Härdle, Wolfgang Karl; Hlávka, Zdeněk; Stahl, Gerhard

Working Paper

On the appropriateness of inappropriate VaR models

SFB 649 Discussion Paper, No. 2006,003

Provided in Cooperation with:
Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

Suggested Citation: Härdle, Wolfgang Karl; Hlávka, Zdeněk; Stahl, Gerhard (2006) : On the appropriateness of inappropriate VaR models, SFB 649 Discussion Paper, No. 2006,003, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:
https://hdl.handle.net/10419/25086

SFB 649 Discussion Paper 2006-003

On the Appropriateness of Inappropriate VaR Models

Wolfgang Härdle*
Zdeněk Hlávka**
Gerhard Stahl***

* CASE - Center for Applied Statistics and Economics,
Humboldt-Universität zu Berlin, Germany
** Department of Statistics, Charles University in Prague,
Czech Republic
*** Bundesanstalt für Finanzdienstleistungsaufsicht, Germany

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

Allgemeines Statistisches Archiv 0,

On the appropriateness of inappropriate VaR models

By Wolfgang Härdle, Zdeněk Hlávka and Gerhard Stahl*

Zusammenfassung Die Berechnung des VaR führt zur Reduktion der Dimension des Raumes der Risikofaktoren. Die vorzunehmenden Vereinfachungen resultieren aus unterschiedlichen Beweggründen, z.B. technische Effizienz, Sachlogik der Ergebnisse und statistische Adäquanz des Modells. Im Kapitel 2 stellen wir drei gängige Mappingverfahren vor: das Marktindexmodell, das Hauptkomponentenmodell und das Modell mit gleichkorrelierten Risikofaktoren. Impulse für Methoden zum Vergleich dieser Modelle im Kapitel 3 kamen vor allem aus der Literatur zur Praxis der Beurteilung von Wetterprognosen (Murphy and Winkler 1992, Murphy 1997). Umfangreiche Überlegungen zu einer quantitativen Analyse werden im vierten Kapitel dieser Arbeit vorgestellt. Die empirische Analyse der DAX Daten wird abschließend mit XploRe durchgeführt.

Summary The Value-at-Risk calculation reduces the dimensionality of the risk factor space. The main reasons for such simplifications are, e.g., technical efficiency, the logic and statistical appropriateness of the model. In Chapter 2 we present three simple mappings: the mapping on the market index, the principal components model and the model with equally correlated risk factors. The comparison of these models in Chapter 3 is based on the literatere on the verification of weather forecasts (Murphy and Winkler 1992, Murphy 1997). Some considerations on the quantitative analysis are presented in the fourth chapter. In the last chapter, we present empirical analysis of the DAX data using XploRe.

Keywords: Value-at-Risk, market index model, principal components, random effects model, probability forecast. JEL C51, C52, G20.

1. Introduction

The well-known G.E.P. Box's remark *All models are wrong, but some are useful* describes one of the main problems arising in each application of Value-at-Risk models: the "useful model" should be as simple as possible. This requires appropriate selection of variables in order to minimize the bias and the variance of the model simultaneously.

Value-at-Risk (VaR), as a standard measure of market risk, predicts the possible loss of a certain portfolio at a specified level α. Mathematically, VaR is the quantile at level α of the random future portfolio value changes. The task of calculating VaR for high dimensional portfolios requires simple and fast algorithms. In the following, we will present some very simple common methods for VaR calculation and we will investigate their appropriateness.

Our desire to choose a simple model leads to dimension reduction techniques applied either in the space of risk-factors or the distributions of the corresponding random variables. In practice, the—from RiskMetrics originating—term *mapping* is used in both cases. The importance of mapping procedures is overwhelming. Many VaR papers focus on one-dimen-

Received: / Revised:

* We acknowledge the support of Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 649 "Economic Risk", MSM 0021620839 and MŠMT 1K04018

sional identically distributed random variables with an insufficient perspective towards the practically more relevant multivariate situation. The question of the interplay between dimensionality- and model reduction and VaR precision is not asked and, consequently, not answered satisfactorily.

Corresponding to the complicated character of the VaR models, the commonly applied simplifications have various motivations. The most important ones are: the technical efficiency, the interpretability of the results, and the statistical appropriateness of the model.

Most often, the superior technical performance of the implemented model is in the focus of the implementation. A lower dimensional model allowing simple and precise handling of the numerical calculations and, at the same time, describing the important dynamic factors, can serve this purpose very well. Futhermore, the choice of the risk-factors depends also on availability of the up-to-date quality approved input parameters such as, for example, market data. Many institutes use the data matrix provided by the RiskMetrics group and containing about 400 risk-factors.

Although the, by mapping simplified, system cannot overcome the inherent rank defficiencies, it can still, at least in a statistical sense, reflect the covariance structure. A carefully parametrized model reduces rank deficiency in the estimated covariance matrix (Davé and Stahl 1998). This is the case especially for the multicollinearity, i.e., the situation often encountered in practice with a very large number of variables observed during a small time window. The dimension reduction techniques remove the insignificant correlations and can be used for clear and concise summarizing reports, Reimers and Zerbs (1998).

The VaR models are applied not only as a measure of risk but also as a means of fulfilling the regulatory office requirements. Special care must be exhibited in the case of capital reserves for open positions since their size depends strongly on the model, Huschens (1998). In the case of Maximum Loss risk measures, Studer (1997) has shown that the dimension is nonlinearly related to the calculated risk.

The following example considers portfolios consisting of long- or short positions. The market value ν_s at time s of the portfolio Π_t is given as

$$\nu_s(\Lambda_t) = \Lambda_t^\top \mathbf{P}_s,$$

where the vector $\Lambda_t \in \mathbb{R}^n$ denotes the nominal volumes at time t and $\mathbf{P}_s$ denotes the corresponding vector of prices. A portfolio such as $\{VW, DC\}$ with $\Lambda_t = (100, 200)^\top$ and $p_s = (40, 60)^\top$ would lead to the value $\nu_s = 16000$ for $s =$ 16.6.2005. The stochastic models for the risk of the value changes for given price $\mathbf{P}_t = p_t$ at time t,

$$\nu_{t+h}(\Lambda_t) - \nu_t(\Lambda_t) = \Lambda_t^\top (\mathbf{P}_{t+h} - p_t) = (\Lambda_t p_t)^\top \frac{(\mathbf{P}_{t+h} - p_t)}{p_t} \tag{1}$$

$$= w_t^\top \mathbf{R}_{t+h}, \tag{2}$$

are not based on the price process, but instead on the vector $\mathbf{R}_t$ of the (discrete) returns. Note that both the multiplication and division by the

vector p_t on the right hand side of (1) are componentwise. The exposure vector, $w_t = (w_{1t}, \ldots, w_{nt})^\top$, is defined as a componentwise product of Λ_t and p_t. In the above example, the exposure is $w_t = (4000, 12000)^\top$.

The *conditional loss distribution*

$$\mathcal{L}(L_{t+h} \mid \mathcal{H}_t) \tag{3}$$

of the possible losses of the portfolio $L_{t+h} = -w_t^\top \mathbf{R}_{t+h}$ in a time horizon h, is the object of our interest. In the following, we will set $h = 1$ (one trading day). The conditioning information $\mathcal{H}_t$, mostly defined by a moving window of fixed length N, consists in the simplest cases only from the observed market prices.

Assuming that the distribution of $\mathbf{R}_{t+h}$ is known, we could assess the distribution of losses (3) via Monte Carlo simulations since the changes of the portfolio value are a known function of the vector

$$L_{t+h} = \Psi_t(\mathbf{R}_{t+h}). \tag{4}$$

In some situations, one can derive explicit solutions that lead to evaluation of an integral involving the loss distribution.

A portfolio is called linear if the function Ψ_t in (4) is linear. For linear portfolios, the returns $\mathbf{R_t}$ are often parametrized by elliptical distributions with

$$E(\mathbf{R}_t) = 0$$

(RiskMetrics 1996, Litterman and Winkelmann 1998). For definiton of elliptical distribution refer to Härdle and Simar (2003). In this situation, the predictive distributions (3) for linear portfolios can be evaluated analytically. The dynamics is mostly analyzed using (I)GARCH- or White Noise models, Gouriéroux (1997).

In a Gaussian model, e.g., a White Noise model with $N(0, \sigma_{it}^2)$ distributed innovations, the VaR of the i-th asset at level α is given by

$$VaR_i(\alpha, h, t) = w_{it} q_\alpha \sigma_{ith},$$

where $\sigma_{ith}^2 = E(R_{i,t+h}^2)$ and q_α denotes the $\alpha-$quantile of the standard Normal distribution. Denoting by

$$\mathbf{VaR} = (VaR_1, \cdots, VaR_n)^\top$$

the VaR-vector (at level α and horizon h), the VaR of a portfolio Π_t can be expressed as:

$$\mathrm{VaR}(\Pi_t) = \mathrm{VaR}(\Pi_t, \alpha, h) = \sqrt{\mathbf{VaR}^\top \; C_t \; \mathbf{VaR}} \tag{5}$$

where C_t is the *correlation matrix* of $\mathbf{R}_{t+h}$ corrected by the signs of our positions, i.e., $c_{ijt} = \rho_{ijt} \; \mathrm{sign}(w_{it}) \; \mathrm{sign}(w_{jt})$. The vector $\mathbf{VaR}$ provides insight

into possible concentrations into single risk factors. Hence, the expression (5) is to be preferred over the similar expression

$$VaR(\Pi_t) = \sqrt{w_t^\top \Sigma_t w_t}.$$

In the next section, we present three common mappings. The third section is devoted to the diagnostics. In the fourth section, we will present our considerations related to the quantitative analysis. The empirical analysis in the last section has been performed with XploRe.

2. Mappings

The mappings that are used in the practice, are mostly based in the space of the portfolio variables and depend on projections (principal components) or regression methods. A prominent example is the mapping of a portfolio onto the market index. Applications based on the simplification of the parameter space are in practice not very common. RiskMetrics (1996) provides an important example: the sparse modelling of IGARCH models using only one smoothing parameter for all risk factors.

A VaR model of a linear portfolio based on (5) would require estimation of altogether $n(n+1)/2$ parameter of the covariance matrix. The amount, n, of the portfolio variables is mostly too large to consider every variable as a risk factor in the VaR model of the general market risk. Considering the fact that realistic values of n are usually much larger than 5000 and, in certain cases, reach values as high as 500000, one tries to keep the number of risk factors as small as possible, i.e.,

$$\dim(\tilde{\mathbf{R}}) = d \ll n.$$

In the case of the RiskMetrics data matrix, the number of risk factors, d, is approximately 400. Models with $d > 5000$ are rarely used in practice.

In order to apply a mapping, we have to select a vector of risk factors:

$$\tilde{\mathbf{R}} = (\tilde{R}_1, \cdots, \tilde{R}_d)^\top. \tag{6}$$

The choice is usually done empirically. Formally, the mapping can be described using a function $\mathcal{M}$ as follows:

$$\mathcal{M} : \mathbf{R} \longrightarrow \mathcal{M}(\mathbf{R}) = \tilde{\mathbf{R}}.$$

For a given $\tilde{\mathbf{R}}$ we have to derive a function ψ such that the distribution of the variable $\tilde{L}_t = \psi(\tilde{\mathbf{R}})$ approximates the distribution of the portfolio losses $L_t = \Psi(\mathbf{R})$ as good as possible. The function ψ maps the exposures w_t implicitly on the risk factors. In the ideal situation, with similarly defined $\tilde{\mathcal{H}}_t$, we would have:

$$\mathcal{L}(L_{t+h} \mid \mathcal{H}_t) = \mathcal{L}(\tilde{L}_{t+h} \mid \tilde{\mathcal{H}}_t)$$

or

$$VaR(\Pi_t) = VaR(\tilde{\Pi}_t).$$

In the simplest case, $\mathcal{M}$ is an inclusion mapping or the risk factors $\tilde{R}_i$ are functions of $\mathbf{R}$. For example, $\tilde{\mathbf{R}}$ could contain a subset of the elements of $\mathbf{R}$ or a linear combination of $\mathbf{R}$.

The common mapping procedures are mostly based on one meta portfolio characterized by the set of all available instruments, without taking into account the characteristics of the specific portfolio, e.g., by using the w_t in the construction of the mapping. In the following, we explore the, by RiskMetrics (1996) publicized, index model, the mapping on synthetic indices (principal components), and we suggest a new method based on a simplified correlation matrix.

2.1. Regression approach: index model. Among all mappings, the market index model is most commonly applied in the practice, Jorion (1997). Cuthbertson (1996) provides an overview of the common variants of the CAPM model as well as the factor models, and he gives a critical assessment of the corresponding model assumptions. The use of the beta factors is based on the assumed relation between the returns of an asset and the returns of a corresponding index, e.g., the DAX. The risk of having an open position in a certain asset is explained using the risk of the index by mapping the position on the index.

Let R_{it} denote the return of the i-th asset and R_{mt} the return of the index. The index model is then specified by the following equations:

$$R_{it} = \beta_i R_{mt} + \varepsilon_{it}, \tag{7}$$

where $E(\varepsilon_{it}) = 0$, $E(\varepsilon_{it}^2) = \sigma_{\varepsilon,it}^2$, $E(\varepsilon_{it} R_{mt}) = 0$, and $E(\varepsilon_{it}\varepsilon_{jt}) = 0$.

These equations are valid for times $t = 1, \ldots, T$ and ε_{it} is called the residual (or idiosyncratic) variable. The above model specification defines seemingly unrelated regression equations. The consequence is that, in this case, the parameters do not have to be estimated jointly.

The variance of R_{it} can be decomposed as

$$\sigma_{it}^2 = \beta_i^2 \sigma_{mt}^2 + \sigma_{\varepsilon,it}^2 \tag{8}$$

and for the covariance we have

$$\sigma_{ijt}^2 = \beta_i \beta_j \sigma_{mt}^2. \tag{9}$$

Under the assumptions of the model (7) and using (8) and (9) we have for the variance of (3) that

$$\begin{aligned} w_t^\top \Sigma_t w_t &= \sum_{j,i=1}^{n} \sigma_{it}\sigma_{jt}\rho_{ijt} w_{it} w_{jt} \\ &= \sum_{j,i=1}^{n} \beta_i \beta_j \sigma_{mt}^2 w_{it} w_{jt} + \sum_{i=1}^{n} \sigma_{\varepsilon,it}^2 w_{it}^2 \\ &= \sigma_{mt}^2 w_t^\top \beta \beta^\top w_t + w_t^\top D_{\varepsilon_t} w_t, \end{aligned} \tag{10}$$

where $\beta = (\beta_1, \cdots, \beta_n)^\top$ denotes the vector of betas. The diagonal matrix $D_{\varepsilon_t} = \text{diag}(\sigma^2_{\varepsilon,1}, \cdots, \sigma^2_{\varepsilon,n})$ denotes the covariance matrix of the vector ε. Equation (10) motivates the following approximations of the covariance matrix Σ_t:

$$\Sigma_{\mathcal{D}} = \sigma^2_m \beta\beta^\top + D_\varepsilon \tag{11}$$
$$\Sigma_\beta = \sigma^2_m \beta\beta^\top. \tag{12}$$

The quality of the approximation strongly depends on the validity of (7). The first term in the sum in (10) allows us to interpret the approximation (12) as a mapping. In this case of a mapping on an index, equation (7) becomes:

$$\mathcal{M}(w_t) = \beta^\top w_t \quad \text{and} \quad \mathcal{M}(\mathbf{R}) = \{R_m\},$$
$$\sigma^2_{mt} w_t^\top \beta\beta^\top w_t + w_t^\top \mathcal{D}_{\varepsilon_t} w_t = \sigma^2_{mt} \mathcal{M}^2(w_t) + w_t^\top \mathcal{D}_{\varepsilon_t} w_t.$$

The advantages of this mapping are obvious: on one hand, we have to estimate only $n+1$, respectively $2n+1$ parameters ($\dim(\beta) = n$, $\dim(\varepsilon) = n$, and σ^2_m) of the matrix Σ_β or, respectively $\Sigma_{\mathcal{D}}$ (instead of $n(n+1)/2$ parameters of Σ_t) and, on the other hand, the use of the risk factor R_m leads to a substantial dimension reduction of the VaR model. These mappings use approximations of $VaR(\Pi_t)$ that follow from equations (11) and (12).

$$\begin{aligned} VaR_{\mathcal{D}}\{\mathcal{M}(\Pi_t)\} &= q_\alpha \sqrt{w_t^\top \Sigma_{\mathcal{D}} w_t} \\ &= \sqrt{\mathbf{VaR}_{\mathcal{M}}^\top \left(\mathbf{1}_{(n\times n)} + B^{-1} D_{\varepsilon_t} B^{-1}\right) \mathbf{VaR}_{\mathcal{M}}} \end{aligned} \tag{13}$$
$$VaR_\beta\{\mathcal{M}(\Pi_t)\} = q_\alpha \sqrt{w_t^\top \Sigma_\beta w_t} = \sqrt{\mathbf{VaR}_{\mathcal{M}}^\top \mathbf{1}_{(n\times n)} \mathbf{VaR}_{\mathcal{M}}} \tag{14}$$

where the i-th element of $\mathbf{VaR}_{\mathcal{M}}$ denotes the VaR of the i-th asset mapped on the index R_m:

$$\mathbf{VaR}_{\mathcal{M}} = VaR_m \beta,$$

B denotes the diagonal matrix $\text{diag}(\mid \beta_1 \mid, \cdots, \mid \beta_n \mid)$ and $\mathbf{1}_{(n\times n)}$ is an $(n \times n)$ matrix of ones. The relations (13) and (14) show clearly how the mappings (11) and (12) simplify the VaR calculation:

1. In (13), C in (5) is replaced by $\mathbf{1}_{(n\times n)} + B^{-1} D_{\varepsilon_t} B^{-1}$.
2. In (14), C in (5) is replaced by $\mathbf{1}_{(n\times n)}$.
3. In (13) and (14), $\mathbf{VaR}$ in (5) is replaced by $\mathbf{VaR}_{\mathcal{M}}$.

Obviously, the effects of the above simplifications vary and their consequences cannot be, in general, predicted. The square VaR of a linear portfolio is a quadratic form of w_t and Σ_t:

$$VaR^2 = w_t^\top \Sigma_t w_t = \mathbf{VaR}^\top C_t \mathbf{VaR}.$$

Obviously,

$$\mathbf{VaR}_{\mathcal{M}}^\top \, C_t \, \mathbf{VaR}_{\mathcal{M}} \leq \mathbf{VaR}^\top \, C_t \, \mathbf{VaR} \leq \mathbf{VaR}^\top \, \mathbf{1}_{(n\times n)} \, \mathbf{VaR}.$$

The effect of a mapping cannot be, in general, fully determined even for portfolios consisting only from long positions. For portfolios, consisting from both long and short positions, is it completely impossible. The substitution of C_t by $\mathbf{1}_{(n\times n)}$ in (5) is conservative, whereas the substitution of $\mathbf{VaR}$ by $\mathbf{VaR}_{\mathcal{M}}$ tends to underestimate the actual risk. The elimination of long and short positions mapped on an index (quasi a stochastic netting) leads to problems that might arise when singular covariance matrices are used (Davé and Stahl 1998). So, $\text{rank}(\mathbf{1}_{(n\times n)}) = 1$ implies that there exists portfolio with exposure $\tilde{w}$ such that

$$\tilde{w}^\top \ \mathbf{1}_{(n\times n)} \ \tilde{w} = 0.$$

Thus the VaRs corresponding to $\tilde{w}$ are equal to zero. Hence, the empirical analyses assessing the validity of the beta factor models are not accessible.

2.2. MAPPING ON PRINCIPAL COMPONENTS. The application of principal components is a well known procedure (Jolliffe 1986, Christensen 1991, Härdle and Simar 2003). The orthogonalization of the variance matrix Σ of the vector $\mathbf{R}$ leads

$$\Sigma = \Gamma\Lambda\Gamma^\top$$

where $\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_n)$ is the diagonal matrix of the eigenvalues and Γ is the matrix consisting of eigenvectors of Σ. Under the usual regularity assumptions ($\lambda_i > 0$), $\mathbf{R}$ can be expressed as

$$\mathbf{R} = \Gamma Y.$$

The elements $(Y_1, \cdots, Y_n)$ of the vectors Y are uncorrelated and are called the principal components of $\mathbf{R}$. The total variability of $\mathbf{R}$ is equal to the trace of Σ:

$$\text{trace}(\Sigma) = \text{trace}(\Gamma\Lambda\Gamma^\top) = \sum_{i=1}^{n} \lambda_i.$$

Thus, the principal components with the largest variances (eigenvalues) explain most of the total variability of $\mathbf{R}$ and, in this sense, can be considered as the most important influence factors. Ordering the eigenvalues so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and denoting by $\Gamma_{(i)}$ the i-th column of the corresponding Γ, the approximation

$$\mathbf{R} \approx \sum_{i=1}^{k} \Gamma_{(i)} Y_i \tag{15}$$

with $k < n$ explains

$$\xi_k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \tag{16}$$

of the total variability of $\mathbf{R}$. Here, the risk factors $\tilde{\mathbf{R}} = \mathcal{M}(\mathbf{R})$ are the principal components $Y_1, \ldots, Y_k$. Typically, the dimension k is chosen as

the minimal number of the principal components that guarantees that $\xi_k > 1 - \xi^*$ where $0 < \xi^* < 1$. In a linear, normally distributed case, we have immediately for the Value-at-Risk of the mapped portfolio at level α that

$$VaR = \Phi^{-1}(\alpha)\sqrt{w_t^\top [\Gamma_{(1)} \cdots \Gamma_{(k)}] \Lambda [\Gamma_{(1)} \cdots \Gamma_{(k)}]^\top w_t}. \tag{17}$$

2.3. Mapping of the parameter space: Equally correlated risk factors. The mappings presented in this section simplify the structure of the parameter space of the stochastic model of the risk factors

$$\Sigma_t = \operatorname{diag}(\sigma_1, \ldots, \sigma_n) C_t \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$$

by the following restriction:

$$C_t = (\rho_{ijt})_{i,j=1,\cdots,n} \text{ with constant } \rho_{ijt} = \rho \in [-1, 1]. \tag{18}$$

Such models are currently applied in the area of credit risk, but similar ideas are also well-known in the portfolio theory, Elton and Gruber (1995).

The estimation of the restricted correlation matrix can be based on the following random effects model:

$$R_{it}/\sigma_i = \alpha_t + \varepsilon_{it},$$

where $E\alpha_t = E\varepsilon_{it} = ER_{it} = 0$, $ER_{it}^2 = \sigma_i^2$, $E\alpha_t^2 = \sigma_\alpha^2$, $E\varepsilon_{it}^2 = \sigma_\varepsilon^2$, $E\alpha_s\varepsilon_{it} = 0$. For the variance and covariance of the returns, we have the following:

$$\begin{aligned} Var(R_{it}) = \sigma_i^2(\sigma_\alpha^2 + \sigma_\varepsilon^2) &= \sigma_i^2 \\ Cov(R_{it}, R_{jt}) &= \sigma_i\sigma_j\sigma_\alpha^2. \end{aligned}$$

We define $R_{it}^* = R_{it}/\sigma_i$. A suitable estimator of σ_i^2 is, for example, the empirical variance of the returns R_{it}, $t = 1, \ldots, T$. Next, we define $\bar{R}_t^* = \sum_{i=1}^n R_{it}^*$ and $\bar{\varepsilon}_t = \sum_{i=1}^n \varepsilon_{it}$. Obviously,

$$R_{it}^* - \bar{R}_t^* = \varepsilon_{it} - \bar{\varepsilon}_t$$

and we apply the following estimator:

$$\begin{aligned} \hat{\sigma}_\varepsilon^2 &= \frac{\sum_{i=1}^n \sum_{t=1}^T (R_{it}^* - \bar{R}_t^*)^2}{n(T-1)}, \\ \hat{\sigma}_\alpha^2 &= 1 - \hat{\sigma}_\varepsilon^2. \end{aligned}$$

The standard deviation in the demonimator of R_{it}^* is estimated by the empirical standard deviation.

The correlation matrix of the returns is then approximated by

$$\hat{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_n) \begin{pmatrix} 1 & \sigma_\alpha^2 & \ldots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & 1 & \ldots & \sigma_\alpha^2 \\ & & \vdots & \\ \sigma_\alpha^2 & \ldots & \sigma_\alpha^2 & 1 \end{pmatrix} \operatorname{diag}(\sigma_1, \ldots, \sigma_n).$$

Comparing the advantages and disadvantages of the mappings presented in the section, it can be seen, that each of the mappings tries to reach, from its own perspective, a compromise between the precision of the VaR estimation, the simplicity of interpretation, the complexity with concern on the application of the procedure in practice, and the simplicity of its technical implementation.

3. Diagnostics

Value-at-Risk models are a special case of statistical prediction models, since the estimated distributions of the risk factors $\tilde{\mathbf{R}}_{t+h}$, describing our uncertainty about the future of the market, allow to access the conditional loss distribution

$$\mathcal{L}(L_{t+h} \mid \mathcal{H}_t)$$

in a horizon h. The empirical basis for judgements on the quality of the predictions of a VaR model is a series of pairs,

$$\{P_t, \nu_t\}_{t=1}^{N}, \tag{19}$$

consisting of the prediction, P_t, and the corresponding realization, ν_t, see Diebold, Gunther, and Tay (1998).

The shorthand notation P_t denotes either the complete estimated predictive distribution or some derived parameter (e.g., VaR, tail-VaR, standard deviation), the letter P stands for "prediction". The use of the symbol $\widehat{P}_t$ in some situation stresses the fact that the distribution we work with is estimated.

The symbol ν_t denotes the corresponding change of the value of the portfolio Π_t in a fixed time horizon, here fixed as one day, $h = 1$. We assume that the observations ν_t are realizations of random variables N_t. We distinguish between l_t and ν_t: the variable L_t refers to the space of the risk factors, i.e., $\mathcal{M}(\mathbf{R})$, whereas N_t refers to the space of the portfolio variables.

3.1. Tools.

The area of the probability forecasts is one of the less explored areas of the mathematical statistics. Substantial part of the theoretical foundations stem from Dawid—his summary papers (Dawid 1986, Dawid 1997) provide extensive overview of the topic. The motivations came from the literature on the evaluation of the weather forecasts. The first publication comes from 1884 (Gilbert 1884, Peirce 1884), the relevant literature begins with Brier's papers in fifties and advances steadily since sixties mainly through Murphy and Winkler (1992) and Murphy (1997).

The above mentioned applied papers focus on discrete random variables. The probability scale $[0, 1]$ is cut into k categories, further l denotes the number of the events, e.g., for weather forecasts $l = 2$ for $E = \{$rain, no rain$\}$. Let us assume that we have N probability forecasts $\{P_{ti}\}_{t=1}^{N}$ as well as the corresponding events $\{\nu_t\}_{t=1}^{N}$. Denoting by P_{ti}, $i = 1, \cdots, l$ the

probabilities predicted for t and by E_{ti} a zero-one variable which is equal to 1 if at time t we observe event from the category i ($\nu_t = i$) and 0 otherwise. Using this notation, the Brier score of a forecast system is defined as:

$$\mathrm{BS} = \frac{1}{N}\sum_{i=1}^{l}\sum_{t=1}^{N}(P_{ti} - E_{ti})^2. \tag{20}$$

In order to assess the quality of a forecast system in more detail, Murphy and Winkler (1992) start, in the context of discrete random variables, from equation (19), ignoring the information of the time dependency, by considering the joint relative frequencies

$$h(p_i, e_j) = \frac{N_{ij}}{N}, \tag{21}$$

where N_{ij} denotes the number of times when the event e_j, $j = 1, \ldots, l$ was predicted with the probability p_i, $i = 1, \ldots, k$.

The factorization of (21)

$$h(p, e) = h(e \mid p)\ h(p) \tag{22}$$

$$= h(p \mid e)\ h(e) \tag{23}$$

into conditional and marginal frequency distributions is the basis for definitions of additional indices and plots.

Important indices and plots. The Brier Score can be interpreted as a Mean Squared Error (MSE) because of its following representation:

$$\mathrm{BS} = \sum_{i=1}^{k}\sum_{j=1}^{l}(p_i - e_j)^2 h(p_i, e_j), \tag{24}$$

$$= (\mu_p - \mu_e)^2 + \sigma_p^2 + \sigma_e^2 - 2\sigma_p\sigma_e\rho_{pe}. \tag{25}$$

The Skill Score is defined as a coefficient of determination of the predictive probabilities seen as a fitted value in the regression model for the events. It can be calculated as a square of the correlation coefficient between the predictions P_{ti} and the zero-one variables E_{ti}.

Another evaluation of the quality of the forecasts results from the comparison of

$$p_i \quad \text{and} \quad \frac{1}{l}\sum_{j=1}^{l} N_{ij}. \tag{26}$$

In a perfect forecast system, the probabilities p_i would correspond to the observed relative frequencies. If this is the case, we say that the forecasts are calibrated. In the calibration plot, where $\frac{1}{N}\sum_{j=1}^{k} N_{ij}$ is plotted against p_i, the points should ideally lie on the diagonal of the unit square.

The forecast method can discriminate between the events if the conditional distributions $h(p|e = 0)$ and $h(p|e = 1)$ are significantly different.

Typically, the distribution of $h(p|e = 0)$ should be concentrated close to zero and the distribution of $h(p|e = 1)$ should lie close to 1. The more different the two distributions are, the better the forecast can discriminate between the two possible values of event e. Some common measures include difference in means or variances of the two distributions. However, most informative are discrimination plots that display the two distributions graphically.

The joint distribution $h(p, e)$ leads to many other measures of calibration, discrimination, refinement, resolution, bias or skill of the forecast procedures. From the many choices we have decided to use the Brier Score and the Skill Score, that can be interpreted as the MSE and the coefficient of determination, respectively (Murphy and Winkler 1992).

For appropriate VaR forecasts, the realizations of the variables $F_{L_t}(N_t)$ should not be distinguishable from independent draws from Uniform $[0, 1]$ distributions, where $F_{L_t}(x)$ denotes the distribution function of P_t. The P-P plot of the transformed observations

$$\{\hat{F}_{L_t}(\nu_t)\}_{t=1}^{N} \tag{27}$$

is called the (absolute) empirical calibration curve.

The weather and VaR forecasts usually differ only in one detail: in VaR models we focus on varying events (quantiles or forecast intervals) with fixed probabilities, by contrast in weather forecasts we obtain probability forecasts of fixed events.

The following indices and plots are motivated by their focus on the events, more precisely on the intervals $(-\infty, VaR]$. To this end, (19) has to be accordingly modified. For an interval forecast based on (5), the model is calibrated if the series:

$$\left\{ s_t^{\alpha} = \frac{\nu_t}{\sigma_t} \right\}_{t=1}^{N} \tag{28}$$

cannot be distinguished from a Gaussian White Noise process. Here, σ_t denotes an estimate of the scale of ν_t. The Quantile-Quantile plots (Q-Q plots) is a convenient graphical device. An important numerical parameter is here

$$\kappa = \hat{\sigma}_{S^{\alpha}} \tag{29}$$

that attempts to quantify the amount of under- and overestimation of the risk forecasted by the model. It is easy to see that in the ideal case, when σ_t reflects perfectly the standard deviation of the process ν_t, the parameter κ would be close to one. Small values of the parameter κ indicate that the model overestimates and large values of the parameter suggest that the model underestimates the true risk.

The empirical calibration curve, i.e., the P-P plot of the time series (27), is a diagnostic plot for checking whether the uniform $U[0, 1]$ distribution is fitting the marginal distributions of the process $\{F_{L_t}(N_t)\}_{t=1}^{N}$.

For the calibrated model, the points in the plot should concentrate close to the diagonal of the unit square. Furthermore, the symmetry properties of the plotted distributions can be easily evaluated. In the context of the

VaR models, the main disadvantage of this plot is its focus on the centre of the distribution (Wilk and Gnanadesikan 1968).

The application of a Q-Q plot as a diagnostics tool for calibration is more subjective since, in this situation, we do not have any fixed reference points (such as the diagonal of the unit square for the P-P plots). We can assess the symmetry and distribution of the process, but the calibration itself cannot be assessed without further specification of the model such as, e.g., (35). The advantage of Q-Q plots lies in its focus on the deviations in the tails of the distribution (Wilk and Gnanadesikan 1968).

The absolute and relative empirical calibration curves investigate only the appropriateness of the marginal distribution of the process L_t. They allow to verify the appropriateness of the choice of the risk factors, but do not allow to draw any conclusion on the temporal dependency structures. The heterogeneity- and independence properties are investigated by the means of time plots. The time series of the indicator function of the VaR exceedances at level α

$$\{I\{\nu_t > VaR_{t-1}(\alpha)\}\}_{t=1}^{N} \tag{30}$$

allows to discover clusters that indicate time dependency of the realizations s^α. Plotting the time series

$$\{\hat{P}_t, \nu_t\}_{t=1}^{N}, \tag{31}$$

the inhomogeneity can be discovered visually. χ^2 statistics allow to identify periods in which the forecasts were not independent.

4. Considerations on the quantitative analysis

The following empirical analysis compares the presented models on simulated portfolios by the means of the diagnostic tools, evaluating the forecast quality and presented in the previous section. We start with the description of the data set and the design of the simulations.

4.1. Design of the study.

The data set. The empirical basis of the following study is provided by the daily discrete returns

$$r_t = \frac{p_{t+1} - p_t}{p_t} \tag{32}$$

obtained from the following 18 German assets and the German market index DAX: Allianz, BASF, Bayer, BMW, Commerzbank, DAX, Deutsche Bank, Degussa-Huels, Dresdner Bank, Hoechst, Hypovereinsbank, Karstadt, Lufthansa, Linde, MAN, Mannesmann, Münchner Rück, Preussag, RWE.

Financial time series typically contain holes corresponding to the national holidays. It is common practice of the data providers to replace the missing values by the values from the previous day. This procedure results, according to the definition in (32), in zero returns. This happens to approximately 5% of the data. Thus, the data set could be described as a mix of two distributions, for example,

$$0.95\ N(0, \Sigma) + 0.05\ \delta_0,$$

where δ_0 is the distribution degenerated in zero. In order to get rid of the possible influence of this kind of model misspecification on the evaluation of the quality of forecasts, these values were removed from the time series.

The time series begins on the January 1-st, 1997 and ends on June 18-th, 1999.

SIMULATION OF THE PORTFOLIO. In order to simulate the exposures, w_t, as realistically as possible, we have set limits—similarly to the capacity load of VaR limits—controlling the behaviour of the simulations:

$$w_{t+1}^{\top} = \frac{w_t^{\top}}{\sigma_{\Pi_t}} \{600\ \mathrm{diag}(1/6 + U_1, \cdots, 1/6 + U_n)\}$$

where U_n is a n-dimensional Uniform distribution and σ_{Π_t} is the volatility of the portfolio. This way of simulating of the portfolio weights guarantees random changes and, at the same time, incorporates natural bounds on the volatility of the portfolio. Notice, that if the portfolio volatility increases, the values of w_t decreases in order to keep the VaR under a specified bound. Simple calculation shows that, if the volatility of the returns does not change, the volatility of this simulated portfolio with varying weights should lie close to 400.

The reasoning behind this approach is based on the fact that the VaR models are, aside to the risk measurement, applied also for risk control. The practical implementations are usually based on some limits on the VaR. It has been observed in practice that these limits exhibit constant capacity over the time within certain borders; this is the motivation and justification for the above assumption.

Due to the simulated portfolio weights that keep automatically the portfolio losses under control, we further assume that the value of the portfolio stays within the fixed interval:

$$N_t \in [-50000, 50000].$$

This interval is split into disjoint sets A_i:

$$[-50000, 50000] = \cup_{i=1}^{5} A_i.$$

In our study, the intervals:

$$A_1 = [-50000, -500), A_2 = [-500, -320), A_3 = [-320, 320),$$

$$A_4 = [320, 500) \text{ and } A_5 = [500, 50000]$$

are used for the evaluation the probability forecasts, i.e., the probabilities of the portfolio changes falling into these intervals are compared with the true events. The intervals A_1–A_5 were selected with respect to the simulated volatility so that the forecasted probabilities span the interval $(0, 1)$ as regularly as possible. This property is important for the methods for the verification of probability forecasts described in Section 5.2. The relative frequencies of the forecasting probabilities resulting from this choice of intervals A_1–A_5 are plotted in Figure 7.

The models. As shown in Sections 1 and 2, the predictive portfolio loss distribution

$$\mathcal{L}(L_{t+h} \mid \mathcal{H}_t)$$

using one class of models, here based on the variance-covariance structure, can lead to different results depending on the used mapping. The following empirical analysis evaluates and compares the forecasting quality of models M_1–M_6, defined as follows:

1. The model M_1 uses all 18 risk variables:
$$\mathcal{L}(\mathbf{R}_{t+1} \mid \mathcal{H}_t) = N_{18}(0, \Sigma_t). \tag{33}$$
The VaR calculation is based on (5).
2. The model M_2 is based on the classical beta factor mapping according to (12). The VaR calculation is based on formula (14).
3. The extended beta factor model, M_3, uses (11). The VaR calculation is based on (13).
4. The model M_4 is based on the principal components (15), the parameter ξ_k, in (16) was fixed at 80%. The VaR calculation is based on (17).
5. The model M_5 is a modification of M_4 such that the weighted returns $(w_{1t}R_{1t}, \cdots, w_{nt}R_{nt})$ instead of $\mathbf{R}_t$ are used for the principal components analysis.
6. The model M_6 uses the simple parametrization, as described in (18). All off-diagonal elements of the correlation matrix are identical, $\rho_{ij} = \rho$ for all $i \neq j$.

The evaluation of the risk is, for all considered models, based on the empirical covariance matrix calculated from the observed returns. After the simplifications provided by the respective models, the risk is evaluated for the simulated portfolios.

5. Empirical analysis

The main focus of our investigation is the evaluation of the quality of the forecasts provided by models M_1–M_6. The VaR forecasts of these models are based on the assumption of a multivariate Gaussian White Noise process as a stochastic model for the returns $\mathbf{R}_t$ of the risk factors. It is known

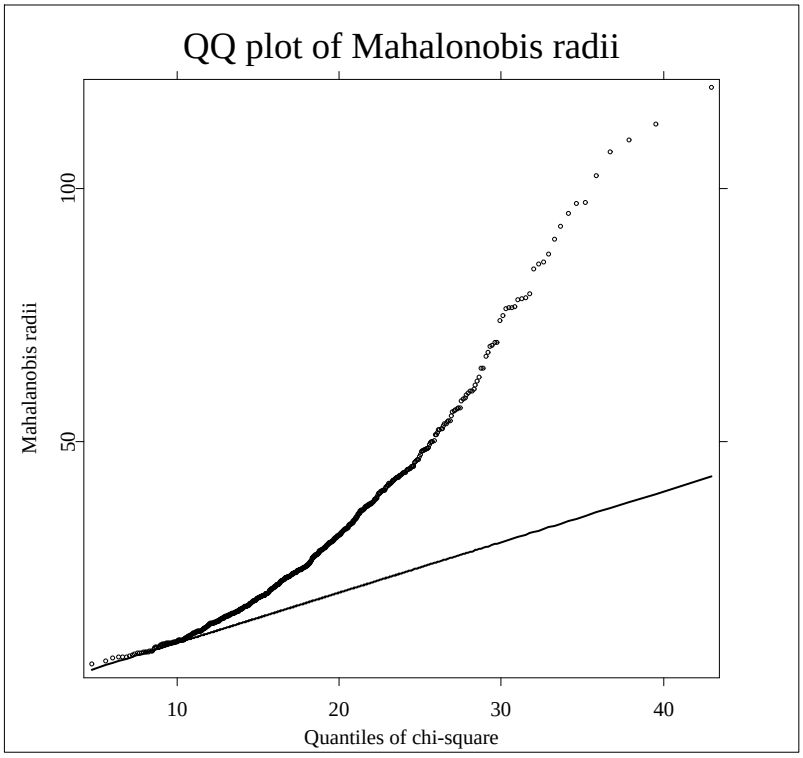

FIGURE 1. Q-Q plot of the squared Mahalanobis radii for the observed returns $\mathbf{r}_t$ of the risk factors. The straight line displays the expected values under normality.

that this process is not able to capture adequately some common characteristics of financial time series such as, e.g., fat tails, leptokurtosis, changes in volatility. Since the national (see, for example, section "Law/Gesetze" at www.bafin.de) as well as international regulatory norms (Basel Committee on Banking Supervision 2005) practically forbid the applications of exponantially weighted observations which enter in the applications of GARCH or IGARCH models, we did not consider these models in this study (Härdle and Stahl 2000). Hence, we do not present any analysis on how well does such stochastic model capture the dynamics of the process.

The verification of whether the observations $\{\mathbf{r_t}\}_{t=1}^{T}$ stem from the p-dimensional Normal $N_p(0, \Sigma)$ distribution, is carried out by means of a Q-Q plot. Assuming that the data are $N(0, \Sigma)$ distributed, we have:

$$\mathbf{r}_t^\top \Sigma^{-1} \mathbf{r}_t \sim \chi^2(p).$$

The sorted observed Mahalanobis radii $\mathbf{r}_t^\top \Sigma^{-1} \mathbf{r}_t$ are then compared with the quantiles of the $\chi^2(p)$ distribution in a Quantile-Quantile plot. The Q-Q plot displayed in Figure 1 deviates largely from the reference line. We conclude

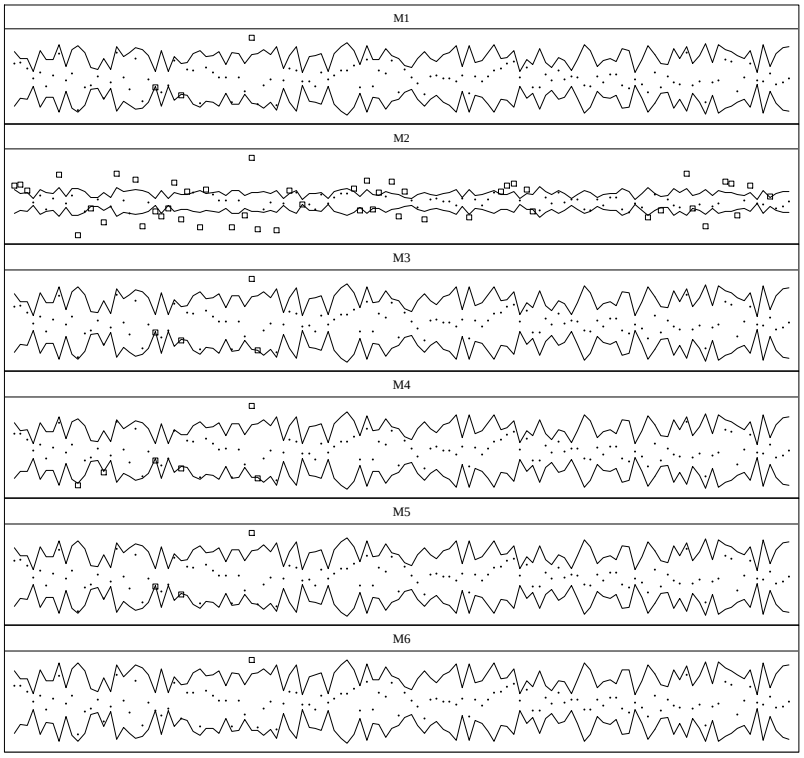

FIGURE 2. Time series of the changes of the value of the portfolio and the VaR forecasts for 249 trading days. The VaR forecasts at level 99% and 1% are plotted as lines. The value changes of the portfolio are plotted as dot if they lie between the two lines, the values of the portfolio falling outside the predicted VaR region are denoted by squares. The plots corresponding to models M_1–M_6 are displayed from top to bottom.

that the data hardly come from a multivariate Normal distribution. It is now of crucial importance whether this fact influences the quality of the forecasts or not.

Figure 2 shows the time series

$$\{P_t, \nu_t\}_{t=1}^{T}$$

of the VaR forecasts at level $\alpha = 99\%$ and $\alpha = 1\%$ and the value changes of the portfolio for $T = 249$ trading days for models M_1–M_6. Here, P_t denotes the VaR predictions and ν_t denotes the observed changes of the portfolio value. The exceedances, i.e., the value changes falling outside the predicted VaR bounds, are marked by squares. With the probability of exceedance set to 2%, we should on average observe 4.98 exceedances.

The forecasting quality of the model M_2 with more than 50 exceedances is obviously insufficient. The VaR bounds for models M_1, M_3, M_4, and M_5

are very similar and the number of exceedances varies between 3 and 6, reasonably close to the expected value 4.98. In our simulation, model M_6 seems to overestimate the risk, leading to only 1 exceedance.

The daily practice of VaR modelling has shown that exploratory diagnostic tools for assessing the quality of the forecasts of the VaR models are sufficient both for sub and full portfolios. Hence, in the following, we will omit the formal statistical inference presented in Davé and Stahl (1998). We will explore the qualities of the models by means of graphical and descriptive statistical tools.

	M_1	M_2	M_3	M_4	M_5	M_6
κ	1.07	2.32	1.06	1.13	1.07	1.07
Brier score	0.15	0.21	0.15	0.16	0.15	0.16
Skill score	0.35	0.15	0.35	0.34	0.35	0.35

TABLE 1. Parameter κ, Brier score and Skill score for the evaluation of the quality of models M_1–M_6. A good model should exhibit κ close to 1, small value of Brier score, and large value of the Skill score.

The graphical analysis of the forecast quality of the VaR models can be based, apart of the timeplot in Figure 2, on various plots of the empirical calibration curves (Dawid 1984). One approach is based on the standardization of the value, L_t, of the portfolio at time t. Since the random variable

$$S_t^\alpha = \frac{L_t}{VaR_{t-1}} \tag{34}$$

has a normal distribution, we obtain that asymptotically

$$\mathcal{L}(L_t/\widehat{VaR}_{t-1}) \approx N(0, 2.33^{-2}), \tag{35}$$

since $\widehat{VaR}_{t-1}$ is in the Gaussian model defined as $2.33\ \hat{\sigma}_{t-1}$. Note also that the inequality $S_t^\alpha < -1$ characterizes the exceedances of the VaR forecast that can be easily identified in the corresponding timeplots. The validity of the VaR model can be verified by checking whether $\{F_{L_t}(N_t)\}_{t=1}^T$ and $\{S_t^\alpha\}_{t=1}^T$ are White Noise processes.

In Figures 3–5, we plot the histograms, the P-P plots and the Q-Q plots for the variables s_t^α defined in (28). Note, that in the above Gaussian context, $2.33\ s_t^\alpha$ is equal to the above defined S_t^α. All histograms in Figure 3, apart of model M_2, look similar to the standard Normal distribution. The histograms for the remaining models, M_1 and M_3–M_6 look almost identical and we can say that all mappings give comparable results. The same conclusion can be derived from the P-P and Q-Q plots in Figures 4 and 5. Again, model M_2 gives much worse results than the remaining models who look again very similar and seem to satisfy our assumptions.

The parameter κ, defined in (29), is tabulated for the six models in Table 1. All models tend to underestimate the risk. Again, the model M_2

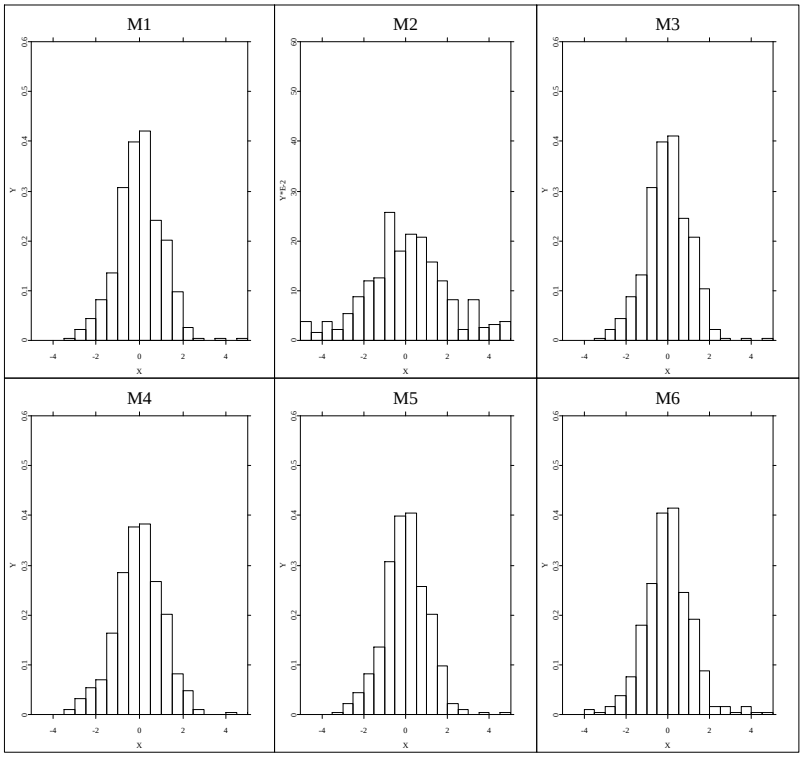

FIGURE 3. Histograms of the variable s_t^α, defined in (28), for models M_1–M_6.

gives unacceptable results. The level of the risk underestimation for the remaining models is much lower and close to 1.

5.1. EXCEEDANCES. The time series of the indicator function of the VaR exceedances (30) at level $\alpha = 80\%$ is plotted in Figure 6. Theoretically, the number of exceedances should lie close to 20%. Visual inspection of such time plots should point out potential time inhomogeneities. Figure 6 is complemented by Table 2 that lists the percentages of the exceedances for each model and for each quarter of the year separately. Most of the time, the percentages are moving rather close ($\pm 6\%$)to the expected 20%. The two exceptions are model M_2 and 1-st quarter of 1999. The bad results for model M_2 are consistent with the results of the previous analyses. The bad behaviour of all models in the beginning of 1999 indicates some model heterogeneities or volatility changes in this period.

5.2. VERIFICATION OF PROBABILITY FORECASTS. The probability forecasts, analyzed in this section, relate to the probabilities of the portfolios

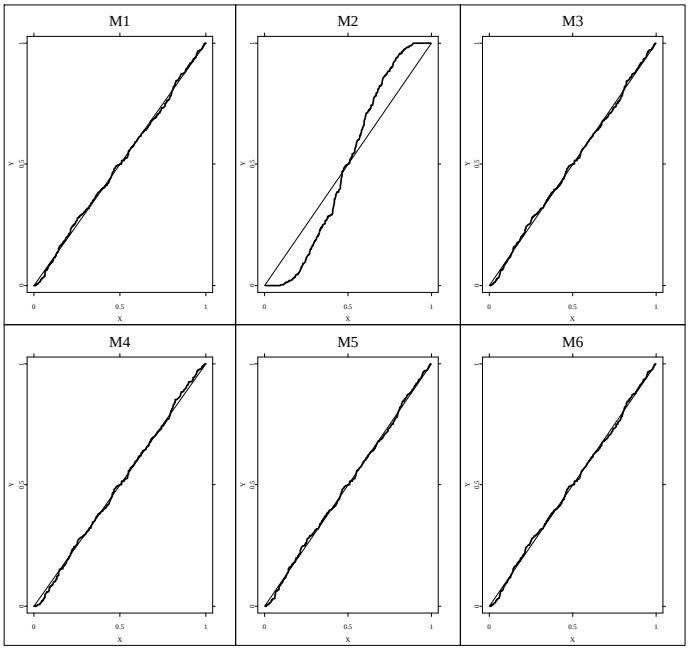

FIGURE 4. P-P plots for the variable s_t^α, defined in (28), for models M_1–M_6.

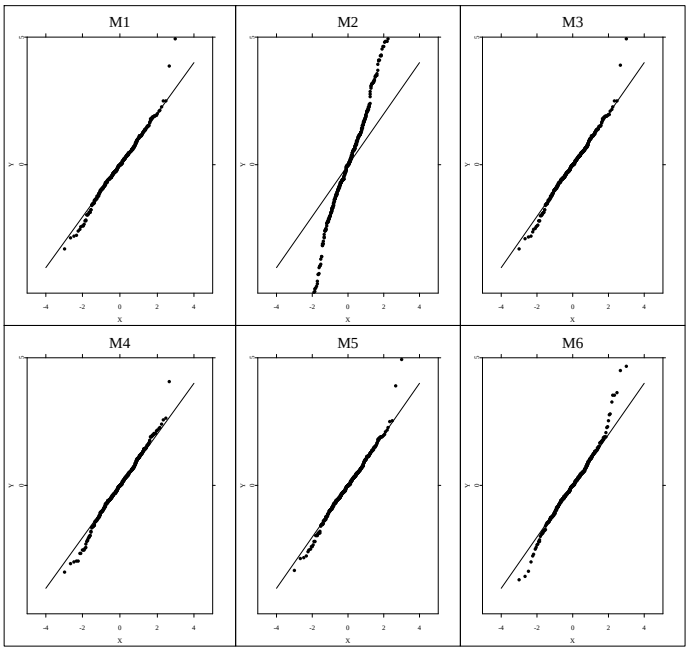

FIGURE 5. Q-Q plots for the variable s_t^α, defined in (28), for models M_1–M_6.

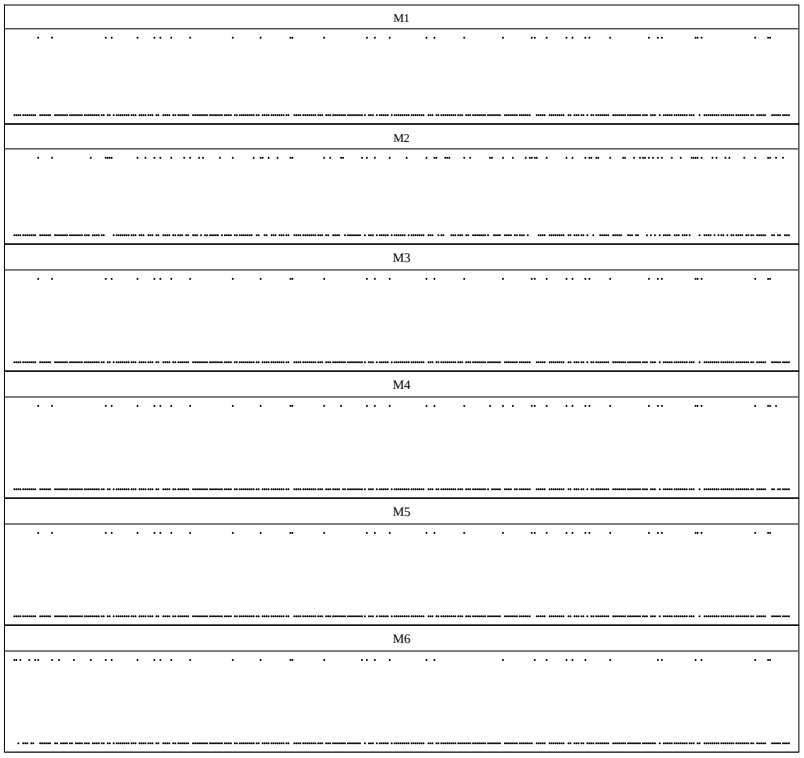

FIGURE 6. The time series of the VaR exceedances at level $\alpha = 0.8$ for models M_1–M_6.

changes falling into intervals A_1–A_5, described in Section 4.1. From the simulation, we have both the predicted probability that the change of the portfolio falls into one of these intervals and we also know the true realization.

The Brier score and Skill score for the models M_1–M_6 are given in the second and third rows of Table 1. From this point of view, we can say that the best model are the full model M_1 and models M_3 and M_5. Models M_4 and M_6 are just a bit worse. Model M_2 shows, once again, worst behaviour.

The relative frequencies of the forecasting probabilities are plotted in the first and in the third row of plots in Figure 7. The intervals A_1–A_5 were selected so that, in this simulations, the forecasting probabilities cover the interval $(0, 1)$ as uniformly as possible. We observe that only model M_2 behaves differently. In comparison to other models, its forecasting probabilities are much larger. Clearly, this is caused by its underestimation of the risk, observed already in the previous analyses, see Figure 2.

The second and the fourth row of plots in Figure 7 show calibration plots for all six models, i.e., each plots displays the forecasting probabilities on the x-axis and the corresponding relative frequencies of the success on the

	M_1	M_2	M_3	M_4	M_5	M_6
Q1 1998	0.14	0.22	0.14	0.14	0.14	0.14
	(1.01)	(0.40)	(1.01)	(1.01)	(1.01)	(1.01)
Q2 1998	0.17	0.37	0.17	0.17	0.17	0.17
	(0.52)	(2.97)	(0.52)	(0.52)	(0.52)	(0.52)
Q3 1998	0.26	0.52	0.24	0.26	0.26	0.26
	(1.05)	(5.73)	(0.78)	(1.05)	(1.05)	(1.05)
Q4 1998	0.14	0.44	0.14	0.14	0.14	0.14
	(1.01)	(4.34)	(1.01)	(1.01)	(1.01)	(1.01)
Q1 1999	0.39	0.70	0.39	0.39	0.39	0.39
	(3.21)	(8.85)	(3.21)	(3.21)	(3.21)	(3.21)
Q2 1999	0.17	0.63	0.17	0.17	0.17	0.17
	(0.43)	(7.00)	(0.40)	(0.43)	(0.43)	(0.43)
χ^2	13.9	188.0	13.4	13.9	13.9	13.9
p-value	0.020	0.000	0.026	0.020	0.020	0.020

TABLE 2. Relative frequencies of the exceedances of the VaR at level $\alpha = 80\%$ for each quart in our data set for models M_1–M_6. The contributions to the χ^2 statistic are given in the parentheses. The last row gives the p-values for the test of the hypothesis $H_0 : p = 0.2$ against the alternative $H_1 : p \neq 0.2$.

y-axis. Clearly, a perfectly calibrated model should lie very close to the diagonal of the unit square. In this case, all models (apart of M_2) provide well calibrated forecasts. Model M_2 underestimates the forecasting probabilities.

In Figure 8, we show the discrimination plots. One curve in each plot is the relative frequency of the forecasting probabilities conditioned on a success, while the other one conditiones on a failure. For good forecasts, one would like to predict success with high probability if it really occures and with low probability if it does not occur. Hence, for a good model, the two curves should be very far from each other. Again, we observe that the behaviour of model M_2 is worse than the behaviour of the other models.

5.3. CONCLUSION. The results of the empirical analyses suggest that all mappings, apart of M_2, lead to results comparable with the full model M_1. The model M_5 using the weighted principal component analysis gives, as expected, slightly better results that model M_4 that uses principal components method without taking the weights into consideration. Surprisingly, the model M_6 seems to give slightly better results than both models based on the principal components.

Our conclusion is that model M_2 should not be used in practice. The remaining models give comparable results. The comparison of the models was based on the indices given in Tables 1 and 2 whereas the graphical

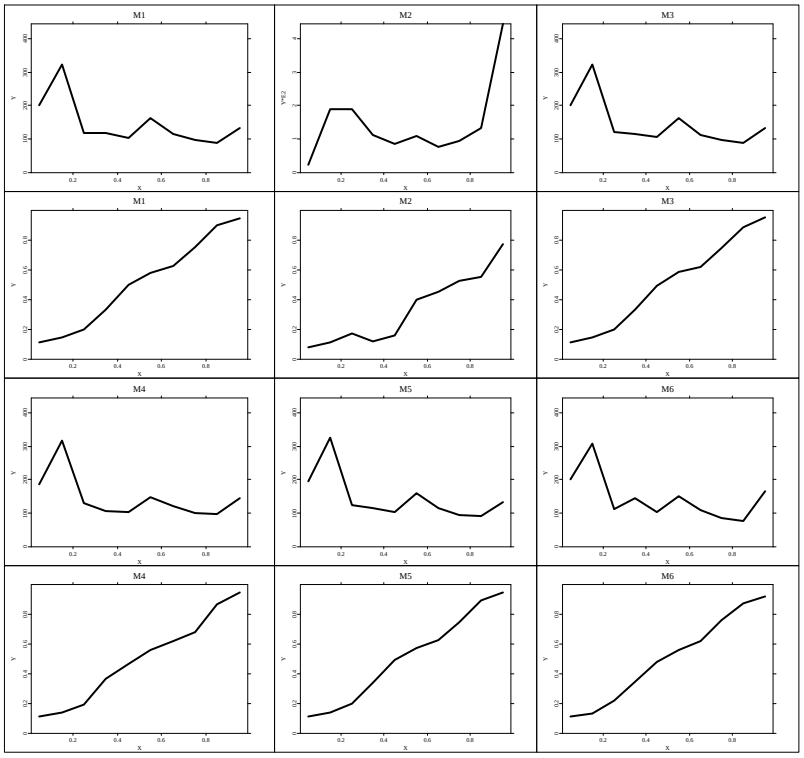

FIGURE 7. The relative frequencies of the forecast probabilities and the discrete calibration curves for models M_1–M_6.

methods (Figures 1–8) help to understand why one method is better than the other one.

REFERENCES

Basel Committee on Banking Supervision (2005). *International Convergence of Capital Measurement and Capital Standards, A Revised Framework (November 2005 Revision)*, Bank of International Settlements, Basel.

Chatfield, C. (1995). Model uncertainty, data mining and statistical inference, *Journal of the Royal Statistical Society, Series A* **158**: 419–466.

Chatfield, C. (1996). Model uncertainty and forecast accuracy, *Journal of Forecasting* **15**: 495-466.

Christensen, R. (1991). *Linear Models for Multivariate Time Series and Spatial Data*, Springer.

Cuthbertson, K. (1996). *Quantitative Financial Economics*, Wiley.

Davé, R. D. and G. Stahl (1998). On the Accuracy of VaR Estimates Based on the Variance-Covariance Approach, in: Bol, Nakhaeizadeh, Vollmer (Eds), *Risk Measurement, Econometrics and Neural Networks*, Heidelberg: Physica, S. 189–232.

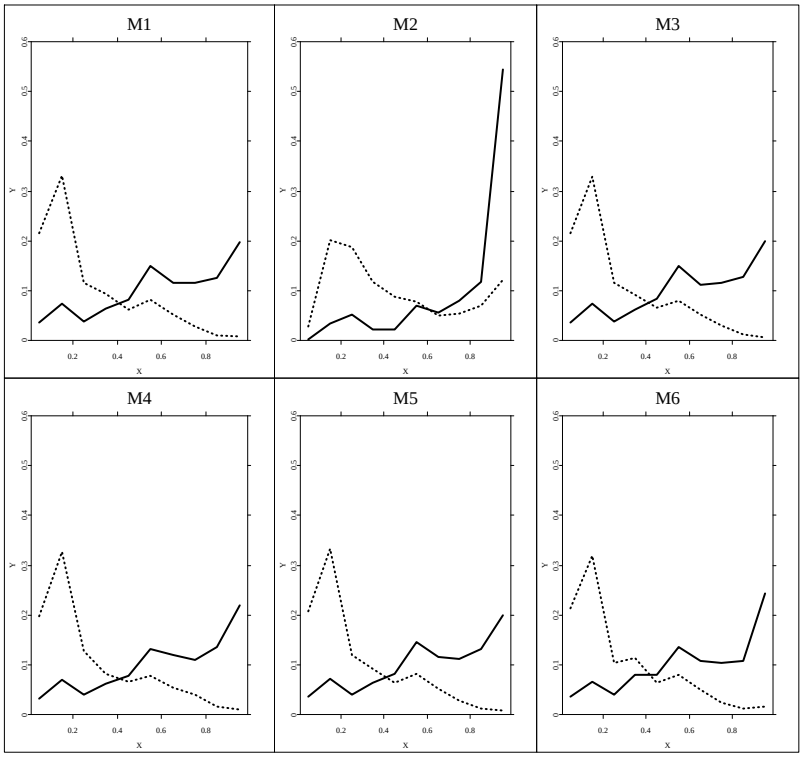

FIGURE 8. Discrimination curves for models M_1–M_6.

Dawid, A. P. (1984). The Prequential Approach, *Journal of the Royal Statistical Society, Series A* **147**: 278–292.

Dawid, A. P. (1986). Probability Forecasting. *Encyclopedia of Statistical Sciences* **7**: 210–218.

Dawid, A. P. (1999). Prequential Analysis. *Encyclopedia of Statistical Sciences* **Update 1**: 464–470

Diebold, F. X., Gunther, T. A., and Tay A. S. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review* **39/4**: 863–883.

Elton, E. J. and M. J. Gruber (1995). Modern Portfolio Theory and Investment Analysis, 5th Ed, Wiley, New York

Gilbert, G. F. (1884). Finley's tornado predictions. *Amer. Meteor. J.* **1**: 166–172.

Gouriéroux, C. (1997). *ARCH Models and Financial Applications*, Springer.

Härdle, W. and Stahl, G. (2000). Backtesting beyond VaR, *In Franke, Härdle, and Stahl (editors): Measuring Risk in Complex Stochastic Systems, Lecture Notes in Statistics 147*, Springer.

Härdle, W. and Simar, L. (2003). *Applied Multivariate Statistical Analysis*, Springer.

Hendry, D. F. (1995). *Dynamic Econometrics*, Oxford.

Hsiao, C. (1986). *Analysis of Panel Data*, Cambridge Univ. Press.

Huschens, S. (1998). Messung des besonderen Kursrisikos durch Varianzzerlegung, *Kredit und Kapital* **4**: 567–591.

Jolliffe, I. T. (1986). *Principal Component Analysis*, Springer.

Jorion, P. (1997). *Value At Risk*, IRWIN, Chicago.

J.P.Morgan/Reuters (1996). *RiskMetrics—Technical Document, Fourth Edition*, Morgan Guarantly Trust Company, New York.

Kreinin, A., Merkoulovitch, L., Rosen, D. and Zerbs, M. (1998). Principal Component Analysis in Quasi Monte Carlo Simulation, *ALGO Research Quarterly* **1/2**: 21–29.

Litterman, R. and Winkelmann,K. (1998). *Estimating Covariance Matrices*, Risk Management Series, Goldman Sachs.

Murphy, A. H. (1997). Forecast Verfication. In: *Economic Value of Weather and Climate Forecasts*, ed. by R. W. Katz and A. H. Murphy, Cambridge University Press

Murphy, A. H. and Winkler, R. L. (1992). Diagnostic verification of probability forecasts, *International Journal of Forecasting* **7**: 435–455.

Peirce, C. S. (1884). The numerical measure of the success of predictions. *Science* **4**: 453–454.

Reimers, M. and Zerbs, M. (1998). Dimension Reduction by Asset Blocks, *ALGO Research Quarterly* **1/2**: 43–55.

Ridder, T. (1998). Basics of Statistical VaR-Estimation, in: Bol, Nakhaeizadeh, Vollmer (Eds), *Risk Measurement, Econometrics and Neural Networks*, Heidelberg: Physica, S. 161–188.

Ridder, T. and Stahl, G. (2000). Flexibles oder starres Cash-Flow Mapping?

Stahl, G. and Traber, U. (2000). Backtesting in Action. In: *Kreditrisikomanagement*, Hrsg.: A. Oehler, Schäffer-Poeschl Verlag, Stuttgart.

Studer, G. (1997). *Maximum Loss for Measurement of Market Risk*, PhD Thesis, ETH Zürich.

Wilk, M. B. and R. Gnanadesikan (1968). Probability Plotting for the Analysis of Data, *Biometrika* **55**: 1–17.

Wolfgang Härdle
C.A.S.E.
Wirtschaftswiss. Fakultät
Institut für Statistik und Ökonometrie
Humboldt-Universität zu Berlin
Spandauer Str. 1, 10178 Berlin

Zdeněk Hlávka
Charles University in Prague
Faculty of Mathematics and Physics
Department of Statistics
Sokolovská 83, 18675 Prague

Gerhard Stahl
BaFin, Bundesanstalt für
Finanzdienstleistungsaufsicht
Graurheindorfer Str. 108, 53117 Bonn

SFB 649 Discussion Paper Series 2006

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Calibration Risk for Exotic Options" by Kai Detlefsen and Wolfgang K. Härdle, January 2006.
002 "Calibration Design of Implied Volatility Surfaces" by Kai Detlefsen and Wolfgang K. Härdle, January 2006.
003 "On the Appropriateness of Inappropriate VaR Models" by Wolfgang Härdle, Zdeněk Hlávka and Gerhard Stahl, January 2006.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".