Netspar Network for Studies on Pensions, Aging and Retirement

Jeroen Kerkhof

Bertrand Melenberg

Hans Schumacher

# Model risk and regulatory capital

**Discussion Papers 2006 – 012**

March 2006

# Model Risk and Regulatory Capital

JEROEN KERKHOF*, BERTRAND MELENBERG†, and HANS SCHUMACHER‡

Tilburg University

PO Box 90153

5000 LE Tilburg

The Netherlands

March 27, 2006

*Fixed Income, Morgan Stanley, London. E-mail: jeroen.kerkhof@morganstanley.com.

†Department of Econometrics and Operations Research, Department of Finance, and CentER, Tilburg University. E-mail: B.Melenberg@TilburgUniversity.nl.

‡Department of Econometrics and Operations Research, and CentER, Tilburg University. E-mail: J.M.Schumacher@TilburgUniversity.nl.

# Model Risk and Regulatory Capital


**ABSTRACT**

In this paper we propose a general framework for the 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. The approach is based on computing worst-case risk measures over sets of models that are in some appropriate sense close to a nominal model. In general, any set of models could be used, but we illustrate how, in particular, past data can be used to construct such model sets in a statistically meaningful way. This empirically based approach allows us to decompose the model risk into estimation risk, misspecification risk, and identification risk. 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.

We present an application to stock portfolios and find that, for usually applied specifications, misspecification risk is much more important than estimation risk. The combination of estimation risk and misspecification risk that we find explains half of the multiplication factors employed by the Bank of International Settlements (BIS). The remaining other half can be attributed to additional misspecification and identification risk, and, possibly, other risk factors.



**JEL codes: G12, G18**

*Key words:* Capital requirements, (Coherent) Risk management, Estimation risk.

# I. 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 modelled. 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 incorrect 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, i.e., the risk of loss from inadequate or failing internal processes, people, and systems, or from external events, 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. But 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 of total market risk and the market risk measure computed from the nominal model.

The set of alternative models can be determined in many different ways. We propose empirically based choices of these alternative models. Our approach allows us to decompose the total model risk into three components: *estimation risk*, which is due to the fact that on the basis of past data we can only *estimate* the future risk; *misspecification risk*, because of the possibility of an incorrect choice of the econometric model; and *identification risk*, which is due to the possibility that different econometric models might describe the same data, and we are not able to choose between these different econometric models due to lack of identification. 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 et al. (1999), Talay and Zheng (2002), and Bossy et al. (2000)). However, our study suggests that the misspecification error is generally a more important factor in model risk than the estimation error, while the identification error has, at least, the potential to be a real leading factor in model risk.

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. In particular, we quantify the estimation and misspecification risk, using nonparametric estimates of the value-at-risk and expected

shortfall. We find that only allowing for estimation risk is not sufficient, but when also taking into account misspecification risk, an appropriate quantification of market risk turns out to become possible. 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 by incorporating estimation and misspecification risk when computing the 1% value-at-risk at a 95% confidence level. The other half can be attributed to a higher required confidence level or to residual risk, including perceived additional misspecification risk, identification risk, and possibly other risk factors.

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 Kerkhof et al. (2003)), 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 first illustrate in a simplified setting model risk, its decomposition in estimation risk, misspecification risk, and identification risk, and we illustrate that misspecification risk might be more important than estimation risk, while identification risk might be the dominant factor. In Section III we first present an overview of market risk measurement, with some emphasis on coherent market risk measurement, and then we propose our general

[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 et al. (1998) fit into the proposed setup.

framework for incorporation of model risk, which fits neatly with coherent market risk measurement methods. In Section IV we specialize our general model risk framework to an empirically based one, assuming having available past data. In this section we also discuss the decomposition of model risk into the three components estimation, misspecification, and identification risks. Section V provides an application to portfolio risk management. We discuss the value-at-risk and the expected shortfall approach. Finally, Section VI concludes.

## II. Illustration of Model Risk

By *market risk* we understand the risk caused by uncertain[2] 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.[3] 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.

To illustrate, suppose that $X_T$ denotes the current (known) position, and that the future position can be described by a random variable $X_{T+1} = X_T \exp(Y_{T+1})$, where the log return $Y_{T+1}$ follows some unknown distribution (conditional upon the information of time $T$, represented by $\mathcal{F}_T$). To proceed, one could choose to model the distribution of the log return $Y_{T+1}$ to be normally distributed with mean $\mu_Y$ and variance $\sigma_Y^2$, i.e.,

$$Y_{T+1} \mid \mathcal{F}_T \sim \mathcal{N}(\mu_Y, \sigma_Y), \tag{1}$$

[2]'Uncertain fluctuations' is usually quantified as 'random fluctuation.'

[3]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.

or, equivalently, $Y_{T+1} = \log(X_{T+1}/X_T) = \mu_Y + \sigma_Y \epsilon_{T+1}$ with $\epsilon_{T+1} \mid \mathcal{F}_T \sim \mathcal{N}(0,1)$. If, in addition, one chooses to measure the market risk of $X_{T+1} - X_T$ using the Value-at-Risk at level $p$, one will find as risk measure,[4]

$$\mathrm{VaR}_T\left(X_{T+1} - X_T\right) = X_T\left(1 - \exp\left(z_p \sigma_Y + \mu_Y\right)\right), \tag{2}$$

where $z_p$ indicates the $p$-th quantile of the standard normal distribution, and where the index $T$ indicates that the VaR is calculated at time $T$, given the information $\mathcal{F}_T$. When $\mathrm{VaR}_T\left(X_{T+1} - X_T\right) = A > 0$, then, by adding the non-risky amount $A$ to the future position, we get

$$\mathrm{VaR}_T\left((X_{T+1} + A) - X_T\right) = \mathrm{VaR}_T\left(X_{T+1} - X_T\right) - A = 0, \tag{3}$$

due to the translation invariance property of the VaR, so that $(A =)\mathrm{VaR}_T(X_{T+1} - X_T)$ is indeed the amount of 'sweetener' needed to make the position $X_{T+1}$ acceptable.

Alternatively, measuring the market risk using the Expected Shortfall (ES) at level $p$, one will find, under the same distributional assumptions, as risk measure[5]

$$\mathrm{ES}_T\left(X_{T+1} - X_T\right) = X_T\left(1 - \frac{1}{p}\exp\left(\mu_Y + \frac{1}{2}\sigma_Y^2\right)\Phi\left(z_p - \sigma_Y\right)\right) \tag{4}$$

where $\Phi$ denotes the standard normal distribution function. The ES also satisfies the translation invariance property, meaning that $\mathrm{ES}_T(X_{T+1} - X_T)$ is the amount of sweetener that makes the position $X_{T+1}$ acceptable in case one uses the ES as risk measure.

However, in a practical situation, VaR and ES are not observable, since $\mu_Y$ and $\sigma_Y$ are unknown parameters. Instead, one has to estimate $\mu_Y$ and $\sigma_Y$, say, by $\widehat{\mu}_Y$ and $\widehat{\sigma}_Y$, respectively, assuming having available, for instance, an *i.i.d.* sample of past values of $Y$. This results in an estimated VaR and estimated ES, replacing $(\mu_Y, \sigma_Y)$ by their estimates $(\widehat{\mu}_Y, \widehat{\sigma}_Y)$. Clearly, $\widehat{\mu}_Y$ and $\widehat{\sigma}_Y$ will contain sampling or estimation error, which will be transferred to the estimated VaR and ES, so that the use of these estimated VaR

[4]The VaR is calculated as $-F^{-1}(p)$, with $F$ the distribution function of $X_{T+1} - X_T$.
[5]The ES is calculated as $-\mathbb{E}\left(X_{T+1} - X_T \mid X_{T+1} - X_T \leq \mathrm{VaR}_T(X_{T+1} - X_T)\right)$.

or ES might result in an underestimation of the actual risk. To take this estimation error into account, one could construct, say, 95%-confidence intervals for the VaR and ES, and use the upper confidence level to quantify the risk with estimation error incorporated. The difference between the upper bound of this confidence interval and the estimated risk measure then gives an indication of −what might be called− the *estimation risk.*

We illustrate this in Table I. This table contains in the second and third columns the estimates of the VaR and ES at levels $p = 0.01$ and $p = 0.025$ and the corresponding estimation risks. These estimation risks are calculated as the difference between the upper bounds of the 95% confidence intervals and the estimated risk measure, based on a two-year sample of 500 observations (= $2 \times 250$ trading days), when the underlying distribution of the log return is normal with annualized mean $\mu = 0$ and variance $\sigma^2 = 0.25$, the initial capital is equal to $X_T = 1$, and the estimation method employed is Maximum Likelihood. The *relative* estimation error, calculated as the estimation error relative to the estimated risk measure, is in all cases around 7%.

However, in addition to estimation risk, we should also take into account the possibility that the log return $Y_{T+1}$ is not distributed according to the $\mathcal{N}(\mu_Y, \sigma_Y)$-distribution, conditional upon $\mathcal{F}_T$. Postulating no more than $Y_{T+1} \mid \mathcal{F}_T \sim F$, with $F$ left unspecified, we easily find for the Value-at-Risk at level $p$

$$\mathrm{VaR}_T\left(X_{T+1} - X_T\right) = X_T(1 - \exp(F^{-1}(p)), \tag{5}$$

and for the Expected Shortfall at level $p$ we get

$$\mathrm{ES}_T\left(X_{T+1} - X_T\right) = X_T(1 - \frac{1}{p}\int \exp(y)\mathbf{1}_{(-\infty, F^{-1}(p)]}(y)dF(y)). \tag{6}$$

Denote by $\widehat{F}$ some non-parametric estimate of $F$ (still assuming an *i.i.d.*-sample). Then a non-parametric estimate of VaR and ES can be obtained by replacing $F$ by $\widehat{F}$. Sampling error in these non-parametric estimates can again be taken into account by constructing, say, 95%-confidence intervals. The difference between the upper bounds of these intervals and the estimated risk measure then provides us with some kind of worst case scenario:

the combination of possible model misspecification and estimation error, which we shall refer to as the combination of *estimation* and *misspecification risk*.

Table I (columns 4-6) presents these combinations of estimation and misspecification risk. These combinations are calculated as the difference between the upper bounds of the 95% confidence intervals and the estimated risk measure, for the cases when the log returns are generated by a normal, a $t_4$, and a mixed normal, in all cases standardized such that the annualized mean equals $\mu = 0$ and $\sigma^2 = 0.25$, using a nonparametric estimation technique.[6] In case of the mixed normal both normals have mean 0 and the first, with probability 0.99, has annualized variance 0.24, and the second, with probability 0.01, has annualized variance 1.24. For the normal distribution we see that the relative model error due to the combination of estimation risk and allowing for the possibility of misspecification risk is around 12%. Since there is no misspecification in this case, allowing for the possibility of misspecification leads to an increase in the estimation risk from around 7% to around 12%. When the distribution is non-normal, the relative combined error increases up to 42% in case of VaR, and even up to 75% in case of ES (in both cases the standardized $t_4$-distribution at level $p = 0.01$), indicating that the misspecification error (up to 63% $(= 75\% - 12\%)$) might be much more serious than the estimation error (around 12%).

Postulating $Y_{T+1} \mid \mathcal{F}_T \sim F$, with $F$ estimated by $\widehat{F}$ under the assumption of a random sample, means that one assumes that the future can be represented accurately by the past. However, whether this is the case or not, cannot be retrieved from past data. Instead, one could also equally well model that $(Y_{T+1} - \alpha)/\beta \mid \mathcal{F}_T \sim F$, postulating that there might be a change in the level or volatility of $Y_{T+1}$ represented by $\alpha$ and $\beta$, respectively. Particularly, this could make sense in a stress testing context. As a consequence, the Value-at-Risk at level $p$ will be given by

$$\mathrm{VaR}_T\left(X_{T+1} - X_T\right) = X_T(1 - \exp(\alpha + \beta F^{-1}\left(p\right)), \tag{7}$$

[6] See the section containing the empirical application for further details.

**Table I**
**Illustration of Model Risk**

This table reports the value-at-risk (VaR) and expected shortfall (ES) market risk at level $p$, assuming i.i.d.-normal returns with annualized mean 0 and variance 0.25, with corresponding estimation risk (Est risk), based on a sample of 500 observations. In addition, estimation and misspecification risk (E&MR) is reported when the data are generated by a normal (N), a $t_4$ (t), and a mixed normal (MixedN) distribution, standardized to have annualized mean 0 and variance 0.25. In case of the mixed normal both normals have mean 0 and the first, with probability 0.99, has annualized variance 0.24, and the second, with probability 0.01, has annualized variance 1.24. For definitions of estimation and misspecification risk, see main text.

| $p$ | VaR | Est. risk | E&MR N | E&MR t | E&MR MixedN. |
|---|---|---|---|---|---|
| 0.01 | 0.0707 | 0.0048 | 0.0095 | 0.0299 | 0.0106 |
| 0.025 | 0.0599 | 0.0043 | 0.0069 | 0.0112 | 0.0067 |
| $p$ | ES | Est. risk | E&MR N | E&MR t | E&MR MixedN. |
| 0.01 | 0.0805 | 0.0053 | 0.0101 | 0.0604 | 0.0197 |
| 0.025 | 0.0709 | 0.0049 | 0.0083 | 0.0303 | 0.0116 |

and for the Expected Shortfall at level $p$ we get

$$\mathrm{ES}_T\left(X_{T+1}-X_T\right)=X_T(1-\frac{1}{p}\int\exp(\alpha+\beta y)\mathbf{1}_{(-\infty,F^{-1}(p)]}(y)dF\left(y\right)). \tag{8}$$

Using the nonparametric estimate $\widehat{F}$ to estimate $F$, assuming that the past is (only) representative for $F$ (but not for $\alpha$ or $\beta$), one can estimate the Value-at-Risk and Expected Shortfall, corresponding to each chosen value of $\alpha$ and $\beta$. Again, sampling error can be taken into account by constructing 95%-confidence intervals. However, the "correct" values of $\alpha$ and $\beta$ cannot be retrieved from past data, since they are not identified. The difference between the upper bounds of the confidence intervals for some range of $\alpha$ and $\beta$ (including $\alpha=0$ and $\beta=1$) and the estimated risk measure then again provide us with some kind of worst case scenario, but now also taking into account the possible error due to the lack of identification. We shall refer to this error as *total model risk*, consisting of *estimation, misspecification*, and *identification risk*, while the sum of

the estimated market risk and this total model risk will be referred to as *total market risk*. The size of the identification risk depends on the selected range of $\alpha$ and $\beta$: the identification risk will be zero in case $\alpha$ is restricted to 0 and $\beta$ is restricted to 1, and the total market risk will become maximal due to the identification risk (i.e., equal to $X_T$, thus, in relative terms 100%) in case $\alpha$ becomes very negative or $\beta$ becomes very large (assuming $F^{-1}(p) < 0$).

# III. Model Risk

In this section we present our quantification of model risk. First, we describe the quantification of market risk given a model class instead of a single model. Value at risk is an important example of a risk measure. But, since nowadays coherency of risk measures plays a crucial role, we also pay special attention to such risk measures. Next, we discuss our proposal for the quantification of model risk, including its characteristics. In this section our approach will be in more or less general terms. In the next section we shall specialize to an empirically based quantification of model risk. For the sake of simplicity we shall focus on single period risk quantification, where the risk of time $t$ is to be quantified at time $t-1$.

## A. Market Risk Measurement

A model will be denoted by $m$, while a model class will be denoted by $\mathcal{M}$. The aim of this subsection is to describe how to quantify the risk of a *product* $\Pi_t$ at some time $t$, using a *risk measurement method* $\rho_t$, given a model $m \in \mathcal{M}$.

Let a model class $\mathcal{M}$ be given. We assume that each model $m \in \mathcal{M}$ describes a probability setting consisting of a probability space $(\Omega_{m,t}, \mathcal{F}_{m,t}, \mathbb{P}_{m,t})$ as viewed from period $t-1$. Given the probability space $(\Omega_{m,t}, \mathcal{F}_{m,t}, \mathbb{P}_{m,t})$, let $\mathcal{R}_t(m) = L_0(\Omega_{m,t}, \mathcal{F}_{m,t}, \mathbb{P}_{m,t})$ denote the space of equivalence classes of measurable real-valued functions on $(\Omega_{m,t}, \mathcal{F}_{m,t})$. Following the terminology of Artzner et al. (1999) and Delbaen (2000), elements of $\mathcal{R}_t(m)$ are called risks (as viewed from time $t-1$).

A *product* $\Pi_t$ at time $t$ and defined on $\mathcal{M}$ is a mapping $\Pi_t$ that assigns to each model

$m \in \mathcal{M}$ a risk $\Pi_{m,t} \in \mathcal{R}_t(m)$. The set of all products $\Pi_t$ defined on $\mathcal{M}$ at time $t$ is denoted by $\mathcal{X}_t(\mathcal{M})$. [7]

Next, we proceed to market risk measurement. First, given a model $m$, a *model m specific risk measure* at time $t-1$ is a map from $\mathcal{R}_t(m)$ to $\mathbb{R} \cup \{\infty\}$.[8] Then a *risk measurement method* at time $t-1$ defined on $\mathcal{M}$ is a mapping $\rho_t$ that assigns to each model $m \in \mathcal{M}$ a model $m$ specific risk measure $\rho_{m,t}$.

Using these concepts, we then have that the (single-period) risk of a *product* $\Pi_t$ using a *risk measurement method* $\rho_t$, given a model $m \in \mathcal{M}$ is given by $\rho_{m,t}(\Pi_{m,t})$. Most risk measures used in practice can be viewed as model specific risk measures and can straightforwardly be extended to risk measurement methods as described here.

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 1** (Value at Risk (VaR)) Let a model class $\mathcal{M}$ be given. The *value-at-risk* method with *reference asset* $N_t \in \mathcal{X}_t(\mathcal{M})$ and *level* $p \in (0,1)$ assigns to a model $m \in \mathcal{M}$ with probability space $(\Omega_{m,t}, \mathcal{F}_{m,t}, \mathbb{P}_{m,t})$ the risk measure $\mathrm{VaR}_{m,t}$ given by

$$\mathrm{VaR}_{m,t} : \mathcal{R}_t(m) \ni X \mapsto -\inf\left\{q \in \mathbb{R} : \mathbb{P}_{m,t}\left(X/N_{m,t} \leq q\right) \geq p\right\} \in \mathbb{R} \cup \{\infty\}. \tag{9}$$

Artzner et al. (1997) introduced the notion of coherent risk measures, which nowadays play a crucial role in regulatory based risk measurement. Their ideas were formalized in Artzner et al. (1999), Artzner (1999), and Delbaen (2000). A *coherent risk measure* is a risk measure that satisfies the so-called axioms of translation invariance, monotonicity, subadditivity, and positive homogeneity. These axioms are in our context

[7] Since $\mathcal{R}_t(m)$ is a vector space, the set of products $\mathcal{X}_t(\mathcal{M})$ has the structure of a vector space as well. For instance, if $\Pi_t^1$ and $\Pi_t^2$ are products defined on $\mathcal{M}$, then $\Pi^1 + \Pi^2$ is the product that associates to a model $m$ in $\mathcal{M}$ the risk $\Pi_{m,t}^1 + \Pi_{m,t}^2$. 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.

[8] Including $\infty$ allows risks to be defined on more general probability spaces, see Delbaen (2000).

model specific, but they can be transferred to risk measurement methods in a straightforward way. We shall say that a risk measurement method $\rho_t$ defined on a model class $\mathcal{M}$ is coherent, if for each $m \in \mathcal{M}$ the risk measure $\rho_{m,t}$ on $m$ satisfies the coherency axioms.[9] The VaR measurement method does not satisfy the subadditivity property (see Artzner et al. (1999) for a counterexample), and, thus, is not coherent. A practically usable coherent regular risk measurement method is the expected shortfall as given in Acerbi and Tasche (2002).

**Definition 2** (Expected Shortfall (ES)) The *expected shortfall method* with *reference asset* $N_t \in \mathcal{X}_t(\mathcal{M})$ and *level* $p \in (0,1)$ assigns to a model $m \in \mathcal{M}$ with probability space $(\Omega_{m,t}, \mathcal{F}_{m,t}, \mathbb{P}_{m,t})$ the risk measure $\mathrm{ES}_{m,t}$ given by

$$\mathrm{ES}_{m,t} : \mathcal{R}_t(m) \ni X \mapsto -\frac{1}{p}\left(\mathbb{E}_{\mathbb{P}_{m,t}} X \mathbf{I}_{(-\infty, Q_{p,m,t}(X/N_{m,t})]}\right.$$
$$\left. + Q_{p,m,t}\left(X/N_{m,t}\right)\left(p - \mathbb{P}_{m,t}\left(X/N_{m,t} \leq Q_{p,m,t}\left(X/N_{m,t}\right)\right)\right)\right) \in I\!R \cup \{\infty\}. \quad (10)$$

where $Q_{p,m,t}\left(X/N_{m,t}\right)$ denotes the (lower) $p$-quantile of $X$.

In the empirical application we shall use both the VaR- and ES-risk measurement methods.

## B. Quantifying model risk

In this subsection we propose a general framework to quantify 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

[9] Notice that in case of the translation invariance axiom, this required that a reference product in $\mathcal{X}_t(\mathcal{M})$ is given.

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.

To quantify, suppose that the financial institution uses a model $m$ based risk measure $\rho_{m,t}$ to assess the acceptability of a product (portfolio) $\Pi_t$. In model $m$, the risk of the product $\Pi_t$ is computed as $\rho_{m,t}(\Pi_{m,t})$. 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 $\sup_{k\in\mathcal{K}} \rho_{m,t}(\Pi_{m,t})$.[10] To define model risk, we need the axiom of translation invariance. A risk measurement method $\rho_t$ satisfies this axiom (with respect to a reference product $N_t \in \mathcal{X}_t(\mathcal{M})$), if, for all $m \in \mathcal{M}$, and for all $X \in \mathcal{R}_t(m)$ and $\tau \in \mathbb{R}$, we have $\rho_{m,t}(X + \tau N_{m,t}) = \rho_{m,t}(X) - \tau$. Model risk may now be quantified as follows.

**Definition 3** (Model risk measure)[11] 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 $\Pi_t$ be a product and let $\rho_t$ be a risk measurement method satisfying the axiom of translation invariance. The *model risk* associated to the risk measure $\rho_t$ of product $\Pi_t$, with respect to the *nominal model* $m$ and the *tolerance set* $\mathcal{K}$, is given by

$$\phi_{\rho_t}(\Pi_t, m, \mathcal{K}) = \sup_{k\in\mathcal{K}} \rho_{k,t}(\Pi_{k,t}) - \rho_{m,t}(\Pi_{m,t}). \tag{11}$$

The intended interpretation that the amount $\phi_{\rho_t}(\Pi_t, m, \mathcal{K})$ can be thought of as a model reserve that should be held to cover the worst-case dynamics of $\mathcal{K}$, depends on the translation invariance axiom of the risk measurement method which is therefore crucial in the definition. We have the following theorem.

**Theorem 1** (Translation Invariance) Let $\rho_t$ be a risk measurement method that is

[10] 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.

[11] The case where $\rho_{m,t}(\Pi_{m,t}) = \infty$ is uninteresting since the financial institution will never accept the product $\Pi_t$ in its portfolio.

translation invariant with respect to a reference product $N_t$. Then the model risk measure associated to $\rho_t$ is invariant in the sense that

$$\phi_{\rho_t}(\Pi_t + \tau N_t, m, \mathcal{K}) = \phi_{\rho_t}(\Pi_t, m, \mathcal{K}) \tag{12}$$

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

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

$$\begin{aligned}
\phi_{\rho_t}(\Pi_t + \tau N_t, m, \mathcal{K}) &= \sup_{k\in\mathcal{K}} \rho_{k,t}\left(\Pi_{k,t} + \tau N_{k,t}\right) - \rho_{m,t}\left(\Pi_{m,t} + \tau N_{m,t}\right) \\
&= \sup_{k\in\mathcal{K}} \rho_{k,t}\left(\Pi_{k,t}\right) - \tau - \rho_{m,t}\left(\Pi_{m,t}\right) + \tau \\
&= \phi_{\rho_t}(\Pi_t, m, \mathcal{K}).
\end{aligned}$$

■

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 a positive homogeneous risk measurement method the model risk measure is positively homogeneous. The model risk measure does not, in general, satisfy the monotonicity and subadditivity axioms of coherence. 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 the nominal market risk and model risk, i.e., $\sup_{k\in\mathcal{K}} \rho_{k,t}(\Pi_{k,t})$.

Thus, when applied to a coherent risk measurement method, our way of measuring model risk can be seen as a coherent approach. Indeed, a coherent risk measure $\rho_{m,t}$ for a model $m$ with corresponding probability space $(\Omega_{m,t}, \mathcal{F}_{m,t}, \mathbb{P}_{m,t})$ can be written in the form[12]

$$\rho_{m,t}\left(\Pi_{m,t}\right) = \sup_{\mathbb{Q}\in\mathcal{P}_{m,t}} \mathbb{E}_{\mathbb{Q}}\left[\Pi_{m,t}\right],$$

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

where $\mathcal{P}_{m,t}$ consist of a set of probability distributions defined on $(\Omega_{m,t}, \mathcal{F}_{m,t})$, where different choices of $\mathcal{P}_{m,t}$ produce different risk measures. The model risk measure for a general coherent risk measure is then given by

$$\phi_{\rho t}(\Pi_t, m, \mathcal{K}) = \sup_{k\in\mathcal{K}} \sup_{\mathbb{Q}\in\mathcal{P}(k,t)} \mathbb{E}_{\mathbb{Q}}\left[\Pi_{k,t}\right] - \sup_{\mathbb{Q}\in\mathcal{P}(m,t)} \mathbb{E}_{\mathbb{Q}}\left[\Pi_{m,t}\right].$$

with corresponding *total market risk*

$$\sup_{k\in\mathcal{K}} \sup_{\mathbb{Q}\in\mathcal{P}(k,t)} \mathbb{E}_{\mathbb{Q}}\left[\Pi_{k,t}\right].$$

Notice, however, the difference between $\mathcal{P}_{m,t}$ and $\mathcal{K}$: $\mathcal{P}_{m,t}$ is a set of probability measures based on *one* base probability measure $\mathbb{P}_{m,t}$ 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.

There are many reasonable choices possible for the set $\mathcal{K}$. In the next section we shall discuss an empirically based one. Model risk then becomes *econometrically* based, allowing for a natural decomposition into estimation risk, misspecification risk, and identification risk. Most of the empirically based model risk investigations so far, are only concerned with estimation risk. But, as we already illustrated in the previous section, an empirically based model risk quantification should also include misspecification risk and take account of potential identification risk.

## IV. Empirically based Risk Measurement Methods

The aim of this section is to present an empirically based procedure to choose both $m$ and $\mathcal{K}$, given some selected model class $\mathcal{M}$, and having available past data. First, we describe in more or less general terms how the selection of a nominal model $m$ takes place, given $\mathcal{M}$, and applying statistical or econometric techniques. Then we describe the potential errors one can make, leading, subsequently, to our proposed empirically based quantification of model risk, decomposed into estimation risk, misspecification

risk, and identification risk. As example, we discuss in some detail the case where the past data are known or postulated to be a random sample from some underlying population distribution.

## A. Selection of a nominal model

In this subsection we describe the econometrically based selection of a nominal model, given some selected model class $\mathcal{M}^s$ (with index $s$ of *selected*) and having available past data, following Heckman (2000). We start by introducing the set of data generating processes $\mathcal{D}_T$, describing the past data. We assume the availability of a time series of $T$ observations $X_1, ..., X_T$ on the actual product $\Pi^a$ (with index $a$ of *actual*). We then postulate that the set $\mathcal{D}_T$ consists of all (relevant) probability distributions on $(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$, such that

$$(X_1, ..., X_T) \sim F_T^a \in \mathcal{D}_T.$$

At most $F_T^a$, or some of its characteristics, can be retrieved on the basis of the data $X_1, ..., X_T$. Many choices for the set $\mathcal{D}_T$ are possible. The only requirement is that the actual $F_T^a$ belongs to $\mathcal{D}_T$. Quite often, however, we do not have much information about $F_T^a$, so that we might have to choose the set $\mathcal{D}_T$ quite large. But in case we would know, for instance, that $X_1, ..., X_T$ are *i.i.d.* or stationary, this might help us in restricting the set $\mathcal{D}_T$.

Next, we assume that, given the model class $\mathcal{M}^s$, with corresponding sets of products $\mathcal{X}_t(\mathcal{M}^s)$, $t = 1, ..., T$, some $\Pi_t^s \in \mathcal{X}_t(\mathcal{M}^s)$ is chosen to represent the actual product $\Pi^a$ at time $t$. Define $\overrightarrow{\Pi}_T^s = (\Pi_1^s, ..., \Pi_T^s)$. Given this selected $\overrightarrow{\Pi}_T^s$ representing $\Pi^a$, a model $m \in \mathcal{M}^s$ describes a potential data generating process for the product $\Pi^a$. This can be formalized by introducing a data transformation $d_T^s$, which assigns, given the vector of products $\overrightarrow{\Pi}_T$, to each model $m \in \mathcal{M}^s$ the *induced* probability distribution $d_T^s(m) \in \mathcal{D}_T$ of $\overrightarrow{\Pi}_{m,T}^s = (\Pi_{m,1}^s, ..., \Pi_{m,T}^s)$, i.e.,

$$(\Pi_{m,1}^s, ...., \Pi_{m,T}^s) \sim d_T^s(m).$$

The set of data generating processes described by $\mathcal{M}^s$ is denoted by $d_T^s(\mathcal{M}^s)$.

In addition, we assume that $\Pi_{T+1}^s \in \mathcal{X}_{T+1}(\mathcal{M}^s)$ represents the actual product in the future period $T+1$. We introduce the target transformation $\tau_T^s$, which, given the product $\Pi_{T+1}^s$ and the risk measurement method $\rho_{T+1}^s$, assigns to each model $m \in \mathcal{M}^s$ the target of modeling, namely the risk measure $\tau_T^s(m) = \rho_{m,T+1}^s(\Pi_{m,T+1}^s) \in \mathbb{R} \cup \{\infty\}$.

The model class $\mathcal{M}^s$ is called *identified at the model side* in case the transformation $d_T^s$ is injective. If not, different models $m \in \mathcal{M}^s$, possibly with different target values $\tau_T^s(m)$, will describe the same data generating process, so that we might not be able to retrieve the target uniquely, at least not on the basis of the available data $X_1, ..., X_T$. The model class $\mathcal{M}^s$ is then called *underidentified at the model side*. In case of underidentification at the model side, the standard procedure is to normalize by choosing a subset $\mathcal{M}_0^s \subset \mathcal{M}^s$, such that $d_T^s$, when restricted to this subset, becomes injective. In case of identification at the model side, we have that $(d_T^s)^{-1}$, the inverse of $d_T^s$, is well defined on $d_T^s(\mathcal{M}^s)$, the set of data generating processes described by $\mathcal{M}^s$, and, thus, for each $\delta \in d_T^s(\mathcal{M}^s)$ we can find the corresponding model $m = (d_T^s)^{-1}(\delta)$, and target $\left(\tau_T^s \circ (d_T^s)^{-1}\right)(\delta) = \tau_T^s(m)$. In the sequel we shall assume that the selected model class $\mathcal{M}^s$ is already identified at the model side.

We have *exact identification at the data side*, in case $d_T^s$ is onto, i.e., $d_T^s(\mathcal{M}^s) = \mathcal{D}_T$, while we have *overidentification at the data side* in case $d_T^s$ is not onto, meaning that the set $d_T^s(\mathcal{M}^s)$ of data generating processes described by model class $\mathcal{M}^s$ is a strict subset of the set $\mathcal{D}_T$ of data generating processes. In the case of overidentification at the data side, one has to account for the possibility that the actual distribution of $(X_1, ..., X_T)$ is not described by the model class $\mathcal{M}^s$, i.e., $F_T^a \notin d_T^s(\mathcal{M}^s)$. The model class describing the log returns by a normal distribution is a typical example of a model class overidentified at the data side.

For the selection of a nominal model, we assume the presence of a data generating process (DGP) selection rule $\Delta_T^s = \Delta_T^s(X_1, ..., X_T)$ being an element of $d_T^s(\mathcal{M}^s)$, depending on the time series of $T$ observations $X_1, ..., X_T$. In case of overidentification at the data side, this selection rule should be model class $\mathcal{M}^s$-specific, in order to

guarantee that the selected DGP $\Delta_T^s$ belongs to $d_T^s(\mathcal{M}^s)$. Only in case of exact identification at the data side, so that $d_T^s(\mathcal{M}^s) = \mathcal{D}_T$, the DGP selection rule can be chosen model-independent, i.e., we could select some $\Delta_T \in \mathcal{D}_T$, and set $\Delta_T^s = \Delta_T$.

Given the DGP selection rule $\Delta_T^s$ combined with the assumption that $\mathcal{M}^s$ is identified at the model side, we can find the nominal selected model, to be denoted by $\widehat{m}_T^s$, and the corresponding target value: $\widehat{m}_T^s = (d_T^s)^{-1}(\Delta_T^s)$ and $\tau_T^s(\widehat{m}_T^s) = \rho^s_{\widehat{m}_T^s,T+1}(\Pi^s_{\widehat{m}_T^s,T+1})$.

The DGP selection rule is usually an econometric or statistical estimator having particular 'good' (model-dependent) characteristics, like unbiasedness or consistency. To describe this case, we assume the existence of a data characterization transformation $\chi_T^s : \mathcal{D}_T \mapsto \mathcal{E}_T^s$, with $\mathcal{E}_T^s$ the characterictics space, such that

- $\widetilde{\chi}_T^s : d_T^s(\mathcal{M}^s) \ni \delta_T \mapsto \chi_T^s(\delta_T) \in \mathcal{E}_T^s$ (i.e., $\chi_T^s$ restricted to $d_T^s(\mathcal{M}^s)$) is injective;
- our modelling approach allows the construction of a sequence $\widehat{e}_T^s = \widehat{e}_T^s(X_1, ..., X_T)$, $T = 1, 2, ...$, that estimates $e_T^{a,s} \equiv \chi_T^s(F_T^a) \in \mathcal{E}_T^s$ in a 'good' way.

The transformation $\chi_T^s$, first of all, characterizes what is *aimed* to be estimated in $\mathcal{D}_T$ in a 'good' way by the data $X_1, ..., X_T$, namely $\Delta_T^{a,s} = \left(\widetilde{\chi}_T^s\right)^{-1}\left(\chi_T^s\left(F_T^a\right)\right)$, which is equal to $F_T^a$ in case of a correctly specified model, i.e., in case $F_T^a \in d_T^s(\mathcal{M}^s)$. For this, we need that $\chi_T^s$ is defined on the whole of $\mathcal{D}_T$, and injective on $d_T^s(\mathcal{M}^s)$. Secondly, $\chi_T^s$ also determines what is *actually* estimated in $\mathcal{D}_T$ by the data, namely $\Delta_T^s = \left(\widetilde{\chi}_T^s\right)^{-1}(\widehat{e}_T)$. For this we need that $\chi_T^s$ restricted to $d_T^s(\mathcal{M}^s)$ is injective.

For instance, in case of a random sample over time, so that $F_T^a$ is the product of $T$ copies of $F^a \equiv F_1^a$, $\chi_T^s(\delta_T)$ might stand for the first two moments of the marginal distribution $\delta$ of $\delta_T = \delta^T \in \mathcal{D}_T$. If the model class generates lognormal distributions, then one could choose for $\widehat{e}_T^s$ the first two sample moments calculated on the basis of $(X_1, ..., X_T)$, so that the estimated model will be the one that generates the lognormal distribution having the first two moments as determined by the sample. This estimation procedure will consistently estimate the lognormal distribution having the same first two moments as $F^a$. In case $F^a$ is also lognormal, $F^a$ itself will be estimated consistently in this way.

## B. Risk measurement errors

In the modeling process, as described in the previous subsection, a number of 'errors' can be made. In this subsection we shall discuss and classify these errors. We start with a model class $\mathcal{M}^s$, identified at the model side, together with corresponding data transformation $d_T^s$, target transformation $t_T^s$, data characterization transformation $\chi_T^s$, and estimator $\widehat{e}_T^s$, with corresponding data selection rule $\Delta_T^s = (\widetilde{\chi}_T^s)^{-1}(\widehat{e}_T^s)$, and estimated model $\widehat{m}_T^s = (d_T^s)^{-1}(\widehat{e}_T^s)$, that estimates in a 'good' way $e_T^{a,s} = \chi_T^s(F_T^a)$, with corresponding DGP $\Delta_T^a = (\widetilde{\chi}_T^s)^{-1}\left(e_T^{a,s}\right)$ and model $m_T^{a,s} = (d_T^s)^{-1}(\Delta_T^{a,s})$

First, we aim to estimate in a 'good' way $m_T^{a,s}$, but we actually do estimate $\widehat{m}_T^s$. As a consequence, we estimate the target risk measures as $\tau_T^s(\widehat{m}_T^s)$ but not as the intended value $\tau_T^s(m_T^{a,s})$, i.e., the nominal (estimated) model is not exactly the same as the model we aim to estimate in a 'good' way. This error will be referred to as *estimation error*.

Secondly, we might choose a model class $\mathcal{M}^s$ that is overidentified at the data side, so that possibly $F_T^a$ does not belong to $d_T^s(\mathcal{M}^s)$. If so, we then postulate that, according to our model, the data is generated by $\Delta_T^{a,s} = d_T^s(m_T^{a,s})$ (estimated by $\Delta_T^s = d_T^s(\widehat{m}_T^s)$), while the actual data is generated by $F_T^a$, possibly unequal to $\Delta_T^{a,s}$. The postulated and actual data generating processes only share their characteristics as described by the transformation $\chi_T^s$, i.e., $\chi_T^s(\Delta_T^{a,s})) = \chi_T^s(F_T^a) = e_T^{a,s}$. As a consequence, the modeled risk measure $\tau_T^s(m_T^{a,s})$ might be misspecified.

We can test for misspecification using standard econometric misspecification test. On the basis of these tests we might conclude that there is potentially misspecification error. To quantify the potential *misspecification error*, we proceed by considering an extended model class $\mathcal{M}^S \supset \mathcal{M}^s$, assumed to be identified at the model side, with corresponding data transformation $d_T^S$, target transformation $t_T^S$, data characterization transformation $\chi_T^S$, and estimator $\widehat{e}_T^S$, that estimates $e_T^{a,S} = \chi_T^S(F_T^a)$ in a 'good' way. The corresponding *misspecification error* is given by the difference between the originally modeled target $\tau_T^s(m_T^{a,s})$, and the target according to the extended model $\tau_T^S(m_T^{a,S})$ (with $m_T^{a,S} = \left(\chi_T^S \circ d_T^S\right)^{-1}(e_T^{a,S})$). If the extended model class is still overidentified at the data side, there remains potential misspecification, which might be discovered by

applying additional misspecification tests. Depending on the outcome of such tests we might extend the model class even further.

Thirdly, we might choose, even in case of exact identification at the data side, the model class $\mathcal{M}^s$, or its extension $\mathcal{M}^S$, incorrectly. Indeed, different (extended) model classes $\mathcal{M}^S$ might describe the same (subset of the) set of data generating processes $\mathcal{D}_T$. If we consider model classes $\mathcal{M}^i$, $i = A, B, ...$ (all identified at the model side, and with corresponding transformations $d_T^i$, $\chi_T^i$, $\tau_T^i$, and estimators $\widehat{e}_T^i$, estimating $e_T^{a,i} = \chi_T^i(F_T^a)$ in a 'good' way, with $d_T^i(\mathcal{M}^i) = d_T^S(\mathcal{M}^S)$, $i = A, B, ...$), as viable alternatives to the (possibly extended) model class $\mathcal{M}^S$, then, in general, $\tau_T^S(m_T^{a,S})$ and $\tau_T^i(m_T^{a,i})$ (with $m_T^{a,i} = \left(\chi_T^i \circ d_T^i\right)^{-1}(e_T^{a,i})$) will be different. Since this difference is due to lack of identification at the model side, i.e., the set of data generating processes $\mathcal{D}_T$ does not allow us to distinguish between $\mathcal{M}^S$ and $\mathcal{M}^i$, $i \in \{A, B, ...\}$, we refer to this difference as *identification error*.

## C. Model risk and its decomposition

In this section we apply the general model risk approach described in section III to the special case where the risk measurement is empirically based. We take for the model class $\mathcal{M}$ used in section III the total model class of the empirically based approach as presented in the previous subsection. To be precise, assuming that the model classes $\mathcal{M}^S$ and $\mathcal{M}^i$, $i = A, B, ...$ do not have models in common, we define the total model class $\overline{\mathcal{M}}$ corresponding to the risk measurement method described in the previous subsection as $\overline{\mathcal{M}} = \mathcal{M}^S \cup \mathcal{M}^A \cup \mathcal{M}^B \cup ...$, thus, we take $\mathcal{M} = \overline{\mathcal{M}}$.

First, we consider the case when we are dealing with the estimation error only. The nominal model employed is $\widehat{m}_T^s$, which is obtained from the 'good' estimator $\widehat{e}_T^s$ through $\widehat{m}_T^s = (\widetilde{\chi}_T^s \circ d_T^s)^{-1}(\widehat{e}_T^s)$. Assuming that we can determine the estimation accuracy of $\widehat{e}_T^s$, we can construct a $(1-\alpha)$-confidence region $\mathcal{E}_T^{s,c}$ in $\mathcal{E}_T^s$ around $\widehat{e}_T^s$. Then, to quantify

the estimation risk, we take for $\mathcal{K}$ the set [13]

$$\mathcal{K}_e = \{m \in \overline{\mathcal{M}} \mid m \in \mathcal{M}^s, (\chi_T^s \circ d_T^s)(m) \in \mathcal{E}_T^{s,c}\}.$$

To deal with the misspecification error, we assume that we can determine the estimation accuracy of $\widehat{e}_T^S$, allowing us to construct a $(1-\beta)$-confidence region $\mathcal{E}_T^{S,c}$ in $\mathcal{E}_T^S$ around $\widehat{e}_T^S$. Then we take for $\mathcal{K}$ the set

$$\mathcal{K}_{e,m} = \{m \in \overline{\mathcal{M}} \mid m \in \mathcal{M}^S, (\chi_T^S \circ d_T^S)(m) \in \mathcal{E}_T^{S,c}\}.$$

Here, it makes sense to choose the confidence region $\mathcal{E}_T^{S,c}$ such that it 'includes' $\mathcal{E}_T^{s,c}$, i.e., such that $\mathcal{E}_T^{S,c} \supset (\chi_T^S \circ (\widetilde{\chi}_T^s)^{-1})(\mathcal{E}_T^{s,c})$. This guarantees that $\mathcal{K}_e \subset \mathcal{K}_{e,m}$, so that $\mathcal{K}_{e,m}$ incorporates both the estimation and misspecification error. To achieve this, we propose to use a family $\{\mathcal{E}_T^{S,c}(\gamma)\}$ of tolerance sets parameterized by the confidence level $\gamma$. For a given confidence level $\alpha$ and a given tolerance set $\mathcal{E}_T^{s,c}$, we then take $\mathcal{E}_T^{S,c} = \mathcal{E}_T^{S,c}(\beta)$ where $\beta$ is defined by[14]

$$\beta = \min(\alpha, \sup\left\{\gamma \in (0,1) : (\chi_T^S \circ (\widetilde{\chi}_T^s)^{-1})(\mathcal{E}_T^{s,c}) \subset \mathcal{E}_T^{S,c}(\gamma)\right\}). \tag{13}$$

To quantify the identification error, we assume that we can also determine the estimation accuracy of $\widehat{e}_T^i$, $i \in \{A, B, ...\}$, so that we can construct $(1-\gamma_i)$-confidence regions $\mathcal{E}_T^{i,c}$ in $\mathcal{E}_T^i$ around $\widehat{e}_T^i$. We take for $\mathcal{K}$ the set

$$\mathcal{K}_{e,m,i} = \{m \in \overline{\mathcal{M}} \mid m \in \mathcal{M}^i, (\chi_T^i \circ d_T^i)(m) \in \mathcal{E}_T^{i,c}, i \in \{S, A, B, ...\}\}.$$

Notice that by construction $\mathcal{K}_{e,m} \subset \mathcal{K}_{e,m,i}$, in case $\gamma_s = \beta$, guaranteeing that $\mathcal{K}_{e,m,i}$ incorporates estimation, misspecification, *and* identification error.

[13] Notice that $\widetilde{\chi}_T^s \circ d_T^s = \chi_T^s \circ d_T^s$.

[14] 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 $\mathcal{K}_{e,m}$ by the convex hull of $\mathcal{K}_e$ and the original $\mathcal{K}_{e,m}$. However, our proposal seems to be more transparent.

Alternatively, instead of constructing a confidence region in $\mathcal{E}_T^i$ and then using the transformation $(\chi_T^i \circ d_T^i)^{-1} : \mathcal{E}_T^i \mapsto \mathcal{M}^i$ to find a set of relevant models $\mathcal{K} \in \mathcal{M}^i$ in order to calculate $\sup_{k \in \mathcal{K}} \rho_{k,T+1}(\Pi_{k,T+1})$ for $i \in \{s, S, A, B, ...\}$, it might also be possible to construct a $(1-\delta_i)$-confidence interval directly around $\tau_T^i(\widehat{m}_T^i) = \rho_{\widehat{m}_T^i, T+1}(\Pi_{\widehat{m}_T^i, T+1}) \in \mathbb{R} \cup \{\infty\}$ and use the supremum over this interval (i.e., the upper boundary of this confidence interval in case the risk is positive) to quantify the corresponding components of model risk. Suppose that we can quantify the estimation inaccuracy by means of the limit distribution of $\tau_T^i(\widehat{m}_T^i)$ given by

$$\sqrt{T}\left(\tau_T^i(\widehat{m}_T^i) - \tau_T^i(m_T^{a,i})\right) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\tau^i}).$$

Then we can estimate the supremum of the $(1-\delta_i)$- confidence interval by

$$\tau_T^i(\widehat{m}_T^i) + z_{\delta_i/2}\widehat{\Sigma}_{\tau^i},$$

with $z_{\delta_i/2}$ denoting the $\delta_i/2$-quantile of the standard normal distribution, and $\widehat{\Sigma}_{\tau^i}$ a consistent estimate of $\Sigma_{\tau^i}$. The corresponding model risk component is then given by

$$\left(\tau_T^i(\widehat{m}_T^i) + z_{\delta_i/2}\widehat{\Sigma}_{\tau^i}\right) - \tau_T^s(\widehat{m}_T^s).$$

For $i = s$ this yields the estimation risk, for $i = S$ also the misspecification risk (assuming that we choose the confidence level large enough, cf. (13)), and for the supremum over $i \in \{S, A, B, ...\}$ we get the total model risk, including also the identification risk.

## D. Example: The i.i.d.-case

As illustration, we consider the case where the observations on the actual product $\Pi_a$ form an i.i.d.-sample from the distribution $F^a \equiv F_1^a$, i.e.,

$$X_1, ..., X_T \overset{i.i.d.}{\sim} F^a.$$

We assume that $F^a \in \mathcal{D} \equiv \mathcal{D}_1 \subset D(\mathbb{R})$, where $D(\mathbb{R})$ is the set of all non-decreasing right continuous functions $z : \mathbb{R} \to \mathbb{R}$ such that $z(-\infty) = 0$, $z(\infty) = 1$. The choice of $\mathcal{D}$ might depend on particular required characteristics. For instance, in case one would like to exploit the first and second moment of $X \equiv X_1$, then it makes sense to restrict the set $D(\mathbb{R})$ to the set $\mathcal{D}$ consisting of distribution functions with bounded first and second moment. We have $(X_1, ..., X_T) \sim F_T^a = (F^a)^T$, i.e., $F_T^a$ is just the product of $T$ independent copies of $F^a$.
The model class $\mathcal{M}^s$ is assumed to be chosen such that

$$\Pi_{m,1}^s, ...., \Pi_{m,T}^s \overset{i.i.d.}{\sim} F_\theta \in \mathcal{D} \subset D(\mathbb{R})$$

with $\theta = \theta_m \in \Theta \subset \mathbb{R}^k$. Then $d_T^s(m) = (F_\theta)^T$ and $\mathcal{D}_T \subset D(\mathbb{R}^T)$ consists of the distribution functions which are the product of $T$ independent copies $F$, such that the marginals satisfy $F \in \mathcal{D}$.

We consider the standard econometric case where the parameter of interest $\theta$ will be estimated by means of the method of moments. For this purpose we assume the existence of a function $\psi : \mathbb{R} \times \Theta \mapsto \mathbb{R}^k$ such that $\mathbb{E}_{\theta_a}(\psi(X, \theta_b))$ is equal to zero if $\theta_a = \theta_b$, and nonzero otherwise, where $\mathbb{E}_\theta(\cdot)$ denotes taking expectations with respect to the distribution $F_\theta$. In addition, we assume that for any $F \in \mathcal{D}$ there exists a unique $\theta \in \Theta$ such that $\mathbb{E}_F(\psi(X, \theta)) = 0$ (with $\mathbb{E}_F(\cdot)$ denoting taking expectations with respect to $F$); we denote this unique $\theta$ by $\theta(F)$. Obviously, if $F = F_\theta$ then we have $\theta(F_\theta) = \theta$. The transformation $\chi_T^s$ is now given by

$$\chi_T^s : d = (F)^T \in \mathcal{D}_T \mapsto \theta = \theta(F) \in \Theta = \mathcal{E}_T^s.$$

The parameter $\theta$ can be estimated by solving the sample analogue of the moment condition $\mathbb{E}_{F^a}(\psi(X, \theta)) = 0$. This yields the estimator $\widehat{e}_T^s = \widehat{\theta}_T$, which will be consistent for $e_T^{a,s} = \theta(F^a)$, and for which the estimation inaccuracy can be quantified under appropriate regularity conditions (see, for instance, Van der Vaart, 1998). So, it is straightforward to construct a $(1 - \alpha)$-confidence region $\mathcal{E}_T^{s,c}$ and corresponding set $\mathcal{K}_e$,

allowing us to quantify the estimation risk.

For instance, in case of the VaR-example of Section II, see equations (1)-(2), we can construct a confidence region around $(\widehat{\mu}_Y, \widehat{\sigma}_Y)$, resulting in $\mathcal{E}_c^s$, and then calculate

$$\sup_{k\in\mathcal{K}_e} \rho_{k,T+1}(\Pi_{k,T+1}) = \sup_{(\mu,\sigma)\in\mathcal{E}_c^s} X_T\left(1-\exp\left(z_p\sigma+\mu\right)\right).$$

Alternatively, we can also easily derive the limit distribution of

$$\tau_T^s(\widehat{m}_T^s) = \rho^s_{\widehat{m}_T^s,T+1}(\Pi^s_{\widehat{m}_T^s,T+1}) = X_T\left(1-\exp\left(z_p\hat{\sigma}_Y+\hat{\mu}_Y\right)\right),$$

and then estimate the supremum of the $(1-\delta_s)$-confidence interval.

To deal with the misspecification risk, we extend $\mathcal{M}^s$ to the model class $\mathcal{M}^S$ such that

$$\Pi^s_{m,1},....,\Pi^s_{m,T} \overset{i.i.d.}{\sim} F \in \mathcal{D} \subset D(\mathbb{R})$$

with $F$ left unspecified. For $\chi^S$ we can take the identity, since we can estimate $F$ consistently by means of the empirical distribution function based on the sample $X_1,...,X_T$. Denote this nonparametric estimate by $\widehat{F}_T$, so that $\widehat{e}_T^S = \widehat{F}_T$. Notice that it yields a consistent estimate for $F^a$. A confidence region of level $\beta$ can be constructed in the following way:

$$\mathcal{E}_T^{S,c} := \left\{ F\left(x\right) \in \left[\hat{F}_T\left(x\right) \pm \frac{k_{\beta/2}}{\sqrt{n}}\right] \quad \forall x \in \mathbb{R}\right\},$$

where $k_{\beta/2}$ is the critical value of the Kolmogorov-Smirnov statistic.[15] This yields in a straightforward way the set $\mathcal{K}_{e,m}$. Alternatively, instead of constructing a confidence region in $\mathcal{E}_T^S$, it is straightforward to construct a confidence interval directly around $\tau_T^S(\widehat{m}_T^S) = \rho^S_{\widehat{m}_T^S,T+1}(\Pi^S_{\widehat{m}_T^S,T+1}) \in \mathbb{R}\cup\{\infty\}$ and use the upper boundary of this confidence interval to quantify the misspecification risk.

In case of the i.i.d.-assumption it makes sense to postulate that this assumption can be extended to the future $(T+1)$-st period as well. However, this is an assumption,

[15] 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)).

which cannot be inferred from the past data up to time $T$. Indeed, it is an identification assumption, possibly leading to an identification error. As illustrated in Section II, this error can be substantial, even be the dominating factor in model risk, depending upon the alternative model classes $\mathcal{M}$, $i \in \{A, B, ...\}$, that one considers to be acceptable alternatives of $\mathcal{M}^s$ and $\mathcal{M}^S$.

Finally, the i.i.d.-assumption might be too strong. In this case, the model class $\mathcal{M}^S$ might be overidentified, and, thus, might be misspecified. In such a case, it could be imbedded in an even larger model class, allowing for non-i.i.d. behavior. The current analysis can straightforwardly be extended to this case. For instance, the model classes $\mathcal{M}^i$, $i \in \{A, B, ...\}$ then describe alternative non-i.i.d. behaviors.

# V. 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 of the previous sections can be used in portfolio risk management, using two financial time series, the Standard and Poor's 500 and the £/ $ exchange rate.

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 how the model risk measurement approach can be taken into account for the (simple) value-at-risk and expected shortfall methodology, as already discussed in subsection D of the previous section and illustated in section II. In addition to the Gaussian model class, used in the illustration in section II, we also investigate a GARCH$(1, 1)$ model class with Gaussian innovations that should be more capable of capturing time varying risk.[16] First, we present a brief description of the data, then we investigate the performance of the model risk approach, and, finally, we present the results in terms of multiplication

[16]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 et al. (2003).

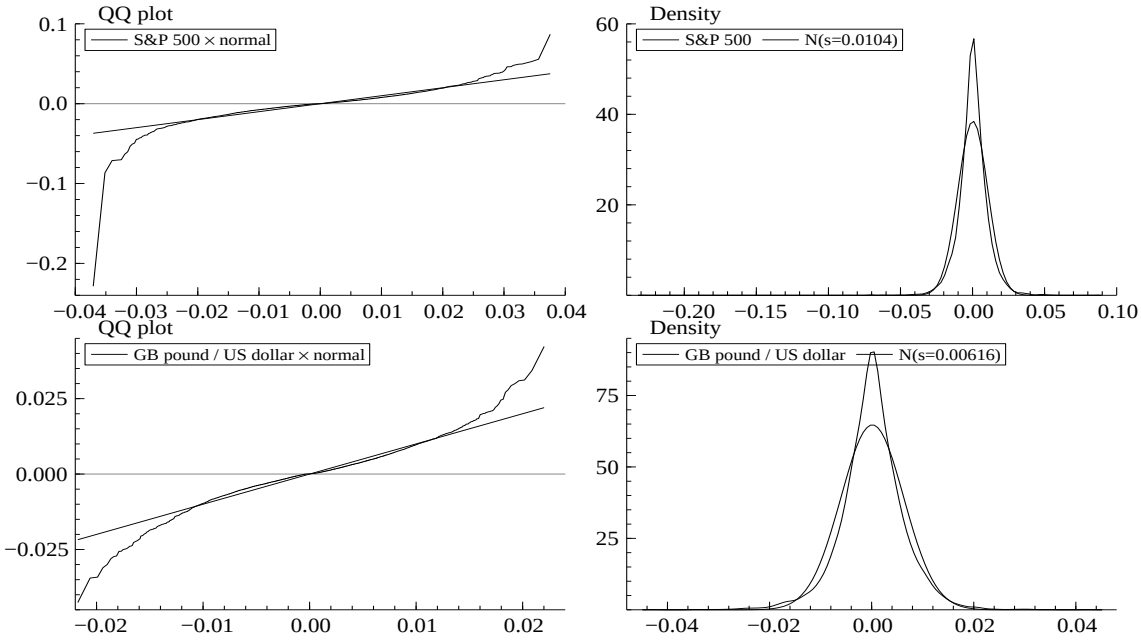


**Figure 1. Data descriptions**

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.

factors. The derivation of the relevant formulas and asymptotic distributions used are presented in Appendix A.

## A. Data Description

We start with investigating the data. The data is obtained from Thomson Datastream (definitions and sources of the data can be found in Appendix B). The data runs from 26-10-'81−29-04-'03 for the S&P 500 market and from 03-01-'86−29-04-'03 for the £/$ exchange rate. Figure 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[17] estimate of the densities of the

[17]We use a normal kernel. 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

S&P 500 and the £/$ exchange rate. So, we expect some misspecification error when calculating the value-at-risk and expected shortfall on the basis of a nominal model assuming normally distributed log-returns.

In the subsequent analysis we shall use a rolling window approach with a window of two years in order to investigate our model risk measurement procedure. Using this rolling window approach we can estimate on the basis of each of these two years of daily data the models under consideration, and estimate the next day's (future) risk measure. This yields a time series of estimated risk measures of length equal to the length of the available time series, except the first two years, that will be used in the statistical investigation.

Figure 2 shows the resulting estimates for the (annualized) mean and volatility of the estimated Gaussian and GARCH$(1,1)$ models. This graph shows that both model classes result in more or less the same predictions of the means. The difference between the model classes is in the predictions of the volatilities. The predictions of the estimated GARCH$(1,1)$ models are, as could be expected, much more erratic than those of the Gaussian models.

## B. Model risk performance

Next, we investigate the performance of our model risk approach. We start by investigating the application to the VaR risk measurement method. Using the rolling window approach, we obtain VaR-estimates for the whole sample range (except the first two years). We do this for the Gaussian model class, the GARCH$(1,1)$-model class, and we also estimate the nonparametric VaRs. 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)). As comparison, we also include the 2.5% level. Denote by $T$ the number of days in the backtesting period, i.e., the sample period (except the first two years), by $f$ the number

data.

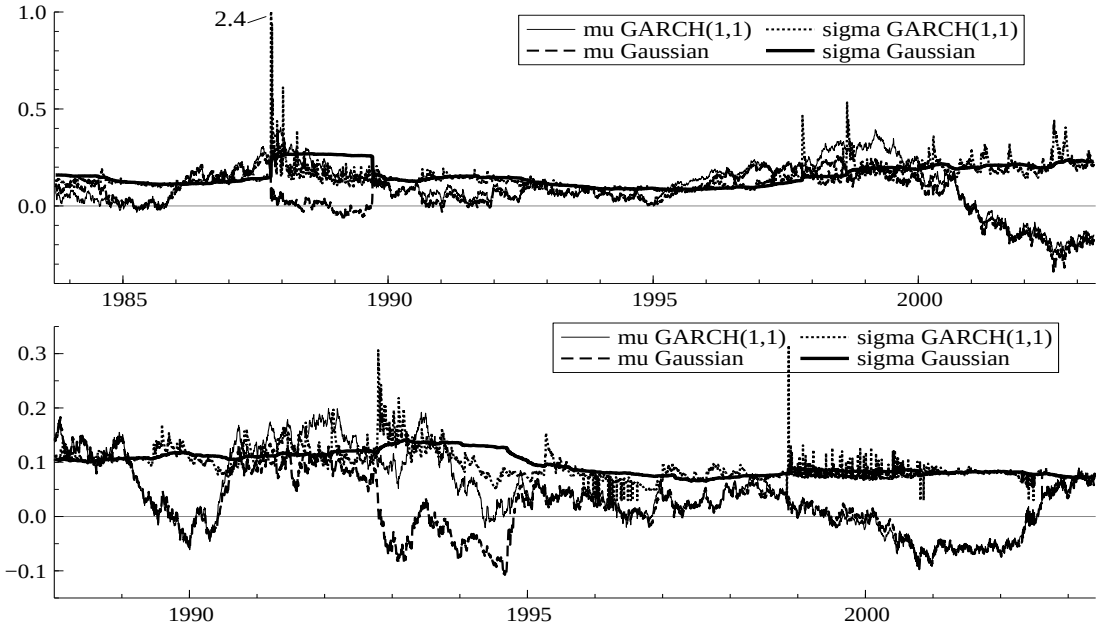


**Figure 2. Parameter estimates**

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 for the £/$ exchange rate.

**Table II**
**FOEL test VaR for S&P 500**

FOEL test for VaR-market (Mkt) risk according to Gaussian, GARCH(1, 1), and non-parametric models and corresponding estimation risk components, yielding estimation and misspecification risk in case of nonparametric model (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 | $F$ | $p$-value | VaR level hypothesis rejected |
|---|---|---|---|---|---|---|
| Gauss. Mkt risk | 2.5% | 2.9% | (2.5%; −) | 1.8 | 0.03 | yes |
| GARCH(1, 1) Mkt risk | 2.5% | 3.1% | (2.7%; −) | 2.8 | 0.00 | yes |
| Nonpar Mkt risk | 2.5% | 2.8% | (2.3%; −) | 1.2 | 0.12 | no |
| Gauss. Est. risk | 2.5% | 2.3% | (1.9%; −) | −0.9 | 0.81 | no |
| GARCH(1, 1) Est. risk | 2.5% | 2.7% | (2.7%; −) | 2.7 | 0.00 | yes |
| Est. & Misspec. risk | 2.5% | 1.9% | (1.5%; −) | −3.5 | 1.00 | no |
| Gauss. Mkt risk | 1% | 1.8% | (1.5%; −) | 4.7 | 0.00 | yes |
| GARCH(1, 1) Mkt risk | 1% | 1.9% | (1.5%; −) | 4.9 | 0.00 | yes |
| Nonpar Mkt risk | 1% | 1.4% | (1.1%; −) | 2.7 | 0.00 | yes |
| Gauss. Est. risk | 1% | 1.4% | (1.1%; −) | 2.8 | 0.00 | yes |
| GARCH(1, 1) Est. risk | 1% | 1.9% | (1.5%; −) | 4.8 | 0.00 | yes |
| Est. & Misspec. risk | 1% | 1.0% | (0.6%; −) | −1.3 | 0.90 | no |

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

$$F = \sqrt{T}\frac{f/T - p}{p(1-p)} \tag{14}$$

In Tables II and III we present the results of a one-sided FOEL test with 95% confidence intervals. The results for the two overidentified model classes, the Gaussian and GARCH(1, 1) model class, indicate that when predicting the VaR-market risk using these model classes leads to an incorrect quantification of the market risk, both in case of the S&P 500 data and in case of the £/$ data. On the other hand, the nonparametric market risk measurement seems to work well in case of the £/$ data, but not in case of the S&P 500 data. Thus, even using an estimated nonparametric nominal model (under

**Table III**
**FOEL test VaR for £/ $ FX rate**

FOEL test for VaR-market (Mkt) risk according to Gaussian, GARCH(1, 1), and nonparametric models and corresponding estimation risk components, yielding estimation and misspecification risk in case of nonparametric model (for definitions, see main text). Daily data on £/ $ from 03-01-'86 to 29-04-'03.

| model | VaR level | FOEL | 1-sided 95% CI | $F$ | $p$-value | VaR level hypothesis rejected |
|---|---|---|---|---|---|---|
| Gauss. Mkt risk | 2.5% | 3.4% | (3.0%; −) | 3.5 | 0.00 | yes |
| GARCH(1, 1) Mkt risk | 2.5% | 3.3% | (2.8%; −) | 3.1 | 0.00 | yes |
| Nonpar Mkt risk | 2.5% | 2.6% | (2.1%; −) | 0.4 | 0.36 | no |
| Gauss. Est. risk | 2.5% | 2.9% | (2.4%; −) | 1.5 | 0.07 | no |
| GARCH(1, 1) Est. risk | 2.5% | 3.2% | (2.7%; −) | 2.8 | 0.00 | yes |
| Misspec. risk | 2.5% | 1.6% | (1.2%; −) | −4.9 | 1.00 | no |
| Gauss. Mkt risk | 1% | 1.9% | (1.5%; −) | 4.4 | 0.00 | yes |
| GARCH(1, 1) Mkt. risk | 1% | 2.0% | (1.6%; −) | 4.9 | 0.00 | yes |
| Nonpar Mkt risk | 1% | 1.1% | (0.8%; −) | 0.8 | 0.21 | no |
| Gauss. Est. risk | 1% | 1.4% | (1.0%; −) | 2.1 | 0.02 | no |
| GARCH(1, 1) Est. risk | 1% | 2.0% | (1.6%; −) | 4.9 | 0.00 | yes |
| Misspec. risk | 1% | 0.6% | (0.4%; −) | −3.6 | 1.00 | no |

the *i.i.d.* assumption may not be appropriate!

However, these results are without taking into account estimation risk. The tables also present the results when estimation risk is included. Including estimation risk in case of the nonparametric model class of course means combining estimation and misspecification risk. For the GARCH$(1,1)$-model class the inclusion of estimation risk is not sufficient to prevent the VaR limit being exceeded too often. For the Gaussian model class taking estimation risk into account seems sufficient for both the S&P 500 and the £/$ exchange rate, but only at the 2.5% level. Thus, only taking account of estimation risk (at the usual 95% confidence level) in these overidentified model classes generally does not seem to suffice for an appropriate quantification of future market risk, particularly not at the 1% level benchmark. However, if we take both estimation and misspecification risk into account, the number of times the VaR limit is crossed does not exceed the level predicted in a statistically significant way. This means that including misspecification risk is needed but also sufficient, when this misspecification risk is quantified at the usual 95% confidence level. There does not seem to be a need for a higher confidence level or for also taking into account additional misspecification risk or identification risk. Formulated alternatively, in extending the overidentified model class $\mathcal{M}^s$ to $\mathcal{M}^S$, we seem to have chosen an appropriate extension, and in choosing $\mathcal{E}_T^{S,c}$ we seem to have selected an appropriate confidence level.

In order to test the performance of the model class for predicting expected shortfall, we use the recently proposed backtesting test for expected shortfall in Kerkhof and Melenberg (2004). This test consists of comparing some risk measure (in our case ES) estimated on the basis of past data, with the same risk measure, but then estimated on the basis of the data for which it is aimed to be an estimate, also taking into account estimation risk. For ES we adopt a higher level than for VaR, namely, the 2.5% and 5% levels, following arguments given in Kerkhof and Melenberg (2004), who motivate that for an appropriate comparison between VaR and ES, the latter should have a higher level. For a detailed description of the test, which also includes a transformation of the return series using a probability integral transform to a standardized return series,

**Table IV**
**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 upper panel presents the results of the S&P 500 and the bottom panel presents the results of the £/ $ FX rate. The column $F$ presents the values of the test statistic presented in Kerkhof and Melenberg (2004). 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 | $F$ | $p$-value | ES model rejected |
|---|---|---|---|---|---|
| S&P 500 | Gauss. | 5% | $-11.3$ | 0.00 | yes |
| | GARCH$(1,1)$ | 5% | $-12.0$ | 0.00 | yes |
| | Nonpar | 5% | 0.5 | 0.69 | no |
| | Gauss. | 2.5% | $-19.5$ | 0.00 | yes |
| | GARCH$(1,1)$ | 2.5% | $-18.1$ | 0.00 | yes |
| | Nonpar | 2.5% | $-0.6$ | 0.28 | no |
| £/ $ | Gauss. | 5% | $-8.4$ | 0.00 | yes |
| | GARCH$(1,1)$ | 5% | $-8.9$ | 0.00 | yes |
| | Nonpar | 5% | 1.0 | 0.84 | no |
| | Gauss. | 2.5% | $-10.4$ | 0.00 | yes |
| | GARCH$(1,1)$ | 2.5% | $-11.5$ | 0.00 | yes |
| | Nonpar | 2.5% | 1.5 | 0.94 | no |

we refer to Kerkhof and Melenberg (2004). As reported by these authors, the rolling window approach employed here can be approximated by their setting.

The results presented in table IV indicate that both the rolling window Gaussian model class and the GARCH$(1,1)$ model class are strongly rejected, while the rolling-window nonparametric model class cannot be rejected for both series. Thus, similar to the VaR-case, the conclusion seems to be that both estimation and misspecification risk at the 95% confidence level are needed, but also sufficient, for an appropriate quantification of expected shortfall. Taking into account additional misspecification risk or identification risk as well does not seem to be required.

## C. Multiplication factors

One way to investigate the relation between the combination of estimation and misspecification risk 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 VaR (ES) upper bound of a 95% confidence interval and the nominal parametric VaR (ES). Plots of the multiplication factors for the Gaussian and GARCH$(1,1)$ model classes are shown in Figure 3. Notice that the difference between the two model classes occurs due to the difference in the nominal models. Our definition of the 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. [18]

We see in the upper panels of Figure 3 that for the Gaussian model class in case of 1% VaR multiplication factors of 2.3 for the S&P 500 and 1.6 for the £/$ exchange rate comfortably cover the estimation and misspecification risks 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 class. Notice that 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 3 we see that for the GARCH$(1,1)$ model class the model risk multiplication factors are much higher than for the Gaussian model class. This can be explained by the fact that the GARCH$(1,1)$ model class responds more quickly to periods of low volatility and then forecasts low values of VaR and ES, contrary to the non-parametric estimates.

In Figures 4 and 5 we give the capital requirements based on the BIS capital re-

[18] 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. 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% confidence interval of a non-parametric VaR (ES) *plus* a penalty, where the penalty is increasing with the degree of rejection (higher penalties for lower $p$-values).

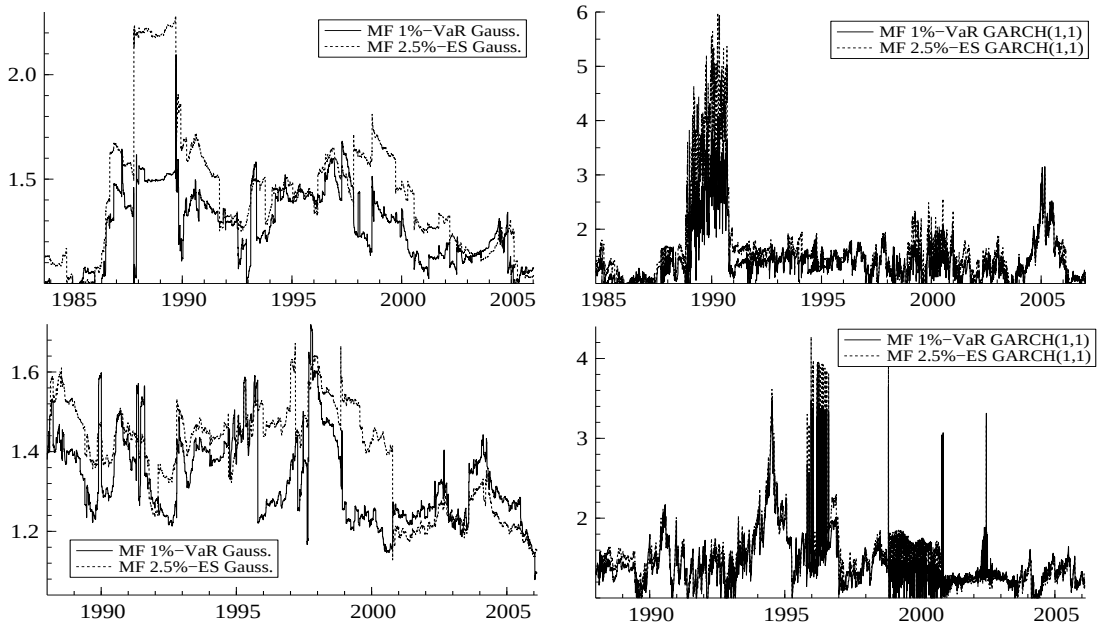


**Figure 3. Model risk multiplication factors**

The upper panels display estimation and misspecification 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 estimation and misspecification 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$(1,1)$).

quirements. 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.[19] The capital requirement can then be determined by multiplying the daily value-at-risks by the multiplication factor and $\sqrt{10}$.[20] 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 3. In Figures 4 and 5 we see that the capital requirements for the GARCH$(1,1)$ model class are much more variable than those of the Gaussian model class . 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 V. In Table V 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 estimation and misspecification 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.

## VI. 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 modelled. 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

[19] 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.

[20] 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 class assumptions this would be correct.

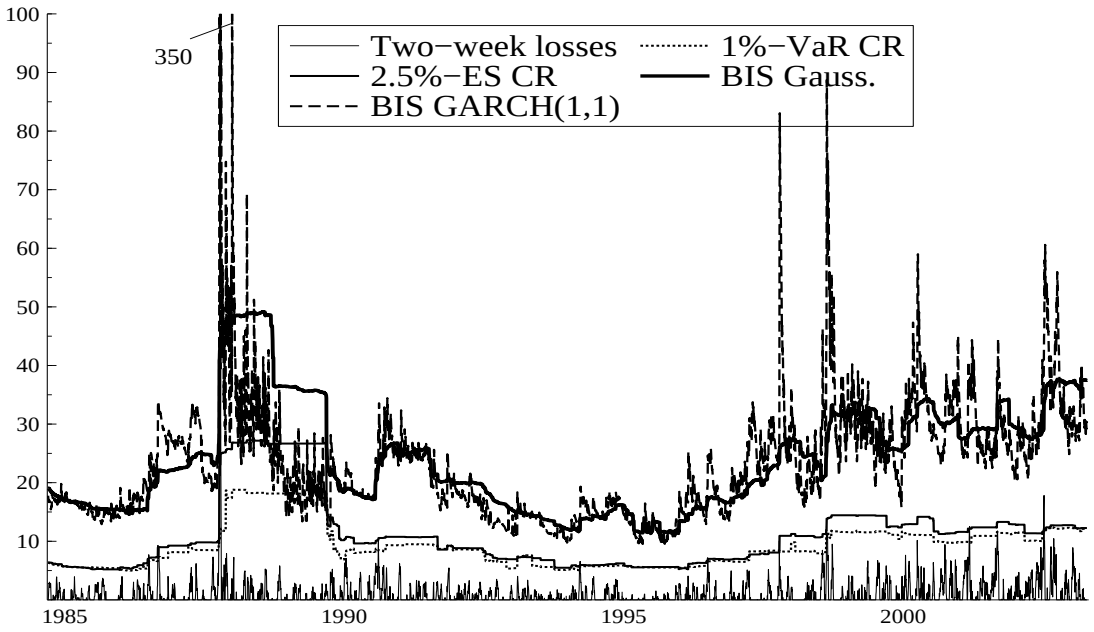


**Figure 4. Capital requirements S&P 500**

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.

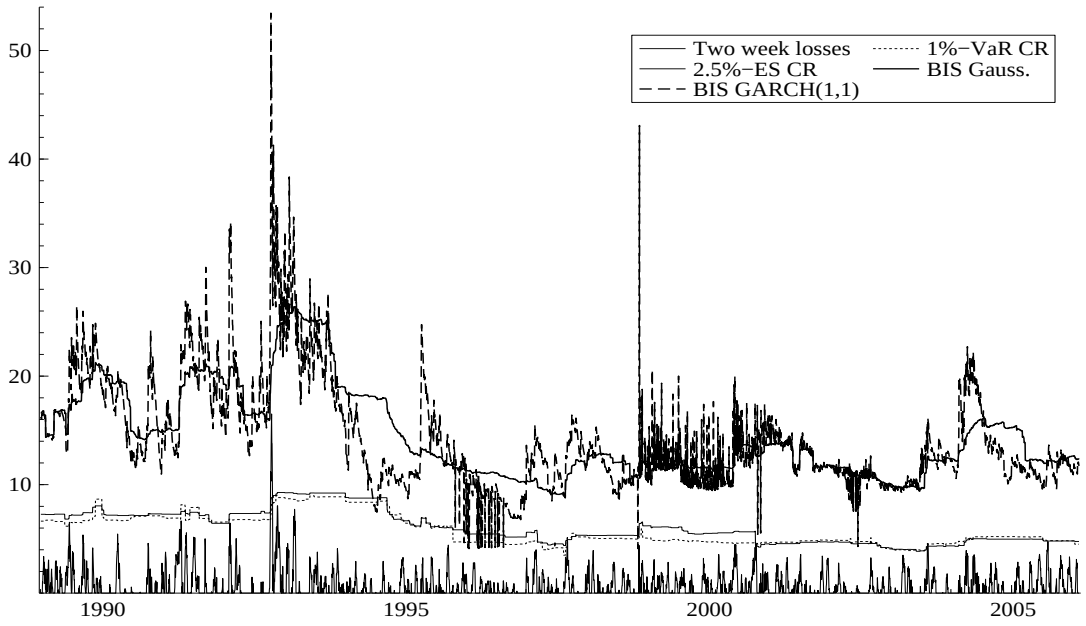


**Figure 5. Capital requirements £/ $ FX rate**

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 V**
**Capital requirement schemes**

This table reports the 1-day average exceedance rate, two-week average exceedance rate, and the average capital requirements (CR). The CR schemes investigated are the BIS CR for the Gaussian model, the BIS CR for the GARCH$(1,1)$ model, the VaR estimation and misspecification risk multiplication (MRMF) factor based CR, and the ES estimation and misspecification 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. | 0.04 | 0.04 | 23.9 |
| | BIS GARCH | 0.09 | 0.00 | 22.0 |
| | MRMF VaR | 2.25 | 1.48 | 9.1 |
| | MRMF ES | 1.30 | 0.94 | 10.7 |
| £/ $ | BIS Gauss. | 0.00 | 0.00 | 14.5 |
| | BIS GARCH | 0.00 | 0.00 | 14.1 |
| | MRMF VaR | 1.51 | 0.83 | 5.8 |
| | MRMF ES | 1.28 | 0.72 | 6.1 |

of a risk measure for one particular model. This allows for a quantification of model risk on top of market risk measurement.

We focus in particular on an empirically based approach to quantify model risk. The 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 more recently proposed coherent risk measures in a natural way.

We decompose the total model risk into three components: estimation risk, misspecification risk, and identification risk. This is established using tolerance sets that are subsets of various model classes, where the model classes depend on the type of model risk to be quantified: for estimation risk we use the *selected*, possibly overidentified, model class, for misspecification risk we use an *extended*, exactly identified model class, while for identification risk we use *alternative* model classes. This approach allows a

division of capital requirements currently used (for example, the multiplication factor of the BIS) in market risk, model risk (estimation, misspecification, and identification risk), and residual risks.

Backtesting might be used to decide whether only estimation risk has to be included, or whether, in addition, also misspecification and possibly even identification risk has to be taken into account. Our empirical results suggest that, for commonly used overidentified model classes, namely a Gaussian and a GARCH$(1,1)$ model class, with a natural extensions to an exactly identified model class, misspecification risk cannot be ignored, while identification risk does not seem to play an additional crucial role. Moreover, 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 capital requirements 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 (other than model 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 modelled 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 modelled and the determination of the size of model risk reserves for different models is left for future empirical research.

# A. Risk measure derivations

## A. Computation of ES

Let $X_T$ denote the current position. Let $m \in \mathcal{M}^s$ be the selected model such that

$$Y_{T+1} = \log\left(\Pi^s_{m,T+1}/X_T\right) \mid \mathcal{F}_T \sim \mathcal{N}(\mu_Y, \sigma_Y^2).$$

Then by definition

$$\begin{aligned}
\mathrm{ES}^s_{m,T}\left(\Pi^s_{m,T+1} - X_T\right) &= -\mathbb{E}_{\mathbb{P}_{m,t}}\left(\Pi^s_{m,T+1} - X_T \mid \Pi^s_{m,T+1} - X_T \leq \mathrm{VaR}^s_{m,T}\left(\Pi^s_{m,T+1} - X_T\right)\right) \\
&= X_T\left(1 - \mathbb{E}_{\mathbb{P}_{m,t}}\left(\exp\left(Y_{T+1}\right) \mid Y_{T+1} \leq \mathrm{VaR}^s_{m,T}\left(Y_{T+1}\right)\right)\right).
\end{aligned}$$

So, writing $\mu = \mu_Y$ and $\sigma = \sigma_Y$,

$$\begin{aligned}
\mathrm{ES}^s_{m,T}\left(\Pi^s_{m,T+1} - X_T\right) &= X_T\left(1 - \frac{1}{p}\int_{-\infty}^{v_p} \exp(x)\,\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)dx\right) \\
&= X_T\left(1 - \frac{1}{p}\exp\left(\mu + \frac{1}{2}\sigma^2\right)\int_{-\infty}^{v_p}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma^2}\left(x - \left(\sigma^2 + \mu\right)\right)^2\right)dx\right) \\
&= X_T\left(1 - \frac{1}{p}\exp\left(\mu + \frac{1}{2}\sigma^2\right)\int_{-\infty}^{\frac{v_p - \mu - \sigma^2}{\sigma}}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y^2\right)dy\right) \\
&= X_T\left(1 - \frac{1}{p}\exp\left(\mu + \frac{1}{2}\sigma^2\right)\Phi\left(z_p - \sigma\right)\right) \qquad (15)
\end{aligned}$$

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

## B. Limit Distributions VaR

We have available data sets of log returns $\left(Y_1^f, ..., Y_T^f\right)$ of the portfolio under consideration each for a period of length $Tf$ years ($Tf = 2$, $f = 1/252$ (one day)). An elementary VaR model assumes that the data is a realization of a random sample where $Y_t^f \sim \mathcal{N}\left(\mu_Y f, \sigma_Y^2 f\right)$ for $t = 1, .., T$ where $\mu_Y$ and $\sigma_Y^2$ denote annualized mean and variance, respectively. We estimate $\mu_Y$ by $\hat{\mu} = \frac{1}{T}\sum_{t=1}^T Y_t^f$ and $\sigma_Y$ by

$\hat{\sigma} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^f - \hat{\mu}\right)}$.

Let $X_T \in \mathbb{R}$ denote the (model independent) initial capital at time $T$, and let $\Pi_{T+1}^s \in \mathcal{X}_{T+1}(\mathcal{M}^s)$ denote the (future) portfolio at time $T+1$. To compute the estimation risk, we follow the approach of focusing directly on the confidence interval of the given risk measurement method (VaR in this case) evaluated at the given product.

First, assume that according to model $m$ the asset log returns are normally distributed and let $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$ be the estimate of the parameter $\theta = (\mu, \sigma)$, with corresponding estimated nominal selected parametric model $\widehat{m}_T^s$. The value-at-risk of portfolio $\Pi_{T+1}^s$ at level $p$ is according to model $\widehat{m}_T^s$ is given by

$$\mathrm{VaR}^s_{\widehat{m}_T^s, T+1}\left(\Pi^s_{\widehat{m}_T^s, T+1}\right) = X_T\left(1 - \exp\left(z_p\hat{\sigma}\sqrt{f} + \hat{\mu}f\right)\right), \tag{16}$$

where $z_p$ denotes the $p^{th}$ quantile of the (standard) normal distribution. Based on an asymptotic approximation, the upper bound of the $(1-\alpha)$ confidence interval around $\mathrm{VaR}^s_{\widehat{m}_T^s, T+1}\left(\Pi^s_{\widehat{m}_t^s, T+1}\right)$ is then given by

$$\mathrm{VaR}^s_{\widehat{m}_T^s, T+1}\left(\Pi^s_{\widehat{m}_T^s, T+1}\right) + z_{\alpha/2}\sqrt{\Sigma_{\mathrm{VaR}}/T}, \tag{17}$$

where

$$\Sigma_{\mathrm{VaR}} = X_T^2 f\, \mathrm{VaR}^s_{\widehat{m}_T^s, T+1}\left(\Pi^s_{\widehat{m}_T^s, T+1}\right)^2\left[\sigma^2\left(1 + f/4\right) + z_p\sigma^3\sqrt{f} + z_p^2\sigma^2/2\right].$$

This follows from a straightforward application of the delta method applied to equation (16), using the well-known limit distribution of $\widehat{\theta}$:

$$\sqrt{n}\left(\begin{bmatrix}\hat{\mu}\\ \hat{\sigma}^2\end{bmatrix} - \begin{bmatrix}\mu_Y\\ \sigma_Y^2\end{bmatrix}\right) \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma_Y^2/f & 0\\ 0 & 2\sigma_Y^4\end{bmatrix}\right), \tag{18}$$

valid under the normality assumption. The results can straightforwardly be generalized to nonnormal distributions.

Nonparametric versions of VaR may be computed on the basis of the empirical distri-

bution function. We denote by $\widehat{m}_T^S$ the corresponding estimated model, with $m \in \mathcal{M}^S$. The nominal nonparametric value-at-risk is given by

$$\text{VaR}^S_{\widehat{m}_T^S}\left(\Pi^S_{\widehat{m}_T^S}\right) = X_T\left(1 - \exp\left(Y^f_{T(\lfloor pn \rfloor + 1)}\right)\right), \tag{19}$$

where $T(t)$ denotes the $t^{th}$ order statistic of $\left(Y_1^f, ..., Y_T^f\right)$ and $\lfloor a \rfloor$ is the largest integer that is less than or equal to $a$. The upper bound of the (nonparametric) confidence interval around $\text{VaR}^S_{\widehat{m}_T^S}\left(\Pi^S_{\widehat{m}_T^S}\right)$ may be computed as

$$\text{VaR}^S_{\widehat{m}_T^S}\left(\Pi^S_{\widehat{m}_T^S}\right) + z_{\beta/2} X_T^2 \sqrt{\frac{p(1-p)}{T\left(F'\left(F_T^{-1}(p)\right)\right)^2}} \tag{20}$$

where $\beta$ is as defined in (13) and where $F'(x)$ can be estimated using, for instance, the Rosenblatt-Parzen kernel estimator.[21]

## C. Limit Distributions Expected Shortfall

Assume the same setting as in case of VaR. The nominal selected parametric ES (at level $p$) under normality of the asset returns is given by

$$\text{ES}^s_{\widehat{m}_T^s}\left(\Pi^s_{\widehat{m}_T^s, T+1}\right) = X_T\left(1 - \frac{1}{p}\exp\left(\hat{\mu} f + \tfrac{1}{2}\hat{\sigma}^2 f\right)\Phi\left(z_p - \hat{\sigma}\sqrt{f}\right)\right). \tag{21}$$

The worst-case parametric ES may now be computed as, applying the delta method,

$$\text{ES}^s_{\widehat{m}_T^s}\left(\Pi_{\widehat{m}_T^s, T+1}\right) + z_{\alpha/2}\sqrt{\Sigma_{ES^s}/T} \tag{22}$$

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

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

$$\Sigma_{ES^s} = X_T^2 f \Psi_{\mathrm{ES}}^2 \left[ \sigma^2 + \left( f\sigma - \sqrt{f}\frac{\phi}{\Phi}\left( z_p - \sigma\sqrt{f} \right) \right) \sigma^3 + \left( \sqrt{f}\sigma - \frac{\phi}{\Phi}\left( z_p - \sigma\sqrt{f} \right) \right)^2 \sigma^2/2 \right], \quad (23)$$

where $\Psi_{\mathrm{ES}} = \mathrm{ES}^s_{\widehat{m}^s_T}\left( \Pi^s_{\widehat{m}^s_T, T+1} \right)$ and $\frac{\phi}{\Phi}(x) = \frac{\phi(x)}{\Phi(x)}$.

The nonparametric ES can be computed by

$$\mathrm{ES}^S_{\widehat{m}^S_T}\left( \Pi^S_{\widehat{m}^s_T, T+1} \right) = X_T \left( 1 - \frac{1}{\lfloor Tp \rfloor + 1} \sum_{t=1}^{T} \exp\left( Y_t^f \right) \mathbf{I}_{\left\{ \left( -\infty, \mathrm{VaR}^S_{\widehat{m}^S_T}\left( \Pi^S_{\widehat{m}^S_T, T+1} \right) \right] \right\}} \left( Y_t^f \right) \right). \quad (24)$$

The upper bound of the corresponding confidence interval can be computed as

$$\mathrm{ES}^S_{\widehat{m}^S_T}\left( \Pi^S_{\widehat{m}^S_T, T+1} \right) + z_{\alpha/2}\sqrt{\Sigma_{\mathrm{ES}^S}/T} \quad (25)$$

where

$$\begin{aligned} \Sigma_{\mathrm{ES}^S} &= X_T^2 \frac{1}{p} \mathbb{E}\left[ Y^2 \mid Y \le \mathrm{VaR}^S_{\widehat{m}^S_T}\left( \Pi^S_{\widehat{m}^S_T, T+1} \right) \right] - \left( \mathrm{ES}^S_{\widehat{m}^S_T}\left( \Pi^S_{\widehat{m}^S_T} \right) \right)^2 \\ &- \left( 1 - \frac{1}{p} \right) \left( \mathrm{VaR}^S_{\widehat{m}^S_T}\left( \Pi^S_{\widehat{m}^S_T, T+1} \right) \right)^2 + \left( 2 - \frac{2}{p} \right) \mathrm{ES}^S_{\widehat{m}^S_T}\left( \Pi^S_{\widehat{m}^S_T, T+1} \right) \mathrm{VaR}^S_{\widehat{m}^S_T}\left( \Pi^S_{\widehat{m}^S_T, T+1} \right). \end{aligned}$$

# 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).

## References


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.

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.

Berkowitz, J. and O'Brien, J.: 2002, How accurate are value-at-risk models at commercial banks?, *Journal of Finance* **57**, 1093–1111.

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.

Delbaen, F.: 2000, Coherent risk measures on general probability spaces, *Working paper ETH* pp. 1–35.

Derman, E.: 1996, Model risk, *Risk* **9**, 34–37.

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.

El Karoui, N., Jeanblanc-Picqué, M. and Shreve, S. E.: 1998, Robustness of the Black-Scholes formula, *Mathematical Finance* **8**, 93–126.

Gibson, R., Lhabitant, F.-S., Pistre, N. and Talay, D.: 1999, Interest rate model risk: An overview, *Journal of Risk* **3**, 37–62.

Green, T. C. and Figlewski, S.: 1999, Market risk and model risk for a financial institution writing options, *Journal of Finance* **54**, 1465–1499.

Heckman, J.: 2000, Causal parameters and policy analysis in economics: A twentieth century retrospective, *The Quarterly Journal of Economics* **15**, 45–98.

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.

Kerkhof, J. and Melenberg, B.: 2004, Backtesting for risk-based regulatory capital, *Journal of Banking and Finance* **28**, 1845–1865.

Kerkhof, J., Melenberg, B. and Schumacher, H.: 2003, How risky are written derivatives positions? a model risk analysis, *Mimeo Tilburg University* .

Kupiec, P. H.: 1995, Techniques for verifying the accuracy of risk management models, *Journal of Derivatives* **3**, 73–84.

Risk Magazine: 1996, Value at risk, *Risk Magazine Special Supplement* pp. 68–71.

Shorack, G. R. and Wellner, J. A.: 1986, *Empirical Processes with Applications to Statistics*, John Wiley & Sons, New York.

Talay, D. and Zheng, Z.: 2002, Worst case model risk management, *Finance and Stochastics* **6**, 517–537.