CONDITIONAL VALUE AT RISK*

Dirk Ormoneit
Department of Statistics
Stanford University
Stanford, CA 94305-4065
ormoneit@stat.stanford.edu

Ralph Neuneier
Siemens AG, ZT IK 4
81730 München
Germany
Ralph.Neuneier@mchp.siemens.de

Abstract

We suggest a new methodology to overcome several well-known deficiencies of Value at Risk computations. Our approach mainly addresses two aspects of Value at Risk: first, to avoid potentially disastrous clustering in predicted tail events we derive a new approach to accurately estimating the *conditional* distribution of asset returns using maximum entropy densities. Second, by the very nature of the maximum entropy model, we account for negative skewness and fat tails in asset returns. In particular, to obtain a robust and scalable estimate of the covariance matrix of the assets in the portfolio we extend an approach by Hull and White to the case of conditional distributions. In extensive experiments with historical stock index data we compare the proposed methodology to alternative estimation approaches using a new, simulation-based statistical testing procedure for serial dependence in the predicted tail events.

1 Introduction

As value at risk ("VaR") has emerged as the dominant measure of market risk implied by a portfolio over the past years, extensive research both within academia and within the financial industry has recently focussed on effective and reliable algorithms to compute the VaR of complicated portfolios. Probably the most influential contribution in this field has been J.P. Morgan's RiskMetrics^{TM} methodology, within which a multivariate normal distribution is employed to model the joint distribution of the assets in a portfolio [Mor96]. While this approach is doubtlessly appealing for its straightforwardness and computational simplicity, several authors have found that the observed frequency of extreme market movements is frequently higher than predicted by RiskMetrics (see, for example, [HW98], Exhibit 3). Obvious explanations for this finding are negative skewness and excess kurtosis ("fat tails") in the true distribution of market returns, which cannot be accounted for using a normal density model as in RiskMetrics.

Hull and White suggest a robust and scalable approach to deal with the problem of non-normality by estimating the covariance matrix in a suitably transformed

*PRELIMINARY AND INCOMPLETE

space within which the asset returns are jointly normal distributed [HW98]. Within this work, we extend their approach to account for potentially disastrous serial dependency in the asset returns. The basic idea is that if Value at Risk is estimated based on the unconditional distribution of portfolio returns, such as in the case of historical simulations or in Hull and White's approach, the predicted tail events will typically reveal undesired serial dependency due to volatility clusters in the asset returns. An example of an unconditional VaR forecast for a sample portfolio consisting of several stock indices is shown in Figure 1. Note that the sequence of indicator variables for the predicted extreme events in the lower part of the figure is not independent but contains several clusters. These clusters are highly undesirable

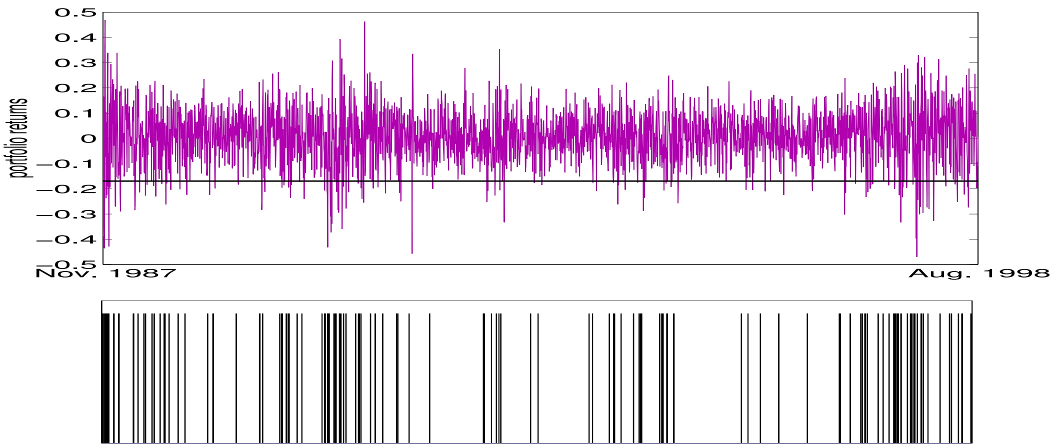

Figure 1: *Daily unconditional VaR forecast for a sample portfolio consisting of the DAX, the DJIA, and the FTSE in the period from November 1987 to August 1998. The lower part shows an indicator variable that takes on the value one if the observed return is below the forecast.*

as they lead to larger reserve requirements than a sequence of independent events with equal probability. Put simply, extreme events should be equally surprising at each date because information on an increased risk level from previous observations should be reflected by a consequently lower VaR forecast. A straightforward way to avoid serial dependency in the predicted events is thus to base one's forecast on the *conditional* distribution of the portfolio returns given past information as it is shown in Figure 2. The conditional VaR forecast, indicated by the solid line, continually adapts in response to the prevailing market volatility in this case and consequently generates an indicator sequence with considerably less dependency. To obtain this forecast, we construct transformations of the individual asset returns to a normalized space within which the returns are jointly normal distributed *and serially independent* using an estimate of the conditional distribution of returns. As a consequence, the covariance matrix characterizing the interaction between the asset returns may be treated as stationary in the transformed space which greatly simplifies its estimation. In particular, to estimate the conditional distributions we use a mixture approach that separately accounts for short-term and long-term dependencies in the

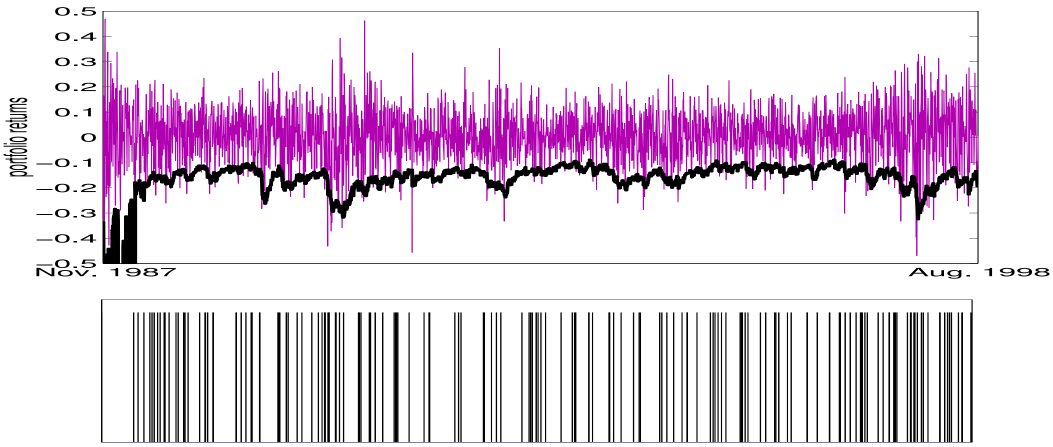

Figure 2: *Daily conditional VaR forecast for the sample portfolio. The sequence of indicator variables in the lower part shows less serial dependency than in the case of the unconditional VaR forecast.*

asset returns. By using a maximum entropy density to model the long-term dependencies, we explicitly account for negative skewness and excess kurtosis.

The organization of the remainder of this work is as follows. In Section 2 we introduce our notation as well as the mixture approach to estimating the conditional density of returns. In Section 3 we discuss our extension of Hull and White's forecasting method based on transformations. In Section 4 we apply the new method to evaluate the risk of a sample portfolio and present a new statistical test procedure to detect serial dependency in the predicted tail events. Section 5 summarizes our results and presents final remarks.

2 Conditional Density Filtration

Consider a portfolio Π consisting of the (possibly time-varying) positions $a_t = (a_{t,1}, \ldots, a_{t,N})'$ in N risky assets during the times $t = 0, \ldots, T$. Our focus is on assessing the risk associated with holding Π over some time period, say from $t-1$ to t. Formally, this risk is completely characterized by the probability distribution of the proportional price changes R_t of Π during this period, defined as

$$R_t = \sum_{i=1}^{n} a_{t,i} \frac{s_{t,i}}{s_{t-1,i}}. \tag{1}$$

Here $s_{t,i}$ denotes the price of asset i at time t and $s_{t,i}/s_{t-1,i}$ is the proportional price change of asset i. Value at risk is concerned with estimating the unknown probability distribution of R_t and evaluating that quantile of this distribution which corresponds to some specified probability α. In the simplest case it is assumed that the vector $(s_{t,1}/s_{t-1,1}, \ldots, s_{t,N}/s_{t-1,N})'$ is governed by a multivariate normal distribution $N(\mu_t, \Sigma_t)$, in which case also R_t is normally distributed with mean $a_t'\mu_t$

and variance $a_t' \Sigma_t a_t$. The problem is thus essentially reduced to the one of estimating μ_t and Σ_t as it is impressively illustrated by the RiskMetrics calculator.

While the assumption of normally distributed price changes is computationally convenient, there is significant empirical evidence that in reality the distribution of asset returns might be more complicated, involving nontrivial characteristics such as negative skewness and "fat tails".[1] As it was shown in the Introduction, additional bias may result from the serial dependence in asset returns when using an unconditional VaR forecast. As a consequence, VaR estimates based on normal distributions may systematically underestimate the risk exposure of a portfolio and have provoked justified criticism in the recent past (for an applied standpoint, see [Tal97]). Below we model VaR based on an estimate of the distribution of the *returns* $r_{t,i} = \log(s_{t,i}/s_{t-1,i})$ rather than directly based on the proportional price changes.[2] In particular, we suggest a novel approach to estimating VaR within which a mixture density is used to model the distribution $p(r_{t,i}|I_{t,i})$ of the returns $r_{t,i}$ conditional on the information set $I_{t,i} = (r_{t-1,i}, \ldots, r_{1,i})$ of past returns. The basic idea underlying the mixture approach is to allow for the separate treatment of *short-term* effects such as volatility clusters and *long-term* events such as market crashes, which both seem highly relevant to predict market risk. However, using standard approaches involving a geometric decay of past information, such as in the case of exponentially weighted averages or GARCH [Bol86], rare extreme events (market crashes) may be forgotten too quickly. On the other hand, using a very long memory as in the unconditional model in Section 1 or in IGARCH [Nel90] prevents the identification of volatility clusters. By using separate short-term and long-term models, the mixture approach accounts for dependencies at different frequencies.

Formally, the proposed mixture is composed of a normal density f_1 to model the short-term dependencies and a *maximum entropy density* f_2 to capture negative skewness and excess kurtosis in the long-term dependencies of $r_{t,i}$:

$$p(x|I_{t,i}) = \kappa_1 f_1(x;\theta_1) + \kappa_2 f_2(x;\theta_2), \tag{2}$$

where

$$\begin{aligned} f_1(x;\theta_1) &= N(x|\mu_{t,i}, \sigma^2_{t,i}), \\ f_2(x;\theta_2) &= \exp\{-(\lambda_0 + \lambda_1 x + \lambda_2 x^2 + \lambda_3 x^3 + \lambda_4 x^4)\}. \end{aligned}$$

κ_1 and κ_2 (satisfying $\kappa_1 \geq 0$, $\kappa_2 \geq 0$, and $\kappa_1 + \kappa_2 = 1$) are mixture weightings reflecting the prior probability of extreme market events. Note that a mixture of the suggested form is consistent with a jump-diffusion model of asset returns as suggested by Merton [Mer76] and Cox and Ross [CR76]. The mean and the variance parameters of the normal density $N(x|\mu_{t,i}, \sigma^2_{t,i})$ are, like in RiskMetrics, computed

[1] Another common source of non-normality are nonlinear dependencies among the components of r_t, e.g. due to derivative assets. For surveys on VaR calculations for derivative portfolios, see [DP97] or Chapter 15 of [Hul98].

[2] In fact, also RiskMetrics uses the distribution of returns rather than of proportional price changes.

as exponentially weighted averages:

$$\mu_{t,i} = \textstyle\sum_{s=1}^{t} w_{t,s} r_{t,i}, \qquad \sigma_{t,i}^2 = \displaystyle\sum_{s=1}^{t} w_{t,s} r_{t,i}^2 - \mu_{t,i},$$

where

$$w_{t,s} = \frac{\exp\{-\delta(t-s)\}}{\sum_{s'=1}^{t} \exp\{-\delta(t-s')\}}.$$

In our experiments, we tried several values for the decay parameter δ and for the mixture weighting κ_1 and found $\delta_1 = 0.05$ and $\kappa_1 = 0.9$ to be a suitable choice. The maximum entropy density $f_2(x;\theta_2)$ has the special property of being least presumptuous with regard to missing information when estimating the density of a random variable whose first four moments are known. Conveniently, the (non-standardized and unweighted) first four moments

$$\begin{aligned} \nu_{t,i,1} &= \textstyle\sum_{s=1}^{t} r_{s,i}, & \nu_{t,i,2} &= \displaystyle\sum_{s=1}^{t} r_{s,i}^2, \\ \nu_{t,i,3} &= \textstyle\sum_{s=1}^{t} r_{s,i}^3, & \nu_{t,i,4} &= \displaystyle\sum_{s=1}^{t} r_{s,i}^4 \end{aligned}$$

operate as sufficient statistics for maximum likelihood estimation using this density, and there exists a one-to-one correspondence between $\nu_{t,i,1}, \nu_{t,i,2}, \nu_{t,i,3}, \nu_{t,i,4}$ and the parameters $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ in the specification of $f_2(x;\theta_2)$. Numerical algorithms for the evaluation of this correspondence have been suggested by Zellner [ZH88] and Ormoneit and White [OW99]. A tabular description sufficient for the practical application of our approach is available on the World Wide Web [OW97].

Below we use the mixture density (2) to construct a nonlinear transformation of the asset returns into a normalized space. Details of this transformation will be described next.

3 Conditional Transformation of Returns

Given an estimate of the conditional probability distribution of the return of each individual asset our main focus is on the construction of a transformation into a space within which the returns are jointly normal distributed and serially independent. Hull and White argue that the variance of an estimator of the covariance matrix characterizing multiple assets may be greatly reduced by transforming into a space with normally distributed returns as there are no outliers in this case.

In this work we argue that also serial dependence increases the variance of the estimate and should be removed by using a transformation based on the conditional distributions. Mathematically, the usual estimate of the covariance matrix is efficient in the transformed space as it coincides with the maximum likelihood estimate. It is well-known that a transformation with the desired properties may be constructed as

$$u_{t,i} = \Phi^{-1}\left[P_{t,i}(r_{t,i})\right], \tag{3}$$

where $P_{t,i}(r) = \int_{-\infty}^{r} p(x|I_t)dx$ is the (conditional) distribution function of $r_{t,i}$ and Φ^{-1} is the inverse of the distribution function (the z-quantile) of the standard normal density [KS77, HW98]. In a second step, the joint distribution of the random vector $u_t = (u_{t,1}, \ldots, u_{t,N})'$ is modeled by a multivariate normal distribution $N(u_t|0, \Sigma_t)$ using the usual covariance estimate

$$\Sigma_t = \frac{1}{t} \sum_{s=1}^{t} u_s u_s'. \tag{4}$$

(Remember that the transformed returns u_t have zero mean.) In order to recover the VaR of the original portfolio from this approximation in principle we have to solve a multivariate integral involving the estimated normal density $N(u_t|0, \Sigma_t)$ and the inverses of the transformations (3). As the numerical evaluation of this integral is computationally extremely demanding, Hull and White suggest to use a stochastic approximation instead. In particular, we generate a large number of random vectors according to

$$\tilde{u}_t \sim N(0, \Sigma_{t-1})$$

and subsequently apply the *inverse transformation*

$$\tilde{r}_{t,i} = P_t^{-1}\left[\Phi(\tilde{u}_{t,i})\right] \tag{5}$$

to obtain a sequence of *pseudo returns*. Here $P_{t,i}^{-1}(z)$ is the inverse of the distribution function (i.e. the z-quantile) of $p(r|I_t)$. Based on the sample price changes $\tilde{r}_{t,i}$ a sample realization of R_t may be computed in analogy to (1), and by repeating the last steps sufficiently often the VaR may be approximated by the appropriate quantile of the empirical distribution function.

4 Evaluating a Sample Portfolio

To assess the merit of our approach we evaluate a sample portfolio consisting of equal positions in three international stock market indices, namely the American DJIA, the German DAX, and the English FTSE index:

$$\Pi_t = a_1 \cdot DAX_t + a_2 \cdot DJ_t + a_3 \cdot FTSE_t. \tag{6}$$

In our example, we chose constant portfolio weights $a_1 = a_2 = a_3 = 1/3$ for simplicity. However, also complicated time-dependent portfolios may easily be evaluated using our approach. To illustrate the effect of the normalization, we compare the raw returns of the DAX to its normalized returns in Figure 3. The histograms on the right show that the skewness and the fat tails of the raw returns were removed after the application of (3). Also, there is less evidence of volatility clusters in the normalized returns. The estimated covariance matrix in the transformed space at the last observation in the data set (July 28, 1998) is

$$\Sigma_{3802} = \begin{pmatrix} 1.0274 & 0.4464 & 0.2108 \\ 0.4464 & 1.0320 & 0.3643 \\ 0.2108 & 0.3643 & 1.0215 \end{pmatrix},$$

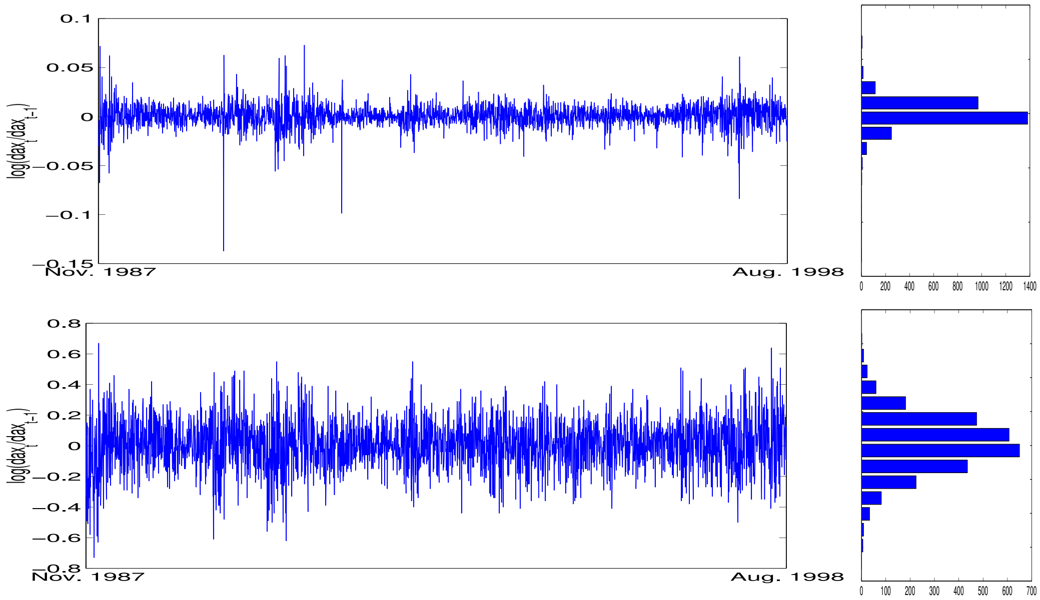

Figure 3: *Raw and transformed daily returns of the DAX. The histograms on the right show the relative frequencies of the returns in the same scale.*

where the rows and returns are ordered as in (6).

A crucial issue for the comparison of VaR approaches is the identification of an appropriate performance measure. A discussion of amenable approaches may be found in [Lop97]. A simple procedure which is commonly applied by both researchers and practitioners is to compare the specified probability of tail events, e.g. $\alpha = 5\%$ in our experiments, directly to their realized relative frequency. For this purpose, let δ_t denote a binary indicator variable taking the value one if the realized return R_t is below the predicted VaR boundary and taking the value zero otherwise. Let T be the number of samples in a test set and let $m \equiv \sum_{t=1}^{T} \delta_t$ and $h \equiv m/T$ denote the predicted number and relative frequency of tail events, respectively. As described in [Lop97], a straightforward likelihood-ratio statistic to assess the difference between h and α is

$$LR = 2\left[\log\left(h^m(1-h)^{T-m}\right) - \log\left(\alpha^m(1-\alpha)^{T-m}\right)\right]. \tag{7}$$

For independently and identically distributed (i.i.d.) δ_t, LR possesses an asymptotic $\chi^2(1)$ distribution which may easily be evaluated.[3] Below we use this asymptotic distribution to test the joint null-hypothesis of no significant difference between h and α and serial independence of δ_t.[4]

Besides obtaining the desired relative frequency h of tail events, the main focus of our experiments is on serial dependence in the indicator variable δ_t. As there exists no standard test statistic which is commonly used in the finance literature,

[3]Under the i.i.d. assumption exact finite sample tests may of course be carried out using the Binomial distribution of h.

[4]More precisely, the null-hypothesis is $H_0 : E[h] = \alpha, \delta_t \sim i.i.d.$.

we suggest a novel test procedure in this work which is based upon the discrete case analogue of the Ljung-Box test for continuous random variables [LB78]. The Ljung-Box statistic is a weighted average of squared autocorrelations at various lags. Formally, it may be written as

$$Q_l = T(T+2)\sum_{k=1}^{l}\frac{\rho^2(k)}{T-k}, \tag{8}$$

where $\rho(k)$ denotes the estimated correlation coefficient of δ_t and δ_{t+k}. As δ_t is discrete in our application, the usual asymptotic properties of Q_l cannot be relied upon. Instead, a Monte-Carlo simulation approach to approximate the distribution of Q_l under the null-hypothesis of i.i.d. events may easily be derived. For this purpose, we generate random sequences of binary variables with expectation α of length T, which may subsequently be used to compute a sample realization of Q_l according to (8). Repeating this procedure sufficiently often one may approximate the true distribution of Q_l under the null-hypothesis using the empirical distribution of the simulated Q-values.

	h	LR	p-value	Q_{20}	p-value
Conditional VaR	6.0%	5.5347	0.0186	0.0420	0.5174
Benchmark	4.998 %	0.00001879	0.9965	0.2908	0.000

Table 1: *Relative frequency and sample statistics of tail events as predicted by Conditional VaR and an unconditional benchmark model. The VaR forecasts are based on a tail probability of $\alpha = 0.05$. The numbers in brackets are the p-values of the statistics. To approximate the p-values for the Q-statistics we used a Monte-Carlo simulation with 1000 samples.*

In Table 1 we report h, the LR statistic, as well as Q_{20} for the VaR estimates of the sample portfolio. To obtain these values, we split the available set of daily prices ranging from January 2, 1984 to July 28, 1998 (3802 observations) into a training and a test set ranging from January 2, 1984 to October 30, 1987 (1000 observations) and from November 2,1987 to July 28, 1998 (3801 observations), respectively. Here the training set was used to determine the optimal transformation parameters $\delta = 0.05$ and $\kappa_1 = 0.9$, and the test set was used to compute the reported statistics. The first row of the table shows the results of Conditional VaR and the second row shows the results of an ideal unconditional forecast computed in hindsight, i.e. it corresponds to the 5% quantile of the returns in the test set. The numbers in brackets denote the p-values for the rejection of the underlying null-hypothesis.[5]

The results in Table 1 show that, with regard to the LR statistic, the benchmark model performs exceptionally well, which is not surprising given that it is based on information that would normally not be available to a risk manager. The p-value

[5]The p-value is the probability that the statistic takes on a value of at least its observed realization under the null-hypothesis. For example, a p-value of 3% means that the null-hypothesis must be rejected in a test using a 95% confidence level.

of the LR statistic for Conditional VaR of 0.0186 is at the limit of rejection using typical confidence levels, and thus indicates a possible systematic deviance of the probability of tail events from the 5% level. Alternative explanations include serial correlation in δ_t, which is also violating the null-hypothesis underlying the LR test, as well as a significant deviation of the true finite sample distribution of LR from its asymptotic approximation. With regard to the Q_{20} statistic, the hypothesis of serial independence in δ_t must be rejected for any reasonable confidence level for the benchmark model, which is also not surprising given the evident serial dependence in the asset returns. For Conditional VaR, the hypothesis of serial independence cannot be rejected using the suggested test procedure.

5 Summary and Conclusion

We discussed a new method to overcome two important shortcomings of common VaR estimates. First, by using a mixture model of the individual asset returns, we allow for the separate treatment of long-term events such as market crashes and short-term events such as volatility clusters which are both essential to VaR. In particular, by using a maximum entropy density as one mixture component, we account for the presence of fat tails and non-zero skewness in asset returns. Second, by transforming the returns into a space within which they are jointly normal distributed and serially independent, we obtain a robust estimate of the covariances between different assets. In particular, as the transformations of the individual asset returns are carried out independently of each other, the danger of "overfitting" the data is drastically reduced in comparison to alternative approaches which aim at building a complete model of the conditional *joint* distribution of asset returns such as multivariate GARCH [BEW88, Chr98]. Our experiments confirm that Conditional VaR leads to serially independent predicted tail events while closely approximating the specified event probabilities.

6 Acknowledgments

The authors thank Alan Gous for his advice regarding the development of the correlation test in Section 4, as well as participants of the Computational Finance'99 conference in New York for helpful comments. The work of Dirk Ormoneit is supported by a grant of the Deutsche Forschungsgemeinschaft (DFG) in the framework of their postdoc program.

References

[BEW88] T. Bollerslev, R. Engle, and J. Wooldridge. A Captial Asset Pricing Model with time varying covariances. *Journal of Political Economy*, 96:116–131, 1988.

[Bol86] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[Chr98] C. Christiansen. Value at Risk using the Factor-ARCH model. Technical report, Department of Finance, Aarkus School of Business, 1998.

[CR76] J. Cox and S. Ross. The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3:145–166, 1976.

[DP97] D. Duffie and J. Pan. An overview of value at risk. *Journal of Derivatives*, 4(3):7–49, 1997.

[Hul98] J. C. Hull. *Introduction to Futures and Options Markets.* Prentice Hall, third edition, 1998.

[HW98] J. Hull and A. White. Value at risk when daily changes in market variables are not normally distributed. *Journal of Derivatives*, 5(3):9–27, 1998.

[KS77] M. Kendall and S. Stuart. *The Advanced Theory of Statistics*, volume 1. Charles Griffin & Cmopany, 1977.

[LB78] G. M. Ljung and G. E. P. Box. On a measure of lack of fit in time series models. *Biometrica*, 65:297–303, 1978.

[Lop97] J. A. Lopez. Regulatory evaluation of Value-at-Risk models. Staff Reports 33, Federal Reserve Bank of New York, November 1997.

[Mer76] R. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3:125–144, 1976.

[Mor96] J. P. Morgan. Statistics of financial market returns, 1996. `http://www.jpmorgan.com/RiskManagement/RiskMetrics/pubs.html`.

[Nel90] D. Nelson. Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6:318–334, 1990.

[OW97] D. Ormoneit and H. White. Table of the correspondence between moments and maximum entropy density parameters, 1997. `http://wwwbrauer.informatik.tu-muenchen.de/~ormoneit/lambda_table.txt`.

[OW99] D. Ormoneit and H. White. An efficient algorithm to compute maximum entropy densities. *Econometric Reviews*, 1999. To appear.

[Tal97] N. Taleb. The world according to Nassim Taleb. *Derivatives Strategy*, January 1997.

[ZH88] A. Zellner and R. A. Highfield. Calculation of maximum entropy distributions and approximation of marginal posterior distributions. *Journal of Econometrics*, 37:195–209, 1988.