Semiparametric Estimation of Value-at-Risk

Jianqing Fan
Juan Gu

Shanghai-Hong Kong Development Institute

Hong Kong Institute of Asia-Pacific Studies

June 2003

Shanghai-Hong Kong Development Institute
Occasional Paper No. 7

About the Authors

Jianqing Fan is Professor of Statistics and Chairman of Department of Statistics, The Chinese University of Hong Kong.

Juan Gu is postdoctor fellow, GF Securities Co., Ltd., Guangzhou, China.

Acknowledgements

This paper was partially supported by the RGC grant CUHK 4299/00P and a direct grant from The Chinese University of Hong Kong.

ISBN 962-441-807-1

Semiparametric Estimation of Value-at-Risk

1. Introduction

Risk management has become an important topic for financial institutions, regulators, nonfinancial corporations and asset managers. Value-at-Risk (VaR) is a measure for gauging the market risks of a particular portfolio. VaR shows the maximum loss over a given time horizon at a given confidence level. The review article by Duffie and Pan (1997) as well as the books edited by Alexander (1998) and written by Dowd (1998) and Jorion (2000) provide a good introduction to the subject.

The field of risk management has evolved very rapidly, and many new techniques have since been developed. Aït-Sahalia and Lo (2000) introduced the concept of the economic valuation of VaR and compared it with the statistical VaR. Other methods include historical simulation approaches and their modifications (Hendricks, 1996; Mahoney, 1996); techniques based on parametric models (Wong and So, 2000), such as GARCH models (Bollerslev, 1986; Engle, 1995) and their approximations; estimates based on extreme value theory (Embrechts, Klüppelberg and Mikosch, 1997) and ideas based on variance-covariance matrices (Davé and Stahl, 1997). The problems of bank capital and VaR were studied in Jackson, Maude and Perraudin (1997). The accuracy of various VaR estimates was compared and studied by Beder (1995) and Davé and Stahl (1997). Engle and Manganelli (1999) introduced a family of VaR estimators, called CAViaR, using the idea of regression quantile.

An important contribution to the calculation of VaR is the RiskMetrics of J. P. Morgan (1996). The method can be regarded as a nonparametric estimation of volatility together with a normality assumption on the return process. The estimator of VaR consists of two steps. The first step is to estimate the volatility of holding a portfolio for one day before converting this into the volatility of multiple days. The second step is to compute the quantile of standardized return processes through the assumption that the processes follow a standard normal distribution. Following this important contribution by J. P. Morgan, many of the subsequent techniques that have been developed share a similar principle.

Many techniques in use are local parametric methods. By using the historical data at a given time interval, parametric models such as GARCH(1,1) or even GARCH(0,1) were built. For example, the historical simulation method can be regarded as a local nonparametric estimation of quantiles. The techniques by Wong and So (2000) can be regarded as modeling a local stretch of data by using a GARCH model. In comparison, the volatility estimated by the RiskMetrics is a kernel estimator of observed square returns, which is essentially an average of the observed volatilities over the past 38 days (see Section 2.1). From the function approximation point of view (Fan and Gijbels, 1996), this method basically assumes that the volatilities of the last 38 days are nearly constant or that the return processes are locally modeled by a GARCH(0,1) model. The latter can be regarded as a discretized version of the geometric Brownian over a short time period for the prices of a held portfolio.

An aim of this paper is to introduce a time-dependent semiparametric model to enhance the flexibility of local approximations. This model is an extension of the time-homogeneous parametric model for term structure dynamics used by Chan et al. (1992). The pseudo-likelihood technique of Fan et al. (forthcoming) will be employed to estimate the local parameters. The volatility estimates of the return processes will then be formed.

The windows over which the local parametric models can be employed are frequently chosen subjectively. For example, in the RiskMetrics, decay factors of 0.94 and 0.97 are recommended by J. P. Morgan for computing daily volatilities and for calculating monthly volatilities (defined as a holding period of 25 days), respectively. It is clear that a large window size will reduce the variability of estimated local parameters. However, this will increase modeling biases (approximation errors). Therefore, a compromise between these two contradictory demands is the art required for the selection of a smoothing parameter in nonparametric techniques (Fan and Gijbels, 1996). Another aim of this paper is to propose new techniques for automatically selecting the window size or, more precisely, the decay parameter. This will allow us to use a different amount of smoothing for different portfolios to better estimate their volatilities.

With estimated volatilities, the standardized returns for a portfolio can be formed and a quantile of this return process is needed for estimating VaR. The RiskMetrics uses the quantile of the standard normal distribution. This can be improved by estimating the quantiles from the standardized return process. In this paper, a new nonparametric technique based on the symmetric assumption of the distribution of the return process is proposed. This increases the statistical efficiency by more than a factor of two as compared with the usual sample quantiles. While it is known that the distribution of asset returns is asymmetric, the asymmetry of the percentiles at a moderate level of percentages α is not very severe. Our experience shows that moderate α efficiency gains can still be made by using symmetric quantile methods. In addition, the proposed technique is robust against the misspecification of parametric models and outliers created by large market movements. By contrast, parametric techniques for estimating quantiles have a higher statistical efficiency for estimated quantiles when the parametric models fit well with the return process. Therefore, in order to ascertain whether this gain will materialize, we also fit parametric t-distributions with an unknown scale and unknown degree of freedom to the standardized return. The method of quantiles and the method of moments are proposed for estimating unknown parameters and, hence, the quantiles. The former approach is more robust, while the latter is more efficient.

Economic and market conditions vary from time to time. It is reasonable to expect that the return process of a portfolio and its stochastic volatility will depend in some way on time. Therefore, a viable VaR estimate should have the ability to self-revise the procedure in order to adapt to changes in market conditions. This includes modifications of the procedures for estimating both volatility and quantiles. A time-dependent procedure is proposed for estimating VaR and has been empirically tested. It shows positive results.

The outline of the paper is as follows: Section 2 revisits the volatility estimation of J. P. Morgan's RiskMetrics before going on to introduce semiparametric models for return processes. Two methods for choosing time-independent and time-dependent decay factors are proposed. The effectiveness of the proposed volatility estimators is evaluated using several measures. Section 3 examines the problems of estimating the quantiles of normalized return processes. A nonparametric technique and two parametric approaches are introduced. Their relative statistical efficiencies are studied. Their efficacies for estimating VaR are compared with J. P. Morgan's method. In Section 4, newly proposed volatility estimators and quantile estimators are combined to yield new estimators for VaR. Their performances are thoroughly tested by using simulated data as well as data from eight stock indices. Section 5 summarizes the conclusions of this paper.

2. Estimation of Volatility

Let S_t be the price of a portfolio at time t. Let

$$r_t = \log(S_t/S_{t-1})$$

be the observed return at time t. The aggregate return at time t for a predetermined holding period τ is

$$R_{t,\tau} = \log(S_{t+\tau-1}/S_{t-1}) = r_t + \cdots + r_{t+\tau-1}.$$

Let Ω_t be the historical information generated by the process $\{S_t\}$; namely, Ω_t is the σ-field generated by $S_t, S_{t-1}, \cdots$. If S_t denotes the current market value of a portfolio, then the value of this portfolio at time $t + \tau$ will be $S_{t+\tau} = S_t \exp(R_{t+1,\tau})$. The VaR measures the extreme loss of a portfolio over a predetermined holding period τ with a prescribed confidence level $1 - \alpha$. More precisely, letting $V_{t+1,\tau}$ be the α-quantile of the conditional distribution of $R_{t+1,\tau}$:

$$P(R_{t+1,\tau} > V_{t+1,\tau} \mid \Omega_t) = 1 - \alpha$$

with probability $1 - \alpha$, the maximum loss of holding this portfolio for a period of τ is $S_t V_{t+1,\tau}$; namely, the VaR is $S_t V_{t+1,\tau}$. See the books by Jorion (2000) and Dowd (1998).

The current value S_t is known at time t. Thus, most efforts in the literature concentrate on estimating $V_{t+1,\tau}$. A popular approach to predicting VaR is to determine, first, the conditional volatility

$$\sigma^2_{t+1,\tau} = \mathrm{Var}(R_{t+1,\tau} \mid \Omega_t)$$

and then the conditional distribution of the scaled variable $R_{t+1,\tau}/\sigma_{t+1,\tau}$. This is also the approach that we follow.

2.1 *Revisiting the RiskMetrics*

An important technique for estimating volatility is the RiskMetrics, which estimates the volatility of a one-period return ($\tau = 1$) $\sigma_t^2 \equiv \sigma_{t,1}^2$ according to

$$\hat{\sigma}_t^2 = (1 - \lambda)\, r_{t-1}^2 + \lambda \hat{\sigma}_{t-1}^2, \tag{2.1}$$

with $\lambda = 0.94$. For a τ-period return, the square-root rule is frequently used in practice:

$$\hat{\sigma}_{t,\tau} = \sqrt{\tau}\, \hat{\sigma}_t. \tag{2.2}$$

J. P. Morgan recommends using (2.2) with $\lambda = 0.97$ to forecast the monthly ($\tau = 25$) volatility of aggregate return. The Bank for International Settlements (Basle Committee on Banking Supervision, 1995) suggests using (2.2) for the capital requirement of a holding period of 10 days. In fact, Wong and So (2000) showed that for the IGARCH(1,1) model defined similarly to (2.1), the square-root rule (2.2) holds. Beltratti and Morana (1999) employed the square-root rule with GARCH models to daily and half-hourly data. By iterating (2.1), it can be easily seen that

$$\hat{\sigma}_t^2 = (1-\lambda)\left\{ r_{t-1}^2 + \lambda r_{t-2}^2 + \lambda^2 r_{t-3}^2 + \cdots \right\}. \tag{2.3}$$

This is an example of exponential smoothing in a time domain (see Fan and Yao, 2003). Figure 1 depicts the weights for several choices of λ.

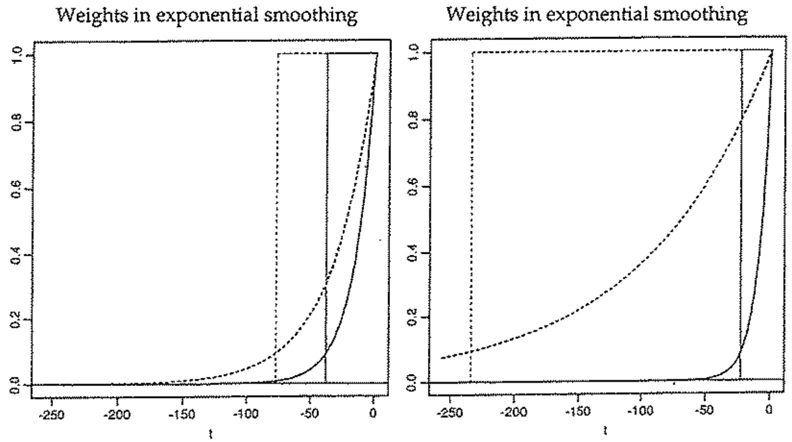
Exponential smoothing can be regarded as a kernel method that uses the one-sided kernel $K_1(x) = b^x I(x > 0)$ with $b < 1$. Assuming $E(r_t \mid \Omega_{t-1}) = 0$, then $\sigma_t^2 = E(r_t^2 \mid \Omega_{t-1})$. The kernel estimator of $\sigma_t^2 = E(r_t^2 \mid \Omega_{t-1})$ is given by

$$\hat{\sigma}_t^2 = \frac{\sum_{i=-\infty}^{t-1} K_1((t-i)/h_1) r_i^2}{\sum_{i=-\infty}^{t-1} K_1((t-i)/h_1)} = \frac{\sum_{i=-\infty}^{t-1} b^{\frac{t-i}{h_1}} r_i^2}{\sum_{i=-\infty}^{t-1} b^{\frac{t-i}{h_1}}},$$

where h_1 is the bandwidth (see Fan and Yao, 2003). It is clear that this is exactly the same as (2.3) with $\lambda = b^{1/h_1}$.

Exponential smoothing has the advantage of gradually, rather than radically, reducing the influence of remote data points. However, the effective number of points used to compute the local average is hard to quantify. If the one-sided uniform kernel $K_2(x) = I[0 < x \leq 1]$ with bandwidth h_2 is used, then it is clear that h_2 data points have been used to compute the local average. According to the equivalent kernel theory (Section 5.4 of Fan and Yao, 2003), the kernel estimator with a kernel function K_1 and a

Figure 1 Weights for the exponential smoothing with several parameters and the weights of their corresponding equivalent uniform kernels



Notes: Left panel, solid curve: $\lambda = 0.94$; left panel, dashed curve: $\lambda = 0.97$; right panel, solid curve: $\lambda = 0.90$; and right panel, dashed curve: $\lambda = 0.99$.

bandwidth h_1 and the kernel estimator with a kernel function K_2 and a bandwidth h_2 conduct approximately the same amount of smoothing when

$$h_2 = \alpha(K_2) h_1 / \alpha(K_1),$$

where

$$\alpha(K) = \left\{ \int_{-\infty}^{\infty} u^2 K(u) du \right\}^{-2/5} \left\{ \int_{-\infty}^{\infty} K^2(u) du \right\}^{1/5}$$

It is clear that $\alpha(K_2) = 3^{0.4} = 1.5518$ and that the exponential smoothing corresponds to the kernel smoothing with $K_1(x) = \lambda^x I(x > 0)$ and $h_1 = 1$. Hence, it uses effectively

$$h_2 = 1.5518/\alpha(K_1).$$

Table 1 records the effective number of data points used in the exponential smoothing.

Assume now the model

$$r_t = \sigma_t \varepsilon_t, \tag{2.4}$$

where ε_t is a sequence of independent random variables with a mean of 0 and a variance of 1. It is well-known that the kernel method can be derived from a local constant approximation (Fan and Gijbels, 1996). Assuming that $\sigma_u \approx \theta$ for u in a neighborhood of a point t, i.e.

$$r_u \approx \theta\varepsilon_u, \quad \text{for } u \approx t \tag{2.5}$$

then the kernel estimator or, specifically, the exponential smoothing estimator (2.1) can be regarded as a solution to the local least-squares problem:

$$\sum_{i=-\infty}^{t-1} (r_i^2 - \theta)^2 \lambda^{(t-i-1)}, \tag{2.6}$$

where λ is a decay factor (smoothing parameter) that controls the size of the local neighborhood (Figure 1).

Table 1 Effective number of data points used in the exponential smoothing

Parameter λ	0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99
Effective number h_2	22.3	24.9	28.2	32.4	38.0	45.8	57.6	77.2	116.4	234.0

From the above function approximation point of view, the J. P. Morgan estimator of volatility assumes that, locally, the return process follows the model (2.5). The model can, therefore, be regarded as a discretized version of the geometric Brownian motion with no drift $d\log(S_u) = \theta dW_u$, for u around t or

$$d\log(S_u) = \theta(u) dW_u, \tag{2.7}$$

when the time unit is small, where W_u is the Wiener process.

2.2 Semiparametric Models

The implicit assumption of J. P. Morgan's estimation of volatility is the local geometric Brownian motion on stock price dynamics. To reduce a possible modeling bias and to enhance the flexibility of the approximation, we enlarge the model (2.7) to the following semiparametric time-dependent model

$$d\log(S_u) = \theta(u) S_u^{\beta(u)}\, dW_u, \tag{2.8}$$

allowing volatility to depend on the value of the asset, where $\theta(u)$ and $\beta(u)$ are the coefficient functions. When $\beta(u) \equiv 0$, the model reduces to (2.7). This time-dependent diffusion model was used for interest rate dynamics by Fan et al. (forthcoming). It is an extension of the time-dependent models previously considered, among others, by Hull and White (1990), Black, Derman and Toy (1990), and Black and Karasinski (1991) and the time-independent model considered by Cox, Ingersoll and Ross (1985) and Chan et al. (1992). Unlike the yields of bonds, the scale of $\{S_u\}$ can be very different over a large time period. However, the model (2.8) is used locally, rather than globally.

Motivated by a continuous-time model (2.8), we model the return process at the discrete time as

$$r_u = \theta(u) S_{u-1}^{\beta(u)}\, \varepsilon_u, \tag{2.9}$$

where ε_u is a sequence of independent random variables with a mean of 0 and a variance of 1. To estimate the parameters $\theta(u)$ and $\beta(u)$, the local pseudo-likelihood technique is employed. For each given t and $u \le t$ in a neighborhood of time t, the functions $\theta(u)$ and $\beta(u)$ are approximated by constants:

$$\theta(u) \approx \theta, \quad \beta(u) \approx \beta .$$

Then, the conditional log-likelihood for r_u given S_{u-1} is

$$-\frac{1}{2}\log(2\pi\theta^2 S_{u-1}^{2\beta}) - \frac{r_u^2}{2\theta^2 S_{u-1}^{2\beta}},$$

when $\varepsilon_u \sim N(0,1)$. In general, the above likelihood is a pseudo-likelihood. Dropping the constant factors and adding the pseudo-likelihood around the point t, we obtain the locally weighted pseudo-likelihood

$$l(\theta, \beta) = -\sum_{i=-\infty}^{t-1}\left\{\log(\theta^2 S_{i-1}^{2\beta}) + \frac{r_i^2}{\theta^2 S_{i-1}^{2\beta}}\right\}\lambda^{t-1-i}, \tag{2.10}$$

where $\lambda < 1$ is the decay factor that makes this pseudo-likelihood use only the local data [see Figure 1 and (2.6)]. Maximizing (2.10) with respect to the local parameters θ and β yields an estimate of the local parameters $\theta(t)$ and $\beta(t)$. Note that for a given β, the maximum is achieved at

$$\hat{\theta}^2(t, \beta) = (1-\lambda)\sum_{i=-\infty}^{t-1}\lambda^{t-1-i} r_i^2 S_{i-1}^{-2\beta} .$$

Substituting this into (2.10), the pseudo-likelihood $l(\hat{\theta}(t, \beta), \beta)$ is obtained. This is a one-dimensional maximization problem, and the maximization can easily be obtained by, for example, searching β over a grid of points or by using other, more advanced, numerical methods. Let $\hat{\beta}(t)$ be the maximizer. Then, the estimated volatility for a one-period return is

$$\hat{\sigma}_t^2 = \hat{\theta}^2(t) S_{t-1}^{2\hat{\beta}(t)}, \tag{2.11}$$

where $\hat{\theta}(t) = \hat{\theta}(t, \hat{\beta}(t))$. In particular, if we let $\beta(t) = 0$, the model (2.9) becomes the model (2.7) and the estimator (2.11) reduces to the J. P. Morgan estimator (2.3).

Our method corresponds to time-domain smoothing, which uses mainly the most recent data. There is also a large literature that postulates models on $\mathrm{Var}(r_t \mid F_{t-1}) = g(r_{t-1}, \cdots, r_{t-p})$. This corresponds to the state-domain smoothing, using mainly the historical data to estimate the function g. See Engle and Manganelli (1999), Yang, Härdle and Nielsen (1999), Härdle and Yatchew (2002) and Fan and Yao (2003). A combination of both time-domain and state-domain smoothing for volatility estimation is an interesting direction for future research.

2.3 *Choice of Decay Factor*

The performance of volatility estimation depends on the choice of decay factor λ. In the J. P. Morgan RiskMetrics, $\lambda = 0.94$ is recommended for the estimation of one-day volatility, while $\lambda = 0.97$ is recommended for the estimation of monthly volatility. In general, the choice of decay factor should depend on the portfolio and holding period, and should be determined from data.

Our idea is related to minimizing the prediction error. In the current pseudo-likelihood estimation context, our aim is to maximize the pseudo-likelihood. For example, suppose that we have observed the price process $S_t, t = 1, \cdots, T$. Note that the pseudo-likelihood estimator $\hat{\sigma}_t^2$ depends on the data up to time $t-1$. This estimated volatility can be used to predict the volatility at time t. The estimated volatility $\hat{\sigma}_t^2$ given by (2.11) can then be compared with the observed volatility r_t^2 for the effectiveness of the estimation. One way to validate the effectiveness of the prediction is to use square prediction errors

$$\mathrm{PE}(\lambda) = \sum_{t=T_0}^{T} (r_t^2 - \hat{\sigma}_t^2)^2, \tag{2.12}$$

where T_0 is an integer such that $\hat{\sigma}_{T_0}^2$ can be estimated with reasonable accuracy. This avoids the boundary problem caused by the exponential smoothing (2.1) or (2.3). The decay factor λ can be chosen to minimize (2.12). Using the model (2.9) and noting that $\hat{\sigma}_t$ is Ω_{t-1} measurable, the expected value can be decomposed as

$$E\{\mathrm{PE}(\lambda)\} = \sum_{t=T_0}^{T} E(\sigma_t^2 - \hat{\sigma}_t^2)^2 + \sum_{t=T_0}^{T} E(r_t^2 - \sigma_t^2)^2. \tag{2.13}$$

Note that the second term is independent of λ. Thus, the point of minimizing $\mathrm{PE}(\lambda)$ is to find an estimator that minimizes the mean-square error

$$\sum_{t=T_0}^{T} E(\sigma_t^2 - \hat{\sigma}_t^2)^2.$$

The question naturally arises why square errors, rather than other types of errors, such as absolute deviation errors, should be used in (2.12). In the current pseudo-likelihood context, a natural alternative is to maximize the pseudo-likelihood defined as

$$\mathrm{PL}(\lambda) = -\sum_{t=T_0}^{T} (\log \hat{\sigma}_t^2 + r_t^2/\hat{\sigma}_t^2), \tag{2.14}$$

compared to (2.10). The likelihood function is a natural measure of the discrepancy between r_t and $\hat{\sigma}_t$ in the current context, and does not depend on an arbitrary choice of distance. The summand in (2.14) is the conditional likelihood, after dropping constant terms, of r_t given S_{t-1} with unknown parameters replaced by their estimated values. The decay factor λ can then be chosen to maximize (2.14). For simplicity in later discussion, we call this procedure the Semiparametric Estimation of Volatility (SEV).

2.4 Choice of Adaptive Smoothing Parameter

The above choice of decay factor remains constant during the post-sample forecasting. It relies heavily on past history and has little flexibility to accommodate changes in stock dynamics over time. Therefore, in order to adapt automatically to changes in stock price dynamics, the decaying parameter λ should be allowed to depend on the time t. A solution to such problems has been explored by Mercurio and Spokoiny (forthcoming) and Härdle, Herwartz and Spokoiny (forthcoming).

To highlight possible changes of the dynamics of $\{S_t\}$, the validation should be localized around the current time t. Let h be a period for which we wish to validate the effectiveness of the volatility estimation. The pseudo-likelihood is then defined as

$$\mathrm{PL}(\lambda, t) = -\sum_{i=t-1-h}^{t-1} (\log \hat{\sigma}_i^2 + r_i^2/\hat{\sigma}_i^2). \tag{2.15}$$

Let $\hat{\lambda}_t$ maximize (2.15). In our implementation, we use $h = 20$, which validates the estimates in a period of about one month. The choice of $\hat{\lambda}_t$ is variable. To reduce this variability, the series $\{\hat{\lambda}_t\}$ can be smoothed further by using the exponential smoothing:

$$\hat{\Lambda}_t = b\hat{\Lambda}_{t-1} + (1-b)\,\hat{\lambda}_t\,. \tag{2.16}$$

In our implementation, we use $b = 0.94$.

To sum up, in order to estimate the volatility $\hat{\sigma}_t$, we first compute $\{\hat{\sigma}_u\}$ and $\{\hat{\Lambda}_u\}$ up to time $t-1$ and obtain $\hat{\lambda}_t$ by minimizing (2.15) and then $\hat{\Lambda}_t$ by (2.16). The value of $\hat{\Lambda}_t$ is then used in (2.10) to estimate the local parameters $\hat{\theta}(t)$ and $\hat{\beta}(t)$, and hence the volatility $\hat{\sigma}_t^2$ using (2.11). The resulting estimator will be referred to as the Adaptive Volatility Estimator (AVE).

The techniques in this section and in Section 2.3 apply directly to the J. P. Morgan type of estimator (2.1). They allow for different decay parameters for different portfolios.

2.5 Numerical Results

In this section, the newly proposed procedures are compared by using three commonly-used methods: J. P. Morgan's RiskMetrics, the historical simulation and the GARCH model using the Quasi-maximum likelihood method (denoted by "GARCH"). See Engle and Gonzalez-Rivera (1991) and Bollerslev and Wooldridge (1992). For the estimation of volatility, the historical simulation method is simply defined as the sample standard deviation of the return process for the past 250 days. For the newly proposed method, we employ the semiparametric estimator (2.11) with $\lambda = 0.94$ (denoted by "Semipara"); the estimator (2.11) with λ chosen by minimizing (2.12) (denoted by "SEV"); and the estimator (2.11) (denoted by "AVE") with the decay factor $\hat{\lambda}_t$ chosen adaptively as in (2.16).

To compare the different procedures for estimating the volatility with a holding period of one day, eight stock indices and two simulated data were used together with the following three performance measures. For other related measures, see Davé and Stahl (1997). For holding period of one day, the error distribution is not very far from normal.

Measure 1: Exceedance Ratio (ER) against Confidence Level

This measure counts the number of events for which the loss of assets exceeds the loss predicted by the normal model at a given confidence α. With estimated volatility, under the normal model, the one-day VaR is estimated by $\Phi^{-1}(\alpha)\hat{\sigma}_t$, where $\Phi^{-1}(\alpha)$ is the α quantile of the standard normal distribution. For each estimated VaR, the ER is computed as

$$\mathrm{ER} = n^{-1} \sum_{t=T+1}^{T+n} I(r_t < \Phi^{-1}(\alpha)\hat{\sigma}_t),$$

for a post sample of size n. This gives an indication of how effectively volatility can be used to estimate a one-period VaR. Note that the Monte Carlo error for this measure has an approximate size $\{\alpha(1-\alpha)/n\}^{1/2}$, even when the true σ_t is used. For example, with $\alpha = 5\%$ and $n = 1000$, the Monte Carlo error is around 0.68%. Thus, unless the post-sample size is large enough, this measure has difficulty in differentiating between various estimators due to the presence of large error margins.

Measure 2: Mean Absolute Deviation Error (MADE)

To motivate this measure, let us first consider the mean-square errors:

$$\mathrm{PE} = n^{-1} \sum_{t=T+1}^{T+n} (r_t^2 - \hat{\sigma}_t^2)^2.$$

Following (2.13), the expected value can be decomposed as

$$E(\mathrm{PE}) = n^{-1} \sum_{t=T+1}^{T+n} E(\sigma_t^2 - \hat{\sigma}_t^2)^2 + n^{-1} \sum_{t=T+1}^{T+n} E(r_t^2 - \sigma_t^2)^2.$$

Note that the first term reflects the effectiveness of the estimated volatility while the second term is the size of the stochastic error, independent of estimators. As in all statistical prediction problems, the second term is usually of a larger order of magnitude than the first term. Thus, a small improvement on PE could mean a substantial improvement over the estimated volatility. However, due to the well-known fact that financial time series contain outliers due to market crashes, the mean-square error is not a robust measure. Therefore, we will use the MADE:

$$\text{MADE} = n^{-1} \sum_{t=T+1}^{T+n} | r_t^2 - \hat{\sigma}_t^2 | .$$

Measure 3: Square-root Absolute Deviation Error (RADE)

An alternative variation to MADE is the RADE, which is defined as

$$\text{RADE} = n^{-1} \sum_{t=T+1}^{T+n} \left| | r_t | - \sqrt{\frac{2}{\pi}} \hat{\sigma}_t \right| .$$

The constant factor comes from the fact that $E|\varepsilon_t| = \sqrt{\frac{2}{\pi}}$ for $\varepsilon_t \sim N(0,1)$.

Measure 4: Test of Independence

A good VaR estimator should have the property that the sequence of the events exceeding VaR behaves like an i.i.d. Bernoulli distribution with a probability of success α. Engle and Manganelli (1999) gave an illuminating example showing that even a bad VaR estimator can have the exceedance ratio α.

Let $I_t = I(r_t < \Phi^{-1}(\alpha)\hat{\sigma}_t)$ be the indicator of the event in which the return exceeds the VaR. Christoffersen (1998) introduced the likelihood ratio test for testing independence and for testing whether the probability $\Pr(I_t = 1) = \alpha$.

Assume $\{I_t\}$ is a first-order Markovian chain. Let $\pi_{ij} = \Pr(I_t = j \mid I_{t-1} = i)$ ($i = 0,1$ and $j = 0,1$) be the transition probability and n_{ij} be the number of events transferring from state i to state j in the post-sample period. The problem is to test

$$H_0 : \pi_{00} = \pi_{10} = \pi, \quad \pi_{01} = \pi_{11} = 1 - \pi$$

Then, the maximum likelihood ratio test for independence is

$$\text{LR1} = 2 \log \left(\frac{\hat{\pi}_{00}^{n_{00}} \hat{\pi}_{01}^{n_{01}} \hat{\pi}_{10}^{n_{10}} \hat{\pi}_{11}^{n_{11}}}{\hat{\pi}^{n_0} (1 - \hat{\pi})^{n_1}} \right), \tag{2.17}$$

where $\hat{\pi}_{ij} = n_{ij}/(n_{ij} + n_{i,1-j})$, $n_j = n_{0j} + n_{1j}$, and $\hat{\pi} = n_0/(n_0 + n_1)$. The test statistic is a measure of deviation from independence. Under the null hypothesis, the test statistic LR1 is distributed approximately according to χ_1^2 when the sample size is large. Thus, reporting the test statistic is equivalent to reporting the p-value.

Measure 5: Testing against a Given Confidence Level

Christoffersen (1998) applied the maximum likelihood ratio test to the problem

$$H_0 : P(I_t = 1) = \alpha \quad \text{versus} \quad H_1 : P(I_t = 1) \neq \alpha$$

under the assumption that $\{I_t\}$ is a sequence of i.i.d. Bernoulli random variables. The test statistic is given by

$$\text{LR2} = 2 \log \left(\frac{\hat{\pi}^{n_0}(1 - \hat{\pi})^{n_1}}{\alpha^{n_0}(1 - \alpha)^{n_1}} \right), \tag{2.18}$$

which follows the χ_1^2-distribution when the sample size is large. Again, the reported p-value is a measure of the deviation from the null hypothesis. This measure is closely related to Measure 1.

Example 1: Stock Indices

We first apply the five volatility estimators to the daily returns of eight stock indices (Table 2). For each stock index, the in-sample period terminated on December 31, 1996 and the post-sample period was from January 1, 1997 to December 31, 2000 ($n = 1014$).

Table 2 Comparisons of several volatility estimation methods

Country	Index	In-sample period	Post-sample period
Australia	AORD	1988-1996	1997-2000
France	CAC 40	1990-1996	1997-2000
Germany	DAX	1990-1996	1997-2000
Hong Kong	HSI	1988-1996	1997-2000
Japan	Nikkei 225	1988-1996	1997-2000
United Kingdom	FTSE 100	1988-1996	1997-2000
United States	S&P 500	1988-1996	1997-2000
United States	Dow Jones	1988-1996	1997-2000

The results are summarized in Table 3. The initial period where σ_t with $t \le T_0$ is not estimated is set to $T_0 = 250$.

From Table 3, the two smallest MADE and RADE are almost always achieved by using semiparametric methods and GARCH methods. In fact, SEV, AVE and GARCH methods are the three best methods in terms of MADE and RADE. Of these, the semiparametric method with a decay parameter (SEV) that is selected automatically by the data performs the best. It achieved the two smallest MADE in a total of eight out eight times, and the two smallest RADE four out eight times. This demonstrates that it is important to allow the algorithm to choose decay factors according to the dynamics of stock prices. The AVE and GARCH methods perform comparably with the SEV in terms of MADE and RADE. The GARCH method slightly outperforms the AVE according to MADE and RADE measures, but AVE outperforms the GARCH method for other measures such as ER and p-value from independence. This demonstrates the advantage of using a time-dependent decay parameter that adapts automatically to any changes in stock price dynamics. These results also indicate that

Table 3 Results of the comparisons of several volatility estimation methods

Index	Method	ER ($\times10^{-2}$)	MADE ($\times10^{-4}$)	RADE ($\times10^{-3}$)	p-value (indep)	p-value (ER=5%)
AORD	Historical	5.23	0.850	4.330	0.01*	0.76
	RiskMetrics	4.93	0.848	4.250	0.01*	0.92
	Semipara	4.93	0.840	4.231	0.01*	0.92
	SEV	5.23	**0.803**	4.300	0.02	0.76
	AVE	4.93	0.835	**4.213**	0.00*	0.92
	GARCH	5.52	**0.786**	**4.168**	0.12	0.46
CAC 40	Historical	6.16	2.177	7.272	0.11	0.10
	RiskMetrics	6.36	2.157	7.102	0.33	0.06
	Semipara	6.55	2.150	7.094	0.20	0.03
	SEV	6.45	**2.077**	**7.035**	0.17	0.04
	AVE	6.36	2.138	7.069	0.91	0.06
	GARCH	8.24	**1.931**	**6.883**	0.20	0.00
DAX	Historical	6.55	2.507	7.814	0.18	0.03
	RiskMetrics	5.46	2.389	7.370	0.59	0.51
	Semipara	5.56	2.389	7.342	0.90	0.43
	SEV	6.06	**2.368**	7.457	0.03	0.14
	AVE	5.96	2.377	**7.330**	0.90	0.18
	GARCH	7.94	**2.200**	**7.174**	0.78	0.00*
HSI	Historical	6.08	5.710	11.167	0.00*	0.12
	RiskMetrics	5.99	5.686	10.818	0.00*	0.15
	Semipara	5.89	5.567	**10.685**	0.00*	0.19
	SEV	5.61	**5.523**	10.743	0.00*	0.37
	AVE	6.55	5.578	10.686	0.00*	0.03
	GARCH	7.30	**5.293**	**10.565**	0.03	0.00*
Nikkei 225	Historical	5.78	2.567	7.824	0.90	0.27
	RiskMetrics	5.78	2.526	7.656	0.35	0.27
	Semipara	6.09	2.507	7.631	0.48	0.13
	SEV	5.68	**2.457**	**7.610**	0.68	0.34
	AVE	6.19	**2.479**	**7.565**	0.25	0.10
	GARCH	5.88	2.563	7.693	0.80	0.22

Table 3 Results of the comparisons of several volatility estimation methods (continued)

Index	Method	ER ($\times 10^{-2}$)	MADE ($\times 10^{-4}$)	RADE ($\times 10^{-3}$)	p-value (indep)	p-value (ER=5%)
FTSE 100	Historical	6.83	1.369	5.761	0.05	0.01
	RiskMetrics	5.94	1.342	5.594	0.45	0.18
	Semipara	6.24	1.328	**5.567**	0.58	0.08
	SEV	6.93	**1.299**	5.598	0.02	0.01
	AVE	6.04	1.328	5.571	0.90	0.14
	GARCH	7.43	**1.256**	**5.497**	0.78	0.00*
S&P 500	Historical	6.34	1.613	6.027	0.65	0.06
	RiskMetrics	5.55	1.647	6.056	0.62	0.44
	Semipara	5.55	1.620	5.995	0.62	0.44
	SEV	5.85	**1.539**	**5.888**	0.46	0.23
	AVE	5.75	**1.611**	**5.984**	0.51	0.29
	GARCH	4.46	1.689	6.163	0.89	0.43
Dow Jones	Historical	6.15	1.493	5.840	0.25	0.11
	RiskMetrics	5.65	1.507	5.784	0.56	0.35
	Semipara	5.75	1.489	**5.739**	0.16	0.29
	SEV	5.75	**1.460**	5.743	0.51	0.29
	AVE	5.85	**1.480**	**5.731**	0.90	0.23
	GARCH	4.46	1.575	5.960	0.68	0.43

Notes: GARCH refers to the GARCH(1,1) model. Numbers with bold face are the two smallest.
* means statistically significant at the 1% level.

our proposed methods for selecting decay parameters are effective. As shown in (2.13), both measures contain a large amount of stochastic errors. A small improvement in MADE and RADE measures indicates a large improvement in terms of estimated volatilities.

Presented in Table 3 are the p-values for testing independence and for testing whether the ER is significantly from 5%. Since the post-sample size is more than 1000, we consider whether the deviations are significant at the 1% level. Most methods have a right ER, except for the GARCH method which tends to underestimate the risk. However, the GARCH method performs particularly well in terms of testing against independence. Its corresponding p-values tend to be large. Other methods perform reasonably well in terms of independence.

As an illustration, Figure 2 presents the estimated volatilities for six stock indices in the post-sample period by using SEV and AVE. The parameters β's in model (2.11) depend on the stock prices and can vary substantially. Since together they predict the volatility, it is more meaningful to present the volatility plots. The volatility predicted by the AVE is more variable than that by the SEV.

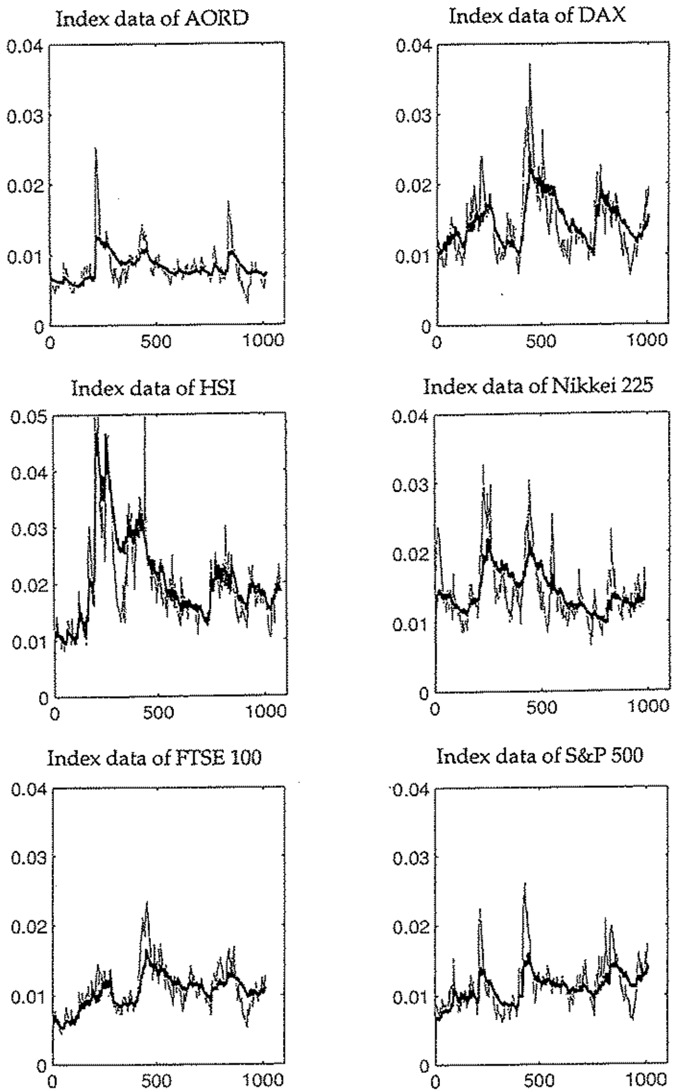
Example 2: GARCH(1,1) Model

Next, consider simulations from the GARCH model:

$$r_t = \sigma_t \varepsilon_t , \quad \sigma_t^2 = c + a\sigma_{t-1}^2 + b r_{t-1}^2 ,$$

where ε_t is the standard Gaussian noise. The first two hundred random series of length 3000 were simulated using the parameters $c = 0.00000038$, $a = 0.957513$ and $b = 0.038455$. These parameters are from the GARCH(1,1) fit to the S&P 500 index from January 4, 1988 to December 29, 2000. The parameter a is reasonably close to the $\lambda = 0.94$ of the RiskMetrics. The second two hundred time series of length 3000 were simulated using the parameters $c_0 = 0.000009025$, $a = 0.9$ and $b = 0.09$. The choice of c is to make the resulting series have approximately the same standard deviation as the returns of the S&P 500. The first 2000 data points were used as the in-sample period, namely $T = 2000$, and the last 1000 data points were used as the post-sample, namely $n = 1000$. The perfor-

Figure 2 Predicted volatility in the post-sample period for several indices



Note: Thin curves: AVE method; thick curves = SEV method.

mance of six volatility estimators of the two models is shown in Tables 4 and 5, respectively.

The performance of each volatility estimator can be summarized by using the average and standard deviation of MADE and RADE over 200 simulations. However, MADE and RADE show quite a large variability from one simulation to another. In order to avoid taking averages over different scales, for each simulated series we first standardize the MADE using the median MADE in that series of the six methods and then average them across 200 simulations. The results are presented as the column "score" in Tables 4 and 5. In addition, the frequency of each method that achieved the best MADE among 200 simulations was recorded, and is presented in the column "best." Further, the frequency of each volatility estimator that achieved the two smallest MADE in each simulation was also counted. More precisely, among 200 simulations, we computed the percentage of a method that performed the best as well as the percentages of the methods ranked in the top two positions. The results are presented in the column "best two" of Tables 4 and 5. For clarity, we omit similar presentations using the RADE measure — the results are nearly the same as with the MADE. The numbers of rejections of null hypotheses are recorded in the columns "reject times (indep)" and "reject times (ER = 5%)."

Using MADE or RADE as a measure, AVE and SEV consistently outperform other methods. GARCH performs quite reasonably in terms of MADE for the second GARCH(1,1) model, but not for the first GARCH(1,1) model. Since the sum of the parameters a and b is close to one, the parameters in GARCH(1,1) cannot be estimated without a large variability. This results in large variances in the computation of a standardized MADE. In terms of ER or Measure 5, which are closely related, RiskMetrics performs consistently well. Since the models used in the simulations are all stationary time-homogeneous models, AVE does not have much of an advantage while SEV performs better in terms of ER and Measure 5. Except for the historical simulation method, all methods behave well in the independence tests (Measure 4).

Table 4 Comparisons of several volatility estimation methods: the first GARCH(1,1) model

Method	ER ($\times 10^{-2}$)	Score ($\times 10^{-2}$)	Best	Best two	Reject times (indep)	Reject times (ER=5%)
Historical	5.32 (0.92)	103.48 (4.87)	13.5	18	14	29
RiskMetrics	5.43 (0.56)	100.17 (0.49)	2	11.5	21	8
Semipara	5.67 (0.61)	99.81 (0.42)	11	35	18	20
SEV	5.44 (0.66)	99.73 (0.50)	16.5	51.5	15	17
AVE	5.94 (0.63)	99.30 (0.58)	51	75	12	53
GARCH(1,1)	4.41 (0.86)	103.67 (4.97)	6	9	23	42

Note: The values in the brackets are their corresponding standard deviations.

Table 5 Comparisons of several volatility estimation methods: the second GARCH(1,1) model

Method	ER ($\times 10^{-2}$)	Score ($\times 10^{-2}$)	Best	Best two	Reject times (indep)	Reject times (ER=5%)
Historical	5.58 (0.98)	109.64 (9.92)	3.5	4	65	49
RiskMetrics	5.54 (0.60)	100.30 (0.60)	1	10.5	14	14
Semipara	5.72 (0.62)	99.75 (0.37)	11.5	47	14	30
SEV	5.75 (0.62)	99.73 (0.42)	14	41	15	33
AVE	6.06 (0.61)	99.53 (0.69)	37	60.5	11	63
GARCH(1,1)	5.03 (0.78)	100.67 (2.65)	33	37	15	14

Note: The values in the brackets are their corresponding standard deviations.

Example 3: Continuous-time Stochastic Volatility (SV) Model

Instead of simulating the data from GARCH(1,1) models, we simulate data from a continuous-time diffusion process with the SV:

$$d\log(S_t) = \alpha dt + \sigma_t dW_t, \qquad d\sigma_t^2 = \kappa(\theta - \sigma_t^2)dt + \omega\sigma_t dB_t,$$

where W_t and B_t are two independent standard Brownian motions. See, for example, Barndorff-Nielsen and Shephard (2001, 2002). The parameters are chosen as $\alpha = 0$, $\kappa = 0.21459$, $\theta = 0.08571$, $\omega = 0.07830$, following Chapman and Pearson (2000) and Fan and Zhang (2003). Two hundred series of 3000 daily data were simulated using the exact simulation method (e.g., Genon-Catalot, Jeantheau and Laredo, 1999; Fan and Zhang, 2003).

This simulation tests the extent to which the six volatility estimators perform when the underlying dynamics differ from GARCH(1,1) and our semiparametric models. The same performance measures as those in Example 2 are used. Table 6 summarizes the results. Similar conclusions to those in Example 2 can be drawn. The AVE and SEV consistently outperform the RiskMetrics using MADE as a measure, even when the model is misspecified. This is due, mainly, to the flexibility of the semiparametric model in approximating the true dynamics, in addition to the data-driven smoothing parameter that enhanced the performance. The historical simulation method performs better in this example than those in the previous example. This is partially due to the fact that the SV model produces more volatile returns. Hence, a larger smoothing parameter in the historical simulation method gives it some advantages.

3. Estimation of Quantiles

The conditional distribution of the multiple period return $R_{t,\tau}$ does not necessarily follow a normal distribution. Indeed, even under

Table 6 Comparisons of several volatility estimation methods: SV model

Method	ER ($\times 10^{-2}$)	Score ($\times 10^{-2}$)	Best	Best two	Reject times (indep)	Reject times (ER=5%)
Historical	5.05 (0.74)	100.02 (1.94)	16.5	34	5	14
RiskMetrics	5.45 (0.59)	100.62 (0.59)	0	3	11	12
Semipara	5.90 (0.65)	100.38 (0.67)	0	5	10	45
SEV	5.64 (0.71)	98.69 (1.42)	34	66	11	30
AVE	6.06 (0.63)	99.02 (0.64)	32	72	8	60
GARCH(1,1)	4.10 (1.77)	108.73 (13.54)	17.5	20	9	116

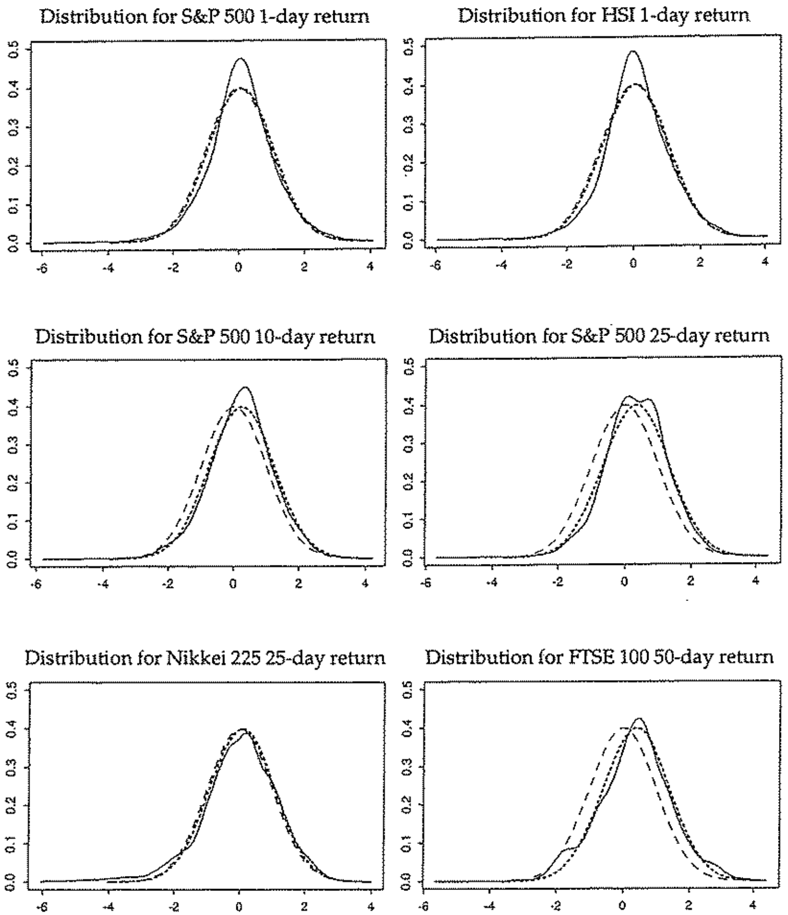
Note: The corresponding standard deviations are in the brackets.

the IGARCH(1,1) model (2.1) with a normal error distribution in (2.4), Wong and So (2000) showed that the conditional distribution $R_{t,\tau}$ given Ω_t is not normal. This was also illustrated numerically by Lucas (2000). Thus, a direct application of $\Phi^{-1}(\alpha)\,\hat{\sigma}_{t+1,\tau}$ will provide an erroneous estimate of the multiple period VaR, where $\hat{\sigma}_{t+1,\tau}$ is an estimated multiple period volatility of returns. In order to provide empirical evidence of the non-normality of the multiple period returns, the distributions $R_{t,\tau}/\hat{\sigma}_{t,\tau}$ for the S&P 500, HSI, Nikkei 225, and FTSE 100 indices are sown in Figure 3. The multiple period volatility is computed using the J. P. Morgan RiskMetrics: $\hat{\sigma}_{t+1,\tau} = \sqrt{\tau}\hat{\sigma}_{t+1}$. The densities are estimated by the kernel density estimator with the rule of thumb bandwidth $h = 1.06n^{-1/5}s$, where n is the sample size and s is the sample standard deviation (taken as 1 to avoid outliers, since the data have already been normalized). See, for example, Chapter 2 of Fan and Gijbels (1996). It is evident that the one-period distributions are basically symmetric and have heavier tails than the standard normal distribution. The deviations from normal are quite substantial for multiple period return processes. Indeed, the distribution is not centered around 0; the centered normal distributions (using medians of the data as the centers and 1 as the standard deviation) fit the data better.

3.1 Nonparametric Estimation of Quantiles

As discussed previously, the distributions of the multiple period returns deviate from normal. Their distributions are generally unknown. In fact, Diebold et al. (1998) reported that converting one-day volatility estimates to τ-day estimates by a scale factor $\sqrt{\tau}$ is inappropriate and produces overestimates of the variability of long time horizon volatility. Danielsson and de Vries (2000) suggested using the scaling factor $\tau^{1/\beta}$, with β being the tail index of extreme value distributions. Nonparametric methods can naturally be used to estimate the distributions of the residuals and correct the biases in the volatility estimation (the issue of whether

Figure 3 Estimated densities of the rescaled multiple period returns for several indices



Notes: Solid curves: estimated densities by using the kernel density estimator;
dashed curves: standard normal densities;
thick dashed curves: normal densities centered at the median of the data with a standard deviation of 1.

the scale factor is correct becomes irrelevant when estimating the distribution of standardized return processes).

Let $\hat{\sigma}_{t,\tau}$ be an estimated τ-period volatility and $\hat{\varepsilon}_{t,\tau} = R_{t,\tau}/\hat{\sigma}_{t,\tau}$ be a residual. Denote by $\hat{q}(\alpha,\tau)$, the sample α-quantile of the residuals $\{\hat{\varepsilon}_{t,\tau}, t = T_0 + 1, \cdots, T - \tau\}$. This yields an estimated multiple period VaR of $\text{VaR}_{t+1,\tau} = \hat{q}(\alpha,\tau)\hat{\sigma}_{t+1,\tau}$. Note that the choice of constant factor $\hat{q}(\alpha,\tau)$ is the same as selecting the constant factor c such that the difference between the ER of the estimated VaR and the confidence level is minimized in the in-sample period. More precisely, $\hat{q}(\alpha,\tau)$ minimizes the function

$$\text{ER}(c) = \left|(T - \tau - T_0 + 1)^{-1} \sum_{t=T_0}^{T-\tau} I(R_{t+1,\tau} < c\hat{\sigma}_{t+1,\tau}) - \alpha\right| .$$

The nonparametric estimates of quantiles are robust against the mis-specification of parametric models and insensitive to a few market movements for a moderate α. Yet they are not as efficient as parametric methods when parametric models are correctly given. To improve the efficiency of nonparametric estimates, we assume that the distribution of $\{\hat{\varepsilon}_{t,\tau}\}$ is symmetric at about the point 0. This implies that

$$q(\alpha,\tau) = -q(1 - \alpha,\tau) ,$$

where $q(\alpha, \tau)$ is the population quantile. Thus, an improved nonparametric estimator is

$$\hat{q}^{[1]}(\alpha,\tau) = 2^{-1}\{\hat{q}(\alpha,\tau) - \hat{q}(1 - \alpha,\tau)\} . \quad (3.1)$$

Denote by

$$\text{VaR}^{[1]}_{t+1,\tau} = \hat{q}(\alpha,\tau)^{[1]}\hat{\sigma}_{t+1,\tau}$$

the corresponding estimated VaR. It is not difficult to show that the estimator $\hat{q}^{[1]}(\alpha,\tau)$ is a factor of $\frac{2-2\alpha}{1-2\alpha}$ that is as efficient as the simple estimate $\hat{q}(\alpha,\tau)$ for $\alpha < 0.5$ (see Appendix A.1 for derivations).

When the distribution of the standardized return process is asymmetric, (3.1) will introduce some biases. For a moderate α where $q(\alpha,\tau) \approx -q(1-\alpha,\tau)$, the biases are offset by the variance gain. As shown in Figure 3, the asymmetry for returns is not very severe for a moderate α. Hence, the gain can still be materialized.

3.2 *Adaptive Estimation of Quantiles*

The above method assumes that the distribution of $\{\hat{\varepsilon}_{t,\tau}\}$ is stationary over time. To accommodate possible nonstationarity, for a given time t, we may only use the local data $\{\hat{\varepsilon}_{i,\tau}, i = t-\tau-h, t-h+1, \cdots, t-\tau\}$. This model was used by several authors, including Wong and So (2000) and Pant and Chang (2001). Let the resulting nonparametric estimator (3.1) be $\hat{q}_t^{[1]}(\alpha,\tau)$. To stabilize the estimated quantiles, we further smooth this quantile series to obtain the adaptive estimator of quantiles $\hat{q}_t^{[2]}(\alpha,\tau)$ via the exponential smoothing:

$$\hat{q}_t^{[2]}(\alpha,\tau) = b\hat{q}_{t-1}^{[2]}(\alpha,\tau) + (1-b)\hat{q}_{t-1}^{[1]}(\alpha,\tau). \tag{3.2}$$

In our implementation, we took $h = 250$ and $b = 0.94$.

3.3 *Parametric Estimation of Quantiles*

Based on empirical observations, one possible parametric model for the observed residuals $\{\hat{\varepsilon}_{t,\tau}, t = T_0+1, \cdots, T-\tau\}$ is to assume that the residuals follow a scaled t-distribution.

$$\hat{\varepsilon}_{t,\tau} = \lambda \varepsilon_t^*, \tag{3.3}$$

where $\varepsilon_t^* \sim t_\nu$, the Student's t-distribution with degree of freedom ν. The parameters λ and ν can be obtained by solving the following equations that are related to the sample quantiles:

$$\begin{cases} \hat{q}(\alpha_1,\tau) = \lambda t(\alpha_1,\nu) \\ \hat{q}(\alpha_2,\tau) = \lambda t(\alpha_2,\nu) \end{cases},$$

where $t(\alpha,\nu)$ is the α quantile of the t-distribution with degree of freedom ν. A better estimator to use is $\hat{q}^{[1]}(\alpha,\tau)$ in (3.1). Using the improved estimator and solving the above equations yield the estimates $\hat{\nu}$ and $\hat{\lambda}$ as follows:

$$\frac{t(\alpha_2,\hat{\nu})}{t(\alpha_1,\hat{\nu})} = \frac{\hat{q}^{[1]}(\alpha_2,\tau)}{\hat{q}^{[1]}(\alpha_1,\tau)}, \quad \hat{\lambda} = \frac{\hat{q}^{[1]}(\alpha_1,\tau)}{t(\alpha_1,\hat{\nu})}. \tag{3.4}$$

Hence, the estimated quantile is given by

$$\hat{q}^{[3]}(\alpha,\tau) = \hat{\lambda} t(\alpha,\hat{\nu}) = \frac{t(\alpha,\hat{\nu})\hat{q}^{[1]}(\alpha_1,\tau)}{t(\alpha_1,\hat{\nu})}, \tag{3.5}$$

and the VaR of a τ-period return is given by

$$\mathrm{VaR}_{t+1,\tau}^{[3]} = \hat{q}^{[3]}(\alpha,\tau)\hat{\sigma}_{t+1,\tau}. \tag{3.6}$$

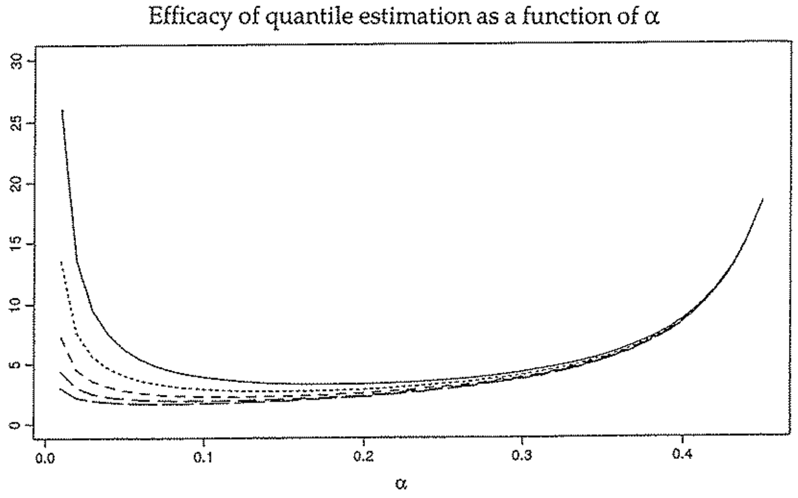
In the implementation, we take $\alpha_1 = 0.15$ and $\alpha_2 = 0.35$. This choice is near optimal in terms of statistical efficiency (Figure 4).

The above method of estimating quantiles is robust against outliers. An alternative approach is to use the method of moments to estimate parameters in (3.3). Note that if $\varepsilon \sim t_\nu$ with $\nu > 4$, then

$$E\varepsilon^2 = \frac{\nu}{\nu-2}, \text{ and } E\varepsilon^4 = \frac{3\nu^2}{(\nu-2)(\nu-4)}.$$

The method of moments yields the following estimates

Figure 4 The efficacy function $g_\nu(\alpha)$ for several degrees of freedom



Notes: From solid, the shortest dash to the longest dash: $\nu = 2$, $\nu = 3$, $\nu = 5$, $\nu = 10$ and $\nu = 40$.
For all νs, the minimum is almost attained at the interval [0.1, 0.2].

$$\begin{cases} \hat{\nu} = (4\hat{\mu}_4 - 6\hat{\mu}_2^2)/(\hat{\mu}_4 - 3\hat{\mu}_2^2) \\ \hat{\lambda} = \{\hat{\mu}_2(\hat{\nu} - 2)/\hat{\nu}\}^{1/2} \end{cases} \tag{3.7}$$

where $\hat{\mu}_j$ is the j^{th} moment, defined as $\hat{\mu}_j = (T - \tau - T_0)^{-1} \sum_{t=T_0+1}^{T-\tau} \hat{\varepsilon}_{t,\tau}^j$. See Pant and Chang (2001) for similar expressions. Using these estimated parameters, we obtain the new estimated quantile and estimated VaR similarly to (3.5) and (3.6). The new estimates are denoted by $\hat{q}^{[4]}(\alpha,\tau)$ and $\mathrm{VaR}_{t+1,\tau}^{[4]}$, respectively. That is,

$$\hat{q}^{[4]}(\alpha,\tau) = \hat{\lambda} t(\alpha,\hat{\nu}), \quad \mathrm{VaR}_{t+1,\tau}^{[4]} = \hat{q}^{[4]}(\alpha,\tau)\hat{\sigma}_{t+1,\tau}.$$

The method of moments is less robust than the method of quantiles. The former also requires the assumption that $\nu > 4$. We will compare their asymptotic efficiency in Section 3.4.

3.4 *Theoretical Comparisons of Estimators for Quantiles*

Of the three methods of estimating quantiles, the estimator $\hat{q}^{[1]}(\alpha,\tau)$ is the most robust method. It imposes very mild assumptions on the distribution of $\hat{\varepsilon}_{t,\tau}$ and, hence, is robust against model mis-specificaton. The two parametric methods rely on the model (3.3), which could lead to erroneous estimations if the model is mis-specified. The estimators $\hat{q}^{[1]}$, $\hat{q}^{[2]}$ and $\hat{q}^{[3]}$ are all robust against outliers, but $\hat{q}^{[4]}$ is not.

In order to give a theoretical study of the properties of the aforementioned three methods for the estimation of quantiles, we assume that $\{\hat{\varepsilon}_{t,\tau}, t = T_0, \cdots, T-\tau\}$ is an independent random sample from the density f. Under this condition, for $0 < \alpha_1 < \cdots < \alpha_k < 1$,

$$\{\sqrt{m}[\hat{q}(\alpha_i,\tau) - q(\alpha_i,\tau)], 1 \le i \le k\} \xrightarrow{L} N(0,\Sigma), \tag{3.8}$$

where $m = T - \tau - T_0 + 1$, $q(\alpha_i,\tau)$ is the population quantile of f, and $\Sigma = (\sigma_{ij})$ with

$$\sigma_{ij} = \alpha_i(1-\alpha_j)/f(q(\alpha_i,\tau))f(q(\alpha_j,\tau)), \text{ for } i > j \text{ and } \sigma_{ji} = \sigma_{ij}.$$

See Prakasa Rao (1987).

To compare this with parametric methods, let us now assume for a moment that the model (3.3) is correct. Using the result in

Appendix A.1, the nonparametric estimator $\hat{q}^{[1]}(\alpha,\tau)$ follows an asymptotically normal distribution with a mean of $\lambda t(\alpha,\nu)$ and a variance (for $\alpha < 1/2$) of

$$V_1(\alpha,\nu,\lambda) = \frac{\lambda^2\alpha(1-2\alpha)}{2f_\nu(t(\alpha,\nu))^2 m}, \tag{3.9}$$

where f_ν is the density of the t-distribution with degree of freedom ν given by

$$f_\nu(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}(1+x^2/\nu)^{-(\nu+1)/2}.$$

Since ν is an integer, any consistent estimator of ν means that it equals to ν with a probability tending to 1. For this reason, ν can be treated as known in the asymptotic study. It follows directly from (3.8) that the estimator $\hat{q}^{[3]}(\alpha,\tau)$ has the asymptotic normal distribution with a mean of $\lambda t(\alpha,\nu)$ and a variance of

$$V_2(\alpha,\alpha_1,\nu,\lambda) = \frac{\lambda^2 t(\alpha,\nu)^2\alpha_1(1-2\alpha_1)}{2f_\nu(t(\alpha_1,\nu))^2 t(\alpha_1,\nu)^2 m}. \tag{3.10}$$

The efficiency of V_2 depends on the choice of α_1 through the function

$$g_\nu(\alpha) = \frac{\alpha(1-2\alpha)}{f_\nu(t(\alpha,\nu))^2 t(\alpha,\nu)^2}.$$

The function $g_\nu(\alpha)$ for several choices of ν is presented in Figure 4. It is clear that the choices of α_1 in the range [0.1, 0.2] are nearly optimal for all values of ν. For this reason, $\alpha_1 = 0.15$ is chosen throughout this paper.

As explained previously, ν can be treated as being known. Under this assumption, as shown in Appendix A.2, the method of moments estimator $\hat{q}^{[4]}(\alpha,\tau)$ is asymptotically normal with a mean of $\lambda t(\alpha,\nu)$ and a variance of

$$V_3(\alpha,\nu,\lambda) = \frac{\lambda^2(\nu-1)t(\alpha,\nu)^2}{2(\nu-4)m}. \tag{3.11}$$

Table 7 depicts the relative efficiency of the three estimators. The nonparametric estimator $\hat{q}^{[1]}$ always performs better than the parametric quantile estimator $\hat{q}^{[2]}$, unless the α and ν are small. The former is more robust against any mis-specification of the model (3.3). The nonparametric estimator $\hat{q}^{[1]}$ is more efficient than the method of moments estimator $\hat{q}^{[3]}$ when the degree of freedom is small; and has reasonable efficiency when ν is large. This, together with the robustness of the nonparametric estimator $\hat{q}^{[1]}$ to mis-specification of models and outliers, indicates that our newly proposed nonparametric estimator is generally preferable than the method of moments. This finding is consistent with our empirical studies.

Table 7 Relative efficiency for three estimators of quantiles

	$\alpha = 5\%$			$\alpha = 1\%$		$\alpha = 10\%$	
ν	V_1/λ^2	V_2/V_1	V_3/V_1	V_2/V_1	V_3/V_1	V_2/V_1	V_3/V_1
2	26.243	1.071	∞	0.253	∞	1.708	∞
3	10.928	1.348	∞	0.394	∞	1.873	∞
4	7.117	1.532	∞	0.518	∞	1.965	∞
5	5.528	1.659	1.469	0.621	0.550	2.024	1.792
6	4.682	1.750	1.008	0.706	0.406	2.064	1.189
7	4.164	1.820	0.862	0.775	0.367	2.093	0.992
10	3.381	1.952	0.729	0.922	0.344	2.146	0.801
20	2.666	2.119	0.663	1.138	0.356	2.210	0.691
40	2.373	2.208	0.647	1.266	0.371	2.242	0.657
100	2.214	2.263	0.642	1.351	0.383	2.261	0.641

In summary, in consideration of the fact that the return series have heavily tails and contain outliers due to large market movements, and in light of its high statistical efficiency even in parametric models, the nonparametric estimator $\hat{q}^{[1]}$ is the most preferred for $\alpha = 5\%$. Among the two parametric methods, the method of moments may be preferred because of its high efficiency when the degree of freedom is large.

3.5 Empirical Comparisons of Quantile Estimators

We compared, empirically, the performances of three different methods for quantile estimation by using the eight stock indices in Example 1. To make the comparison easier, the same volatility estimator to all indices and holding periods was applied. The multi-period estimation of volatility in the RiskMetrics is employed for $\tau = 10$, $\tau = 25$ and $\tau = 50$. The estimated quantiles $\hat{q}^{[j]}(\alpha,\tau)$ (j = 1, 2, 3, 4) are used to obtain the estimated VaR. The effectiveness of the estimated VaR is measured by the ER

$$\text{ER} = (n-\tau)^{-1} \sum_{t=T+1}^{T+n-\tau} I(R_{t,\tau} < \text{VaR}_{t,\tau})$$

in the post-sample period (January 1, 1997 to December 31, 2000 with n = 1014). Note that due to the insufficient number of non-overlapping τ-day intervals in the post-sample period, overlapping intervals are used. This has two advantages: (1) it increases the number of intervals by a factor of approximately τ and (2) the results are insensitive to the starting date of the post-sample period (not the case for non-overlapping intervals). The confidence level $1 - \alpha = 95\%$ is used. To compare with the performance of the RiskMetrics, we also form $\text{VaR}^{[0]}_{t+1,\tau} = \Phi^{-1}(\alpha)\hat{\sigma}_{t+1,\tau}$. This follows exactly the recommendations of J. P. Morgan's RiskMetrics. The results are presented in Table 8.

Table 8 Comparisons of the ERs of several VaR estimators ($\times 10^{-2}$)

Index	Holding period	$\text{VaR}^{[0]}$	$\text{VaR}^{[1]}$	$\text{VaR}^{[2]}$	$\text{VaR}^{[3]}$	$\text{VaR}^{[4]}$
AORD	$\tau = 10$	5.42	3.55	5.52	2.37	1.28
	$\tau = 25$	5.52	2.56	5.23	1.08	0.99
	$\tau = 50$	1.28	0.10	3.75	0.10	0.00
CAC 40	$\tau = 10$	3.67	3.67	3.48	3.48	2.88
	$\tau = 25$	4.47	4.57	4.07	3.57	3.48
	$\tau = 50$	2.78	3.38	2.48	2.78	2.38
DAX	$\tau = 10$	4.87	4.27	5.06	5.06	4.67
	$\tau = 25$	5.66	4.97	5.06	6.95	5.36
	$\tau = 50$	4.07	3.97	3.77	4.27	3.97
HSI	$\tau = 10$	6.55	4.49	4.96	5.52	4.40
	$\tau = 25$	9.17	6.08	8.04	5.89	5.71
	$\tau = 50$	7.11	2.99	5.14	3.46	3.55
Nikkei 225	$\tau = 10$	4.97	4.26	5.98	3.65	4.06
	$\tau = 25$	6.59	4.56	7.20	4.46	4.67
	$\tau = 50$	11.76	2.54	11.56	8.22	5.17
FTSE 100	$\tau = 10$	4.75	4.06	5.05	3.47	4.06
	$\tau = 25$	3.96	2.48	5.35	3.17	2.77
	$\tau = 50$	2.67	2.48	4.95	2.08	2.67
S&P 500	$\tau = 10$	3.77	3.67	5.05	2.38	4.36
	$\tau = 25$	3.17	3.57	5.05	3.96	3.67
	$\tau = 50$	2.78	3.77	5.95	3.87	3.77
Dow Jones	$\tau = 10$	5.06	4.96	5.95	3.77	5.16
	$\tau = 25$	4.86	5.16	7.24	6.65	5.75
	$\tau = 50$	3.47	3.87	4.66	3.67	4.07

To summarize the performance presented in Table 8, we computed the average and standard deviation of the ERs for a holding period $\tau = 10, 25, 50$ for each of the eight stock indices. In addition, the MADEs from the nominal level $\alpha = 5\%$ were also computed. Table 9 depicts the results of these computations.

For a holding period of 10 days, the adaptive nonparametric estimator $VaR^{[2]}$ has the smallest bias as well as the second-smallest standard deviation, of the five competing methods. The RiskMetrics has a comparable amount of biases, yet its variability is larger. The nonparametric estimator $VaR^{[1]}$ has the smallest standard deviation, though its bias is the third-largest. The adaptive nonparametric method has, also, the smallest MADE from 5%. For a holding period of 25 days, the RiskMetrics has the smallest biases, but its reliability is quite poor. Its variability is ranked third out of the five competing methods. The overall deviation (measured by MADE) from the nominal level is, again, achieved

Table 9 Summary of the performance of several VaR estimators ($\times 10^{-2}$)

Measure	Holding period	$VaR^{[0]}$	$VaR^{[1]}$	$VaR^{[2]}$	$VaR^{[3]}$	$VaR^{[4]}$
Average	$\tau = 10$	4.88	4.12	5.13	3.71	3.86
Standard	$\tau = 10$	0.91	0.48	0.79	1.12	1.23
MADE	$\tau = 10$	0.63	0.88	0.52	1.43	1.18
Average	$\tau = 25$	5.42	4.24	5.90	4.47	4.05
Standard	$\tau = 25$	1.85	1.27	1.39	1.97	1.66
MADE	$\tau = 25$	1.31	1.07	1.18	1.66	1.41
Average	$\tau = 50$	4.49	2.89	5.28	3.56	3.20
Standard	$\tau = 50$	3.39	1.27	2.75	2.30	1.55
MADE	$\tau = 50$	2.73	2.11	1.63	2.25	1.85

by the two nonparametric methods of estimation of quantiles. For a holding period of 50 days, the variability of the RiskMetrics is, again, very large. The adaptive nonparametric method is the best in terms of bias, variance and overall deviation from the target level of 5%. It is also worthwhile to note that the variability increases as the holding period lengthens. This is understandable, since the prediction involves a longer time horizon.

The above performance comparisons show convincingly that the newly proposed nonparametric methods for the estimation of quantiles outperform the two parametric methods and the RiskMetrics.

4. Estimation of Value-at-Risk

We have proposed three new volatility estimators based on the semiparametric model (2.8). These, together with the J. P. Morgan RiskMetrics and the GARCH model estimator, give rise to five volatility estimators. Further, we have introduced four new quantile estimators, based on parametric and nonparametric models. These and the normal quantile give five quantile estimators. Combinations of these volatility estimators and quantile estimators yield 25 methods for estimating VaR. To make comparisons easier, we eliminate a few unpromising combinations. For example, our previous studies indicate that the semiparametric volatility estimator (SVE) with $\lambda = 0.94$ does not work very well and that the parametric methods for quantile estimation do not perform as well as their nonparametric counterparts. Therefore, these methods are not considered. Instead, a few promising methods are considered to highlight the points that we advocate. Namely, that the decay parameter should be determined by data and that the time-dependent decay parameter should have a better ability to adapt to changes in market conditions. In particular, we select the following procedures:

RiskMetrics:	Normal quantile and volatility estimator (2.1)
Nonparametric RiskMetrics (NRM):	Nonparametric quantile $\hat{q}^{[1]}$ and (2.1)
Semiparametric Risk Estimator (SRE):	Nonparametric quantile $\hat{q}^{[1]}$ and SVE
Adaptive Risk Estimator (ARE):	Adaptive quantile estimator $\hat{q}^{[2]}$ and AVE
GARCH Estimator (GARCH):	Normal quantile and GARCH(1,1) model

The SRE and ARE are included in the study because they are promising. The former has time-independent decay parameters and quantiles, while the latter has time-dependent parameters and quantiles. NRM is also included in our study because of its simplicity. It possesses a very similar spirit to the RiskMetrics. GARCH is included because of its popularity in analyzing financial data.

To compare these five methods, we use simulated data sets and the eight stock indices. We begin with the simulated data. Two hundred series of length 3000 were simulated from the continuous-time SV model in Example 3. As in Example 2, the first 2000 data were regarded as the in-sample period and the last 1000 data points were treated as the post-sample period. The ERs were computed for each series for the holding periods $\tau = 1$, 10, 25 and 50. The results are summarized in Table 10. Using the MADE from the nominal confidence level 5% as an overall measure, the RiskMetrics and ARE are the best performers. One reason for the RiskMetrics to perform well is that the stochastic noises are generated from a normal distribution. It is easy to understand that if the noise does not follow a normal distribution, the RiskMetrics will not perform well.

We now apply the five VaR estimators to the eight stock indices depicted in Example 1. As in Example 1 and shown in Table 2, the post-sample period was from January 1, 1997 to December 31, 2000. The ERs were computed for each method. The results are shown in Table 11. To make the comparison easier, Table 12 shows the summary statistics of Table 11.

Table 12 shows that the ARE is the best procedure among the five VaR estimators for all holding periods. For the one-period, SRE outperforms the RiskMetrics, but for the multi-period, the RiskMetrics outperforms the SRE. NRM improves somewhat on the performance of the RiskMetrics. For the real data sets, it is clear that it is worthwhile to use time-dependent methods such as ARE. Indeed, the gain is more than the price that we have to pay

Table 10 Summary of the performance of five VaR estimators

Measure	Holding period	RiskMetrics	NRM	SRE	ARE	GARCH
Average	$\tau = 1$	5.45	4.95	4.99	4.99	4.29
Standard	$\tau = 1$	0.55	0.58	0.66	0.62	2.16
MADE	$\tau = 1$	0.61	0.47	0.54	0.49	1.81
Average	$\tau = 10$	5.19	5.12	5.16	5.12	4.29
Standard	$\tau = 10$	1.60	1.82	1.83	1.75	2.52
MADE	$\tau = 10$	1.24	1.45	1.41	1.37	2.12
Average	$\tau = 25$	5.00	5.18	5.18	5.18	4.17
Standard	$\tau = 25$	2.53	2.76	2.80	2.71	3.16
MADE	$\tau = 25$	1.96	2.14	2.16	2.04	2.73
Average	$\tau = 50$	5.41	5.41	5.37	5.33	4.21
Standard	$\tau = 50$	3.89	3.89	3.94	3.77	3.70
MADE	$\tau = 50$	3.07	3.07	3.10	2.98	3.16

Note: MADE from the nominal confidence level of 5%.

Table 11 Comparisons of the ERs of five VaR estimators

Index	Holding period	RiskMetrics	NRM	SRE	ARE	GARCH
AORD	$\tau = 1$	4.93	4.93	4.73	5.23	5.42
	$\tau = 10$	5.42	3.85	4.34	4.73	5.13
	$\tau = 25$	5.52	3.25	3.55	5.33	4.93
	$\tau = 50$	1.28	0.39	1.28	4.34	1.38
CAC 40	$\tau = 1$	6.36	6.45	6.06	5.56	8.24
	$\tau = 10$	3.67	3.67	2.68	3.28	4.57
	$\tau = 25$	4.47	4.57	3.87	3.67	5.46
	$\tau = 50$	2.78	3.38	2.78	2.28	3.28
DAX	$\tau = 1$	5.46	5.46	5.56	5.76	7.85
	$\tau = 10$	4.87	4.27	4.47	4.47	6.26
	$\tau = 25$	5.66	4.97	4.47	5.06	6.95
	$\tau = 50$	4.07	3.97	3.48	3.58	4.27
HSI	$\tau = 1$	5.99	5.71	6.27	4.96	7.30
	$\tau = 10$	6.55	4.49	5.43	5.05	9.17
	$\tau = 25$	9.17	6.08	7.95	8.98	12.25
	$\tau = 50$	7.11	2.99	4.40	5.15	11.23
Nikkei 225	$\tau = 1$	5.78	5.68	5.68	5.27	5.88
	$\tau = 10$	4.97	4.26	3.96	5.58	4.97
	$\tau = 25$	6.59	4.56	3.96	6.90	6.69
	$\tau = 50$	11.76	2.54	1.22	10.75	10.85
FTSE 100	$\tau = 1$	5.94	6.04	6.04	6.14	7.43
	$\tau = 10$	4.75	4.06	3.96	5.05	4.95
	$\tau = 25$	3.96	2.48	2.97	6.63	4.65
	$\tau = 50$	2.67	2.48	2.38	5.15	2.87
S&P 500	$\tau = 1$	5.55	5.25	5.15	4.96	4.46
	$\tau = 10$	3.77	3.67	3.87	5.35	2.78
	$\tau = 25$	3.17	3.57	3.47	4.96	1.98
	$\tau = 50$	2.78	3.77	4.36	6.64	2.28
Dow Jones	$\tau = 1$	5.65	5.65	5.65	5.65	4.96
	$\tau = 10$	5.06	4.96	4.96	5.85	4.17
	$\tau = 25$	4.86	5.16	4.46	6.15	4.37
	$\tau = 50$	3.47	3.87	4.27	5.16	3.37

Table 12 Summary of the performance of five VaR estimators

Measure	Holding period	RiskMetrics	NRM	SRE	ARE	GARCH
Average	$\tau = 1$	5.71	5.65	5.64	5.44	6.44
Standard	$\tau = 1$	0.42	0.46	0.51	0.41	1.43
MADE	$\tau = 1$	0.73	0.66	0.71	0.46	1.59
Average	$\tau = 10$	4.88	4.15	4.21	4.92	5.25
Standard	$\tau = 10$	0.91	0.44	0.82	0.80	1.86
MADE	$\tau = 10$	0.63	0.85	0.90	0.55	1.14
Average	$\tau = 25$	5.43	4.33	4.34	5.96	5.91
Standard	$\tau = 25$	1.85	1.16	1.54	1.60	2.99
MADE	$\tau = 25$	1.31	0.98	1.40	1.30	1.93
Average	$\tau = 50$	4.49	2.92	3.02	5.38	4.94
Standard	$\tau = 50$	3.39	1.17	1.32	2.52	3.86
MADE	$\tau = 50$	2.73	2.08	1.98	1.58	3.08

for the adaptation to changes in market conditions. The results also provide stark evidence that the quantile of standardized returns should be estimated and the decay parameters should be determined from data.

Table 13 shows the results for hypothesis testing, as in the performance measures 4 and 5. The results are shown for a one-day holding period. As indicated before, for a multiple-day VaR prediction, overlapping intervals were used. Hence, hypothesis testing could not be applied. Except for the HSI, there is little evidence against the hypothesis of independence and the hypothesis that ER = 5%. The p-values for AORD also tend to be small.

Table 13 Comparisons of the ERs and test results of five VaR estimators

Index	Holding period	RiskMetrics	NRM	SRE	ARE	GARCH
AORD	ER	4.93	4.93	4.73	5.23	5.42
	p-value (indep)	0.01	0.01	0.01	0.02	0.12
	p-value (ER=5%)	0.92	0.92	0.71	0.76	0.46
CAC 40	ER	6.36	6.45	6.06	5.56	8.24
	p-value (indep)	0.33	0.37	0.90	0.12	0.36
	p-value (ER=5%)	0.06	0.04	0.14	0.43	0.00*
DAX	ER	5.46	5.46	5.56	5.76	7.85
	p-value (indep)	0.59	0.59	0.90	0.16	0.78
	p-value (ER=5%)	0.51	0.51	0.43	0.28	0.00*
HSI	ER	5.99	5.71	6.27	4.96	7.30
	p-value (indep)	0.00*	0.00*	0.00*	0.00*	0.03
	p-value (ER=5%)	0.15	0.30	0.07	0.96	0.00*
Nikkei 225	ER	5.78	5.68	5.68	5.27	5.88
	p-value (indep)	0.35	0.32	0.32	0.19	0.80
	p-value (ER=5%)	0.27	0.34	0.34	0.71	0.22
FTSE 100	ER	5.94	6.04	6.04	6.14	7.43
	p-value (indep)	0.45	0.49	0.82	0.11	0.78
	p-value (ER=5%)	0.18	0.14	0.14	0.11	0.00*
S&P 500	ER	5.55	5.25	5.15	4.96	4.46
	p-value (indep)	0.62	0.88	0.32	0.89	0.89
	p-value (ER=5%)	0.44	0.73	0.85	0.96	0.43
Dow Jones	ER	5.65	5.65	5.65	5.65	4.96
	p-value (indep)	0.56	0.56	0.18	0.56	0.68
	p-value (ER=5%)	0.35	0.35	0.35	0.35	0.43

Note: * means statistically significant at the 1% level.

5. Conclusions

We have proposed semiparametric methods for estimating volatility, as well as nonparametric and parametric methods for estimating the quantiles of scaled residuals. The performance comparisons were studied both empirically and theoretically. We have shown that the proposed semiparametric model is flexible in approximating stock price dynamics.

For volatility estimation, it is evident from our study that the decay parameter should be chosen from data. Our proposed method of choosing the decay parameter has been demonstrated to be quite effective. An adaptive procedure has also been proposed, which allows the automatic adaptation of periodic changes in market conditions. The AVE has been shown to outperform the other procedures, while the SVE also performs competitively.

For the quantile estimation, our study shows that the nonparametric method has a very high efficiency compared with its parametric counterparts. Furthermore, it is robust against misspecifications of models. An adaptive procedure was introduced to accommodate the changes in market conditions over time, which allows it to outperform other competing approaches.

For the VaR estimation, it is natural to combine the AVE with the adaptive quantile estimator (i.e. ARE), and to combine the semiparametric estimator of volatility with the nonparametric estimator of quantiles (i.e. SRE), to yield effective estimators for VaR. The former is designed to accommodate changes in market conditions over time while the latter is introduced for situations where the market conditions do not change abruptly. Both methods perform outstandingly, although preference is given to the ARE method. This is due partially to the changes in market conditions over time.

Some parameters in the adaptive volatility and adaptive nonparametric quantile estimators were chosen arbitrarily. The performance of our proposed procedure can be further ameliorated if these parameters are optimized. An advantage of our procedure is that it can be combined with other volatility estimators and quantile estimators to yield new and more powerful procedures for estimating VaR.

References

Aït-Sahalia, Y. and A. W. Lo (2000). Nonparametric risk management and implied risk aversion. *Journal of Econometrics*, 94(1-2): 9-51.

Alexander, C. (ed.) (1998). *Risk Management and Analysis*, Vols. I and II. West Sussex: Wiley.

Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society Series B*, 63(2):167-241.

Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society Series B*, 64(2):253-280.

Basle Committee on Banking Supervision (1995). Public disclosure of the trading and derivatives activities of banks and securities firms. Basle: Bank for International Settlements.

Beder, T. S. (1995). VaR: Seductive but dangerous. *Financial Analysts Journal*, 51(5):12-24.

Beltratti, A. and C. Morana (1999). Computing value at risk with high frequency data. *Journal of Empirical Finance*, 6(5):431-455.

Black, F. and P. Karasinski (1991). Bond and option pricing when short rates are lognormal. *Financial Analysts Journal*, 47(4):52-59.

Black, F., E. Derman and W. Toy (1990). A one-factor model of interest rates and its application to treasury bond options. *Financial Analysts Journal*, 46(1):33-39.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307-327.

Bollerslev, T. and J. M. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews*, 11:143-172.

Chan, K. C., G. A. Karolyi, F. A. Longstaff and A. B. Sanders (1992). An empirical comparison of alternative models of the short-term interest rate. *The Journal of Finance*, 47(3):1209-1227.

Chapman, D. A. and N. D. Pearson (2000). Is the short rate drift actually nonlinear? *The Journal of Finance*, 55(1):355-388.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841-862.

Cox, J. C., J. E. Ingersoll Jr. and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385-408.

Danielsson, J. and C. G. de Vries (2000). Value-at-risk and extreme returns. *Annals d'Economie et de Statistique*, 60:239-269.

Davé, R. D. and G. Stahl (1997). On the accuracy of VaR estimates based on the variance-covariance approach. Working paper, Olshen and Associates.

Diebold, F. X., A. Hickman, A. Inoue and T. Schuermann (1998). Converting 1-day volatility to h-day volatility: Scaling by $\sqrt{h}$ is worse than you think. *Risk*, 11:104-107.

Dowd, K. (1998). *Beyond Value at Risk: The New Science of Risk Management*. New York: Wiley.

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

Embrechts, P., C. Klüppelberg and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer.

Engle, R. F. (1995). *ARCH: Selected Readings*. Oxford: Oxford University Press.

Engle, R. F. and G. Gonzalez-Rivera (1991). Semiparametric ARCH models. *Journal of Business and Economic Statistics*, 9(4):345-359.

Engle, R. F. and S. Manganelli (1999). CAViaR: Conditional autoregressive value at risk by regression quantile. Cambridge, MA: NBER Working Paper 7341.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.

Fan, J. and Q. Yao (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer-Verlag.

Fan, J. and C. Zhang (2003). A Reexamination of diffusion estimators with applications to financial model validation. *Journal of the American Statistical Association*, 98(461):118-134.

Fan, J., J. Jiang, C. Zhang and Z. Zhou (forthcoming). Time-dependent diffusion models for term structure dynamics and the stock price volatility. *Statistica Sinica*.

Genon-Catalot, V., T. Jeantheau and C. Laredo (1999). Parameter estimation for discretely observed stochastic volatility models. *Bernoulli*, 5(5):855-872.

Härdle, W. and A. Yatchew (2002). Dynamic nonparametric state price density estimation using constrained least-squares and the bootstrap. Berlin: Discussion paper 16, Quantification and Simulation of Economics Processes, Humboldt-Universität zu Berlin.

Härdle, W., H. Herwartz and V. Spokoiny (forthcoming). Time inhomogeneous multiple volatility modelling. *Journal of Financial Econometrics*.

Hendricks, D. (1996). Evaluation of value-at-risk models using historical data. *Economic Policy Review (Federal Reserve Bank of New York)*, 2(1):39-69.

Hull, J. and A. White (1990). Pricing interest-rate-derivative securities. *The Review of Financial Studies*, 3:573-592.

J. P. Morgan (1996). *RiskMetrics Technical Document* (4th ed.). New York: Morgan Guaranty Trust Company.

Jackson, P., D. J. Maude and W. Perraudin (1997). Bank capital and value at risk. *Journal of Derivatives*, 4(3):73-89.

Jorion, P. (2000). *Value at Risk: The New Benchmark for Managing Financial Risk* (2nd ed.). New York: McGraw-Hill.

Lucas, A. (2000). A note on optimal estimation from a risk-management perspective under possibly misspecified tail behavior. *Journal of Business and Economic Statistics*, 18(1):31-39.

Mahoney, J. M. (1996). Forecast biases in value-at-risk estimations: Evidence from foreign exchange and global equity portfolios. New York: Working paper, Federal Reserve Bank of New York.

Mercurio, D. and V. Spokoiny (forthcoming). Statistical inference for time-inhomogeneous volatility models. *The Annals of Statistics*.

Pant, V. and W. Chang (2001). An empirical comparison of methods for incorporating fat tails into value-at-risk models. *The Journal of Risk*, 3(3):1-21.

Prakasa Rao, B. L. S. (1987). *Asymptotic Theory of Statistical Inference*. New York: John Wiley & Sons.

Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *The Journal of Finance*, 52(5):1973-2002.

Wong, C. M. and M. K. P. So (2000). On conditional moments of GARCH models, with applications to multiple period value at risk estimation. Unpublished manuscript.

Yang, L. J., W. Härdle and J. P. Nielsen (1999). Nonparametric autoregression with multiplicative volatility and additive mean. *Journal of Time Series Analysis*, 20(5):579-604.

Appendix

In this appendix, we give some theoretical derivations on the relative efficiencies of several nonparametric and parametric estimators of quantiles. The basic assumption is that $\{\hat{\varepsilon}_{t,\tau}, t = T_0, \cdots, T-\tau\}$ is an independent random sample from a population with a probability density f.

A.1 Relative Efficiency of Estimator (3.1)

By the symmetry assumption, $f(q(\alpha,\tau)) = f(q(1-\alpha,\tau))$. By using (3.8), we obtain that the estimators $\hat{q}(\alpha,\tau)$ and $\hat{q}(1-\alpha,\tau)$ are jointly asymptotically normal with a mean of $(q(\alpha,\tau), -q(\alpha,\tau))^T$ and a covariance matrix of

$$f(q(\alpha,\tau))^{-2}\begin{pmatrix} \alpha(1-\alpha) & \alpha^2 \\ \alpha^2 & \alpha(1-\alpha) \end{pmatrix}.$$

It follows that $\hat{q}^{[3]}(\alpha,\tau)$ has an asymptotically normal distribution with a mean of $q(\alpha,\tau)$ and a variance of

$$4^{-1}f(q(\alpha,\tau))^{-2}[\alpha(1-\alpha) - 2\alpha^2 + \alpha(1-\alpha)] = 2^{-1}f(q(\alpha,\tau))^{-2}\alpha(1-2\alpha)$$

and that $\hat{q}(\alpha,\tau)$ is asymptotically normal with a mean of $q(\alpha\tau)$ and a variance of

$$f(q(\alpha,\tau))^{-2}\alpha(1-\alpha).$$

Consequently, the estimator $\hat{q}^{[3]}(\alpha,\tau)$ is a factor of $2(1-\alpha)/(1-2\alpha)$ that is as efficient as $\hat{q}(\alpha,\tau)$.

A.2 Asymptotic Normality for the Method of Moments Estimator

Recall that the variance of the squared t-random variable with degree of freedom ν is given by $2\nu^2(\nu-1)/(\nu-2)^2(\nu-4)$. By the central limit theorem,

$$\sqrt{m}\left(\hat{\mu}_2 - \frac{\nu}{\nu-2}\lambda^2\right) \xrightarrow{L} N\left(0, \frac{2\nu^2(\nu-1)}{(\nu-2)^2(\nu-4)}\lambda^4\right).$$

Using the delta-method,we deduce that

$$\sqrt{m}\left(\sqrt{\hat{\mu}_2} - \sqrt{\frac{\nu}{\nu-2}}\lambda\right) \xrightarrow{L} N\left(0, \frac{\nu(\nu-1)}{2(\nu-2)(\nu-4)}\lambda^2\right).$$

Hence, the estimated quantile

$$\hat{q}^{[4]}(\alpha,\tau) = \sqrt{\hat{\mu}_2}\sqrt{\frac{\nu-2}{\nu}}\,t(\alpha,\tau)$$

has an asymptotically normal distribution with a mean of $\lambda t(\alpha,\tau)$ and a variance of

$$\frac{\lambda(\nu-1)t(\alpha,\tau)^2}{2(\nu-4)}.$$