On the Economic Costs of Value at Risk Forecasts

Tatiana Miazhynskaia
Engelbert J. Dockner
Georg Dorffner

Report No. 85
November, 2003
November, 2003

SFB
‘Adaptive Information Systems and Modelling in Economics and Management Science’

Vienna University of Economics and Business Administration
Augasse 2–6, 1090 Wien, Austria

in cooperation with
University of Vienna
Vienna University of Technology

http://www.wu-wien.ac.at/am

Papers published in this report series are preliminary versions of journal articles and not for quotations.

This piece of research was supported by the Austrian Science Foundation (FWF) under grant SFB#010 (‘Adaptive Information Systems and Modelling in Economics and Management Science’).
On the Economic Costs of Value at Risk Forecasts

Tatiana Miazhynskaia
Austrian Research Institute for Artificial Intelligence, Vienna, Austria
and Engelbert J. Dockner
Department of Business Studies, University of Vienna, Austria
and Georg Dorffner
Austrian Research Institute for Artificial Intelligence
and Department of Medical Cybernetics and Artificial Intelligence,
University of Vienna, Vienna, Austria

Abstract
We specify a class of non-linear and non-Gaussian models for which we estimate and forecast the conditional distributions with daily frequency. We use these forecasts to calculate VaR measures for three different equity markets (US, GB and Japan). These forecasts are evaluated on the basis of different statistical performance measures as well as on the basis of their economic costs that go along with the forecasted capital requirements. The results indicate that different performance measures generate different rankings of the models even within one financial market. We also find that for the three markets the improvement in the forecast by non-linear models over linear ones is negligible, while non-gaussian models significantly dominate the gaussian models.

JEL Classification: C32, C45, C52
Keywords: neural network models, Value-at-Risk, capital requirement

1 Introduction

Modelling and forecasting volatility of financial time series has been an important research topic for the last several years. There are two main reasons for the strong interest in volatility estimates. Since the prices of derivative products do depend on the volatility of the underlying instrument any pricing of these products requires volatility forecast. The second reason is related to the concept of volatility as a measure of market risk. Since the modern banking industry requires an efficient management of all risks in today’s new global financial architecture, heavy emphasis must be placed on financial market risks. As a consequence many regulatory requirements (e.g. those initiated by the Bank for International Settlements) are by now standardized and have introduced many novel concepts and tools into the management of market, credit and operational risk. In the case of market risk these developments have led to an uniformly accepted and applied risk measure called Value-at-Risk (VaR). The VaR of a portfolio position is defined as the maximum potential loss for this position for a given holding period and a given confidence level. Alternative specifications of financial products, increasing availability of financial data and rapid advances in
computer technology have led to the introduction and formulation of various VaR models that can currently be applied to measure the market risk of a portfolio analysis.

The VaR concept can be viewed as a generalization of the risk sensitivities related to different risk factors. As an example let us quickly look at the market risk of a simple European call option. If we ignore higher order approximations the option’s delta is the sensitivity of the call price with respect to the risk resulting from a change in the price of the underlying. Hence the delta linearly measures market risk. This measure, however, is incomplete as long as we do not know what the volatility of the risk factor is. If we multiply the sensitivity of the position with the volatility of the risk factor we end up with the VaR, a therefore a comprehensive measure of market risk. This simple description points out that the calculation of the VaR is directly related to forecasting volatility of a position. Only if we have full knowledge about the conditional density it is not necessary to express percentiles of distributions as multiples of the standard deviation. In that case we can directly calculate the value at risk. VaR models that are based on standard distributions (e.g. normal distribution) first estimate the standard deviation (or covariance matrix) in order to calculate the VaR for a given confidence level. For that reason good volatility forecasts are an integral part of sound VaR models.

One of the most widely used volatility models is the GARCH model (Bollerslev, 1986) for which the conditional variance is governed by a linear autoregressive process of past squared returns and variances. The standard GARCH model based on a normal distribution captures several ”stylized facts” of asset return series, like heteroskedasticity (time-dependent conditional variance), volatility clustering and excess kurtosis. Recent empirical research, however, has found that there are additional empirical regularities in return data such as negative and autocorrelated skewness (asymmetry), fat tails and time dependent kurtosis that can not be described by the classical GARCH model. For that reason several alternative specifications have been formulated in the literature.

We take into account the latest developments in conditional volatility research and propose a generalized model that extends the existing literature in two directions: the first one is to allow for non-linear dependencies in the conditional mean and variance, and the second one concerns a non-standard specification of the conditional density. To estimate nonlinear conditional second moments we use a neural network-based approach (i.e., so called recurrent mixture density networks) for which the conditional mean and variance are modelled by a multi-layer perceptrons (see, e.g., Schittenkopf et al. (2000)).

With regard to the specification of the conditional distributions, we compare three different density specifications: 1) a standard GARCH model and its non-linear generalization with a conditional normal distribution (heteroskedastic, but neither skewed nor leptokurtic); 2) a non-linear recurrent GARCH model with a Student’s t-distribution (heteroskedastic, not skewed but leptokurtic); and 3) linear and non-linear recurrent mixture density models, for which the conditional distributions are approximated by a mixture of gaussians (two components) (heteroskedastic, skewed and leptokurtic in a time-dependent manner).

These model specifications make clear that our point of interest in this study is twofold. On the one hand we are interested in forecasting volatilities in order to accurately estimate the value at risk of a portfolio. On the other hand we are concerned with the forecast of conditional distributions that allows the calculation of VaR directly. Based on these two objectives we empirically evaluate the forecasting performance of alternative volatility models and apply statistical tests to discriminate between alternative VaR models. For the latter we apply the Basle traffic light test, the proportion of failure test and interval tests. All these tests evaluate the accuracy of a VaR model on the basis of statistical procedures. Since it is very likely that the statistical criteria do not single out one model as the best, we alternatively calculate the costs of capital requirements as induced by a specific VaR model. The rationale behind this approach is the following. Assume that out of several competing models there are two that perform equally well with respect
to forecasting the value at risk of a portfolio position, i.e. these two models have two similar statistical characteristics. The two models, however, can lead to very different costs, as far as the capital requirements are concerned. Form a banks point of view it is not only necessary to have a risk management model that correctly predicts the market risk, but one that additionally uses the least capital possible. Since any capital requirement incurs opportunity costs for the bank (i.e. capital that is in an unproductive, regulatory use), it has an interest to cut this requirement down as much as possible. Hence, VaR models should not only be judged on the basis of their forecasting power, but also on the basis of their capital costs.

This discussion motivates the structure of our empirical analysis. It is based on return series of stock indices from three different financial markets. We use return series of the Dow Jones Industrial Average (USA), the FTSE 100 (Great Britain) and the NIKKEI 225 index (Japan) over a period of more than 13 years in order to evaluate in detail the out-of-sample predictive performance of our models. Our empirical analysis has the following structure. We predict conditional distributions and calculate the VaR for each of our models for three different homogeneous portfolios based on the same stock indices. To evaluate the quality and accuracy of the VaR models we apply a number of statistical tests specifically designed to interval forecasts. Among those are regulatory backtesting required as a part of the capital-adequacy framework (the Basle Committee’s traffic light); exceptions testing which examines the frequency with which losses greater than the VaR estimate are observed together with independence of these events; statistical test on the accuracy of point estimation of the VaR significance level. The advantage of these tests is given by the fact that the actual loss of any portfolio can be measured exactly and hence the VaR forecasts can be evaluated on the basis of actual observations.

As pointed out above, our central focus is also related to the analysis of the efficiency of VaR measures, as measured by the costs of capital associated with VaR based regulatory capital requirements (calculation of the lost interest yield connected with the dynamically computed model-based capital reserves).

The paper is organized as follows. In the next section we present the models we are working with. Section 3 discusses the data that is used in the empirical analysis. Section 4 empirically evaluates six alternative VaR models. Finally, Section 5 concludes the paper.

2 Description of Models

The usual approach for modeling return series is to split the returns into a predictable deterministic component \( \mu_t \) (mean) and a stochastic error process \( e_t \) with independent realizations and with \( \mathbb{E}(e_t | I_{t-1}) = 0, \mathbb{E}(e_t^2 | I_{t-1}) = \sigma_t^2 \), where \( I_{t-1} \) denotes series history up to time \( t-1 \). \( \sigma_t \) is an estimate of the volatility of the return series at time \( t \).

The most prominent model of time-varying volatility is GARCH\((p,q)\) (introduced in Bollerslev (1986)), where conditional variances are governed by a linear autoregressive process of past squared errors and variances, i.e.

\[
\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i e_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2, \tag{1}
\]

with the restrictions \( \alpha_0 > 0, \alpha_i \geq 0, \beta_i \geq 0 \) to ensure positive variances. Stationarity in variance imposes the condition \( \sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1 \).

In this paper we consider only the GARCH\((1,1)\) model which is \((1)\) with \( p = q = 1 \). The GARCH\((1,1)\) specification has proven attractive for models of returns. It typically dominates other GARCH models using Akaike or Schwarz information criteria (see Bollerslev et al. (1992)).

Due to the significant autocorrelation found in many return series, we chose the autoregressive process
of the order 1 for the mean equation, i.e.
\[ \mu_t = a_1 \mu_{t-1} + a_0 \]

One possible extension of the GARCH model is to substitute the conditional normal distribution by a Student’s-$t$ distribution with $\nu$ degrees of freedom in order to allow for excess kurtosis in the conditional distribution (see Bollerslev (1987) for details). The conditional variance is again given by the specification (1) and the parameter restrictions for stationarity of the model are the same as those for the GARCH model. Since the conditional density of $t$-distribution is symmetric, the conditional skewness is again 0. The new parameter degrees of freedom $\nu$ determines, among other characteristics, the kurtosis of the conditional distribution. For $\nu > 4$, the conditional kurtosis is given by $3(\nu - 2)/(\nu - 4)$ which is always larger than 3. Therefore, GARCH-$t$ models exhibit fat tails in the unconditional and in the conditional distribution. As for GARCH models, the higher-order moments of the distribution are not time-dependent.

The second direction of the extension of the classical GARCH model is to allow for non-linear dependencies in the conditional mean and in the conditional variance. As a tool for non-linear regression we used neural network-based modeling, so called recurrent mixture density networks, describing conditional mean and variance by multi-layer perceptrons (MLP) (Schittenkopf et al., 2000).

In the simplest case an MLP with one input unit, one layer of hidden units and one output unit realizes the mapping
\[ \tilde{f}(x_t) = g \left( \sum_{j=1}^{H} v_j h(w_j x_t + c_j) + s x_t + b \right), \]
where $H$ denotes the number of hidden units, $w_j$ and $v_j$ the weights of the first and second layer, $s$ the shortcut weight, and $c_j$ and $b$ the bias weights of the first and second layer. In general, the activation function $h$ of the hidden units is chosen to be bounded, non-linear, and increasing as, e.g., the hyperbolic tangent. The activation function of the output unit may be unrestricted, e.g. $g(x) = x$. Hornik et al. (1989) showed that MLP can approximate any smooth, non-linear function with arbitrary accuracy as the number of hidden units tends to infinity. In such a way, MLP can be interpreted as a non-linear autoregressive model of first order and can be applied to predict the parameters of conditional density of the return series.

Recurrent mixture density network models RMDN(n) approximate the conditional distributions of returns by a mixture of $n$ Gaussians:
\[ \rho(r_t | I_{t-1}) = \sum_{i=1}^{n} \pi_{i,t} k(\mu_{i,t}, \sigma_{i,t}^2), \]
where $k(\mu_{i,t}, \sigma_{i,t}^2)$ is the gaussian density and the parameters $\pi_{i,t}$, $\mu_{i,t}$, and $\sigma_{i,t}^2$ of the $n$ gaussian components are estimated by three MLPs:
\[ \pi_{i,t} = s(\tilde{\pi}_{i,t}) = \frac{\exp(\tilde{\pi}_{i,t})}{\sum_{j=1}^{n} \exp(\tilde{\pi}_{j,t})} \]
\[ \tilde{\pi}_{i,t} = \tilde{f}_{1,i}(r_{t-1}) \]
\[ \mu_{i,t} = \tilde{f}_{2,i}(r_{t-1}) \]
\[ \sigma_{i,t}^2 = \tilde{f}_{3,i}(\sigma_{1,t-1}^2, \sigma_{2,t-1}^2, \ldots, \sigma_{n,t-1}^2, \epsilon_{t-1}^2). \]
where $\tilde{f}_{m,i}$ denotes the $i$th component of the output of the $m$th MLP. The softmax function $s(\tilde{\pi}_{i,t})$ in (4) ensures that the priors $\pi_{i,t}$ are positive and that they sum up to 1 which makes the right-hand side of (3) a density. The MLPs $\tilde{f}_{1,i}$ and $\tilde{f}_{2,i}$ estimating the priors and the centers are standard MLPs (2). The MLP $\tilde{f}_{3,i}$
In our numerical experiments we used three data sets related to different financial markets:

1. daily closing values of the American stock index Dow Jones Industrial Average (DJIA);

3 Data Sets and Estimation Procedure

In our numerical experiments we used three data sets related to different financial markets:
2. daily closing values of the FTSE 100 traded at the London Stock Exchange;
3. daily closing values of the Japan index NIKKEI 225.

The taken time interval for all data sets was 13 years from 1985 to 1997.

All data were transformed into continuously compounded returns $r_t$ (in percent) in the standard way by the natural logarithm of the ratio of consecutive daily closing levels.

In order to take care of stationarity issues and increase the reliability of the empirical analysis, all time series were divided into overlapping segments of a fixed length of 700 trading days, where the first 500 returns of each segment form a training set, the next 100 points form a validation set and the remaining 100 returns are used for testing (see Fig.1). The first segment starts on trading day 1 and ends on day 700, the second segment begins on day 101 and ends on trading day 800 and so on. The training sets are used to optimize the parameters of each model. The validation sets are used for an “early stopping” strategy to avoid overfitting for the neural networks models and independent test sets are reserved for out-of-sample model performance evaluation. The test sets are not overlapping.

In such a way, according to the available data, we got 26 segments for DJIA and FTSE 100 and 25 for NIKKEI 225 return series. As an example, we plotted time series of NIKKEI 225 returns with their unconditional kernel density approximations as well as the summary of the descriptive statistics in Fig.2-3. It can be seen that all logarithmic series exhibit significant skewness and excess kurtosis indicating non-normality of the unconditional distributions. Enormous skewness and kurtosis on segments 1-6 can be explained by the influence of the October 1987 default.

We fitted GARCH(1,1), RMDN(1), GARCH(1,1)-$t$, RMDN(1)-$t$, LRMDN(2) and RMDN(2) models to each of the training sets separately. The number of optimized parameters of a particular model is 5 for the GARCH(1,1) model, 26 for RMDN(1), 6 for GARCH(1,1)-$t$, 27 for RMDN(1)-$t$, 16 for LRMDN(2) and 54 for RMDN(2). The number of hidden units of the MLPs in the RMDN-models was chosen to be $H = 3$. The parameters of all models were optimized with respect to the average negative loglikelihood of the sample (loss function)

$$\mathcal{L} = -\frac{1}{N} \sum_{t=1}^{N} \log \rho(r_t | I_{t-1}),$$

where $N$ denotes the sample size and $\rho(r_t | I_{t-1})$ is the conditional probability density function of the corresponding distribution. We refer to $\mathcal{L}$ as the loss function of a data set, since we will make use of values of $\mathcal{L}$ calculated for data sets which were not used to estimate the model parameters.

The optimization routine was a scaled conjugate gradient algorithm. We performed optimization of RMDN models with several parameter initializations in an attempt to approach a global optimum. For the models with $t$-distribution, the degrees-of-freedom parameter was additionally optimized by a one-dimensional search routine.

Since the main goal of this work is out-of-sample diagnostic, i.e., comparison of model performance on a future data set (test set), we are interested in obtaining models with optimal generalization performance. However, all standard neural network architectures such as the fully connected multi-layer perceptron are prone to overfitting (see, e.g., Geman et al. (1992)): while the network seems to become better and better , i.e., the error (in our case - the value of the loss function) on the training set decreases, beginning with some point during training the error on an unseen sample increases. In order to prevent the RMDN models from overfitting the training data, the generalization error is estimated by the performance of the model on a validation set and an “early stopping” strategy is applied (see, e.g., Reed (1993), and Prechelt (1998)). More precisely, the model parameters are optimized with respect to the loss function on the training set and
after each iteration the loss function on the validation set is calculated. Finally, the RMDN model on the optimization iteration \( t^* \) is selected, where

\[
t^* = \arg \min_{t_0 < t < T} \mathcal{L}_{\text{validation}}(t),
\]

where \( t \) is an iteration number;

\( T \) is the number of all iterations performed;

\( t_0 \) - minimal iteration number chosen to avoid artefact behaviour of the loss function on the validation set in such a way that the parallel value of the loss function on the training set of simpler (less parametrized) model is beaten.

In the next section we are going to apply these alternative models to forecast the value at risk of an equity position and evaluate the forecasting performance on the basis of four statistical procedures as well as one economic criterion.

### 4 Forecasting the VaR of Equity Positions

As pointed out in the introduction a widely applied tool for financial risk assessment is the concept of Value at Risk. Statistically the VaR can be associated with a percentile of the returns distribution. As soon as this probability distribution is specified, the VaR can be calculated by using the \((100-p)\%\) percentile, \( r^*_p \), of the distribution as

\[
\text{VaR} = \text{today’s price} \cdot (\exp(r^*_p) - 1).
\]

Let \( F(\cdot) \) be the corresponding cumulative distribution function of the return model considered now as the VaR model. Then \( r^*_p = F^{-1}(p) \), where \( F^{-1}(\cdot) \) denotes the inverse of \( F \). For more information on VaR see, for instance, Dowd (1998), Duffie and Pan (1997) and Wilson (1998).

We proceeded in the following way:

The parameters of each model were fixed within every segment and we computed a forecast of tomorrow’s return distribution as well as VaR estimates given the past data for every point in the test part of this segment. As usual in bank policy, we chose \( p = 1\% \) and considered the 99% level on one-day VaR measures of the investment in the portfolio of stocks, corresponding to DJIA, FTSE 100 or NIKKEI 225 stock indices. In such a way, we got VaR estimate series for the whole data samples omitting the first 600 points used for the training and validation sets of the first segment. The examples of dynamical VaR estimates for NIKKEI return series are plotted in Fig. 4-6 in Appendix. Comparing the realization of the indices with VaR estimates, we determined the indicator variable \( \theta_t \) as the outcome of a binomial event: either one-day actual loss \( L_t \) on trading activity is less than VaR estimates (a success), or the loss on the activity exceeds the potential loss estimates (a failure), i.e.

\[
\theta_t = \begin{cases} 
1, & \text{if } L_t < \text{VaR}_t, \\
0, & \text{otherwise.}
\end{cases}
\]

(12)

We now assess the quality and accuracy of our six alternative VaR models by applying them to the three different data sets. The approach that we use is frequently referred to as “backtesting” (see, e.g., Crouhy et al. (1998), Eberlein et al. (2001), Lopez (1998)). In order to discriminate between the different models we make use of four statistical test procedures. They are comprised of

- a point estimator for correctly evaluated VaR levels;
- the regulatory backtest required as a part of the capital-adequacy framework ("Basle traffic light" test);
- exceptions testing which examines the frequency with which losses greater than the VaR estimates are observed ("Proportion of failures" test) and
- conditional coverage methodology which tests exception frequency together with the independence hypothesis.

Moreover we apply a method that ranks the different models on the basis of an economic principle. We

- estimate the economic costs that result from the forecasted VaR and its corresponding level of capital requirements.

4.1 Test 1: Point estimator for $p$.

To compare the predictive power of the models considered we estimated the empirical probability of the portfolio losses being less than VaR as the relative frequency of correctly evaluated VaR, i.e.

$$PE = \frac{1}{N} \sum_{t=1}^{N} \theta_t,$$

where $N$ is the number of VaR evaluations on the test set.

We computed statistics PE for every test set. To get statistically significant conclusions we tested the hypothesis that the mean value of the relative frequency of correctly evaluated 99% VaR is exactly 0.99. The resulting p-values are reported in Table 1.

The first column in Table 1 for every data set gives the mean value of the PE statistic over all test sets and the second column summarizes $p$-values of the mean $t$-test. Obviously, non-gaussian models dominate the gaussian models with respect to the accuracy of VaR estimations except FTSE 100 series. For DJIA the models GARCH(1,1)-$t$ and LRMDN(2) show the best performance with significant $p$-value, while for NIKKEI 225 indices mixture density models tend to be better. The gaussian models GARCH(1,1) and RMDN(1) clearly underestimate VaR in comparison with fat-tailed distributions. According to this test, for FTSE 100 data all the models seem to give correct VaR predictions.

4.2 Test 2: Basle traffic light.

This is the backtesting framework developed by the Basle Committee on Banking Supervision. Since 1996, the Committee enforces banks (and others, like insurance companies) to develop their own “internal” risk models to evaluate their portfolio risk. The capital rules cover all assets in a bank’s trading account as well as all foreign exchange and commodity positions. Any bank or other financial institution whose trading activity accounts for more than ten percent of its total assets or is more than $1$ billion must hold regulatory capital against their market risk exposure. These capital charges are based on VaR estimates generated by the banks’ own VaR models and a multiplication factor defined by supervising authorities according to the traffic light concept of the Basle Committee on Banking Supervision. According to this concept, internal banks’ models are classified into three zones. This classification into green, yellow or red zones depends on how often the actual losses exceed the daily 99% VaR predictions over a period of $n$ trading days. The boundaries of the zones are chosen based on the binomial probabilities of obtaining a particular number of exceptions from the sample under the assumption of true 99% coverage. The yellow zone begins at the
point such that the probability of obtaining that number or fewer exceptions equals or exceeds 95%. The beginning of the red zone is defined at the point where the binomial probability equals or exceeds 99.99%.

Based on such classification, the necessary capital reserves are assigned. Green zone means that the multiplication factor of 3 is applied to the VaR value, yellow results in a higher (add-on) factor between 3 and 4, whereas red normally means rejection of the model.

Our backtesting period covers the whole sample period with \( n = 2300 \) that exceed the 250 days that are typically used in practice.

The results of the hypothetical classification are listed in Table 2. As before, all the models for FTSE 100 data are in acceptable zones. For other markets the models with \( f \)-distribution together with mixture density networks yield the most reliable setup. Among the non-gaussian models, the linear mixture density network performs best. The gaussian models are again mostly rejected. Moreover, the test indicates the bad performance of the RMDN(2) model on DJIA data set.

4.3 Test 3: Proportion of failures.

Kupiec (1995) presents a more sophisticated approach to the analysis of exceptions based on the observation that a comparison between daily profit or loss outcomes and the corresponding VaR measures gives rise to a binomial experiment. The verification scheme is based on the whole sample period and determines the outcome of a binomial event according to (12). These performance data are distributed as a series of draws from an independent Bernoulli distribution and the verification test is based on the proportion of failures (PF) in the sample. Ideally, the frequency of failures, i.e. of days where the actual loss exceeds the predicted VaR level, should be close to \( p \% \). Following Kupiec, we apply a LR test to examine whether the observed frequency deviates substantially from the predicted level. Under the null hypothesis \( \text{H}_0: p = p^* \), the probability of observing \( x \) failures in a sample of size \( n \) is:

\[
\binom{n}{x} (1 - p)^{n-x} p^x,
\]

where \( \binom{n}{x} \) denotes the binomial coefficient and \( p \) is the probability of a failure on any one of the independent trials. The likelihood ratio test statistic is given by

\[
P F = 2 \log((1 - \frac{x}{n})^{n-x}(\frac{x}{n})^x) - 2 \log((1 - p^*)^{n-x}(p^*)^x),
\]

where \( p^* \) is the probability of a failure under the null hypothesis. Under \( \text{H}_0 \) the PF test has a chi-square distribution with 1 degree of freedom. In such a way, the critical number of failures that could be observed in a sample with a size equal to the considered samples (2300 points) without rejecting the null \( \text{H}_0: p^* = 0.01 \) at the 5% confidence level is \( 15 \leq x \leq 32 \).

Table 3 summarizes the performance of our models with respect to the proportion of failures test. The column denoted by "failures" give the number of failures in the whole sample for the corresponding index. The next column shows whether the null hypothesis \( \text{H}_0: p^* = 0.01 \) can be rejected at the 5% significance level. The results from the table are consistent with the previous tests.

4.4 Test 4: Conditional coverage test.

In the test above the forecast performance is examined over the sample period without reference to the information available at each point in time (unconditional evaluation). However, in the presence of the
higher-moment dynamics, testing for conditional accuracy becomes important since forecasts ignoring such dynamics may have correct unconditional coverage, characterized by clustered outliers.

The $L_{cc}$ test proposed by Christoffersen (1998) is specifically a test of correct conditional coverage. The analysis is based on the indicator sequence $\theta_t$ (12). Accurate VaR forecasts should exhibit the property of correct conditional coverage, which implies that the $\theta_t$ series must exhibit both correct unconditional coverage and serial independence. The $L_{cc}$ test of this joint hypothesis is formed by combining tests of each property. In fact, the null of the unconditional coverage test will be tested against the alternative of the independence test (first order Markov dependence). Then the relevant test statistic is

$$L_{cc} = -2 \log \left( \frac{L(p^*)}{L(\Pi_1)} \right),$$

where $L(p^*)$ denotes the likelihood function under the hypothesis of correct unconditional coverage, i.e.

$$L(p^*) = (1 - p^*)^{n-x} (p^*)^x,$$

with $x$ for the number of failures, $n$ - sample size, $p^* = 0.01$;

$L(\Pi_1)$ is the likelihood for the binary first-order Markov chain with the estimated transition probability matrix

$$\Pi_1 = \begin{bmatrix} 1 - \tilde{\pi}_{01} & \tilde{\pi}_{01} \\ 1 - \tilde{\pi}_{11} & \tilde{\pi}_{11} \end{bmatrix},$$

where

$$\tilde{\pi}_{01} = \frac{n_{01}}{n_{00} + n_{01}}, \quad \tilde{\pi}_{01} = \frac{n_{11}}{n_{10} + n_{11}},$$

and $n_{ij}$ is the number of observations in indicator sequence with value $i$ followed by $j$.

Then the approximate likelihood function is

$$L(\Pi_1) = (1 - \tilde{\pi}_{01})^{n_{00}} \cdot \tilde{\pi}_{01}^{n_{01}} \cdot (1 - \tilde{\pi}_{11})^{n_{10}} \cdot \tilde{\pi}_{11}^{n_{11}}.$$  

The distribution of the main test statistic $LR_{cc}$ is asymptotically $\chi^2(2)$.

The $p$-values of this conditional coverage test are listed in Table 4. In general, the results repeat ones from the unconditional proportion of failures test, slightly extending the number of accepted models for NIKKEI series. This test again support only linear non-gaussian models for American stock index, while for the Japan index both linear and non-linear non-gaussian models can not be rejected at the 5% significance level and the linear mixture density model dominates over all others.

4.5 Test 5: Lost interest yield.

The major disadvantage of the statistical tests above is that they all put emphasis on the number of how many time a model underestimated the actual VaR. Therefore a model that underestimates the VaR frequently but only by a very small amount will be rejected while a model with few but very large underestimates will be accepted. Since financial institutions prefer VaR models that are not only able to pass a back-testing procedure but that provide small VaR predictions, and hence low levels of regulatory capital, are certainly preferred.

Therefore, to check the efficiency of the VaR measures we developed a new test, providing a quantitative basis for the incorporation of VaR prediction in regulatory capital requirements.

Any financial institution must hold regulatory capital to cover its potential market risk exposure. We used dynamically computed daily VaR estimates generated by our VaR models and assumed for simplicity
that capital reserves equal to the VaR estimates will be held for 1 day. When actual the portfolio loss $L_t$ does not exceed the predicted loss (i.e. the VaR) for this day, we considered the required capital lost for investment for 1 day and computed the continuously compounded lost interest yield as $\text{VaR}_t \cdot (e^{i/250 \cdot 1} - 1)$ with a given fixed interest rate $i$. In the case of an underestimated loss (the portfolio loss $L_t$ is greater than $\text{VaR}_t$) banks face additional capital charges and have to reallocate their assets which we model by a penalty term. In this case lost interest yield $= \text{VaR}_t \cdot (e^{i/250 \cdot 1} - 1) + \text{Penalty}$, where, e.g., $\text{Penalty} = 1.2 \cdot (L_t - \text{VaR}_t) \cdot e^{i/250 \cdot 1}$ to cover higher transaction costs of capital reallocations.

According to this strategy, we calculated lost yield summed up over all test points and then scaled the result to remove the dependency on the portfolio size. These relative lost interest yields for all our data in the sample are depicted in Fig 7.

For all the markets the best model is the linear mixture density network. In general, this test clearly rejects linear and non-linear gaussian models and favors the mixture density models over the models with $t$-distributions.

5 Discussion and Conclusions

We analyzed the impact of non-linearity and of non-gaussian distributions on the estimation of values at risk in three different equity portfolios. We divided our data set into a number of segments in order to take into account stationarity issues. The parameters of all models were first estimated on the training part of every segment by a maximum likelihood methodology and then we performed out-of-sample forecasts and forecast evaluations on the test sets. The models were evaluated with respect to the ability to forecast the daily VaR models.

Summing up, we derived the following conclusions:

- Different performance measures generate different ranking of the models even within one financial market and, hence, there is no consistent ordering to determine one "best" model over all markets and for every market separately.

- We found no significant differences in performance between linear and non-linear models for all data series except DJIA, where the tests support only linear GARCH-$t$ and linear mixture density network.

- All tests in general confirmed the conclusion that non-gaussian models significantly dominated the gaussian ones with respect to the most performance measures for all stock indices considered. But within non-gaussian models themselves there was some difference across the markets and while both the models with $t$-distribution and mixture density networks are capable to capture fat tails elements in the conditional distribution, only mixture density networks allow for time-varying skewness and kurtosis which are common in financial markets.

Acknowledgements

This work was funded by the Austrian Science Fund under grant SFB#010: “Adaptive Information Systems and Modeling in Economics and Management Science”. The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry of Education, Science and Culture.
References


6 Tables

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean p-value</th>
<th>Mean p-value</th>
<th>Mean p-value</th>
<th>Mean p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH(1,1)</td>
<td>0.981 0.0044</td>
<td>0.989 0.5891</td>
<td>0.984 0.0796</td>
<td></td>
</tr>
<tr>
<td>RMDN(1)</td>
<td>0.981 0.0045</td>
<td>0.988 0.3069</td>
<td>0.981 0.0138</td>
<td></td>
</tr>
<tr>
<td>GARCH(1,1)-D8</td>
<td>0.988 0.2865</td>
<td>0.990 0.9557</td>
<td>0.985 0.1162</td>
<td></td>
</tr>
<tr>
<td>RMDN(1)-D8</td>
<td>0.985 0.1663</td>
<td>0.989 0.5308</td>
<td>0.986 0.1635</td>
<td></td>
</tr>
<tr>
<td>LRMDN(2)</td>
<td>0.988 0.4199</td>
<td>0.990 0.9638</td>
<td>0.988 0.5235</td>
<td></td>
</tr>
<tr>
<td>RMDN(2)</td>
<td>0.982 0.0645</td>
<td>0.988 0.3798</td>
<td>0.987 0.3850</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: p-values of the mean $t$-test for PE test sample

<table>
<thead>
<tr>
<th>Model</th>
<th>DJIA failures</th>
<th>DJIA zone</th>
<th>FTSE 100 failures</th>
<th>FTSE 100 zone</th>
<th>NIKKEI 225 failures</th>
<th>NIKKEI 225 zone</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH(1,1)</td>
<td>45 red</td>
<td>32 yellow</td>
<td>39 yellow</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMDN(1)</td>
<td>48 red</td>
<td>26 green</td>
<td>43 red</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH(1,1)-D8</td>
<td>30 green</td>
<td>29 green</td>
<td>35 yellow</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMDN(1)-D8</td>
<td>39 yellow</td>
<td>25 green</td>
<td>32 yellow</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LRMDN(2)</td>
<td>30 green</td>
<td>27 green</td>
<td>27 green</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMDN(2)</td>
<td>44 red</td>
<td>30 green</td>
<td>34 yellow</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Classification according to the Basle rules. The column "failures" give the number of failures in the whole sample for the corresponding index.
<table>
<thead>
<tr>
<th>Model</th>
<th>DJIA failures</th>
<th>H0: $p^* = 0.01$</th>
<th>FTSE 100 failures</th>
<th>H0: $p^* = 0.01$</th>
<th>NIKKEI 225 failures</th>
<th>H0: $p^* = 0.01$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH(1,1)</td>
<td>45 rejected</td>
<td>32 not rejected</td>
<td>39 rejected</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMDN(1)</td>
<td>48 rejected</td>
<td>26 not rejected</td>
<td>43 rejected</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH(1,1)-t</td>
<td>30 not rejected</td>
<td>29 not rejected</td>
<td>35 rejected</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMDN(1)-t</td>
<td>39 rejected</td>
<td>25 not rejected</td>
<td>32 not rejected</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LRMDN(2)</td>
<td>30 not rejected</td>
<td>27 not rejected</td>
<td>27 not rejected</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMDN(2)</td>
<td>44 rejected</td>
<td>30 not rejected</td>
<td>34 rejected</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: PF test

<table>
<thead>
<tr>
<th>Model</th>
<th>DJIA p-value</th>
<th>Hypothesis</th>
<th>FTSE 100 p-value</th>
<th>Hypothesis</th>
<th>NIKKEI 225 p-value</th>
<th>Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH(1,1)</td>
<td>0.0001</td>
<td>rejected</td>
<td>0.1551</td>
<td>not rejected</td>
<td>0.0037 rejected</td>
<td></td>
</tr>
<tr>
<td>RMDN(1)</td>
<td>0.0000</td>
<td>rejected</td>
<td>0.6075</td>
<td>not rejected</td>
<td>0.0009 rejected</td>
<td></td>
</tr>
<tr>
<td>GARCH(1,1)-t</td>
<td>0.2487</td>
<td>not rejected</td>
<td>0.3285</td>
<td>not rejected</td>
<td>0.0544 not rejected</td>
<td></td>
</tr>
<tr>
<td>RMDN(1)-t</td>
<td>0.0009</td>
<td>rejected</td>
<td>0.6899</td>
<td>not rejected</td>
<td>0.1551 not rejected</td>
<td></td>
</tr>
<tr>
<td>LRMDN(2)</td>
<td>0.2626</td>
<td>not rejected</td>
<td>0.5140</td>
<td>not rejected</td>
<td>0.5140 not rejected</td>
<td></td>
</tr>
<tr>
<td>RMDN(2)</td>
<td>0.0003</td>
<td>rejected</td>
<td>0.2487</td>
<td>not rejected</td>
<td>0.0798 not rejected</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Christoffersen test of conditional coverage
7 Figures

Figure 1: Segment structure of the data sets
Figure 2: The plot (a) and kernel density estimation (b) of NIKKEI 225 return time series.

Figure 3: The basic statistics of NIKKEI 225 daily returns for all segments in the analysis: (a) - mean values; (b) - standard deviation; (c) - skewness; (d) - kurtosis.
Figure 4: VaR estimates for NIKKEI 225 series, models GARCH(1,1) and RMDN(1).
Figure 5: VaR estimates for NIKKEI 225 series, models GARCH(1,1)-t and RMDN(1)-t.
Figure 6: VaR estimates for NIKKEI 225 series, models LRMDN(2) and RMDN(2).
Figure 7: Lost Interest Yield (relative values).