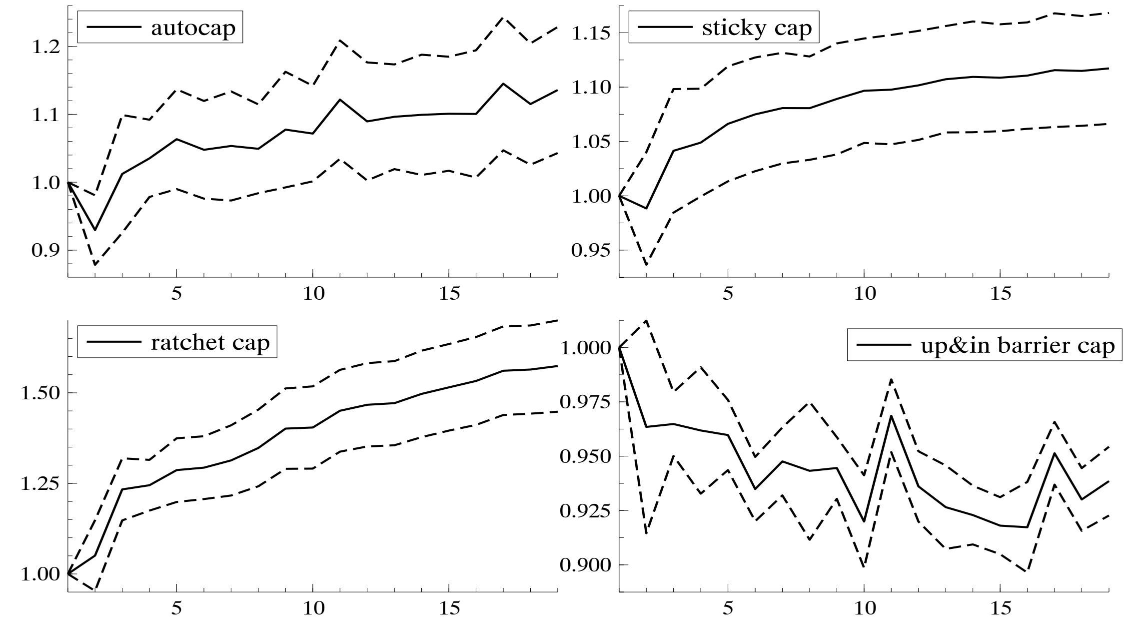


Figure 7.5: **Estimation risk hump shaped volatility**

This figure gives the prices of the $K-$factor model (number of factors on the horizontal axis) relative to the one-factor model including the 95% confidence regions. The term structure is assumed flat at 4% and the volatility term structure is assumed to be hump shaped. All products are ATM, the barrier for the up-and-in barrier cap equals 5%, the start strike for the sticky cap and ratchet cap equals 4% and the number of caps for the autocap equals half the number of settings. The upper panels present the results of a 3 year deal and the lower panels present the results of a 5 year deal. The number of simulations equals 10,000.

7.6 Conclusions

In this chapter, we provide a method to take estimation risk into account when computing exotic interest rate derivatives prices. This provides traders and risk managers with a price range of values for the exotic product contrary to a single estimate. We find that this estimation risk is very much product dependent. Autocaps, sticky caps, and especially ratchet caps are very sensitive to correlation resulting in considerable standard errors for the products. For some products risk managers are advised to set reserves up to the price of the product.

Chapter 8

Conclusions and Directions for Further Research

8.1 Summary and conclusions

Since the mid nineties the quantitative risk management literature surged. The first part of this thesis adds to that literature from both a theoretical and empirical point of view on model risk.

In Chapter 2 we introduce a backtest framework for, among other things, the most popular risk measurement methods, value-at-risk and expected shortfall. We provide ample simulation evidence that for appropriately adjusted levels (in case of the Gaussian distribution this means that a 1% value-at-risk about equals a 2.5% expected shortfall) our expected shortfall test has equal size, but considerably better power. Since the probability of detecting a misspecified model is higher for a given value of the test statistic, this allows the regulator to set lower multiplication factors. We suggested a scheme for determining multiplication factors. This scheme results in less severe penalties for the backtest based on expected shortfall compared to backtests based on value-at-risk, and compared to the current Basel Accord backtesting scheme in case the test incorrectly rejects the model. Therefore, we conclude that the prospects for setting up viable capital determination schemes based on expected shortfall are promising.

In Chapter 3 we apply the backtest framework set out in Chapter 2 to positions containing derivatives. Where value-at-risk can be tested using a binomial test, this is not the case for expected shortfall and we need information of the distribution in the tail. By nature, the characteristics of derivatives positions change over time. To overcome this problem, we present a transformation procedure. We tested several risk management models for computing expected shortfall and value-at-risk for one-period hedge errors of hedged derivatives positions. We found that the practically popular method of historical simulation provides the reasonably accurate estimates of the risk of a derivative portfolio.

Chapter 4 presents a model risk measurement framework. Our framework extends the (market) risk framework set out by Artzner et al. (1999) and Delbaen (2000) by considering risk measurement methods for a class of models instead of a risk measure for one particular model. This allows for a quantification of model risk on top of market risk measurement. This allows regulators to set capital requirements for trading activities in a market, based on the extent to which this market can be reliably modeled. The general framework presented is elaborated in such a manner that it fits well into the capital adequacy framework set out by the Basel Committee and that of many internal risk management divisions. Our results suggest that, for commonly used models, a Gaussian and a GARCH$(1,1)$ model, misspecification risk dominates estimation risk. The analysis indicates that the multiplication factor set by the BIS is conservative if it would only be intended to cover model risk.

In Chapter 5 we have empirically investigated the model risk associated with writing plain vanilla straddles in the S&P 500 equity derivatives market and the \$/£, £/¥, and the \$/¥ FX derivatives markets. We found that in our samples for the S&P 500 market considerable estimation and misspecification risk is present. Estimation risk is found to be significant for all products, while misspecification risk is significant for all three months options. Furthermore, we find that in the S&P 500 market a risk premium is demanded for bearing the misspecification risk and this premium increases towards the end of our sample. For the FX markets we also find substantial misspecification risk, which is found to be statistically significant for the \$/¥ and £/¥ markets. Interestingly, in our sample there does not appear to be a risk premium for bearing the misspecification risk.

In the second part of the thesis we investigate some models and products in the interest rate derivatives markets. Chapter 6 refutes a claim by Longstaff et al.(2001a, 2001b) that their discrete string model is more parsimonious than the Libor market model. It is shown that discrete string models are observationally equivalent to market models. We derive that the number of parameters needed for the estimation these models equals $NK - K(K-1)/2$ in case of N Libor rates and a $K-$factor model. As a consequence of the observational equivalence discrete string models are a special case of the HJM framework.

Chapter 7 investigates the factor dependence and estimation risk for some commonly used exotic interest rate derivatives. We suggest the (stationary) bootstrap for computing the estimation risk for the exotics. We find that autocaps, sticky caps, and ratchet caps are sensitive to the number of factors used and have considerable estimation risk.

8.2 Directions for Further Research

The research conducted in this thesis can be extended in various ways. The discussion between value-at-risk and expected shortfall addressed in Chapters 2 to 4 can be continued by investigating both capital requirement schemes in market crises; which method has the most prevention power? Furthermore, more empirical evidence is needed to identify the extent to which markets can be reliably modeled for different markets.

In order to investigate the model risk of writing derivatives one can investigate more advance option pricing models such as, for example, stochastic volatility models. Furthermore, alternative hedge strategies can be conducted. An important property of hedge strategies related to model risk is robustness. Exposure to model risk can be reduced by use of robust hedging strategies. A hedging strategy is called "robust" if it performs well under a wide range of model assumptions. For example, if a bank does not want any exposure to model risk for a certain type of derivative it can get rid off the model risk by selling the security (we abstract here from credit risk). However, such a hedge, a perfect static hedge, is rarely possible in practice and banks need to rely on quasi-static and dynamic hedges. In general one expects that the performance of a hedging strategy will depend on how far reality is from the nominal model that has served as a basis for the construction of the strategy. A hedging strategy that is designed for a particular model cannot be expected to do well in a world that is totally different from the model assumptions. Therefore, robustness has to be measured in terms of performance degradation with respect to models that are in some sense close to the nominal one. The value-at-risk and expected shortfall measures used in this thesis could serve the role of a distance measure for the degradation in performance.

Obviously the choice of instruments to be used in hedging plays an important role. In cases in which there is a fairly wide choice of possible hedging instruments, such as in the fixed-income markets, one may optimize for robustness of the hedging strategies over the various possible choices of instruments. Using, for instance, value-at-risk or expected shortfall as distance measures, one can define a strategy to be "optimally robust" with respect to a certain tolerance if it optimizes the performance that it guarantees over the set of models that lie within the specified tolerance (in the sense of the chosen distance measure) of a nominal model.

Often, model parameters are calibrated on the basis of observed prices rather than estimated from historical data. There are several ways in which the study of robust hedging is connected to calibration. Robustness studies could help to identify the model parameters that are critical to hedge performance and that therefore must be determined accurately. Even more importantly perhaps, robustness considerations may lead to adaptations in pricing models and will in this way have an impact on calibration.

Most of the suggestions for further research addressed above can be investigated both from a theoretical and empirical point of view. However, a number of financial markets are not mature enough (for example, the strongly growing and important credit derivatives market) to provide enough data for empirical analysis. Finding relevant sets of alternative models for these markets poses an interesting challenge.

Bibliography

Acerbi, C. and Tasche, D.: 2002, On the coherence of expected shortfall, *Journal of Banking and Finance* **26**, 1487–1503.

Artzner, P.: 1999, Application of coherent risk measures to capital requirements in insurance, *North American Actuarial Journal* **3**, 11–26.

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D.: 1997, Thinking coherently, *Risk* **10**, 68–71.

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D.: 1999, Coherent measures of risk, *Mathematical Finance* **9**, 203–228.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D. and Ku, H.: 2002, Coherent multiperiod risk measurement, *working paper*.

Barndorff-Nielsen, O. E.: 1996, Normal inverse Gaussian distributions and stochastic volatility modelling, *Scandinavian Journal of Statistics* **24**, 1–13.

Basel Committee on Banking Supervision: 1996a, *Amendment to the Capital Accord to Incorporate Market Risks,* Bank for International Settlements, Basel.

Basel Committee on Banking Supervision: 1996b, *Supervisory Framework for the Use of "Backtesting" in Conjunction with the Internal Models Approach to Market Risk Capital Requirements,* Bank for International Settlements, Basel.

Basel Committee on Banking Supervision: 1999, *A New Capital Adequacy Framework,* Bank for International Settlements, Basel.

Basel Committee on Banking Supervision: 2003, *Basel II: Third Consultative Paper,* Bank for International Settlements, Basel.

Basilevsky, A.: 1995, *Statistical Factor Analysis and Related Methods: Theory and Applications*, John Wiley & Sons, Inc., New York.

Berkowitz, J.: 2001, Testing density forecasts, with applications to risk management, *Journal of Business and Economic Statistics* **19**, 465–474.

Berkowitz, J. and O'Brien, J.: 2002, How accurate are value-at-risk models at commercial banks?, *Journal of Finance* **57**, 1093–1111.

Black, F.: 1976, The pricing of commodity contracts, *Journal of Financial Economics* **3**, 167–179.

Black, F. and Scholes, M.: 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* **81**, 637–659.

Bollerslev, T.: 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**, 307–327.

Bossy, M., Gibson, R., Lhabitant, F.-S., Pistre, N., Talay, D. and Zheng, Z.: 2000, Volatility model risk measurement and strategies against worst case volatilities, *Journal de la Société Française de Statistiques* **141**, 73–86.

Boyle, P. and Emanuel, D.: 1980, Discretely adjusted option hedges, *Journal of Financial Economics* **8**, 259–282.

Brace, A., Gatarek, D. and Musiela, M.: 1997, The market model of interest rate dynamics, *Mathematical Finance* **7**, 127–155.

Brigo, D. and Mercurio, F.: 2001, *Interest Rate Models: Theory and Practice*, Springer-Verlag, Berlin Heidelberg New York.

Canina, L. and Figlewski, S.: 1993, The informational content of implied volatility, *Review of Financial Studies* **6**, 659–681.

Carr, P., Geman, H. and Madan, D.: 2001, Pricing and hedging in incomplete markets, *Journal of Financial Economics* **32**, 131–167.

Carr, P., Geman, H., Madan, D. and Yor, M.: 2002, The fine structure of asset returns: An empirical investigation, *Journal of Business* **75**, 305–332.

Christensen, B. and Prabhala, N.: 1998, The relation between implied and realized volatility, *Journal of Financial Economics* **50**, 125–150.

Christoffersen, P., Hahn, J. and Inoue, A.: 2001, Testing and comparing value-at-risk measures, *Journal of Empirical Finance* **8**, 325–342.

Cochrane, J. H. and Saá-Requejo, J.: 2000, Beyond arbitrage: Good-deal asset price bounds in incomplete markets, *Journal of Political Economy* **108**, 79–119.

de Jong, F., Driessen, J. and Pelsser, A.: 2001, Libor market models versus swap market models for the pricing of interest rate derivatives: An empirical analysis, *European Finance Review* **5**, 201–237.

de Jong, F., Driessen, J. and Pelsser, A.: 2002, On the information in the interest rate term structure and option prices, *mimeo University of Amsterdam*.

Delbaen, F.: 2000, Coherent risk measures on general probability spaces, *Working paper ETH* pp. 1–35.

Delbaen, F. and Schachermayer, W.: 1994, A general version of the fundamental theorem of asset pricing, *Mathematische Annalen* **300**, 463–520.

Derman, E.: 1996, Model risk, *Risk* **9**, 34–37.

Derman, E.: 2001, Markets and models, *Risk* **14**, 48–50.

Diebold, F. X., Gunther, T. A. and Tay, A. S.: 1998, Evaluating density forecasts, *Internation Economic Review* **39**, 863–883.

Dowd, K.: 1998, *Beyond Value at Risk: The New Science of Risk Management*, John Wiley & Sons, England.

Duffie, D. and Pan, J.: 1997, An overview of value at risk, *Journal of Derivatives* **4**, 7–49.

Eberlein, E., Kallsen, J. and Kristen, J.: 2003, Risk management based on stochastic volatility, *Journal of Risk* **5**, 19 – 44.

Efron, B.: 1979, Bootstrap methods: Another look at the jackknife, *Annals of Statistics* **7**, 1–26.

El-Jahel, L., Perraudin, W. and Sellin, P.: 1999, Value at risk for derivatives, *Journal of Derivatives* **6**, 7–26.

El Karoui, N., Jeanblanc-Picqué, M. and Shreve, S. E.: 1998, Robustness of the Black-Scholes formula, *Mathematical Finance* **8**, 93–126.

Föllmer, H. and Schied, A.: 2002, Convex measures of risk and trading constraints, *Finance and Stochastics* **6**, 429 – 447.

Galai, D.: 1977, Tests of market effficiency of the Chicago Board Options Exchange, *Journal of Business* **50**, 167–197.

Garman, M. B. and Kohlhagen, S. W.: 1983, Foreign currency option values, *Journal of International Money and Finance* **2**, 231–237.

Gibson, R., Lhabitant, F.-S., Pistre, N. and Talay, D.: 1999, Interest rate model risk: An overview, *Journal of Risk* **3**, 37–62.

Goldstein, R.: 2000, The term structure of interest rates as a random field, *Review of Financial Studies* **13**, 365–384.

Green, T. C. and Figlewski, S.: 1999, Market risk and model risk for a financial institution writing options, *Journal of Finance* **54**, 1465–1499.

Heath, D., Jarrow, R. and Morton, A.: 1992, Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation, *Econometrica* **60**, 77–105.

Heston, S.: 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* **6**, 327–343.

Hirsa, A., Courtadon, G. and Madan, D.: 2002, The effect of model risk on the valuation of barrier options, *Journal of Risk Finance* **4**, 47–55.

Horowitz, J. L.: 1999, The bootstrap, *Handbook of Econometrics, vol. 5* pp. 1–98.

Hull, J. C.: 2002, *Options, Futures, & Other Derivatives*, Prentice Hall, Toronto.

Hull, J. and Suo, W.: 2002, A methodology for assessing model risk and its application to the implied volatility function model, *Journal of Financial and Quantitative Analysis* **37**, 297–318.

Hull, J. and White, A.: 1987, The pricing of options on assets with stochastic volatilities, *Journal of Finance* **42**, 281–300.

Hunt, P. and Kennedy, J.: 2000, *Financial Derivatives in Theory and Practice*, John Wiley & Sons, Chichester.

Hunter, C., Jäckel, P. and Joshi, M. S.: 2001, Getting the drift, *Risk* pp. 81–84.

Jameson, A. and Gadanecz, B.: 2003, OTC derivatives market activity in the second half of 2002, Bank for International Settlements, Basel, *Technical report*.

Jamshidian, F.: 1997, Libor and swap market models and measures, *Finance and Stochastics* **1**, 293–330.

Jaschke, S. and Küchler, U.: 2001, Coherent risk measures and good-deal bounds, *Finance and Stochastics* **5**, 181–200.

Jeanneau, S.: 2002, Derivative markets, *BIS Quarterly Review* pp. 35–42.

Jorion, P.: 2000, *Value at Risk: The New Benchmark for Managing Financial Risk*, 2 edn, McGraw-Hill, New York.

Kennedy, D. P.: 1994, The term structure of interest rates as a Gaussian random field, *Mathematical Finance* **4**, 247–258.

Kennedy, D. P.: 1997, Characterizing Gaussian models of the term structure of interest rates, *Mathematical Finance* **7**, 107–118.

Kerkhof, J. and Melenberg, B.: 2003, Backtesting for risk-based regulatory capital, *Forthcoming Journal of Banking and Finance.*

Kerkhof, J., Melenberg, B. and Schumacher, H.: 2002, Model risk and regulatory capital, *CentER discussion paper 2002-27* pp. 1–56.

Kloeden, P. E. and Platen, E.: 1999, *Numerical Solution of Stochastic Differential Equations*, 3^{rd} edn, Springer Verlag, Berlin.

Kupiec, P. H.: 1995, Techniques for verifying the accuracy of risk management models, *Journal of Derivatives* **3**, 73–84.

Longstaff, F., Santa-Clara, P. and Schwartz, E.: 2001a, The relative valuation of caps and swaptions: Theory and empirical evidence, *Journal of Finance* **56**, 2067–2109.

Longstaff, F., Santa-Clara, P. and Schwartz, E.: 2001b, Throwing away a billion dollars: The cost of suboptimal exercise strategies in the swaptions market, *Journal of Financial Economics* **62**, 39–66.

Magnus, J. and Neudecker, H.: 1999, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, revised edn, John Wiley & Sons, Chichester, UK.

McNeil, A. and Frey, R.: 2000, Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach, *Journal of Empirical Finance* **7**, 271–300.

Merton, R. C.: 1973, Theory of rational option pricing, *Bell Journal of Economics and Management Science* **4**, 141–183.

Merton, R. C., Scholes, M. S. and Gladstein, M. L.: 1978, The returns and risk of alternative call option portfolio investment strategies, *Journal of Business* **51**, 183–242.

Merton, R. C., Scholes, M. S. and Gladstein, M. L.: 1982, The returns and risks characteristics of alternative put option strategies, *Journal of Business* **55**, 1–56.

Miltersen, K. R., Sandmann, K. and Sondermann, D.: 1997, Closed form solutions for term structure derivatives with lognormal interest rates, *Journal of Finance* **52**, 409–430.

Miron, P. and Swannell, P.: 1992, *Pricing and Hedging Swaps*, Euromoney Books, London.

Newey, W. and West, K.: 1987, A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix, *Econometrica* **55**, 703–708.

Pelsser, A.: 2000, *Efficient Methods for Valuing Interest Rate Derivatives*, Springer Verlag, Berlin.

Politis, D. N. and Romano, J. P.: 1994, The stationary bootstrap, *Journal of the American Statistical Association* **89**, 1303–1313.

Politis, D. N. and White, H.: 2003, Automatic block-length selection for the dependent bootstrap, *forthcoming Econometric Reviews*.

Politis, D., Romano, J. and Wolf, M.: 1999, *Subsampling*, Sringer-Verlag, New York.

Rebonato, R.: 2002, *Modern Pricing of Interest-Rate Derivatives: The Libor Market Model and Beyond*, Princeton University Press, New Jersey.

Risk Magazine: 1996, Value at risk, *Risk Magazine Special Supplement* pp. 68–71.

RiskMetrics: 1996, *Technical Document*, 4^{th} edn, JP Morgan.

Santa-Clara, P. and Sornette, D.: 2001, The dynamics of the forward interest rate curve with stochastic string shocks, *Review of Financial Studies* **14**, 149–185.

Shorack, G. R. and Wellner, J. A.: 1986, *Empirical Processes with Applications to Statistics*, John Wiley & Sons, New York.

SPAN: 1995, *Standard Portfolio Analysis of Risk*, Chicago Mercantile Exchange, Chicago.

Talay, D. and Zheng, Z.: 2002, Worst case model risk management, *Finance and Stochastics* **6**, 517–537.

Tasche, D.: 2002, Expected shortfall and beyond, *Journal of Banking and Finance* **26**, 1519–1533.

Van der Vaart, A. W.: 1998, *Asymptotic Statistics*, Cambridge University Press.

Van der Vaart, A. W. and Wellner, J. A.: 1996, *Weak Convergence and Empirical Processes*, Springer-Verlag, New York.

West, K. D.: 1996, Asymptotic inference about predictive ability, *Econometrica* **64**, 1067–1084.

Yamai, Y. and Yoshiba, T.: 2002a, Comparative analyses of expected shortfall and value-at-risk: Their estimation error, decomposition, and optimization, *Monetary and Economic Studies* **20**, 87–122.

Yamai, Y. and Yoshiba, T.: 2002b, On the validity of value-at-risk: Comparative analyses with expected shortfall, *Monetary and Economic Studies* **20**, 57–86.

Zhang, Z. and Wu, L.: 2001, Optimal low-rank approximation to a correlation matrix, *Working paper University of Science and Technology, Hong Kong.*

Nederlandse Samenvatting

Gedurende de laatste decennia zijn we getuige geweest van een toenemende complexiteit van producten in de financiële markten. Dit heeft ertoe geleid dat financiële instellingen steeds meer gebruik zijn gaan maken van kwantitatieve modellen. Naast een toenemende complexiteit, is ook de omvang van de financiele markten spectaculair toegenomen.

De groeiende complexiteit en handelsvolume van financiële markten maakt de taak voor regulerende instanties lastiger en belangrijker. Dezer dagen bestaat de portefeuille van een bank ook voor een groot gedeelte uit afgeleide instrumenten (derivaten) wiens waarde afhangt van traditionelere instrumenten zoals aandelen, obligaties, en wisselkoersen, maar ook exotischere onderliggende waarden zoals volatiliteit en kredietrisico. Een van de meest in het oog springende verschillen tussen derivaten en de meer fundamentele waarden, zoals aandelen en obligaties, is het belang van theoretische waarderingsmodellen. Deze modellen worden gebruikt om het gedrag van de rentetermijnstructuur, volatiliteitstermijnstructuur, aandelen, etc te voorspellen. Hierdoor wordt een nieuw soort risico geïntroduceerd; modelrisico.

Hoewel modelrisico het meest prominent is in derivatenmarkten is het zeker niet beperkt tot derivatenmarkten. Zo gebruiken risicomanagers allerlei modellen om (neerwaartse) risicomaten te berekenen om het risico van portefeuilles in te schatten. Ze gebruiken modellen die de dynamiek van de portefeuille moeten beschrijven. Indien deze modellen de dynamiek niet accuraat beschrijven volgt een troebele inschatting van het risico. Deze inschatting is van belang voor de bank zelf, alsmede voor de regelgevers die op basis ervan de reserves voor een bank bepalen. Met behulp van een kwaliteitstoets wordt de kwaliteit van het model onderzocht en afhankelijk van de resultaten wordt de reserve berekend. In hoofdstuk 2 presenteren we een raamwerk voor deze kwaliteitstoetsen. Dit raamwerk is toepasbaar op alle hedendaags populaire risicomaatstaven. Het geeft een vergelijking tussen de value-at-risk, de huidige risicomaatstaf van het Basel akkoord, en expected shortfall, de in de academische literatuur populaire tegenhanger. We vinden dat de kwaliteit van modellen gebruikt voor het bepalen van expected shortfall beter gemeten kan worden dan die van modellen voor het bepalen van value-at-risk.

Hoofdstuk 3 gebruikt het raamwerk van hoofdstuk 2 om het risico van portefeuilles met derivaten te bepalen. Het behandelt met name het probleem dat derivatenportefeuilles van samenstelling veranderen over de tijd zonder dat de portefeuille wordt aangepast, aangezien looptijd en moneyness over de tijd veranderen. Benaderingsformules worden bepaald waardoor derivatenportefeuilles in het raamwerk van hoofdstuk 2 kunnen worden geplaatst.

Handelaren zijn zich wel degelijk bewust van het feit dat de modellen die ze gebruiken niet geheel correct zijn en proberen ze aan te passen aan de marktsituatie. Zo gebruiken ze nog steeds overvloedig het Black-Scholes model ondanks dat de gebreken van dit model veelvuldig zijn aangetoond. Met vuistregels gebaseerd op marktkennis bereiken ze echter bevredigende resultaten.

Risicomanagers zijn in het algemeen verder van de markt verwijderd en moeilijker is staat subjectieve correcties aan modellen te maken. Daardoor is modelrisico voor risicomanagers wat lastiger te interpreteren en kwantificeren dan de traditionele risicos, zoals marktrisico, kredietrisico, etc. Een mogelijke oplossing is het zetten van modelreserves voor handelaren. Idealiter zijn deze modelreserves afhankelijk van de markt en het product dat verhandeld wordt, aangezien sommige markten/producten gemakkelijker te modelleren zijn dan anderen. In hoofdstuk 4 presenteren we een raamwerk voor de kwantificering van modelrisico. Het raamwerk is gebaseerd op het berekenen van een risicomaatstaf in een slechtst mogelijk geval in een verzameling van modellen in de buurt van een referentiemodel. Het maakt een opslitsing tussen schattingsrisico en misspecificatierisico. In de behandelde empirische toepassing blijkt het misspecificatierisico veel belangrijker dan het schattingsrisico.

Met de toenemende liquiditeit van optiemarkten is het prijzen van standaardproducten, zoals call- en putopties, niet echt meer een uitdaging. Door marktprijzen op een handige manier te interpoleren kan al gauw een goede inschatting voor een standaardproduct gevonden worden. Het modelrisico van het prijzen van derivaten is daardoor te verwaarlozen (dit is niet het geval voor exotische derivaten, wat in hoofdstuk 7 aan de orde komt). De risico's van de standaardproducten moeten echter ook worden afgedekt en de resultaten hiervan hangen af van de risicoafdekkingsstrategie. In hoofdstuk 5 bekijken we empirisch het modelrisico van risicoafdekken. We bestuderen de S&P 500, en de belangrijkste wisselkoersen. We gebuiken hiervoor de bootstraptechniek en beargumenteren waarom deze te preferen is boven de tot nu toe gebruikte techniek van historische simulatie. We vinden ook hier dat het misspecificatierisico aanzienlijk is.

In hoofdstuk 6 laten we zien dat 2 modellen, het libor marktmodel en het discrete stringmodel, die tot nu toe als verschillend werden verondersteld eigenlijk hetzelfde zijn. Deze modellen dienen dus voor risicomanagementdoeleinden als gelijk te worden behandeld. Hoofdstuk 7 behandelt de factorafhankelijkheid en het schattingsrisico van deze modellen voor het waarderen van exotische producten. Het laat zien dat het modelrisico voor het prijzen van derivaten zeker niet te verwaarlozen is voor exotische derivaten.

Center for Economic Research, Tilburg University, The Netherlands
Dissertation Series

No.	Author	Title	Published
1	P.J.J. Herings	Static and Dynamic Aspects of General Disequilibrium Theory; ISBN 90 5668 001 3	June 1995
2[1]	Erwin van der Krabben	Urban Dynamics: A Real Estate Perspective - An institutional analysis of the production of the built environment; ISBN 90 5170 390 2	August 1995
3	Arjan Lejour	Integrating or Desintegrating Welfare States? - a qualitative study to the consequences of economic integration on social insurance; ISBN 90 5668 003 X	September 1995
4	Bas J.M. Werker	Statistical Methods in Financial Econometrics; ISBN 90 5668 002 1	September 1995
5	Rudy Douven	Policy Coordination and Convergence in the EU; ISBN 90 5668 004 8	September 1995
6	Arie J.T.M. Weeren	Coordination in Hierarchical Control; ISBN 90 5668 006 4	September 1995
7	Herbert Hamers	Sequencing and Delivery Situations: a Game Theoretic Approach; ISBN 90 5668 005 6	September 1995
8	Annemarie ter Veer	Strategic Decision Making in Politics; ISBN 90 5668 007 2	October 1995
9	Zaifu Yang	Simplicial Fixed Point Algorithms and Applications; ISBN 90 5668 008 0	January 1996
10	William Verkooijen	Neural Networks in Economic Modelling - An Empirical Study; ISBN 90 5668 010 2	February 1996
11	Henny Romijn	Acquisition of Technological Capability in Small Firms in Developing Countries; ISBN 90 5668 009 9	March 1996
12	W.B. van den Hout	The Power-Series Algorithm - A Numerical Approach to Markov Processes; ISBN 90 5668 011 0	March 1996
13	Paul W.J. de Bijl	Essays in Industrial Organization and Management Strategy; ISBN 90 5668 012 9	April 1996

No.	Author	Title	Published
14	Martijn van de Ven	Intergenerational Redistribution in Representative Democracies; ISBN 90 5668 013 7	May 1996
15	Eline van der Heijden	Altruism, Fairness and Public Pensions: An Investigation of Survey and Experimental Data; ISBN 90 5668 014 5	May 1996
16	H.M. Webers	Competition in Spatial Location Models; ISBN 90 5668 015 3	June 1996
17	Jan Bouckaert	Essays in Competition with Product Differentiation and Bargaining in Markets; ISBN 90 5668 016 1	June 1996
18	Zafar Iqbal	Three-Gap Analysis of Structural Adjustment in Pakistan; ISBN 90 5668 017 X	September 1996
19	Jimmy Miller	A Treatise on Labour: A Matching-Model Analysis of Labour-Market Programmes; ISBN 90 5668 018 8	September 1996
20	Edwin van Dam	Graphs with Few Eigenvalues - An interplay between combinatorics and algebra; ISBN 90 5668 019 6	October 1996
21	Henk Oosterhout	Takeover Barriers: the good, the bad, and the ugly; ISBN 90 5668 020 X	November 1996
22	Jan Lemmen	Financial Integration in the European Union: Measurement and Determination; ISBN 90 5668 021 8	December 1996
23	Chris van Raalte	Market Formation and Market Selection; ISBN 90 5668 022 6	December 1996
24	Bas van Aarle	Essays on Monetary and Fiscal Policy Interaction: Applications to EMU and Eastern Europe; ISBN 90 5668 023 4	December 1996
25	Francis Y. Kumah	Common Stochastic Trends and Policy Shocks in the Open Economy: Empirical Essays in International Finance and Monetary Policy; ISBN 90 5668 024 2	May 1997
26	Erik Canton	Economic Growth and Business Cycles; ISBN 90 5668 025 0	September 1997

No.	Author	Title	Published
27	Jeroen Hoppenbrouwers	Conceptual Modeling and the Lexicon; ISBN 90 5668 027 7	October 1997
28	Paul Smit	Numerical Analysis of Eigenvalue Algorithms Based on Subspace Iterations; ISBN 90 5668 026 9	October 1997
29	Uri Gneezy	Essays in Behavioral Economics; ISBN 90 5668 028 5	October 1997
30	Erwin Charlier	Limited Dependent Variable Models for Panel Data; ISBN 90 5668 029 3	November 1997
31	Rob Euwals	Empirical Studies on Individual Labour Market Behaviour; ISBN 90 5668 030 7	December 1997
32	Anurag N. Banerjee	The Sensitivity of Estimates, Inferences, and Forecasts of Linear Models; ISBN 90 5668 031 5	December 1997
33	Frans A. de Roon	Essays on Testing for Spanning and on Modeling Futures Risk Premia; ISBN 90 5668 032 3	December 1997
34	Xiangzhu Han	Product Differentiation, Collusion and Standardization; ISBN 90 5668 033 1	January 1998
35	Marcel Das	On Income Expectations and Other Subjective Data: A Micro-Econometric Analysis; ISBN 90 5668 034 X	January 1998
36	Jeroen Suijs	Cooperative Decision Making in a Stochastic Environment; ISBN 90 5668 035 8	March 1998
37	Talitha Feenstra	Environmental Policy Instruments and International Rivalry: A Dynamic Analysis; ISBN 90 5668 036 6	May 1998
38	Jan Bouwens	The Use of Management Accounting Systems in Functionally Differentiated Organizations; ISBN 90 5668 037 4	June 1998
39	Stefan Hochguertel	Households' Portfolio Choices; ISBN 90 5668 038 2	June 1998
40	Henk van Houtum	The Development of Cross-Border Economic Relations; ISBN 90 5668 039 0	July 1998
41	Jorg Jansen	Service and Inventory Models subject to a Delay-Limit; ISBN 90 5668 040 4	September 1998

No.	Author	Title	Published
42	F.B.S.L.P. Janssen	Inventory Management Systems: control and information issues; ISBN 90 5668 041 2	September 1998
43	Henri L.F. de Groot	Economic Growth, Sectoral Structure and Unemployment; ISBN 90 5668 042 0	October 1998
44	Jenke R. ter Horst	Longitudinal Analysis of Mutual Fund Performance; ISBN 90 5668 043 9	November 1998
45	Marco Hoeberichts	The Design of Monetary Institutions; ISBN 90 5668 044 7	December 1998
46	Adriaan Kalwij	Household Consumption, Female Employment and Fertility Decisions: A microeconometric analysis; ISBN 90 5668 045 5	February 1999
47	Ursula Glunk	Realizing High Performance on Multiple Stakeholder Domains: A Resource-based Analysis of Professional Service Firms in the Netherlands and Germany; ISBN 90 5668 046 3	April 1999
48	Freek Vermeulen	Shifting Ground: Studies on the Intersection of Organizational Expansion, Internationalization, and Learning; ISBN 90 5668 047 1	February 1999
49	Haoran Pan	Competitive Pressures on Income Distribution in China; ISBN 90 5668 048 X	March 1999
50	Laurence van Lent	Incomplete Contracting Theory in Empirical Accounting Research; ISBN 90 5668 049 8	April 1999
51	Rob Aalbers	On the Implications of Thresholds for Economic Science and Environmental Policy; ISBN 90 5668 050 1	May 1999
52	Abe de Jong	An Empirical Analysis of Capital Structure Decisions in Dutch Firms; ISBN 90 5668 051 X	June 1999
53	Trea Aldershof	Female Labor Supply and Housing Decisions; ISBN 90 5668 052 8	June 1999
54	Jan Fidrmuc	The Political Economy of Reforms in Central and Eastern Europe; ISBN 90 5668 053 6	June 1999
55	Michael Kosfeld	Individual Decision-Making and Social Interaction; ISBN 90 5668 054 4	June 1999

No.	Author	Title	Published
56	Aldo de Moor	Empowering Communities - A method for the legitimate user-driven specification of network information systems; ISBN 90 5668 055 2	October 1999
57	Yohane A. Khamfula	Essays on Exchange Rate Policy in Developing Countries; ISBN 90 5668 056 0	September 1999
58	Maurice Koster	Cost Sharing in Production Situations and Network Exploitation; ISBN 90 5668 057 9	December 1999
59	Miguel Rosellón Cifuentes	Essays on Financial Policy, Liquidation Values and Product Markets; ISBN 90 5668 058 7	December 1999
60	Sharon Schalk	Equilibrium Theory: A salient approach; ISBN 90 5668 059 5	December 1999
61	Mark Voorneveld	Potential Games and Interactive Decisions with Multiple Criteria; ISBN 90 5668 060 9	December 1999
62	Edward Droste	Adaptive Behavior in Economic and Social Environments; ISBN 90 5668 061 7	December 1999
63	Jos Jansen	Essays on Incentives in Regulation and Innovation; ISBN 90 5668 062 5	January 2000
64	Franc J.G.M. Klaassen	Exchange Rates and their Effects on International Trade; ISBN 90 5668 063 3	January 2000
65	Radislav Semenov	Cross-country Differences in Economic Governance: Culture as a major explanatory factor; ISBN 90 5668 065 X	February 2000
66	Alexandre Possajennikov	Learning and Evolution in Games and Oligopoly Models; ISBN 90 5668 064 1	March 2000
67	Eric Gaury	Designing Pull Production Control Systems: Customization and Robustness; ISBN 90 5668 066 8	March 2000
68	Marco Slikker	Decision Making and Cooperation Restrictions; ISBN 90 5668 067 6	April 2000
69	Ruud Brekelmans	Stochastic Models in Risk Theory and Management Accounting; ISBN 90 5668 068 4	June 2000
70	Flip Klijn	A Game Theoretic Approach to Assignment Problems; ISBN 90 5668 069 2	June 2000

No.	Author	Title	Published
71	Maria Montero	Endogenous Coalition Formation and Bargaining; ISBN 90 5668 070 6	June 2000
72	Bas Donkers	Subjective Information in Economic Decision Making; ISBN 90 5668 071 4	June 2000
73	Richard Nahuis	Knowledge and Economic Growth; ISBN 90 5668 072 2	October 2000
74	Kuno Huisman	Technology Investment: A Game Theoretic Real Options Approach; ISBN 90 5668 073 0	October 2000
75	Gijsbert van Lomwel	Essays on Labour Economics; ISBN 90 5668 074 9	November 2000
76	Tina Girndt	Cultural Diversity and Work-Group Performance: Detecting the Rules; ISBN 90 5668 075 7	December 2000
77	Teye A. Marra	The Influence of Proprietary Disclosure Costs on the Decision to Go Public: Theory and Evidence; ISBN 90 5668 076 5	January 2001
78	Xiaodong Gong	Empirical Studies on the Labour Market and on Consumer Demand; ISBN 90 5668 078 1	February 2001
79	Kanat Çamlibel	Complementarity Methods in the Analysis of Piecewise Linear Dynamical Systems; ISBN 90 5668 079 X	May 2001
80	Joost Driessen	Empirical Studies on the Pricing of Bonds and Interest Rate Derivatives; ISBN 90 5668 080 3	June 2001
81	Bram van den Broek	Uncertainty in Differential Games; ISBN 90 5668 081 1	June 2001
82	Ricardo Recht	Organisational Culture and Privatisation: A Case Study of the Argentinean Railway Sector; ISBN 90 5668 082 X	June 2001
83	Willem Verhagen	Inflation Targeting and Interest Rate Policy; ISBN 90 5668 083 8	June 2001
84	Vincent Verouden	Essays in Antitrust Economics; ISBN: 90 5668 086 2	June 2001
85	Rosalia Vazquez Alvarez	A Nonparametric Approach to the Sample Selection Problem in Survey Data; ISBN 90 5668 087 0	June 2001
86	Theo Leers	Public Pensions and Population Ageing: An Economic Analysis of Fertility, Migration, and Social-Security Policy; ISBN 90 5668 084 6	June 2001

No.	Author	Title	Published
87	Bernard Conlon	Consumer Rationality in Choice; ISBN 90 5668 085 4	June 2001
88	Judith Timmer	Cooperative behaviour, uncertainty and operations research; ISBN 90 5668 088 9	September 2001
89	Enrico Diecidue	Nonexpected Utility and Coherence; ISBN 90 5668 089 7	November 2001
90	Alexander Konovalov	Essays in General Equilibrium Theory; ISBN 90 5668 091 9	November 2001
91	Stefan Stremersch	Essays on Marketing Strategy in Technology-Intensive Markets; ISBN 90 5668 090 0	December 2001
92	Rian Drogendijk	Expansion Patterns of Dutch Firms in Central and Eastern Europe: Learning to Internationalize; ISBN 90 5668 092 7	December 2001
93	Han Donker	Takeovers. An Empirical Study of Ownership Structure and the Wealth to Shareholders in Dutch Takeovers; ISBN 90 5668 093 5	December 2001
94	Erica van Herpen	Perceptions and Evaluations of Assortment Variety; ISBN 90 5668 094 3	December 2001
95	Jana Vyrastekova	Cheap Talk and Spiteful Preferences in Ultimatum Games: Experiments and Evolutionary Rationale; ISBN 90 5668 095 1	March 2002
96	Tao Lin	Statistics of Extremes in the Space of Continuous Functions; ISBN 90 5668 096 X	May 2002
97	Sander Onderstal	Papers in Auction Theory; ISBN 90 5668 097 8	May 2002
98	Wolf Wagner	Risk Sharing under Incentive Constraints; ISBN 90 5668 098 6	June 2002
99	Willem-Jan van den Heuvel	Integrating Modern Business Applications with Objectified Legacy Systems; ISBN 90 5668 099 4	June 2002
100	Jacco Wielhouwer	Optimal Tax Depreciation and its Effects on Optimal Firm Investments; ISBN 90 5668 100 1	June 2002
101	Peter Roosenboom	Corporate Governance Mechanisms in IPO Firms; ISBN 90 5668 101 X	September 2002

No.	Author	Title	Published
102	Victoria Shestalova	Essays in Productivity and Efficiency; ISBN 90 5668 103 6	September 2002
103	Paula van Veen-Dirks	Flexibility and Control: An Empirical Study Relating Production Flexibility to the Design of Performance Management Systems; ISBN 90 5668 102 8	September 2002
104	Michal Matějka	Management Accounting in Organizational Design: Three Essays; ISBN 90 5668 105 2	November 2002
105	Maarten Pronk	Market Liquidity around Earnings Announcements; ISBN 90 5668 106 0	November 2002
106	Nancy Kamp-Roelands	Towards a framework for auditing environmental reports; ISBN 90 5668 108 7	November 2002
107	Ana Ferreira	Statistics of extremes: estimation and optimality; ISBN 90 5668 107 9	November 2002
108	Alexei Goriaev	On the behaviour of Mutual Fund Investors and Managers; ISBN 90 5668 110 9	November 2002
109	Paul Ingenbleek	Money for Value Pricing from a Resource-Advantage Perspective; ISBN 90 5668 109 5	December 2002
110	Frank Wijen	Stakeholder influence and organizational learning in environmental management; ISBN 90 5668 111 7	December 2002
111	Michele Belot	Labor market institutions in OECD countries: Origins and consequences; ISBN 90 5668 113 3	January 2003
112	Federica Teppa	Estimating preference parameters in individual financial decision making; ISBN 90 5668 112 5	January 2003
113	Dmitry Danilov	The effects of pretesting in econometrics with applications in finance; ISBN 90 5668 104 4	February 2003
114	Jacco Thijssen	Investment under uncertainty, market evolution and coalition spillovers in a game theoretic perspective; ISBN 90 5668 114 1	May 2003

No.	Author	Title	Published
115	Laura Spierdijk	Empirical Studies of Market Microstructure; ISBN 90 5668 115 X	June 2003
116	Nicole van Beeck	A New Decision Support Method for Local Energy Planning in Developing Countries; ISBN 90 5668 116 8	June 2003
117	Grzegorz Pawlina	Corporate Investment under Uncertainty and Competition: A Real Options Approach; ISBN 90 5668 118 4	June 2003
118	Jana P. Fidrmuc	The Impact of Blockholders on Information Signalling, Productivity, and Managerial Disciplining; ISBN 90 5668 117 6	June 2003
119	Sjoerd Beugelsdijk	Culture and Economic Development in Europe; ISBN 90 5668 121 4	September 2003
120	Bas van Groezen	The Wealth of Generations: Ageing, Pensions and Welfare from a Macroeconomic Perspective; ISBN 90 5668 120 6	October 2003
121	Pham Do Kim Hang	Essays in Game Theory and Natural Resource Management; ISBN 90 5668 122 2	September 2003
122	Jeroen van de Ven	Psychological Sentiments and Economic Behaviour; ISBN 90 5668 123 0	October 2003
123	Jeroen Kerkhof	Model Risk Analysis for Risk Management and Option Pricing; ISBN 90 5668 124 9	November 2003

[1] Copies can be ordered from Thela Thesis, Prinseneiland 305, 1013 LP Amsterdam, The Netherlands, phone: + 31 20 6255429; fax: +31 20 6203395; e-mail: office@thelathesis.nl