Journal of Banking & Finance 34 (2010) 267–279

Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

Model risk and capital reserves

Jeroen Kerkhof [a], Bertrand Melenberg [b,c,*], Hans Schumacher [b]

[a] VAR Strategies, Belgium
[b] Department of Econometrics and Operations Research, CentER, Netspar, Tilburg University, Netherlands
[c] Department of Finance, Tilburg University, Netherlands

ARTICLE INFO

Article history:
Received 24 September 2008
Accepted 24 July 2009
Available online 6 August 2009

JEL classification:
G12
G18

Keywords:
Capital requirements
Model risk
Risk measures
Estimation risk
Misspecification risk
Identification risk

ABSTRACT

We propose a procedure to take model risk into account in the computation of capital reserves. This addresses the need to make the allocation of capital reserves to positions in given markets dependent on the extent to which reliable models are available. The proposed procedure can be used in combination with any of the standard risk measures, such as Value-at-Risk and expected shortfall.

We assume that models are obtained by usual econometric methods, which allows us to distinguish between estimation risk and misspecification risk. We discuss an additional source of risk which we refer to as identification risk. By way of illustration, we carry out calculations for equity and FX data sets. In both markets, estimation risk and misspecification risk together explain about half of the multiplication factors employed by the Bank for International Settlements (BIS).

1. Introduction

Due to the growing complexity of financial markets, financial institutions rely more and more on models to assess the risks to which they are exposed. The accuracy of risk assessment depends crucially on the reliability of these models. In spite of the very substantial efforts made both by practitioners and academics to improve the quality of market models, one needs to recognize that there is no such thing as a perfect model. The hazard of working with a potentially incorrect model is referred to as *model risk*. Methods for the quantification of this type of risk are not nearly as well developed as methods for the quantification of market risk given a model, and the view is widely held that better methods to deal with model risk are essential to improve risk management.[1]

While model risk is present in each market, the significance of its role may be different in different markets, depending on such factors as the amount of experience that has been built up and the complexity of products that are being traded. Additionally, it may be important to distinguish between the various purposes for which models are used, such as pricing, hedging, and the computation of capital reserves. There can be several sources of model risk, including for instance the risk of human error in modeling; various factors, and ways to guard against them, are discussed by Derman (1996).

The main question that we want to answer in this paper is how to incorporate model risk associated to the application of econometric methods into the computation of required levels of capital reserves. While it is of course possible and useful to investigate the effect of parameter variations on risk measures that are computed in a particular parametric framework, as for instance in Bongaerts and Charlier (2009), our aim in this paper is to take model risk explicitly into account as a separate risk factor. The paper is similar in spirit to the work of West (1996), who discusses when and how to adjust critical values for tests of predictive ability in order to take parameter estimation uncertainty into account. Here we adjust levels of risk measures, such as VaR, rather than critical values. There are parallels as well with the work of Brock et al. (2003, 2007) on the role of model risk in policy evaluation. However, the model averaging method proposed by Brock et al.

* Corresponding author. Address: Department of Econometrics and Operations Research, CentER, Netspar, Tilburg University, Netherlands. Tel.: +31 13 466 2730; fax: +31 13 466 3280.

E-mail addresses: jeroen@varstrategies.com (J. Kerkhof), b.melenberg@uvt.nl (B. Melenberg), j.m.schumacher@uvt.nl (H. Schumacher).

[1] For instance, the chief executive of the British Financial Services Authority, Hector Sants, writes in a letter to financial industry CEOs (August 13, 2008; http://www.fsa.gov.uk/pubs/ceo/valuation.pdf): "Few firms have sufficiently developed frameworks for articulating model risk tolerance, and measuring and controlling risks within that tolerance. We believe a better defined and implemented model risk management framework could therefore feed into a better defined and implemented valuation risk management framework".

0378-4266/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbankfin.2009.07.025

(2003) is difficult to use in the applications we have in mind, since it requires specification of a prior over the model space. The dispersion view expounded by Brock et al. (2007) is useful for robustness analysis of policies; risk assessment, however, typically calls for quantification of risk by means of a single number expressing the required capital reserve.

Hull and Suo (2002) quantify model risk by comparing the pricing and hedging performance of a simple model within the context of a more complicated model, which is interpreted as representing the true data generating process. This is a way of judging whether a proposed model simplification is feasible. In this paper we do not assume that we know the true data generating process. Closer to the present paper is the work by Cont (2006), who studies the impact of model risk on pricing. Our paper is different in the sense that we are concerned with risk measures rather than prices; moreover, we discuss a specific procedure to arrive at classes of alternative models, whereas the discussion of Cont (2006) is abstract in this respect. Approaches to the design of policies that are robust with respect to model uncertainty have been developed for instance by Hansen and Sargent (2007) and ElKaroui et al. (1998). Robust hedging strategies can play an important role in mitigation of model risk. Here we assume that the effect of hedging is already incorporated in the definition of a given position.

In this paper, we incorporate model risk into risk measure calculations by constructing classes of models on the basis of standard econometric procedures. We then arrive at adjustments of a nominal risk measure, such as VaR, by computing the worst case across such classes. We distinguish between several stages of modeling which each give rise to different model classes and hence to different adjustments of a chosen risk measure. In this way, we define several components of model risk which we refer to as *estimation risk*, *misspecification risk,* and *identification risk*.

Estimation risk is the risk associated with inaccurate estimation of parameters. This type of model risk has perhaps most frequently been discussed in the literature; see for instance Gibson et al. (1999), Talay and Zheng (2002) and Bossy et al. (2000). Misspecification risk is associated with incorrect model specification. The presence of misspecification risk may be detected by the use of standard econometric methods. Identification risk arises when observationally indistinguishable models have different consequences for capital reserves.

For an example of identification risk, consider the following situation. The value of a mortgage portfolio depends on the delinquency behavior of home owners, which in turn may be influenced by house price appreciation rates. If in the available data set the house price appreciation rate is constant, or nearly so, then on the basis of statistical procedures it is not possible to make statements about the dependence of model parameters on the appreciation rate. Of course, one might construct models under the assumption that parameters do not depend on the house price appreciation rate. However, such models may not provide adequate risk assessment in situations in which the appreciation rate does move considerably.[2]

As noted by Cont (2006), the notion of components of model risk does not come up in an abstract framework, since in such a setting one can work with a class of alternative models which is assumed given and which in principle does not need to have any particular structure. However, in applications we must have a way to construct the class of alternative models, and this may lead to the presence of more structure than can be or needs to be supposed at the abstract level. The situation we consider in this paper is of that type. The distinctions that we construct are meaningful within the context of a given econometric framework; of course, they do depend on the specific framework that is chosen and we do not claim that model risk in general should be or even can be decomposed in such a way.

Currently, no explicit capital requirements are imposed by regulators in connection with model risk, save perhaps risk factors due to human errors which are covered under operational risk. However, the Basel Committee does apply the so called *multiplication factors* which could be motivated in part as a way of taking model risk into account. In our empirical applications, we consider time series data of the S&P 500 index and the USD/GBP exchange rate. We consider two simple models (a Gaussian i.i.d. model and a GARCH(1,1) model) which are estimated on the basis of rolling-window data, and we find that both are rejected when used as nominal models. However, adding misspecification risk and estimation risk at the usual 95% confidence level leads to risk measure levels that pass the standard backtests in all cases we consider. 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 for estimation risk. Of course there are other considerations that may play a role as well in the determination of multiplication factors, such as macroeconomic effects (procyclicality) and the effects on the behavior of regulated institutions; see for instance Heid (2007) and Kaplanski and Levy (2007).

The remainder of the paper is structured as follows. In Section 2 we first present a simple example based on simulated data to illustrate the approach to model risk that we take in this paper. Then, in Section 3, we discuss the formulation of risk measures in an environment in which we work with multiple models that may employ different probability spaces. Our method of quantifying model risk is presented in general terms in Section 4. Then we turn to empirical applications in Section 5. Section 6 concludes.

2. Illustration of model risk

By *market risk* we understand the risk caused by fluctuations in asset prices. Market risk for a given position may be quantified by a risk measure such as Value-at-Risk (VaR) or Expected Shortfall (ES). The purpose of this section is to illustrate in a simple example the possible impact of several forms of model risk on market risk assessment.

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 available at time T, represented by a σ-algebra $\mathscr{F}_T$. To form a risk assessment, one could for instance choose to model the distribution of the log return Y_{T+1} as a normal distribution with mean μ and variance σ^2, i.e.,

$$Y_{T+1}|\mathscr{F}_T \sim \mathscr{N}(\mu, \sigma^2) \tag{1}$$

or, equivalently, $Y_{T+1} = \log(X_{T+1}/X_T) = \mu + \sigma\epsilon_{T+1}$ with $\epsilon_{T+1}|\mathscr{F}_T \sim \mathscr{N}(0, 1)$. The Value-at-Risk at level p is then given by

$$\mathrm{VaR}_T(X_{T+1} - X_T) = X_T(1 - \exp(z_p\sigma + \mu)) \tag{2}$$

where z_p denotes the pth quantile of the standard normal distribution, and where the index T indicates that the VaR is calculated at time T, given the information $\mathscr{F}_T$. One may also choose an alterna-

[2] This example is modeled on the 2007 subprime mortgage crisis, but is stated in an idealized form to bring out more clearly the notion of identification risk. Whether in the case of the subprime crisis it was really not possible to predict the changes in delinquency behavior on the basis of data is debatable; see for instance Demyanyk and Van Hemert (forthcoming).

Table 1
Illustration of model risk.

p	VaR	Est.risk	E&MR N	E&MR t	E&MR MixedN
0.01	7.07	0.48 (0.07)	0.95 (0.13)	2.99 (0.43)	1.06 (0.15)
0.025	5.99	0.43 (0.07)	0.69 (0.12)	1.12 (0.19)	0.67 (0.11)
p	ES	Est. risk	E&MR N	E&MR t	E&MR MixedN
0.01	8.05	0.53 (0.07)	1.01 (0.13)	6.04 (0.75)	1.97 (0.24)
0.025	7.09	0.49 (0.07)	0.83 (0.12)	3.03 (0.43)	1.16 (0.16)

The table reports the Value-at-Risk (VaR) and expected shortfall (ES) market risk at level p, together with corresponding estimation risk (Est risk), assuming i.i.d. normal returns with annualized mean 0 and variance 0.25, with relative estimation risk in brackets. The sample size is 500; the initial capital is 100. In addition, the sum of estimation and misspecification risk (E&MR) is reported in experiments in which the data set is generated by a normal (N), a t_4 (t), or a mixed normal (MixedN) distribution, all 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, while the second, with probability 0.01, has annualized variance 1.24. Relative risks (with respect to VaR or ES) are shown in brackets.

tive risk measure such as Expected Shortfall (ES) at level p, which under the normality assumption is given by

$$\mathrm{ES}_T(X_{T+1}-X_T) = X_T\left(1-\frac{1}{p}\exp\left(\mu+\frac{1}{2}\sigma^2\right)\Phi(z_p-\sigma)\right) \tag{3}$$

where Φ denotes the standard normal cumulative distribution function.

However, in a practical situation, VaR and ES are not observable, since μ and σ are unknown parameters. Instead, one has to estimate μ and σ, say, by $\widehat{\mu}$ and $\widehat{\sigma}$ respectively, assuming, for instance, the availability of an i.i.d. sample of past values of Y. Replacing (μ,σ) by their estimates $(\widehat{\mu},\widehat{\sigma})$, one obtains an estimated VaR and estimated ES. The quantity obtained in this way will be referred to as the *nominal market risk* as measured by VaR or ES.

Clearly, the estimated parameters $\widehat{\mu}$ and $\widehat{\sigma}$ in general contain sampling (estimation) error, which is transferred to the estimated VaR and ES, so that the use of the estimated VaR or ES may 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 nominal market risk then gives an indication of what might be called the *estimation risk.*[3]

This is illustrated in Table 1. The second column of the table shows estimates of VaR and ES at levels $p = 0.01$ and $p = 0.025$ which have been computed on the basis of simulated data. Specifically, we used 500 observations of a model whose log returns are i.i.d. normally distributed with annualized mean $\mu = 0$ and variance $\sigma^2 = 0.25$; the initial capital is 100 and the estimation method is Maximum Likelihood. The third column shows corresponding estimation risks which have been calculated as the difference between the upper bounds of the 95% confidence intervals and the nominal market risk, which is given by the point estimates of VaR and ES. The *relative* estimation error, calculated as the estimation error relative to the nominal market risk, and included between brackets, takes values around 7% in all cases.

We now include nonnormal data generating processes in the analysis, while holding on to the i.i.d. assumption. In practice, inadequacy of the normality assumption for the conditional distribution of log returns may be detected by misspecification tests. In such a case we might postulate no more than

$$Y_{T+1}|\mathscr{F}_T \sim G \tag{4}$$

where G is an arbitrary continuous and strictly increasing cumulative distribution function. We then find for the Value-at-Risk at level p

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

while for the Expected Shortfall at level p we get

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

Let $\widehat{G}$ denote a nonparametric estimate of G; then a nonparametric estimate of VaR and ES can be obtained by replacing G by $\widehat{G}$. Sampling error in this nonparametric estimate can again be taken into account by constructing, say, a 95% confidence interval. The difference between the upper bound of this intervals and the nominal market risk then provides us with a measure that may be viewed as incorporating a combination of possible model misspecification and estimation error. We shall refer to this as the combination of *estimation* and *misspecification risk.*

Table 1 (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 nominal market risk, for the cases when the log returns are generated by a normal, a t_4, or a mixed normal distribution, in all cases standardized such that the annualized mean and variance are given by $\mu = 0$ and $\sigma^2 = 0.25$. The mixed normal distribution is composed of two zero-mean normal distributions of which the first, with probability 0.99, has variance 0.24, while the second, with probability 0.01, has variance 1.24. A nonparametric estimation technique (cf. Section 5) is used to compute both the nominal market risk and the 95% confidence interval. The table shows that, when the normal distribution is used to generate the data, the combination of estimation and misspecification risk slightly exceeds the estimation risk that we found when estimation was done under the (correct) assumption of normality of the log returns (12%, rather than 7%). Much larger figures are found when the true distribution is nonnormal, in particular when the t_4 distribution is used with $p = 0.01$; in relation to the estimated risk measures, the combination of estimation and misspecification risk amounts to 42% in the VaR case and even 75% in the ES case.

Even if $Y_{T+1}|\mathscr{F}_T \sim G$ cannot be rejected on the basis of misspecification tests using the available past data, measures quantifying future risk employing this model may still fail. Indeed, this may happen when past and present are insufficiently representative of the future (for risk assessment purposes). Such contingencies can be explored by stress testing. For instance, one may propose a model of the form $(Y_{T+1}-\alpha(Z_{T+1}))/\beta(Z_{T+1})|\mathscr{F}_T \sim G$ where Z_t is a variable that has been constant or nearly constant for $t \leqslant T$ but that for some reason experiences a shock between time T and time $T+1$. When α and β are included, the Value-at-Risk at level p is given by

$$\mathrm{VaR}_T(X_{T+1}-X_T) = X_T(1-\exp(\alpha+\beta G^{-1}(p))),$$

and for the Expected Shortfall at level p we get

$$\mathrm{ES}_T(X_{T+1}-X_T) = X_T\left(1-\frac{1}{p}\int \exp(\alpha+\beta y)\mathbb{1}_{(-\infty,G^{-1}(p)]}(y)\,dG(y)\right).$$

[3] We employ the assumption of i.i.d. returns. However, this assumption might not be satisfied in a practical application. Non-i.i.d. returns require, in particular, an adaptation of the calculation of the estimation risk. Grané and Veiga (2008) indicate how to proceed in case of non-i.i.d. returns.

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

3. Model risk

In the literature, model risk is usually formalized in terms of a collection of different probability measures on the same probability space (as for instance in Cont (2006)). However, financial institutions may carry out risk assessment studies on the basis of models of different natures. For instance, models based on a finite number of economic scenarios may be used alongside continuous-time models based on stochastic differential equations or discrete-time (VAR, GARCH, ...) models. In this section we present a worst-case formulation of model risk extending across models that may represent uncertainty in different ways. We do assume that (as is usual) risk assessment is done on the basis of a probabilistic model, concentrating in particular on Value-at-Risk and Expected Shortfall. The discussion in this section is abstract; in the next section, we discuss a specific construction of a class of alternative models within the context of a given econometric framework. For the sake of simplicity we focus on the quantification of risk involved with a payoff taking place at a fixed date rather than with a payoff stream. The assessment of risk is to be made at the current time. In the notation we suppress the time indices referring to the current time and the time at which the payoff takes place.

We start with a model class $\mathscr{M}$. Models may be of different nature so that we prefer to consider $\mathscr{M}$ as an abstract set, but we do assume that to each model $m \in \mathscr{M}$ there is an associated probability triple $(\Omega_m, \mathscr{F}_m, P_m)$ which allows a probabilistic formulation of the risks that we want to quantify. Associated to the triple $(\Omega_m, \mathscr{F}_m, P_m)$ is the space of random variables $L_0(\Omega_m, \mathscr{F}_m, P_m)$ (the P_m-equivalence classes of $\mathscr{F}_m$-measurable functions on Ω_m); for brevity, this space will be denoted by $L(m)$. Extending the terminology of Artzner et al. (1999), we shall say that a *risk* is a mapping Π that assigns to each $m \in \mathscr{M}$ a random variable $\Pi(m) \in L(m)$. The set of all such mappings is denoted by $\mathscr{X}(\mathscr{M})$. A "risk" in this sense involves probabilistic uncertainty through the probabilistic nature of each $\Pi(m)$ as well as ambiguity through the parametrization in terms of $\mathscr{M}$. Elements of $\mathscr{X}(\mathscr{M})$ will sometimes be referred to simply as "products", "assets", or "positions", even though this usage neglects the distinction that should be made between an actual product, such as a Google share, and its representation as a mathematical object. The values that $\Pi(m)$ take may be interpreted as representing payoffs in terms of units of currency, or payoffs in terms of units of a chosen reference asset.

To extend the notion of risk measure to a multimodel setting, we use a description that takes the probability space as an argument, rather than as a given environment. Value-at-Risk at a given level p, for instance, is redefined as a mapping that takes two arguments, namely a model m and a random variable X defined on $L(m)$, and that assigns to m and X the number $\mathrm{VaR}_p(m, X)$ defined by

$$\mathrm{VaR}_p(m, X) = -\inf\{x \in \mathbb{R} | P_m(X \leqslant x) \geqslant p\}. \tag{7}$$

Likewise, we can define (employing the description of expected shortfall in Acerbi and Tasche (2002)), with $Q_p = -\mathrm{VaR}_p(m, X)$,

$$\mathrm{ES}_p(m, X) = -\frac{1}{p}\left[E_{P_m} X \mathbb{1}_{X \leqslant Q_p} + Q_p(p - P_m(X \leqslant Q_p))\right]. \tag{8}$$

In general, we may define a *risk measurement method* for the model class $\mathscr{M}$ as a mapping ρ that assigns a number $\rho(m, X) \in \mathbb{R}' := \mathbb{R} \cup \{\infty\}$ to each pair (m, X) where $m \in \mathscr{M}$ and $X \in L(m)$. Given such a risk measurement method ρ, we may then define another mapping which (by abuse of notation) is also denoted by ρ, and which assigns to each element of $\mathscr{X}(\mathscr{M})$ a function from $\mathscr{M}$ to $\mathbb{R}'$, defined by

$$(\rho(\Pi))(m) = \rho(m, \Pi(m)).$$

In this way we define a risk measure that can be used for multimodel representations; we will call it a *multimodel risk measure*.

In the above we did not discuss which properties a risk measure ρ should have in order to qualify for that title. In the recent literature there has been a renewed interest in axioms for risk measures (cf., for instance, Artzner et al. (1999), Föllmer and Schied (2002), and Frittelli and Rossazza Gianin (2002), to mention only a few contributions). Here, we just note that the multimodel risk measures introduced above can be looked at as parametrized versions of risk measures defined on a given probability space, so that notions such as monotonicity, translation invariance etc. carry over in a straightforward way. In particular, if axioms are imposed, such as to support the interpretation of the risk measure ρ as the amount of required capital, then the function $(\rho(\Pi))(m)$ defined on a given model class $\mathscr{M}$ can be viewed as giving, for each $m \in \mathscr{M}$, the amount of required capital under the assumptions of model m.

A multimodel risk measure as defined above is a function on a given model class $\mathscr{M}$. To arrive at a single number which represents an overall assessment of risk, we shall take a worst-case approach and define

$$\mathrm{RISK}_{\rho,\mathscr{M}}(\Pi) = \sup_{m \in \mathscr{M}} \rho(m, \Pi(m)). \tag{9}$$

An alternative would be a Bayesian approach, in which the market risk measure is a weighted average of risk measures according to some prior. Depending on its risk attitude, a financial institution may then assign larger weights to unfavorable models. However, the choice of a prior is difficult and to a large extent arbitrary. In a worst-case approach, one only needs to specify the model set $\mathscr{M}$; this may be seen as an acknowledgement of the restrictions of statistical modeling in the face of limited data and limited understanding of the true dynamics.

The risk defined in (9) depends on the model class $\mathscr{M}$. If the model class consists of a single element (the case of no ambiguity), then the definition of risk measurement method, as given above, coincides with the usual one. In the general case, given a risk measurement method ρ, a *nominal model* m_0 and a *tolerance set* $\mathscr{K}$ with $m_0 \in \mathscr{K}$, we can define the *model risk* of product Π by[4]

$$\phi_\rho(\Pi, m_0, \mathscr{K}) = \mathrm{RISK}_{\rho,\mathscr{K}}(\Pi) - \mathrm{RISK}_{\rho,m_0}(\Pi). \tag{10}$$

In this way one obtains a decomposition of the *total market risk* $\mathrm{RISK}_{\rho,\mathscr{M}}(\Pi)$ as the sum of *nominal market risk* $\mathrm{RISK}_{\rho,m_0}(\Pi)$ and model risk. Obviously, similar definitions may be stated in case one has several levels of ambiguity expressed by nested model classes; we define, assuming $\mathscr{K}_1 \subset \mathscr{K}_2$,

$$\phi_\rho(\Pi, \mathscr{K}_1, \mathscr{K}_2) = \mathrm{RISK}_{\rho,\mathscr{K}_2}(\Pi) - \mathrm{RISK}_{\rho,\mathscr{K}_1}(\Pi). \tag{11}$$

[4] We assume here that the nominal risk $\mathrm{RISK}_{\rho,m_0}(\Pi)$ is finite; if not, we leave model risk undefined.

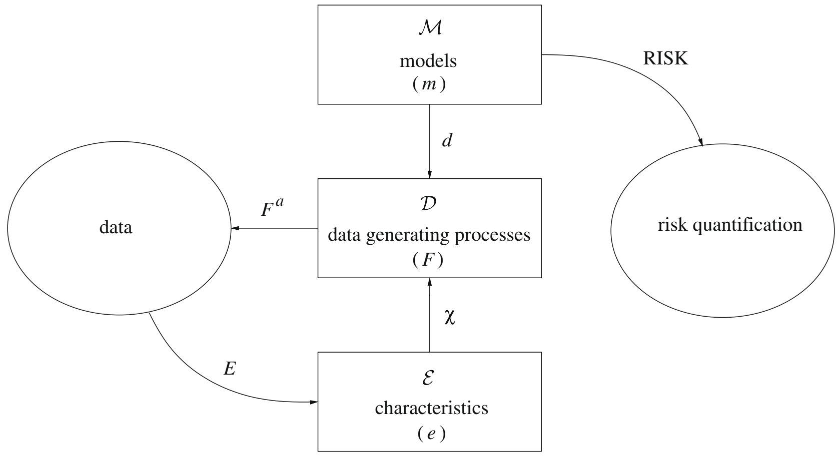

Fig. 1. Risk assessment on the basis of econometric modeling.

Below we discuss how such model classes come up naturally in the context of econometric modeling.

4. Model classes from econometric modeling

In addition to the framework of the previous section, we now assume that we have data available and that econometric procedures will be applied to extract suitable models from the data. In addition to the probability space $(\Omega_m, \mathscr{F}_m, P_m)$ and the random variable $\Pi(m)$ that are used for risk assessment and that therefore serve a forward-looking purpose, we now assume that to each model m under consideration there is also associated a joint distribution function F_m from which the observed data may be supposed to be drawn. This joint distribution function is referred to as the "data generating process" of model m. For the purposes of the discussion in this section we do not need a specification of the relation between the forward-looking part of the model and the data generating process; in practice, of course, these components are typically closely related.

We assume that a basic (low-dimensional) model class $\mathscr{M}$ is given and that we also have an extension model class $\mathscr{M}_1$ which may for instance be a nonparametric class. Working within the class $\mathscr{M}$, econometric methods provide us typically with a point estimate as well as a tolerance (confidence) set for the class of models that may describe a given product Π. It may happen, however, that these models fail to satisfy misspecification tests. The econometrician can then switch to the broader class $\mathscr{M}_1$ and obtain a new tolerance set $\mathscr{K}_1$ which typically is wider than the original set $\mathscr{K}$. For the purposes of risk management, it may be desirable to apply further tests that are not data-based ("stress tests") and these tests may suggest a further extension of the tolerance set. In this way we obtain a hierarchy of tolerance sets which leads to a decomposition of total risk in various components.

4.1. Selection of nominal models

In this subsection we describe an econometrically based selection of the set of nominal models $\mathscr{K}$, given some selected model class $\mathscr{M}$, and having available past and present data.[5] The setup is illustrated in Fig. 1.

Assume that the available (past and present) data can be represented by a vector z. This data set may consist of past and present observations concerning the product Π under consideration, but observations on other variables may be included as well. We think of the data z as being a realization of a random vector Z which is distributed according to a joint cumulative distribution function which we call the *actual data generating process*. In general, the actual data generating process F^a is unknown, but we can assume that it belongs to a (large) set of possible data generating processes which is denoted by $\mathscr{D}$ (the box "data generating processes" in Fig. 1, where the "data" are generated by F^a). The data generating process associated to a model m is given by a mapping d, which assigns to each model $m \in \mathscr{M}$ the *induced* probability distribution $d(m) \in \mathscr{D}$ (see "d", linking the boxes "models" and "data generating processes" in Fig. 1). Thus, given model m, we postulate $Z \sim d(m)$. The set of all data generating processes induced by $\mathscr{M}$ is denoted by $d(\mathscr{M})$. We have $d(\mathscr{M}) \subset \mathscr{D}$, where typically the inclusion is strict.

VaR example. To illustrate, consider the case where the data set consists of observations on a product Π forming a time series sample of length T. Suppose one postulates $Z = (Z_1, \ldots, Z_T)' \sim F^a = G^T$, i.e., $Z_1, \ldots, Z_T \overset{\text{i.i.d.}}{\sim} G$. Then a model class may be given by $\mathscr{M} = \{m_\theta | \theta \in \Theta \subset \mathscr{R}^k\}$, with $d(m_\theta) = (G_\theta)^T$, and $d(\mathscr{M}) = \{(G_\theta)^T | \theta \in \Theta \subset \mathbb{R}^k\}$. The example in Section 2 corresponds to the case where G_θ is the normal distribution function with parameter vector $\theta = (\mu, \sigma)'$, cf. (1).

For purposes of estimation, data generating processes are typically represented in terms of *characteristics* such as moments, cumulative distribution functions, transformations, copulas, and so on. Let $\mathscr{E}$ denote the *characteristics space* and let the *data characterization mapping* be given by $\chi : \mathscr{E} \to d(\mathscr{M})$ (see "χ", linking "characteristics" and "data generating processes" in Fig. 1). For instance, in terms of our illustration when the model class is given by $\mathscr{M} = \{m_\theta | \theta \in \Theta \subset \mathbb{R}^k\}$, with $d(m_\theta) = (G_\theta)^T$, we can take $\mathscr{E} = \Theta$, with $\chi(\theta) = (G_\theta)^T$.

The selection of the set of nominal models now proceeds as follows. First, let $\widehat{E} = E(z) \subset \mathscr{E}$ represent the *selected characteristics* based on the employed econometric method (for instance ML, GMM, ...) and using the available data z (indicated by "E", linking "data" and "characteristics" in Fig. 1). In case we only work with a point estimate, we have $\widehat{E} = \{\widehat{e}\}$, where $\widehat{e} = \widehat{e}(z)$ represents an *estimate* of the characteristics. Alternatively, when allowing for estimation inaccuracy the estimation procedure results in a confidence set $\widehat{E}$. Next, given the data characterization mapping χ and the selected characteristics $\widehat{E}$, the selected (or estimated) data generating processes in $\mathscr{D}$ are given by $\Delta := \chi(\widehat{E}) \subset d(\mathscr{M})$. Finally, given Δ, we find as set of nominal models $\mathscr{K} = d^{-1}(\Delta)$.[6]

[5] We follow to some extent Heckman (2000).

[6] By definition, $d^{-1}(\Delta) = \{m \in \mathscr{M} | d(m) \in \Delta\}$.

This results in the quantification of market risk given by $\text{RISK}_{\rho,d^{-1}(\Delta)}(\Pi)$.

VaR example (cont'd). We have, with $\hat{e} = \hat{\theta} = (\hat{\mu}, \hat{\sigma})'$, and $\Delta = \chi(\hat{e})$,

$$\text{RISK}_{\text{VaR},d^{-1}(\Delta)}(\Pi) = X_T(1 - \exp(z_p\hat{\sigma} + \hat{\mu})), \tag{12}$$

the sample analogue of (2). In addition, we can construct a confidence region around $\hat{e}$, resulting in $\hat{E}$, and then calculate, with $\Delta = \chi(\hat{E})$,

$$\text{RISK}_{\text{VaR},d^{-1}(\Delta)}(\Pi) = \sup_{(\mu,\sigma)\in\hat{E}} X_T(1 - \exp(z_p\sigma + \mu)). \tag{13}$$

In case $d^{-1}(\Delta) = \{m_0\}$ for some $m_0 \in \mathscr{M}$, the market risk is quantified by $\text{RISK}_{\rho,m_0}(\Pi) = \rho(m_0, \Pi_{m_0})$, so that there is no ambiguity about the appropriate model in $\mathscr{M}$ that is used to quantify the market risk. However, when $d^{-1}(\Delta)$ is not a singleton, there is ambiguity. This situation may arise for two reasons. First, the selected characteristics that define Δ may not be unique, for instance, due to taking into account estimation inaccuracy. We refer to this type of ambiguity as *data ambiguity*, since there is ambiguity about the appropriate data generating process in $d(\mathscr{M})$. Secondly, $d^{-1}(\Delta)$ may not be a singleton, even if Δ is, because the transformation d is not one-to-one. In this case multiple models $m \in \mathscr{M}$ describe the same data generating process in $\mathscr{D}$, so that knowing this data generating process does not suffice to retrieve the underlying model. We refer to this type of ambiguity as *model ambiguity*. In econometric terms, model ambiguity arises in case of *lack of identification* or *underidentification*: in case d is not one-to-one different models in $\mathscr{M}$ describe the same data generating processes in $\mathscr{D}$, so that on the basis of a data generating process one is unable to identify a unique corresponding model $m \in \mathscr{M}$. Inasmuch as models that are mapped to the same data generating process are observationally equivalent, model ambiguity may be resolved by introducing appropriate identifying assumptions.

4.2. Estimation risk

Econometric methods typically deliver both a point estimate $\hat{e}$ and a confidence interval $\hat{E}$. This allows the quantification of *estimation risk* as follows (11):

$$\phi_\rho(\Pi, d^{-1}(\chi(\hat{e})), d^{-1}(\chi(\hat{E}))). \tag{14}$$

For instance, the difference between (12) and (13) yields the quantification of the estimation risk in the VaR example under the normality assumption.

Under the assumption that the mapping d is one-to-one so that there is no model ambiguity, we can for any given $e \in \mathscr{E}$ define a model $m = m(e) \in \mathscr{M}$ uniquely by requiring $d(m(e)) = \chi(e)$. We can then define a mapping g from the characteristics space $\mathscr{E}$ to the extended real line $\mathbb{R}'$ by $g(e) = \rho(m(e), \Pi(m(e)))$, and the estimation risk can be quantified as

$$\sup_{e\in\hat{E}} g(e) - g(\hat{e}). \tag{15}$$

If the characteristics space $\mathscr{E}$ admits a differentiable structure and the mapping g is smooth, it may be possible to construct an approximate quantification of estimation risk directly. For instance, in the VaR example, the nominal risk based on characteristics e, $g(e)$, is given by (12). Under appropriate regularity conditions, the limit distribution of this variable can be obtained by the functional delta method (see for instance Van der Vaart (1998)). Then an approximation to $\sup_{e\in\hat{E}} g(e)$ can easily be constructed from the distribution of $g(e)$.

4.3. Misspecification risk

We assume now that the result of the econometric procedure discussed in the previous subsection is subjected to a series of data-based misspecification tests; these may consist of standard backtests or of a more extensive suite of tests in which the assumptions of model set $\mathscr{M}$ are confronted with the data. We also assume that the econometrician has a larger model class available, denoted by $\mathscr{M}_1$.

For simplicity of terminology, we shall refer to the collection of tests alluded to above as the *set of backtests*, even though this set may contain tests that would not normally go under this name. The purpose of the backtests is to test, on the basis of past and present data, whether the risk assessment obtained on the basis of the estimation procedure described in the previous subsection is reliable. In particular, one may test the hypothesis $F^a \in \Delta$ where $\Delta = \chi(E(z))$ is the set of data generating processes obtained from the estimation procedure on the basis of the data z. When the hypothesis is rejected, the following possibilities exist. First, it may of course happen, as in any statistical procedure, that by chance the test produces an incorrect solution and that in fact F^a does belong to Δ. Secondly, again by chance, it might happen that F^a does not belong to the estimated set Δ, even though F^a does belong to $d(\mathscr{M})$; however, the probability of this should be small when the data set is sufficiently large and the estimation procedure is valid. Finally, it may be the case that $F^a \notin d(\mathscr{M})$. In this case the relation $F^a \in \Delta$ cannot hold for any choice of the selected set of characteristics $\hat{E}$. We shall henceforth focus on this possibility and speak of a *misspecification error* in case the hypothesis $F^a \in \Delta$ is rejected.

If a misspecification error is indicated, an extended model class $\mathscr{M}_1$ may be invoked. Correspondingly we extend all elements of the structure in Fig. 1. In the extended setting, we obtain a new estimated set of data generating processes $\Delta_1 = \chi_1(E_1(z))$. If now the hypothesis $F^a \in \Delta_1$ is not rejected on the basis of the applied backtests, we can quantify the misspecification risk according to (11) by

$$\phi_\rho(\Pi, d^{-1}(\Delta), d_1^{-1}(\Delta_1)). \tag{16}$$

VaR example (cont'd). When misspecification tests using the observations $z_1, \ldots, z_T$ reveal that the normality assumption is too strong, we can consider as an extension a model class $\mathscr{M}_1$ such that $d_1(\mathscr{M}_1) = \chi_1(\mathscr{G})$, where $\mathscr{G}$ is a subset of the set $D(\mathbb{R})$ of all non-decreasing right continuous functions $h : \mathbb{R} \to \mathbb{R}$ such that $h(-\infty) = 0$ and $h(\infty) = 1$. The choice of $\mathscr{G}$ may be motivated by particular required characteristics; for instance, in case one would like to exploit the first and second moment of Z_1, then it makes sense to let the set $\mathscr{G}$ consist of distribution functions with bounded first and second moments. Let $\chi_1(G) = G^T$ so that $\chi_1(\mathscr{G}) = \{G^T | G \in \mathscr{G} \subset D(\mathbb{R})\}$. We can estimate G consistently by means of the empirical distribution function based on the sample $z_1, \ldots, z_T$. Denote this nonparametric estimate by $\hat{G}$, so that $\hat{e}_1 = \hat{G}$. To obtain a confidence set $\hat{E}_1$, we may take a confidence region of level α that can be constructed in the following way:

$$\hat{E}_1 = \left\{ G \in \mathscr{G} | G(x) \in \left[\hat{G}(x) \pm \frac{k_{\alpha/2}}{\sqrt{n}} \right] \ \forall x \in \mathbb{R} \right\},$$

where $k_{\alpha/2}$ is the critical value of the Kolmogorov–Smirnov statistic.[7]

Just as in the case of estimation risk, a simplified approximate procedure may be used to determine misspecification risk if sufficient smoothness is present to apply the functional delta method.

[7] 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)).

VaR example (cont'd). It is straightforward to construct a confidence interval directly around the following nonparametric estimate of (5)

$$g_1(\hat{e}_1) = X_T(1-\exp(\hat{G}^{-1}(p)) = X_T(1-\exp(z_{T(\lfloor pn\rfloor+1)}))),$$

where $T(t)$ denotes the tth order statistic of $(z_1,\ldots,z_T)'$ 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 $g_1(\hat{e}_1)$ may be computed as

$$g_1(\hat{e}_1) + z_{\alpha/2}X_T\exp(\hat{G}^{-1}(p))\sqrt{\frac{p(1-p)}{T(G'(G^{-1}(p)))^2}}$$

where α is the level of the confidence interval and where $G\prime(x)$, the derivative of G evaluated at x, can be estimated using, for instance, the Rosenblatt–Parzen kernel estimator.

4.4. Identification risk

For purposes of risk assessment it may not be sufficient to apply only tests based on past and present data. After all, risk management is concerned with the *future*. Therefore, additional tests may be applied which we shall refer to as *stress tests*. Such tests serve in particular to capture the risk associated with changes of a type that is not represented in past data. For instance, one may consider the effects of changes in certain parameters which have been constant in the period that is covered by the available data (relating to central bank policy, for instance), or one may look at the impact of rare events which have not occurred in the data period. To accommodate such effects, it may be necessary to extend the model class further, beyond what may have been needed to arrive at models that are not rejected by backtests. The extended model set then typically contains parameters that are not identified on the basis of the available data, such as a parameter which expresses the influence of a factor that has been constant in the data period, so that in this connection we speak of *identification risk*.

Associated with the model class $\mathscr{M}_2$, we assume the presence of all elements indicated in Fig. 1. In particular, an estimated set $\hat{E}_2$ of characteristics is constructed which typically contains $\hat{E}_1$ and contains characteristics that are not distinguishable from those in $\hat{E}_1$ on the basis of the available data. Correspondingly, a tolerance set $\mathscr{K}_2 = d_2^{-1}(\chi_2(\hat{E}_2))$ can be constructed. A quantitative measure of identification risk is then given by

$$\phi_\rho\Big(\Pi, d_1^{-1}(\chi_1(\hat{E}_1)), d_2^{-1}(\chi_2(\hat{E}_2))\Big). \tag{17}$$

It is often possible to describe the relation between the model class $\mathscr{M}_2$ relating to stress tests and the model class $\mathscr{M}_1$ relating to backtests by means of parameters a which can vary within a set A in the model class $\mathscr{M}_2$ but which are fixed, say at value $a_1 \in A$, in the model class $\mathscr{M}_1$. The estimation procedure applied in the model $\mathscr{M}_2$ can then be constructed as a parametrized version of the estimation procedure applied within $\mathscr{M}_1$. In particular, the identification risk can then be represented as

$$\sup_{a\in A}\mathrm{RISK}_{\rho,m(a)}(\Pi) - \mathrm{RISK}_{\rho,m(a_1)}(\Pi). \tag{18}$$

VaR example (cont'd). In case the i.i.d. assumption cannot be rejected, 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, which cannot be inferred from the past and present data up to time T. An extended model class may be constructed for instance as in Section 2.

By definition, identification risk cannot be quantified on the basis of information extracted from the data by means of econometric procedures. Instead, its quantification must take place on the basis of the risk manager's insight in the functioning of a particular market. In the empirical applications that will be discussed below, we will focus on estimation and misspecification risk.

4.5. Total model risk

Given the quantification of estimation, misspecification, and identification risk, we can quantify the total model risk. We assume that we start with a (low-dimensional) model class $\mathscr{M}$ which allows the construction of a point estimate as well as a confidence interval. From this we obtain a quantification of estimation risk from (14) or (15). It may happen that the model estimated within the class $\mathscr{M}$ does not pass backtests, and in this case we assume that extension to a larger class $\mathscr{M}_1$ is possible so that a model can be constructed that is no longer rejected by the applied backtests. We can then quantify the misspecification risk using (16) or the alternative procedure based on the functional delta method. Finally, for purposes of risk management it may be deemed necessary to use an even larger model class $\mathscr{M}_2$ which may not be identified on the basis of the available data. The identification risk is quantified on the basis of (17) or (18). Summing the estimation, misspecification, and identification risks, we get the *total model risk*. Adding to the total model risk the market risk, as quantified on the basis of $\hat{e}$ by $\mathrm{RISK}_{\rho,d^{-1}(\chi(\hat{e}))}(\Pi)$, we get the *total market risk*, i.e., $\mathrm{RISK}_{\rho,\mathscr{K}}(\Pi)$ with $\mathscr{K} = d_2^{-1}(\Delta_2)$.

5. Empirical applications

In this section we demonstrate how the methodology of the previous sections can be used in portfolio risk management. We use data sets obtained from time series of the Standard and Poor's 500 equity index and the £/$ exchange rate. As already mentioned above, since our analysis in this section is based on return data, we do not consider identification risk.

The Bank for International Settlements (BIS) has suggested 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, as already illustrated in Section 2 on the basis of simulated data. First, we present the data and our basic model class $\mathscr{M}$. Then we discuss the selection of an appropriate extended model class $\mathscr{M}_1$. Finally, we present the model risk results in terms of multiplication factors.

5.1. Data and basic model class

The data set that we work with has been obtained from Thomson Datastream.[8] The periods covered by our data set are 26-10-'81–12-01-'06 in the case of the S&P 500 market and 03-01-'86–12-01-'06 in the case of the £/$ exchange rate. Fig. 2 shows the normal density with variance equal to the sample variances of the S&P 500 data and the British pound/ US dollar (£/$) exchange rate data and compares this to a nonparametric density[9] estimate of the densities of the S&P 500 and the £/$ 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.

[8] For the S&P 500 series we use the total return series from Thomson Datastream code: S&PCOMP(RI). The £/$ exchange rate 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).

[9] We use the Rosenblatt–Parzen nonparametric kernel estimate, with 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 data.

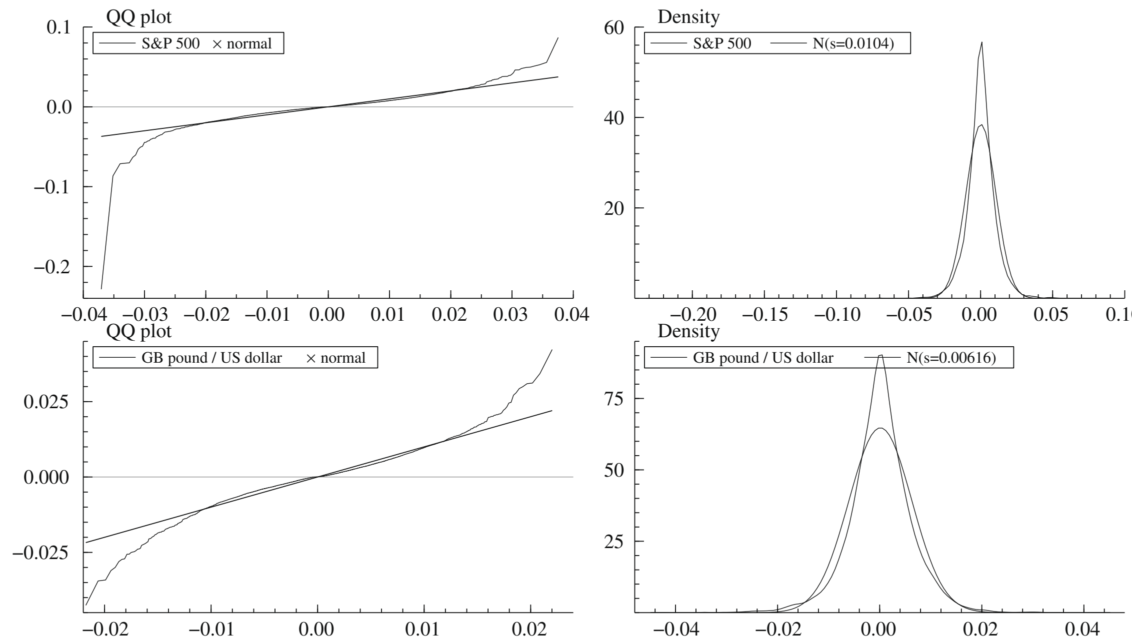

Fig. 2. QQ plot and density comparison of the normal density with nonparametric density estimate (using the Rosenblatt–Parzen kernel estimator with Gaussian kernel and bandwidth $h = 1.06sn^{-1/5}$) of the daily (total) returns of the S&P 500 and £/$ exchange rate. The data periods are 26-10-'81–12-01-'06 (S&P 500) and 03-01-'86–12-01-'06 (£/$ exchange rate).

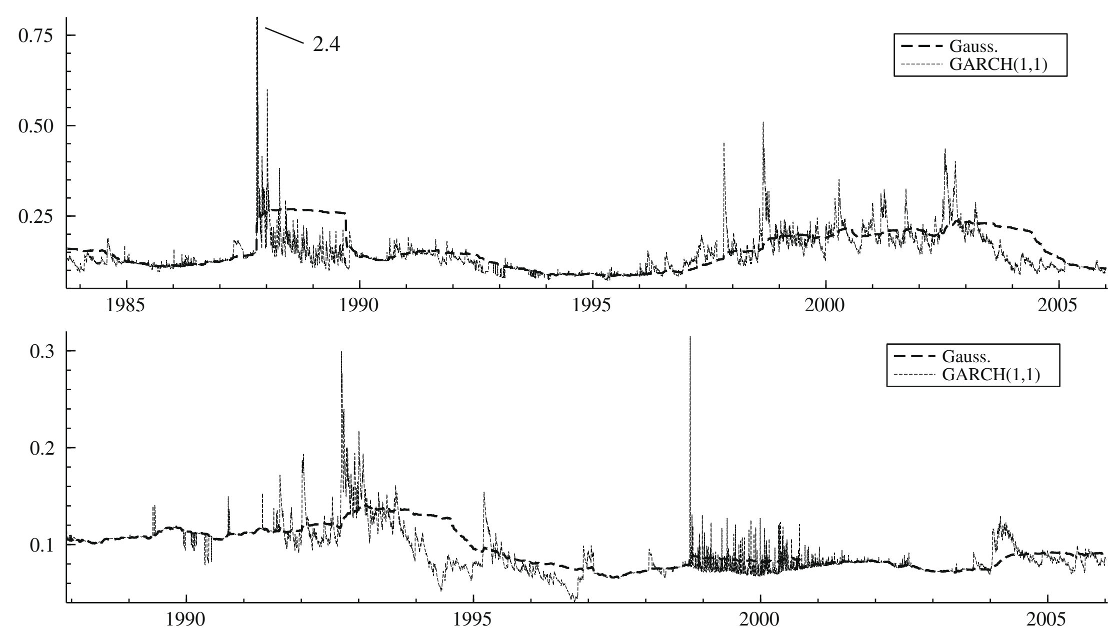

Fig. 3. The upper panel displays the parameter estimates of the volatility in the S&P 500 market for both the Gaussian and the GARCH(1,1) model (with $\mu = 0$). The lower panel displays the parameter estimates of the volatility in the £/$ FX rate market for both the Gaussian and the GARCH(1,1) model (with $\mu = 0$). For both markets the estimates use two-year rolling window models. The data sets run from 26-10-'81 to 12-01-'06 in case of the S&P 500 market and from 03-01-'86 to 12-01-'06 in case of the £/$ exchange rate.

In the subsequent analysis we use a rolling window approach with a window of two years in order to apply 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 obtain an estimate of nominal daily risk. This yields a time series of estimated risk measures of length equal to the length of the available time series minus the first two years. This time series is used in the statistical investigation.

As basic model class $\mathscr{M}$ we use the Gaussian i.i.d. model class (1), with $\mu = 0$.[10] Fig. 3 shows the resulting estimates (corresponding to $\widehat{e}$) for the (annualized) volatilities,[11] on the basis of which the

[10] Estimating μ as well yields somewhat less stable parameter estimates. However, in terms of the model risk analysis the outcomes with μ estimated and $\mu = 0$ are almost the same.

[11] The other estimates presented in this figure will be discussed in the next subsection.

Table 2
Coverage tests VaR for S&P 500.

Model	VaR level (%)	FOEL (%)	1-Sided 95% CI	p-Value FOEL	p-Value Chris.
Gauss. Mkt risk	2.5	2.9	(2.5%; –)	0.03	0.00
Gauss. Est. risk	2.5	2.3	(2.0%; –)	0.82	0.00
GARCH(1,1) Missp. risk	2.5	3.0	(2.7%; –)	0.01	0.00
Nonpar Missp. risk	2.5	1.9	(1.6%; –)	1.00	0.00
Gauss. Mkt risk	1	1.8	(1.5%; –)	0.00	0.00
Gauss. Est. risk	1	1.4	(1.2%; –)	0.00	0.00
GARCH(1,1) Missp. risk	1	1.8	(1.5%; –)	0.00	0.00
Nonpar Missp. risk	1	0.9	(0.6%; –)	0.91	0.70

The table shows the results of FOEL and Christoffersen (1998) (Chris.) tests for VaR market (Mkt) risk according to the Gaussian model, and VaR market risk including the corresponding estimation risk components. See main text for definitions. Daily data on S&P 500 (total return) index from 26-10-'81 to 12-01-'06.

market risk (i.e., $\textsc{risk}_{\rho,d^{-1}(\chi(\widehat{e}))}(\Pi)$) can be quantified, with $\textsc{risk}$ equal to VaR or ES. Taking into account the corresponding estimation inaccuracy, we quantify the corresponding estimation risk via (15).

5.2. Model class selection

In addition to $\mathscr{M}$, the Gaussian i.i.d. model class (1), we consider two possible extensions to deal with potential model misspecification. The first one, $\mathscr{M}_1^{\text{np}}$, is the nonparametric extension (4). The second one, $\mathscr{M}_1^{\text{G}}$, is a GARCH(1,1) model class with Gaussian innovations, which has more potential to capture time-varying risk.[12] In case of the nonparametric model class $\mathscr{M}_1^{\text{np}}$, we proceed as described in Subsection 4.3, where we use the Rosenblatt–Parzen kernel estimate (with normal kernel) as nonparametric estimate.[13] In case of the GARCH(1,1) model class $\mathscr{M}_1^{\text{G}}$, we proceed as in case of the basic model class $\mathscr{M}$. Fig. 3 also shows the estimates for the (annualized) volatilities according to the GARCH(1,1) models. The predictions of the estimated GARCH(1,1) models are, as could be expected, much more erratic than those of the Gaussian models. In case the basic Gaussian model class $\mathscr{M}$ is rejected, but the model class $\mathscr{M}_1^{\text{np}}$ or $\mathscr{M}_1^{\text{G}}$ turns out to be an appropriate extension, we quantify the corresponding misspecification risk, using (16),[14] also taking into account the corresponding estimation risk. We investigate the selection of an appropriate model class $\mathscr{M}$, with corresponding selected Δ, focusing in particular on the case where backtests are applied based on the rolling window 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 consider four cases: Gaussian i.i.d. without and with estimation risk, nonparametric i.i.d. with estimation risk, and GARCH(1, 1) with estimation risk. 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. Thus, we consider the case where backtesting is done based on the FOEL. 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)). For purposes of comparison, we also include the 2.5% level. Denote by T the number of days in the backtesting period, i.e., the sample period with exception of the first two years, by f the number of times the VaR level has been exceeded, and by $1-p$ the 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

$$F=\sqrt{T}\,\frac{f/T-p}{p(1-p)}. \tag{19}$$

The FOEL test is an *unconditional* coverage test: it compares the *unconditional* predicted frequency according to the model with the empirical frequency observed in the data. However, as discussed by Kupiec (1995), the FOEL test might not be very powerful. When the return distribution changes over time, for instance, from periods of low volatility to periods of high volatility, and *vice versa*, a model that does not take this into account might still have correct unconditional coverage, while at any given time it may have an incorrect *conditional* coverage. Christoffersen (1998) proposes a test for correct conditional coverage, testing both for the appropriate distributional assumptions *and* absence of clustering. We also use this conditional coverage test in our model class selection. The FOEL test can be seen as testing for the appropriate distributional assumption, while the Christoffersen (1998) test combines correct distributional assumptions with absence of clustering.

In Tables 2 and 3 we present the results of the two coverage tests with 95% confidence intervals. The results for the basic Gaussian model without estimation risk indicate a failure to pass the backtesting requirement in terms of appropriate distributional assumptions, both in case of the S&P 500 data and in case of the £/$ data. Taking estimation risk into account (at the usual 95% confidence level) seems sufficient for both the S&P 500 and the £/$ exchange rate to pass the FOEL test, but only at the 2.5% level. However, the Christoffersen (1998) test reveals a failure of appropriate conditional coverage. Thus, only taking account of estimation risk in the Gaussian model class seems insufficient for passing the investigated backtest requirements.

As a consequence of this outcome, we include misspecification risk by passing from the Gaussian i.i.d. model to the GARCH(1,1) and nonparametric i.i.d. model classes (including estimation risk at the 95% confidence level). The GARCH(1,1) model class is not sufficient to prevent the VaR limit from being exceeded too often unconditionally. In fact, it performs even worse than the basic Gaussian model class: the extra flexibility of the GARCH(1,1) model class does not seem to be helpful in achieving better out-of-sample VaR predictions. On the other hand, in case of the nonparametric i.i.d. model class the number of times the 1% VaR limit is crossed (conditionally or unconditionally) does not exceed the required level in a statistically significant way, as follows from the FOEL and Christoffersen tests. However, in case of the S&P 500 at the 2.5% level, clustering of excessive losses results in a rejection on the basis of the Christoffersen (1998) test. This suggests that the

[12] Berkowitz and O'Brien (2002) find that an ARMA(1,1)–GARCH(1,1) model with Gaussian innovations does a good job in forecasting Value-at-Risk for their portfolios of investment banks. Since we do not find any statistically significant ARMA structure in our data, we restrict the model to the GARCH(1,1) type. Just like in the basic model class we set the mean equal to 0, yielding more stable estimates than when the means are estimated as well, but without affecting the outcomes of the risk analysis. We also tried other GARCH models classes, like the HYGARCH model class, which is able to capture long memory. The results in terms of our risk analysis are quite similar to the GARCH models. Therefore, we only report the results in terms of the GARCH model class. For alternative (advanced) volatility estimation methods see, for example, Eberlein et al. (2003).

[13] The bandwidth h has been set equal to $h=1.06\hat{\sigma}n^{-1/5}$, with $\hat{\sigma}$ the estimated standard deviation of the returns or exchange rates in the corresponding rolling window sample. This is the optimal bandwidth selection for normally distributed data.

[14] Or the analogue of (15), with $\widehat{e}_1$ replacing $\widehat{e}$, and $\widehat{E}_1$ replacing $\widehat{E}$.

Table 3
Coverage tests VaR for £/ $ FX rate.

Model	VaR level (%)	FOEL (%)	1-Sided 95% CI	*p*-Value FOEL	*p*-Value Chris.
Gauss. Mkt risk	2.5	3.4	(3.0%; –)	0.00	0.00
Gauss. Est. risk	2.5	2.9	(2.5%; –)	0.07	0.03
GARCH(1,1) Missp. risk	2.5	3.3	(2.8%; –)	0.00	0.00
Nonpar Missp. risk	2.5	1.6	(1.3%; –)	1.00	0.49
Gauss. Mkt risk	1	1.9	(1.5%; –)	0.00	0.00
Gauss. Est. risk	1	1.4	(1.1%; –)	0.02	0.00
GARCH(1,1) Missp. risk	1	1.9	(1.6%; –)	0.00	0.00
Nonpar Missp. risk	1	0.6	(0.4%; –)	1.00	1.00

The table shows the results of FOEL and Christoffersen (1998) (Chris.) tests for VaR market (Mkt) risk according to the Gaussian model, and VaR market risk including the corresponding estimation risk components. See main text for definitions. Daily data on £/$ from 03-01-'86 to 12-01-'06.

Table 4
ES tests.

Market	Model	ES level (%)	*F*	*p*-Value	ES model rejected
S&P 500	Gauss.	5	−10.0	0.00	Yes
	GARCH(1,1)	5	−10.9	0.00	Yes
	Nonpar	5	0.5	0.69	No
	Gauss.	2.5	−16.4	0.00	Yes
	GARCH(1,1)	2.5	−16.1	0.00	Yes
	Nonpar	2.5	0.4	0.66	No
£/$	Gauss.	5	−7.5	0.00	Yes
	GARCH(1,1)	5	−7.7	0.00	Yes
	Nonpar	5	1.0	0.84	No
	Gauss.	2.5	−9.6	0.00	Yes
	GARCH(1,1)	2.5	−10.3	0.00	Yes
	Nonpar	2.5	1.5	0.94	No

Test of expected shortfall for the Gaussian and GARCH(1,1) and the nonparametric 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 12-01-'06. Daily data on the £/$ FX rate index from 03-01-'86 to 12-01-'06.

nonparametric i.i.d. model class suffices at low VaR levels (like 1%), but might fail at higher levels (like 2.5%). To conclude, we reject the basic Gaussian i.i.d. model class $\mathscr{M}$, but accept the nonparametric i.i.d. model class $\mathscr{M}_1^{\mathrm{np}}$ as an appropriate extension, at least as far as the VaR risk measurement method at the 1% benchmark level is concerned. However, at the higher 2.5% VaR level a further extension of the model class, also allowing for non-i.i.d. behavior, would be needed.

Next, we consider expected shortfall. In this case we apply the backtest for expected shortfall proposed by Kerkhof and Melenberg (2004). This test uses the functional delta method and compares an estimate of the risk measure under consideration (in our case ES) obtained from the test sample to an estimate of the same risk measure obtained from a reference sample. This test already accounts for estimation risk. For ES we adopt higher levels 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.[15]

The results presented in Table 4[16] indicate that both the rolling-window Gaussian model class and the GARCH(1,1) model class, both with estimation risk (at the usual 95% confidence level) included, are strongly rejected, while the rolling-window nonparametric model class, with estimation risk included, cannot be rejected for both series. Thus, similar to the VaR case, we reject the basic model class $\mathscr{M}$, but if we include misspecification risk by means of the nonparametric i.i.d. model class $\mathscr{M}_1^{\mathrm{np}}$, also including the corresponding estimation risk at the 95% confidence level, the computed ES levels pass the considered backtest requirement, so that we accept the model class $\mathscr{M}_1^{\mathrm{np}}$.

5.3. Model risk via multiplication factors

The relation between on the one hand the misspecification risk (with corresponding estimation risk included) and on the other hand the nominal (Gaussian i.i.d.) market risk may be represented in terms of a multiplication factor.[17] The rolling-window estimation procedure described above generates point estimates as well as confidence intervals in the three model classes that we consider. We define for each day the nonparametric multiplication factor for VaR at level p as the quotient $\mathrm{VaR}^{\mathrm{np}} + 0.95_p/\mathrm{VaR}_p^{\mathrm{N}}$, where $\mathrm{VaR}^{\mathrm{np}} + 0.95_p$ denotes the upper bound of the 95% confidence interval in the nonparametric i.i.d. model class, and $\mathrm{VaR}_p^{\mathrm{N}}$ denotes the VaR at level p computed on the basis of the estimated nominal Gaussian i.i.d. model. In this way, the multiplication factor takes into account both estimation risk and misspecification risk. Plots of the multiplication factors for VaR at level 1% are shown in the left panels of Fig. 4 for the two markets that we consider; corresponding plots for ES at level 2.5% are shown as well. These panels show that the estimation and misspecification risks at the 95% confidence level are covered comfortably during the full sample period by fixed multiplication factors

[15] We refer to Kerkhof and Melenberg (2004) for a detailed description of the test, which also includes a transformation of the return series to a standardized return series using a probability integral transform. The rolling window approach that we use here can be accommodated within the setting of the same reference, as noted in the cited paper.

[16] For the nonparametric case, we apply the method of Chen (2008), but we simplify by assuming independence over time.

[17] Since both VaR and ES are not necessarily positive it may be more appropriate in principle to use differences rather than ratios, as we did in Section 4. To take position size into account, differences might be expressed as a percentage of portfolio value. The use of multiplication factors applied directly to a risk measure is standard in BIS regulation, however, and therefore we phrase the results in these terms.

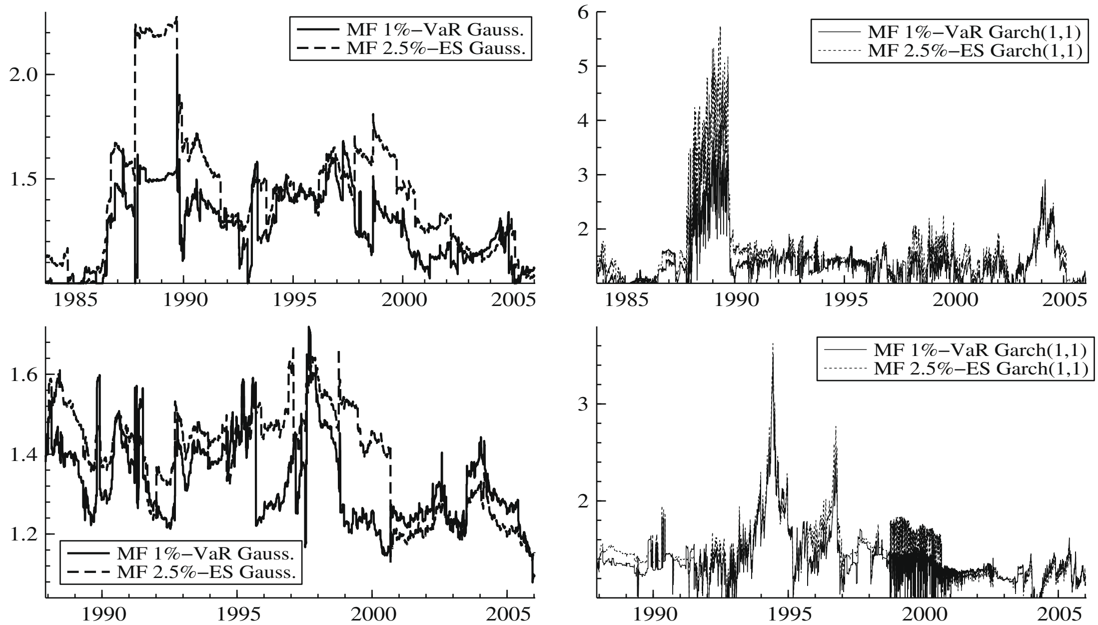

Fig. 4. The upper panels display estimation and misspecification risk multiplication factors (on the vertical axis) of the 1% VaR and 2.5% ES for the S&P 500 index during the period 26-10-'83–12-01-'06 (left is Gaussian and right GARCH(1,1)). The lower panels display estimation and misspecification risk multiplication factors of the 1% VaR and 2.5% ES for the £/ $ FX rate during the period 03-01-'88–12-01-'06 (Gaussian on the left, GARCH(1,1) on the right).

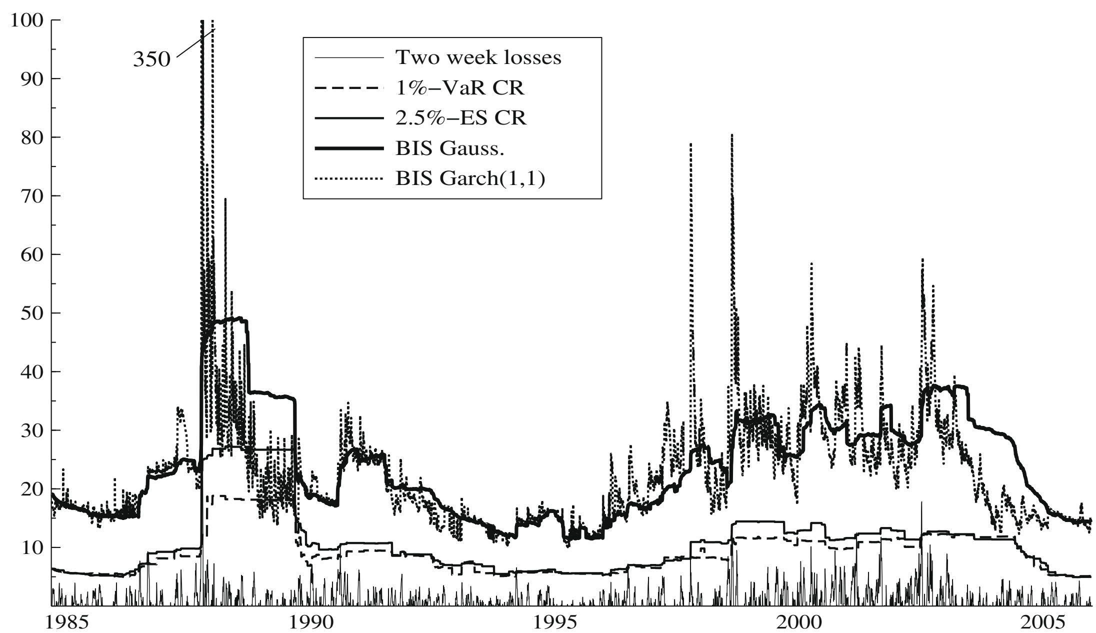

Fig. 5. The figure compares two-week losses on the S&P 500 to the capital requirements (on the vertical axis) for an investment in the S&P 500. Given are the capital requirements using the BIS regulation and the capital requirements based on 1% VaR and 2.5% ES model risk multiplication factors. The graph has been truncated: in the GARCH(1,1) model, the BIS capital requirement reaches 350 at the time of the 1987 stock market crash.

at the levels 2.3 (S&P 500) and 1.6 (£/$ exchange rate). In the case of the 2.5% ES we find that multiplication factors 1.7 (S&P 500) and 1.5 (£/$ exchange rate) are sufficient for the Gaussian model class. These levels are lower than the BIS multiplication factor which is normally 3, but of course the BIS factor must also cover other sources of risk, possibly including identification risk.[18] The analysis of the previous subsection revealed that the GARCH(1,1)-model class, although being an extension of the basic Gaussian model class, in terms of risk measurement actually seemed to perform worse than the Gaussian model class. To provide some further understanding, we plot in the right panels of Fig. 4 GARCH multiplication factors, defined as $\mathrm{VaR}^{np}+0.95_p/\mathrm{VaR}_p^G$ where VaR_p^G denotes the VaR at level p computed on the basis of the estimated nominal GARCH(1,1) model. GARCH multiplication factors for ES are defined similarly. The right panels indicate that for the GARCH(1,1) model class the model risk

[18] The value 3 of the multiplication factor may also be obtained from estimation risk at a confidence level of about 99.99%. This interpretation may however not be suitable in light of the emphasis usually placed on standard confidence levels.

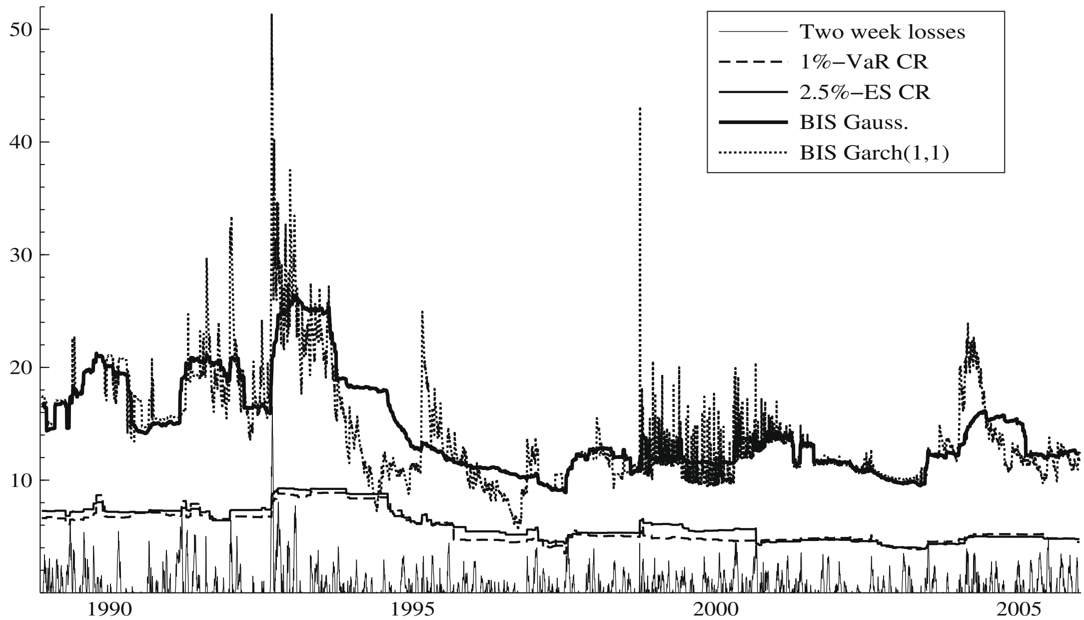

Fig. 6. The figure compares two-week losses on the £/$ FX rate to the capital requirements (on the vertical axis) are given for a firm trading in the £/$ FX rate. Shown are the capital requirements using the BIS regulation and the capital requirements based on 1% VaR and 2.5% ES model risk multiplication factors.

Table 5
Capital requirement schemes.

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
	MRMF VaR	2.25	1.48	9.1
	MRMF ES	1.30	0.94	10.7
£/$	BIS Gauss.	0.00	0.00	14.5
	MRMF VaR	1.51	0.84	5.8
	MRMF ES	1.29	0.73	6.1

The 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 VaR misspecification risk multiplication (MRMF) factor based CR (with estimation risk included), and the ES misspecification risk multiplication factor based CR (with estimation risk included). Data sets: S&P 500 (26-10-'83–12-01-'06) and £/$ FX rate (03-01-'88–12-01-'06).

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 much more quickly to periods of low volatility and then forecasts low values of VaR and ES, contrary to the nonparametric estimates. The relative stability of the multiplication factor for the i.i.d. Gaussian models may be interpreted as an indication of better robustness properties of this model class compared to the GARCH(1,1) models.

Figs. 5 and 6 show the capital requirements based on the BIS capital requirements. All requirements are based on investments of $100 in the market; results can therefore be interpreted as percentages. We have used the BIS backtest procedure (see Basel Committee on Banking Supervision (1996b)) to backtest the Gaussian model (and also the GARCH(1,1) model for comparison) and to determine the multiplication factors according to the BIS rules.[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 Fig. 4. In Figs. 5 and 6 we see that under normal market conditions the model reserves based on the model risk measures, including the misspecification risk with corresponding estimation risk, cover the losses safely. These figures also show again that the capital requirements for the nominal GARCH(1,1) model class (without estimation risk) are much more variable than those of the Gaussian model class.

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 5. Table 5 shows that the number of exceedances of the two-week VaR and ES is very small for all capital requirement schemes. Of course, the capital requirements set by the BIS are exceeded least, but they are also quite large compared to the estimation and misspecification

[19] Multiplication factors are adapted on the basis of backtesting performance. Such adaptation is done on a quarterly basis. To avoid possible random effects related to the timing of the three-month periods, we have actually implemented it on a daily basis. 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.

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.

6. Conclusions

In this paper we have presented a method to adjust capital requirements for trading activities in a market, based on the extent to which this market can be reliably modeled. The framework extends standard frameworks for the determination of market risk by considering risk measurement methods for a class of models rather than for one particular model. This allows the quantification of model risk on top of nominal market risk.

We focus in particular on model risk associated to uncertainties in econometric modeling. This leads in a natural way to the construction of total model risk from three components: estimation risk, misspecification risk, and identification risk. These distinctions are related to different levels of ambiguity which are reflected by model classes that are typically used in econometric modeling of financial markets.

We illustrate how backtesting may be used to decide whether only estimation risk has to be included, or whether, in addition, also misspecification risk needs to be taken into account. Our empirical results suggest that, for commonly used model classes, namely an i.i.d. Gaussian and a GARCH(1, 1) model class, misspecification risk cannot be ignored. Satisfactory results were obtained from a nonparametric i.i.d. model by including estimation risk at a standard confidence level. However, in case of the S&P 500 these satisfactory results only apply to the 1% VaR level. For the higher 2.5% VaR level the assumption of i.i.d. returns seems to be too strong, requiring as topic of future research an extension of the nonparametric approach allowing for non-i.i.d. returns. We also have found higher volatility of multiplication factors in the GARCH models, which may be indicative of a certain lack of robustness of these models. Moreover, the analysis indicates that, in the markets we have investigated, the multiplication factor set by the BIS is conservative by a factor of about two, if it would only be intended to cover estimation and misspecification risk. Of course it has to be taken into account that other risk factors play a role as well, and that in our empirical applications we have only considered well established markets.

Concluding, the framework presented here allows regulators to differentiate their capital requirements on the basis of the extent to which a market can be reliably modeled on the basis of current technology. Depending on the performance of available models for market risk assessment, model risk reserves can be determined. The empirical analysis in this paper was limited to investments of rather standard types. Further research would be needed to look into the role of model risk for more complicated products and in less liquid markets, where estimation and identification risk may play a larger role. We have only considered risk assessment; a further step is to investigate consequences for risk management, for instance in terms of robust hedging strategies.

References

Acerbi, C., Tasche, D., 2002. On the coherence of expected shortfall. Journal of Banking and Finance 26, 1487–1503.
Artzner, Ph., Delbaen, F., Eber, J.-M., 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.
Berkowitz, J., O'Brien, J., 2002. How accurate are value-at-risk models at commercial banks? Journal of Finance 57, 1093–1111.
Bongaerts, D., Charlier, E., 2009. Private equity and regulatory capital. Journal of Banking and Finance 33, 1211–1220.
Bossy, M., Gibson, R., Lhabitant, F.-S., Pistre, N., Talay, D., Zheng, Z., 2000. Volatility model risk measurement and strategies against worst case volatilities. Journal de la Société Française de Statistique 141, 73–86.
Brock, W.A., Durlauf, S.N., West, K.D., 2003. Policy analysis in uncertain economic environments (with discussion). Brookings Papers on Economic Activity 2003/1, pp. 235–322.
Brock, W.A., Durlauf, S.N., West, K.D., 2007. Model uncertainty and policy evaluation: Some theory and empirics. Journal of Econometrics 136, 629–664.
Chen, S.X., 2008. Nonparametric estimation of expected shortfall. Journal of Financial Econometrics 6, 87–107.
Christoffersen, P.F., 1998. Evaluating interval forecasts. International Economic Review 39, 841–862.
Cont, R., 2006. Model uncertainty and its impact on the pricing of derivative instruments. Mathematical Finance 16, 519–547.
Demyanyk, Y., Van Hemert, O., forthcoming. Understanding the subprime mortgage crisis. Review of Financial Studies. doi:10.1093/rfs/hhp033.
Derman, E., 1996. Model risk. Risk 9, 34–37.
Eberlein, E., Kallsen, J., Kristen, J., 2003. Risk management based on stochastic volatility. Journal of Risk 5, 19–44.
El Karoui, N., Jeanblanc-Picqué, M., Shreve, S.E., 1998. Robustness of the Black–Scholes formula. Mathematical Finance 8, 93–126.
Föllmer, H., Schied, A., 2002. Convex measures of risk and trading constraints. Finance and Stochastics 6, 429–447.
Frittelli, M., Rossazza Gianin, E., 2002. Putting order in risk measures. Journal of Banking and Finance 26, 1473–1486.
Gibson, R., Lhabitant, F.-S., Pistre, N., Talay, D., 1999. Interest rate model risk: An overview. Journal of Risk 3, 37–62.
Grané, A., Veiga, H., 2008. Accurate minimum capital risk requirements: A comparison of several approaches. Journal of Banking and Finance 32, 2482–2492.
Hansen, L.P., Sargent, T.J., 2007. Robustness. Princeton University Press, Princeton, NJ.
Heckman, J., 2000. Causal parameters and policy analysis in economics: A twentieth century retrospective. The Quarterly Journal of Economics 15, 45–98.
Heid, F., 2007. The cyclical effects of the Basel II capital requirements. Journal of Banking and Finance 31, 3885–3990.
Hull, J., 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.
Kaplanski, G., Levy, H., 2007. Basels value-at-risk capital requirement regulation: An efficiency analysis. Journal of Banking and Finance 31, 1887–1906.
Kerkhof, J., Melenberg, B., 2004. Backtesting for risk-based regulatory capital. Journal of Banking and Finance 28, 1845–1865.
Kupiec, P.H., 1995. Techniques for verifying the accuracy of risk management models. Journal of Derivatives 3, 73–84.
Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics. John Wiley & Sons, New York.
Talay, D., Zheng, Z., 2002. Worst case model risk management. Finance and Stochastics 6, 517–537.
Van der Vaart, A.W., 1998. Asymptotic Statistics. Cambridge University Press, Cambridge.
West, K.D., 1996. Asymptotic inference about predictive ability. Econometrica 64, 1067–1084.