Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Finance Research Letters 4 (2007) 137–145
Finance Research Letters
www.elsevier.com/locate/frl

Optimality of the RiskMetrics VaR model

Gloria González-Rivera *, Tae-Hwy Lee, Emre Yoldas

Department of Economics, University of California, Riverside, CA 92521, USA

Received 29 March 2007; accepted 1 June 2007

Available online 13 June 2007

Abstract

Using different loss functions in estimation and forecast evaluation of econometric models can cause sub-optimal parameter estimates and inaccurate assessment of predictive ability. Though there are not general guidelines on how to choose the loss function, the modeling of Value-at-Risk is a rare instance is which the loss function for forecasting evaluation is well defined. Within the context of the RiskMetrics™ methodology, which is the most popular to calculate Value-at-Risk, we investigate the implications of considering different loss functions in estimation and forecasting evaluation. Based on U.S. equity, exchange rates, and bond market data we find that there can be substantial differences on the estimates under alternative loss functions. On calculating the 99% VaR for a 10-day horizon, the RiskMetrics™ model for equity markets overestimates substantially the decay factor. However, the out-of-sample performance is not systematically superior by using the estimates under the correct loss function.

JEL classification: C22; C52; C53; G0

Keywords: Estimation; Forecasting; Loss functions; Optimality; VaR

1. Introduction

In applied time series it is often the case that the loss function used for estimation of the parameters of the model is different from the one(s) used in the evaluation of the model. This is a logical inconsistency that is difficult to justify. If the researcher knows her loss function and her objective is to minimize the loss, then optimality—either optimal parameter estimates

* Corresponding author. Fax: +1 951 827 5685.
E-mail addresses: gloria.gonzalez@ucr.edu (G. González-Rivera), taelee@ucr.edu (T.-H. Lee), emre.yoldas@email.ucr.edu (E. Yoldas).

1544-6123/$ – see front matter
doi:10.1016/j.frl.2007.06.001

and/or optimal forecasts—must be understood with respect to that loss function. The reality is that there is little knowledge on what loss function we should choose since this choice is specific to the particular economic questions under analysis. Mostly we resort to symmetric quadratic loss functions because of their simplicity and easy mathematical treatment. The optimal forecast associated with this type of loss function is simply the conditional mean, but if the loss function is asymmetric, the optimal forecast is more complex as it depends not only on the loss function but also on the probability density function of the forecast error (Granger, 1999), which may be also unknown.

The aforementioned logical inconsistency—using different loss functions in estimation and evaluation—has also significant statistical consequences with regard to the assessment of the predictive ability of competing models, e.g. Christoffersen and Jacobs (2004). When a single loss function is used in both stages (estimation and forecasting), innovation uncertainty is asymptotically the only source of variability in the forecast error if the model is correctly specified. Under correct specification, estimation uncertainty vanishes asymptotically. For misspecified models, the uncertainty associated with parameter estimation becomes also relevant, and not taking it into account may result in an incorrect assessment of the predictive ability of the models.[1]

In financial econometrics, the modeling of Value-at-Risk (VaR) is a rare instance in which the loss function of the forecaster is well defined as opposed to many applications in which there are no clear guidelines as to which loss functions to use for estimation and/or model validation. The VaR is a conditional quantile and so can be directly estimated from minimizing the check function of Koenker and Bassett (1978), which in turn is also the loss function for evaluation of the conditional quantile forecast. However, in many studies of VaR models, different loss functions in estimation and forecasting are still chosen. Our objective is to investigate the implications of this practice by focusing on the RiskMetrics (henceforth RM) model, which due to its simplicity is widely accepted among practitioners.[2]

Using U.S. equity, exchange rate, and bond data, we estimate the RM model under the two loss functions—the squared error loss and the check loss functions. Under correct model specification, a suitable loss function will lead to consistent estimation. However, when misspecification is present, the choice of the loss function for estimation is most important. It is very likely that the RM model is misspecified, especially when it is indiscriminately used in a variety of data sets. We investigate how much we may gain by using the same loss function for estimation and for validation of the RM model. Our findings indicate that for a 10-day horizon there can be substantial differences especially for equities under Student-t distribution.

The rest of the paper is organized as follows. In Section 2 we review the optimality of a forecast for a squared loss function and for a check loss function. In this context, we also review the RM model. We show that the relevant loss function for VaR calculations is the check function and discuss the estimation of the RM model under the two aforementioned loss functions. In

[1] West (1996) studies the effect of parameter estimation uncertainty on the asymptotic inference in comparing the loss values of competing models. As long as the parameter values are consistently estimated the parameter estimation uncertainty vanishes asymptotically in the loss. For the consistency under squared error loss see White (1994), and for the consistency under the check loss see Komunjer (2005).

[2] Perignon and Smith (2006) find that the most commonly used approach for VaR computation among commercial banks is Historical Simulation, which is essentially a parameter free method based on the *unconditional* empirical distribution. Andersen et al. (2006) emphasize the importance of the *conditional* perspective and the exclusive modeling of dependence among various assets for effective risk management. In this regard, RM provides a simple yet informative solution.

Section 3, we implement our approach with an extensive data set and we discuss the empirical results. Finally, in Section 4 we conclude.

2. Optimal forecast and the RiskMetrics model

Consider a stochastic process $Z_t \equiv (Y_t, X_t')$ where Y_t is the variable of interest and X_t is a vector of other variables. Suppose that there is a model for Y_t, i.e. $Y_t = m(X_t, \lambda)$ where λ needs to be estimated, and assume that there are $T \equiv (R + P)$ observations. We generate P forecasts using the model. For each time $t + 1$ in the prediction period, we use either a rolling sample $\{Z_{t-R+1}, \ldots, Z_t\}$ of size R or the whole past sample $\{Z_1, \ldots, Z_t\}$ to estimate the parameter λ. Let $\hat{\lambda}_t$ denote this estimator, then we can generate a sequence of one-step-ahead forecasts $\{f(Z_t, \hat{\lambda}_t)\}_{t=R}^{T-1}$.

Suppose that there is a decision maker who takes the one-step point forecast $f_{t,1} \equiv f(Z_t, \hat{\lambda}_t)$ of Y_t and uses it in some relevant decision. The one-step forecast error $e_{t+1} \equiv Y_{t+1} - f_{t,1}$ will result in a cost of $c(e_{t+1})$ where the function $c(e)$ will increase as e increases in size, but not necessarily symmetrically or continuously. The optimal forecast $f_{t,1}^*$ will be chosen to produce the forecast errors that minimize the expected loss

$$\min_{f_{t,1}} \int_{-\infty}^{\infty} c(y - f_{t,1})\, \mathrm{d}F_t(y), \tag{1}$$

where $F_t(y) \equiv \Pr(Y_{t+1} \leqslant y | I_t)$ is the conditional cumulative distribution function of Y_{t+1} and I_t is some proper information set at time t including Z_{t-j}, $j \geqslant 0$. The corresponding optimal forecast error is $e_{t+1}^* = Y_{t+1} - f_{t,1}^*$. From the first order condition of (1), it can be shown that the condition for forecast optimality is that the "generalized forecast error," $g_{t+1} \equiv \partial c(y - f_{t,1}^*)/\partial f_{t,1}$, has the martingale difference (MD) property, i.e. $E(g_{t+1}|I_t) = 0$ a.s. (Granger, 1999).

Now let us consider the two functions for $c(e)$. The first is the squared error loss $c(e) = e^2$. In this case, the generalized forecast error is $g_{t+1} = -2e_{t+1}^*$, which is an MD process. Therefore, the optimal forecast $f_{t,1}^*$ is the conditional mean $f_{t,1}^* = E(Y_{t+1}|I_t)$. The second loss function to consider is the check function $c(e) = \rho_\alpha(e)$ where $\rho_\alpha(e) \equiv [\alpha - 1(e < 0)] \times e$, for $\alpha \in (0, 1)$, and $1(\cdot)$ is the indicator function. The optimal forecast $f_{t,1}^*$ for a given α may be estimated from $\min_{f_{t,1}} \int_{-\infty}^{\infty} \rho_\alpha(y - f_{t,1})\, \mathrm{d}F_t(y)$. It can be shown (see, e.g., Giacomini and Komunjer, 2005, Lemma 1) that $E(\alpha - 1(Y_{t+1} < f_{t,1}^*)|I_t) = 0$ a.s. In this case, the generalized forecast error is $g_{t+1} = \alpha - 1(Y_{t+1} < f_{t,1}^*)$, which is an MD process. Therefore, the optimal forecast is the conditional α-quantile $f_{t,1}^* \equiv q_\alpha(Y_{t+1}|I_t) \equiv q_{\alpha,t}$.

Let $\{Y_t\}_{t=1}^T$ be a time series of continuously compounded daily returns on a financial asset. The VaR is simply a conditional quantile return which serves as an indicator of the maximum loss associated with a financial asset (or a portfolio) at a given confidence level. That is, $f_{t,1}^* \equiv q_\alpha(Y_{t+1}|I_t) \equiv q_{\alpha,t}$ is the VaR at $(1 - \alpha)$ confidence level, which means that with probability $(1 - \alpha)$, the return will not be lower than the VaR.

The RM model is based on the assumption that the return series follows a stochastic process of the form

$$Y_t = \mu_t(\lambda) + \varepsilon_t = \mu_t(\lambda) + \sigma_t(\lambda) z_t, \tag{2}$$

where λ is the vector that fully parameterizes the process, $\{z_t\}$ is an i.i.d. Gaussian sequence with zero mean and unit variance, $\mu_t(\lambda) = E[Y_t|I_{t-1}]$ and $\sigma_t^2(\lambda) = E[\varepsilon_t^2|I_{t-1}]$. The RM model sets

$\mu_t(\lambda) \equiv 0\ \forall t$ due to difficulties associated with precise estimation of expected returns. Then, the conditional α-quantile is

$$\mathrm{VaR}_\alpha \equiv q_{\alpha,t} = \Phi^{-1}(\alpha)\sigma_t(\lambda), \tag{3}$$

where $\Phi(\cdot)$ is the Gaussian distribution function, and the conditional variance is modeled as an exponentially weighted moving average (EWMA) of squared past returns

$$\sigma_t^2(\lambda) = (1-\lambda)\sum_{j=1}^{t-1}\lambda^{j-1}Y_{t-j}^2 \tag{4}$$

with decay factor $\lambda \in (0,1)$. For large t, (4) is well approximated by $\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)Y_{t-1}^2$, which allows easy incorporation of new information into volatility and VaR forecasts.

Regulators usually require the calculation of VaR for a 10-day horizon. RM uses the "square root of time" rule to extend daily VaR forecasts to horizons containing multiple trading days. Let $q_{\alpha,t}^h$ denote the h-day VaR at $(1-\alpha)$ confidence level, then we have $q_{\alpha,t}^h = \sqrt{h}q_{\alpha,t}$ where $q_{\alpha,t}$ is given above. Hence, the 10-day VaR is simply obtained by multiplying the daily VaR by $\sqrt{10}$.[3]

J.P. Morgan (1996) originally estimated the decay factor by minimizing the squared error loss for the conditional variance

$$\hat{\lambda}_t^h = \operatorname*{arg\,min}_{\lambda\in(0,1)} \frac{1}{t-h+1}\sum_{j=1}^{t-h+1}\left[h\sigma_j^2(\lambda) - \left(Y_j^h\right)^2\right], \tag{5}$$

where $Y_j^h = \sum_{l=0}^{h-1} Y_{j+l}$ (i.e. continuously compounded h-day return). Since the VaR $q_{\alpha,t}$ is a constant multiple of the estimated conditional volatility under the location-scale assumption (3), the loss function associated with the square of the VaR forecast is also the squared error loss function. However, the purpose of VaR analysis is forecasting the conditional quantile rather than volatility, thus the loss function for the estimation of λ should reflect this fact. In other words, the check loss function $\rho_\alpha(e) \equiv [\alpha - 1(e<0)] \times e$ should be used to estimate λ instead of the squared error loss $c(e) = e^2$. The optimal forecast $f_{t,1}^* \equiv q_\alpha(Y_{t+1}|I_t) \equiv q_{\alpha,t}$ is derived from the martingale property of the generalized forecast error.

Therefore, we consider estimating the RM model by minimizing the check loss, in which case the optimal decay factor is given by

$$\tilde{\lambda}_{t,\alpha}^h = \operatorname*{arg\,min}_{\lambda\in(0,1)} \frac{1}{t-h+1}\sum_{j=1}^{t-h+1}\rho_\alpha\left(e_j^h\right), \tag{6}$$

where $e_j^h = Y_j^h - q_{\alpha,j}^h$.

[3] Note that the square root of time rule is exact only when RM model is correctly specified and normality holds as noted by Tsay (2005). When we consider other specifications for conditional volatility and/or heavy tailed distributions, this rule serves as an approximation. There is not a generally agreed method to construct multiple day VaR forecasts under heavy tailed distributions. Danielsson and Vries (1997) suggest a scaling rule of the form $h^{1/\tau}$ where τ is the tail index. However, McNeil and Frey (2000) question the usefulness of this rule and argue that the scaling factor depends on the level of conditional volatility and provide some estimates. According to their results using $h^{1/2}$ is a very good approximation when volatility is high and is not too far from optimal for average volatility levels. Thus, when we construct the 10-day VaR forecasts under Student-t distribution below, we follow the square root of time rule. The interested reader is referred to Diebold et al. (1998) for a detailed discussion of the scaling issue in VaR modeling.

3. Empirical implementation

We have two aims in this note. The first aim is to quantify the differences in the estimates of the optimal decay factor using the check loss as opposed to the squared error loss function. The second aim is to examine how much we may lose in forecasting VaR by estimating the decay factor under squared error loss instead of using the check loss, or simply fixing it at 0.94, the value originally proposed by RiskMetrics.

As for the first aim, we estimate λ for the RM model by a simple grid search based on the aforementioned two loss functions.[4] We use the first 500 observations to get an initial conditional variance estimate from (4) and then use the recursive approximation to (4) to obtain the conditional variance (and thus VaR) forecasts for the rest of the sample. This yields $T-500$ conditional variances and VaR forecasts which are then used to evaluate the respective loss functions. This exercise is repeated for all 99 values of $\lambda \in \Lambda = \{0.01, 0.02, \ldots, 0.99\}$ to estimate λ (i.e., choose λ that minimizes the loss function in Eq. (5) or Eq. (6)). Note that for 10-day horizon we use overlapping 10-day returns.

As for the second aim of the paper, we compare the out-of-sample performance of VaR forecasts based on the squared error loss and the forecasts based on the check loss function, which is the appropriate criterion for estimation and forecast evaluation. We also consider the alternative of simply fixing λ at 0.94, the value proposed by RM and commonly used in practice. We estimate λ by minimizing the squared error loss and check loss functions using the most recent R observations based on the procedure described above in Eqs. (5) and (6). Then, $P = 1000$ VaR forecasts are obtained from (4) by using the most recent R observations for each prediction (i.e. a rolling window scheme is adopted).[5] These alternative strategies deliver three estimates of the decay factor: $\{0.94, \hat{\lambda}_t^h, \tilde{\lambda}_{t,\alpha}^h\}$. For each estimate, we compute the out-of-sample mean forecast check loss (MFCL) given by

$$MFCL_\alpha^h(\lambda_t) = \frac{1}{P-h+1} \sum_{t=R}^{T-h} \rho_\alpha\big(\hat{e}_{t+1}^h\big), \quad \hat{e}_{t+1}^h = Y_{t+1}^h - k_\alpha \sqrt{h}\hat{\sigma}_{t+1}(\lambda_t), \tag{7}$$

where $P = T - R$, $\lambda_t \in \{0.94, \hat{\lambda}_t^h, \tilde{\lambda}_{t,\alpha}^h\}$, $\hat{\sigma}_{t+1}(\lambda_t) = \lambda_t \hat{\sigma}_t(\lambda_t) + (1-\lambda_t)Y_t^2$, and the scaling factor k_α is either $\Phi^{-1}(\alpha)$ (conditional normality) or $k_\alpha = H^{-1}(\alpha)\sqrt{(\nu-2)/\nu}$ (conditional Student-t).[6] Note also that in the estimation of $\hat{\lambda}_t^h$ and $\tilde{\lambda}_{t,\alpha}^h$, described in (5) and (6), we use a rolling window of observations $\{Y_{t-R+1}, \ldots, Y_t\}$. We measure the departure of optimality (or "non-optimality") of the RM's VaR forecasts with respect to the following criteria:

[4] Estimation can also be performed by gradient based optimization methods under squared error loss or any other loss function that is continuous and differentiable. However, the check loss is not fully differentiable and we need to use more complicated methods like integer programming or the generic algorithm. Since the parameter of interest λ is bounded, a grid search is the simplest procedure that is viable under both loss functions. This also allows us to observe the variation in loss with respect to the entire range of the decay factor.

[5] Note that here we use the first 250 observations, instead of 500, to get an initial variance estimate from (4) due to the reduction in the size of the estimation sample.

[6] The original RM model assumes normal distribution. In our analysis we also consider the Student-t distribution to account for the well-known stylized fact that asset returns have leptokurtic distributions. We set the degrees of freedom parameter equal to 6. We have also experimented with different degrees of freedom such as 5 and 7 and we have obtained almost identical results. One can also estimate the degrees of freedom at the cost of implementing a more complex estimation procedure.

$$L_1(\alpha, h) = MFCL_\alpha^h(0.94)/MFCL_\alpha^h(\tilde{\lambda}_{t,\alpha}^h), \tag{8}$$

$$L_2(\alpha, h) = MFCL_\alpha^h(\hat{\lambda}_t^h)/MFCL_\alpha^h(\tilde{\lambda}_{t,\alpha}^h). \tag{9}$$

A value of $L_1(\alpha, h)$ or $L_2(\alpha, h)$ greater than unity indicates the non-optimality of the RM model for VaR forecasting, which is due to the estimation of λ using the squared error loss instead of the check loss.

We use daily return series from the three sets of data: (i) U.S. equity markets, (ii) foreign exchange (FX) markets, and (iii) U.S. bond markets. For each of these three sets, we consider the following three series. The three U.S. equity return series are the Dow-Jones Industrial index (DJ), NYSE Composite index (NYSE), and NASDAQ Composite index (NASDAQ). The three FX series considered are Swiss Frank (CHF), British Pound (GBP), and Japanese Yen (JPY). The three U.S. bond market yield series are obtained for the T-Bills with three month (3M), six month (6M), and one year (1Y) maturities.

The equity data starts on June 1, 1995 and ends on December 31, 2004 with a total of $T = 2411$ daily returns. For the FX data, the sample period is from January 3, 1995 to December 31, 2004 providing a total of $T = 2514$ observations. The bond data runs from January 3, 1995 to December 31, 2004 providing $T = 2501$ observations.

In Table 1 we report the estimation results of λ. In Panel A, we present the estimates for a 1-day horizon and in Panel B for a 10-day horizon. For the check loss, we consider $\alpha \in \{0.01, 0.05, 0.10\}$. Overall, there are discrepancies between the RM estimates and those based on the check loss function mostly for equities and bonds. The discrepancies are more notorious for the estimates based on the check loss function with $\alpha = 0.01$, which is the most relevant because the financial regulators require the calculation of VaR at the 99% confidence level. For exchange rates, both estimates are virtually identical with the exception of JPY at 1-day horizon. For the 1-day horizon, the largest discrepancy occurs for the NASDAQ returns under the Normal and the Student-t distributions. There are also substantial differences for 3M T-bill under both distributions. For the 10-day horizon, the discrepancies get larger for the equities returns either under normality or under Student-t with a tendency for the estimates of the RM model to overestimate the decay factor. For the NASDAQ returns, the RM model provides an estimate of 0.98 but the estimate based in the check loss ($\alpha = 0.01$) is only 0.83, and therefore attaching a much lower weight to the most recent variance.

In Table 2 we report the out-of-sample performance of VaR forecasts by comparing the MFCLs defined in Eqs. (8) and (9). Given the results of Table 1, we exclusively consider $\alpha = 0.01$. As before, in Panel A we present the results for a 1-day horizon and in Panel B for a 10-day horizon. Theoretically the ratios should be greater than or equal to one, but in practice issues of model uncertainty, parameter uncertainty, and model instability may deliver ratios smaller than one. It is very likely that the RM specification is not the best model for *all* the data sets considered in this paper so that the possibility of dealing with a misspecified model is not negligible. Under model misspecification it can be proven that the uncertainty of the forecast error depends not only on the variance of the innovation but also on the form of the misspecification, i.e. omitted variables, wrong functional form, structural breaks, etc. and consequently parameter uncertainty does not vanish even asymptotically. The weight of all these factors on the forecast error and on its associated loss is data-dependent. Therefore, the issue of out-of-sample predictive ability requires an empirical investigation. Roughly speaking, for the data sets considered we find that the MFCLs provided by the estimates based on the check loss are smaller than those based on the RM estimates (ratios (8) and (9) larger than one) in those instances for which we find the largest estimation discrepancies in Table 1, that is, for NASDAQ, JPY, and one-year T-bills. However an

Table 1
Decay factor estimates

Panel A. 1-day horizon

	Normal distribution				Student-*t* distribution			
	$\hat{\lambda}^1_T$	$\tilde{\lambda}^1_{T,0.01}$	$\tilde{\lambda}^1_{T,0.05}$	$\tilde{\lambda}^1_{T,0.10}$	$\hat{\lambda}^1_T$	$\tilde{\lambda}^1_{T,0.01}$	$\tilde{\lambda}^1_{T,0.05}$	$\tilde{\lambda}^1_{T,0.10}$
DJ	0.92	0.94	0.95	0.93	0.92	0.90	0.95	0.95
NYSE	0.91	**0.95**	0.93	**0.95**	0.91	**0.96**	**0.95**	**0.96**
NASDAQ	0.92	**0.85**	**0.96**	**0.97**	0.92	**0.81**	**0.96**	**0.97**
CHF	0.99	0.97	0.98	0.98	0.99	0.99	0.98	0.98
GBP	0.96	0.96	0.95	0.96	0.96	0.98	0.96	0.96
JPY	0.94	**0.98**	0.97	0.97	0.94	**0.98**	0.97	0.97
3M	0.91	0.88	0.91	**0.84**	0.91	**0.86**	0.91	**0.85**
6M	0.93	**0.99**	0.91	**0.89**	0.93	0.93	0.92	0.90
1Y	0.97	0.99	0.98	0.95	0.97	0.99	0.98	0.96

Panel B. 10-day horizon

	Normal distribution				Student-*t* distribution			
	$\hat{\lambda}^{10}_T$	$\tilde{\lambda}^{10}_{T,0.01}$	$\tilde{\lambda}^{10}_{T,0.05}$	$\tilde{\lambda}^{10}_{T,0.10}$	$\hat{\lambda}^{10}_T$	$\tilde{\lambda}^{10}_{T,0.01}$	$\tilde{\lambda}^{10}_{T,0.05}$	$\tilde{\lambda}^{10}_{T,0.10}$
DJ	0.99	0.99	0.99	0.99	0.99	**0.88**	0.99	0.99
NYSE	0.98	**0.92**	0.99	0.95	0.98	**0.89**	0.99	0.99
NASDAQ	0.98	**0.89**	0.95	0.96	0.98	**0.83**	0.96	0.96
CHF	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.97
GBP	0.99	0.99	0.98	0.99	0.99	0.99	0.98	0.98
JPY	0.98	0.99	0.99	0.99	0.98	0.99	0.99	0.99
3M	0.99	0.98	0.96	**0.88**	0.99	0.97	0.96	**0.89**
6M	0.92	**0.98**	**0.97**	0.94	0.92	**0.97**	**0.97**	0.93
1Y	0.99	0.96	0.98	0.99	0.99	**0.95**	0.98	0.99

Notes. This table reports decay factor estimates based on Eqs. (5) and (6). We use daily return series from the three sets of data: (i) U.S. equity markets (DJ, NYSE, NASDAQ), (ii) FX markets (CHF, GBP, JPY), and (iii) U.S. bond markets (3M, 6M, 1Y). The sample period for the three assets is approximately 10 years, from the beginning of 1995 to the end of 2004, with a sample size of about $T = 2500$. The scaling factor, $k_\alpha \equiv q_{\alpha,t}/\sigma_t(\lambda)$, is computed from the α-quantile of the normal distribution [$k_\alpha = \Phi^{-1}(\alpha)$], and from the α-quantile of the Student-t distribution [$k_\alpha = H^{-1}(\alpha)\sqrt{(\nu-2)/\nu}$ with $\nu = 6$]. The bold-font values indicate the largest discrepancies between the estimates provided by the RM model and those based on the check loss function.

overall assessment of the results of Table 2 indicates that the out-of-sample forecasting performance of VaR forecasts based on the RM specification under both loss functions, check loss and the squared error loss, is very similar, that is, the mean statistical losses are roughly equivalent.

4. Concluding remarks

Our objective has been to investigate the implications from considering different loss functions in the estimation and the forecasting stages. Though there are not general guidelines regarding the choice of the loss function, in the VaR modeling we find an instance in which the loss function for forecasting evaluation is well defined. Focusing on the popular RM model we have addressed two issues in this paper. First, we have quantified the differences in the estimates of the optimal decay factor using the check loss as opposed to the squared error loss. Second, we have compared the out-of-sample performance of VaR forecasts for which the decay

Table 2
Out-of-sample performance of VaR forecasts comparison of loss functions

Panel A. 1-day horizon

	Normal distribution		Student-t distribution	
	$L_1(0.01, 1)$	$L_2(0.01, 1)$	$L_1(0.01, 1)$	$L_2(0.01, 1)$
DJ	0.997	0.997	1.002	0.999
NYSE	0.969	0.989	0.955	0.961
NASDAQ	1.001	0.998	**1.019**	1.011
CHF	1.010	0.981	1.012	0.999
GBP	0.957	0.975	0.950	0.967
JPY	**1.037**	**1.021**	**1.047**	**1.035**
3M	0.939	0.942	0.930	0.927
6M	0.981	0.977	0.981	0.976
1Y	**1.072**	**1.054**	**1.061**	**1.040**

Panel B. 10-day horizon

	Normal distribution		Student-t distribution	
	$L_1(0.01, 10)$	$L_2(0.01, 10)$	$L_1(0.01, 10)$	$L_2(0.01, 10)$
DJ	**1.022**	1.005	**1.065**	**1.061**
NYSE	0.992	1.007	0.910	0.909
NASDAQ	1.009	0.969	**1.019**	1.007
CHF	1.002	0.975	1.001	0.990
GBP	1.003	1.000	0.988	1.000
JPY	0.993	0.888	0.972	0.884
3M	0.941	**1.081**	0.896	**1.076**
6M	0.858	0.910	0.874	0.912
1Y	0.948	1.003	0.981	**1.023**

Notes. This table reports the ratios of the out-of-sample mean forecast check loss (MFCL) defined in Eqs. (8) and (9). A value greater than one indicates that the RM model is suboptimal. The scaling factor, $k_\alpha \equiv q_{\alpha,t}/\sigma_t(\lambda)$, is computed from the α-quantile of the normal distribution [$k_\alpha = \Phi^{-1}(\alpha)$], and from the α-quantile of the Student-t distribution [$k_\alpha = H^{-1}(\alpha)\sqrt{(\nu-2)/\nu}$ with $\nu = 6$]. The bold-font values indicate the largest discrepancies between the MFCLs provided by the estimates of the RM model and by those based on the check loss function.

factor is estimated under squared error loss or simply fixed at 0.94 instead of estimating it using the check loss. Our empirical results indicate that on calculating the 99% VaR, the RM model for the equity markets tends to overestimate the decay factor for a 10-day horizon. However, the out-of-sample forecasting results show that one may not necessarily gain in terms of predictive ability on a systematic basis by estimating λ using the check loss function.

Acknowledgment

We thank an anonymous referee for useful comments.

References

Andersen, T.G., Bollerslev, T., Christoffersen, P., Diebold, F.X., 2006. Practical volatility and correlation modeling for financial risk management. In: Carey, M., Stulz, R. (Eds.), Risks of Financial Institutions. University of Chicago Press for NBER, pp. 513–548.

Christoffersen, P., Jacobs, K., 2004. The Importance of the loss function in option valuation. Journal of Financial Economics 72, 291–318.

Danielsson, J., de Vries, C.G., 1997. Value-at-Risk and extreme returns. FMG discussion paper No. 273, Financial Markets Group, London School of Economics.

Diebold, F.X., Hickman, A., Inoue, A., Schuermann, T., 1998. Converting 1-day volatility to h-day volatility: Scaling by square root-h is worse than you think. Wharton Financial Institutions Center, Working Paper 97-34. Published in condensed form as: Scale models, Risk 11, 104–107.

Giacomini, R., Komunjer, I., 2005. Evaluation and combination of conditional quantile forecasts. Journal of Business and Economic Statistics 23, 416–431.

Granger, C.W.J., 1999. Outline of forecast theory using generalized cost functions. Spanish Economic Review 1, 161–174.

J.P. Morgan Guaranty Trust Company (1996). RiskMetrics—Technical Document, fourth ed.

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46, 33–50.

Komunjer, I., 2005. Quasi-maximum likelihood estimation for conditional quantiles. Journal of Econometrics 128, 137–164.

McNeil, A.J., Frey, R., 2000. Estimation of tail related risk measures for heteroskedastic financial time series: An extreme value approach. Journal of Empirical Finance 7, 271–300.

Perignon, C., Smith, D., 2006. The Level and Quality of Value-at-Risk Disclosure by Commercial Banks. Available at SSRN: http://ssrn.com/abstract=952595.

Tsay, R., 2005. Analysis of Financial Time Series, second ed. Wiley, New York.

West, K.D., 1996. Asymptotic inference about prediction ability. Econometrica 64, 1067–1084.

White, H., 1994. Estimation, Inference, and Specification Analysis. Cambridge Univ. Press, Cambridge.