ELSEVIER

Japan and the World Economy
14 (2002) 63–85

Japan and the WORLD ECONOMY

www.elsevier.com/locate/econbase

Assessing the risk forecasts for Japanese stock market

Tae-Hwy Lee[a,*], Burak Saltoğlu[b,1]

[a]*Department of Economics, University of California, Riverside, CA 92521, USA*
[b]*Department of Economics, Marmara University, Istanbul 81040, Turkey*

Received 12 March 2001; received in revised form 19 September 2001; accepted 01 October 2001

Abstract

We evaluate predictive performance of a selection of value-at-risk (VaR) models for Japanese stock market data. We consider traditional VaR models such as Riskmetrics method, historical simulation, variance–covariance method, Monte Carlo method, and their variants which are integrated with various ARCH models. Also considered are more recent models based on non-parametric quantile regression and extreme value theory (EVT). We apply these methods to the Japanese stock market index (1984–2000) and compare their performances in terms of various evaluation criteria using the method of White [Econometrica 68 (5) (2000) 1097–1126] for three out-of-sample periods of 1995–1996, 1997–1998, and 1999–2000. © 2002 Elsevier Science B.V. All rights reserved.

JEL classification: C3; C5; G0

Keywords: VaR; ARCH; Historical simulation; Variance–covariance method; Monte Carlo method; Non-parametric quantile regression; Extreme value theory; GEV; GPD; Hill estimator; Data snooping; Predictive ability; Reality check; Loss functions

1. Introduction

The financial turmoil hit Japan in 1997, destroyed the optimism about this model economy and had a negative spill over effect on the world financial system. Economists and policy analysts attempting to explain the causes of this unexpected episode are faced with a significant challenge. Possible causes of the crisis in Japan and Asia have been extensively studied by Dornbusch (1998a,b), Krugman (1998), Mishkin (1999, 2000), Corsetti et al. (1999), Goldstein et al. (2000), Mikitani and Posen (2000), Haggard (2000), and Beim and

* Corresponding author. Tel.: +1-909-7875037; fax: +1-909-7875685.
E-mail addresses: taelee@mail.ucr.edu (T.-H. Lee), saltoglu@marun.edu.tr (B. Saltoğlu).
[1] Tel.: +90-216-3368487; fax: +90-216-3464356.

PII: S0922-1425(01)00080-9

Calomiris (2001, Chapter 8). Some effects of these crisis on US are also studied in several articles in FRBNY (2000). Most of these researchers have agreed on the fact that lack of prudent supervision and control of risk management in the banking sector constitute one of the major causes of this crisis. Similarly, Fischer (1998) argues that the need for transparency and better banking supervision are the key lessons that are drawn from the crisis which hit Japan. Moreover, Dornbusch (1998a,b) and Mishkin (2000) both assert that the lack of quality and control in risk management systems in banking sector is one of the major culprits of the financial crisis. Goldstein et al. (2000) claim that like many other crisis the banking sector crisis in Japan stems from the asset side of the balance sheet. In measuring market risks, Japan have been following the regulations set by Basle Committee on Banking Supervision (1996). The regulations require Japanese banks to report their daily risk measures called value-at-risk (VaR) to their regulator financial supervisory agency (FSA). Hence, the relative success and the use of these VaR models for the Japanese economy during the crisis needs further investigation.

The existing literature on VaR models have been evolved over the last decade mainly focusing on the US market data. An extensive review of the literature on the conventional VaR modelling can be found in Dowd (1998). Diebold and Santomera (1999), Christoffersen and Errunza (2000) and Stulz (2000) point out the problems arising from the normality assumption which is commonly used in conventional risk measurement methodologies. In supporting these claims, traditional VaR models are criticized for their inability to capture the extreme price movements that can take place during financial turmoil. Danielsson (2000) argued that the statistical analysis made in times of stability does not provide much guidance in times of crisis. He further claims that the use of VaR models for regulatory purposes is questionable for this reason. However, the validity of such claims require a careful examination of a wide selection of VaR models with special reference to crisis period.

Only very few studies can be seen on evaluating the risk forecasts for the Japanese economy despite the significant need and interest in both academic and policy purposes. The only exception is Danielsson and Morimoto (2000), who analyzed the forecasting ability of extreme value theory (EVT)-based VaR model in the Japanese economy. Given the importance and its implications on the many other emerging markets, a more comprehensive study is necessary in this field. The purpose this paper is to investigate thoroughly the relative predictive performance of various alternative VaR models with special reference to Japanese stock market. This information that will be extracted from this study will have significant policy implications for understanding the past crisis and shed some light for future turmoils. To this end, a broad variety of conventional VaR models (which were available during the crisis) and some recent EVT-based methods (which became popular after 1997) are covered in this study. The conventional VaR models studied here include variance–covariance, historical simulation, and Monte Carlo simulation methods. More details can be found in some recent books on risk management such as Jorion (2000), Alexander (1998), or Dowd (1998). Some of these conventional models are also modified by integrating various ARCH tools such as generalized ARCH (GARCH) of Bollerslev (1986), exponential GARCH (EGARCH) of Nelson (1991), and threshold GARCH (TGARCH) of Glosten et al. (1993). In order to see whether any efficiency gain that can be obtained by lifting the critical normality assumption, some non-parametric alternatives are also studied. A hybrid historical simulation method developed by Boudoukh et al. (1998) and a non-parametric quantile estimation based on kernel density (see Xiang, 1996; ?, for applications) are also employed in the present paper.

Furthermore, in order to investigate the claim that conventional VaR models are unable to capture market risks three popular EVT-based VaR models, namely, generalized extreme value (GEV) distribution, generalized Pareto distribution and the tail index estimator by Hill (1975) are studied. As is known, EVT models are specially designed to model extreme price movements. The theoretical and applied works on the EVT models can be found in Longin (1996, 2000), Ho et al. (2000), Neftçi (2000), Danielsson et al. (2001), Danielsson and deVries (1997) and Danielsson et al. (2001).

Altogether, 27 alternative VaR models are studied in this paper. All of these models are compared for their out-of-sample predictive ability via White (2000) "reality check" method which is a data snooping robust methodology. In implementing the reality check we have used three different objective functions, namely, quasi-log likelihood defined in Bertail et al. (2000), the tail mean return (defined in Section 3), and the coverage likelihood ratio. Christoffersen (1998) likelihood ratio tests for conditional and unconditional coverage probabilities are also employed. Daily Japanese stock market index from 1984 to 2000, are used. The VaR forecasts generated from these models for three out-of-sample periods of 1995–1996, 1997–1998, and 1999–2000 are studied.

Unlike the recent results obtained in the literature, the findings of our analysis suggest that the predictive performance of the EVT models are less than satisfactory for various loss functions. In contrast, the predictive performance of some of the traditional methods such as TGARCH and Monte Carlo models with alternative volatility structures appear to be more successful than the benchmark Riskmetrics model. However, none of the available methods produce a uniformly superior risk forecasts for all loss functions and all periods. Therefore, our findings further reveal difficulties and challenges faced by the policy analysts, practitioners and academics, who want to use risk forecasts to understand and prevent potential future crises.

The organization of the paper is as follows. In Section 2, various VaR models are discussed. In Section 3, forecast evaluation criteria and the reality check are discussed. Section 4 presents the empirical results and Section 5 concludes.

2. VaR models

Consider the return series $\{y_t\}_{t=1}^T$ of a financial asset. The value-at-risk, denoted as $\mathrm{VaR}_t(\alpha)$, can be defined as the conditional quantile

$$\Pr(y_t \le \mathrm{VaR}_t(\alpha)|\mathscr{F}_{t-1}) = \alpha. \tag{1}$$

To establish some notation, suppose $\{y_t\}_{t=1}^T$ follows the stochastic process

$$y_t = \mu_t + \varepsilon_t, \tag{2}$$

where $E(\varepsilon_t|\mathscr{F}_{t-1}) = 0$ and $E(\varepsilon_t^2|\mathscr{F}_{t-1}) = \sigma_t^2$ given the information set $\mathscr{F}_{t-1}$ (σ-field) at time $t-1$. Let $z_t \equiv \varepsilon_t/\sigma_t$ have the conditional distribution Φ_t with zero conditional mean and unit conditional variance, i.e. $z_t|\mathscr{F}_{t-1} \sim \Phi_t(0, 1)$.

We now turn to various methods of estimating $\mathrm{VaR}_t(\alpha)$. In this section, we present various VaR models developed and used over the last 6 years. These models may be divided into four main categories: (1) variance–covariance methods, (2) non-parametric methods, (3) Monte Carlo simulation methods, (4) the EVT-based VaR method (also known as XVaR).

2.1. Variance–covariance methods

The first method is the most standard approach, often called variance–covariance methods. In this paper, because we consider a single stock index instead of portfolio we do not consider covariances and, thus, it may be called as variance methods. In this method, $\mathrm{VaR}_t(\alpha)$ can be estimated by

$$\mathrm{VaR}_t(\alpha) = \mu_t + \Phi_t^{-1}(\alpha)\sigma_t. \tag{3}$$

Hence, estimation of the VaR involves the estimation of $\Phi_t(\cdot)$, μ_t, and σ_t. We consider various estimation methods of VaR, which may be labeled with different methods of estimating $\Phi_t(\cdot)$ and σ_t.

We either assume a certain parametric distribution for $\Phi_t(\cdot)$ (e.g. normal distribution, Student-t distribution, generalized error distribution (GED), etc.) or estimate it non-parametrically. The conditional distribution $\Phi_t(\cdot)$ is assumed to be constant over time or simply assumed as Gaussian $N(0,1)$ in which case $\Phi_t^{-1}(0.05) = 1.645$ and $\Phi_t^{-1}(0.01) = 2.326$. To take care of the fat tail distributions of financial returns series, it is also often assumed as Student-$t(v)$ with v degrees of freedom. For $t(6)$, $\Phi_t^{-1}(0.05) = 1.943\sqrt{(v-2)/v} = 1.586$ and $\Phi_t^{-1}(0.01) = 3.143\sqrt{(v-2)/v} = 2.566$. We use $t(6)$ in this paper.

The conditional variance σ_t^2 is estimated with various volatility methods such as a simple moving average model (Alexander, 1998), an exponentially weighted moving average (EWMA) model of Riskmetrics, and ARCH models of Engle (1982), Bollerslev (1986), Nelson (1991), and Glosten et al. (1993).

The simplest method to calculate the VaR is to estimate the volatility of the asset return by historical moving average variance. In this method, we estimate the volatility

$$\sigma_t^2 = \frac{1}{m-1}\sum_{j=1}^{m}(y_{t-j} - \hat{\mu}_t^m)^2 \tag{4}$$

where $\hat{\mu}_t^m = (1/m)\sum_{j=1}^{m} y_{t-j}$. See Alexander (1998) for its empirical advantages and disadvantages. This method will be denoted as $\mathrm{MA}(m)$. In our empirical part, we use MA(200).

One of the most popular volatility model in risk management framework is the Riskmetrics model of Morgan (1995), which is an IGARCH specification of the following form:

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)(y_{t-1} - \hat{\mu}_t)^2, \tag{5}$$

where $\hat{\mu}_t = 1/(t-1)\sum_{j=1}^{t-1} y_{t-j}$. Riskmetrics methodology assumes a fixed constant $\lambda = 0.94$ which substantially reduces the volatility computations. This method will be denoted as $\mathrm{RM}(\lambda)$. For our empirical analysis in Section 4 we consider RM(0.94), RM(0.97) and RM(0.90).

We consider the following three ARCH models. First, the standard GARCH model is

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon_{t-1}^2. \qquad (6)$$

The second ARCH model is TGARCH of Glosten et al. (1993):

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon_{t-1}^2 + \gamma\varepsilon_{t-1}^2 1(\cdot) \quad (\varepsilon_{t-1} \geq 0), \qquad (7)$$

where $1(\cdot)$ is an indicator function. The third one is EGARCH of Nelson (1991):

$$\ln \sigma_t^2 = \omega + \beta \ln \sigma_{t-1}^2 + \alpha[|z_{t-1}| - cz_{t-1}]. \qquad (8)$$

We consider three distributions of z_t, which are $N(0, 1)$, Student-t, and generalized error distribution (GED). Mittnik et al. (1998) analyze the Nikkei index by various ARCH models with various distributions of Laplace, double Weibull, generalized exponential, Student-t, α-stable distributions. The asymmetric α-stable distribution is their preferred model. They use weekly returns of the Nikkei 225 index for a pre-crisis period from 31 July 1983 to 9 April 1995. Although we note here that the ARCH model with α-stable distribution could be a serious candidate, we do not employ it here. In our empirical analysis in Section 4, each of the three ARCH models are estimated with three innovation distributions, and will be denoted as GARCH_i, EGARCH_i, and TGARCH_i with $i = N(0, 1)$, t, or GED, so that there are total nine ARCH models considered.

2.2. *Non-parametric methods*

VaR estimates can also be obtained by using non-parametric methods. We use two different non-parametric methods in this paper.

2.2.1. *Historical simulation methods*

The main idea behind the historical simulation approach is the assumption that historical distribution of returns will remain the same over the next periods, therefore, the empirical distributions of portfolio returns will be used in estimating VaR. In other words, this method uses the empirical quantiles of the historical distribution of return series, $\{y_{t-j}\}_{t=1}^{t-1}$, to estimate $\text{VaR}_t(\alpha)$ for a given confidence level α. See Jorion (2000, p. 221) for more details on HS. We will denote this method as HS.

Boudoukh et al. (1998) suggested a hybrid version of the historical simulation approach. They use exponentially declining weights to discount the distant past in obtaining VaR, i.e. $\text{VaR}_t(\alpha)$ is estimated from the empirical percentile of the historical distribution of $\{w_j^m y_{t-j}\}_{t=1}^{m}$ with weights $w_j^m = (1-\lambda)\lambda^j/(1-\lambda)^m$. See Boudoukh et al. (1998) for more details. The hybrid HS will be denoted as $\text{HS}(m, \lambda)$. We will use HS(200, 0.99) and HS(200, 0.97) in our empirical analysis.

2.2.2. *Non-parametric quantile regressions*

VaR, which is basically a quantile estimator, can be estimated within the context of non-parametric quantile regression methods. There has been various studies made on parametric quantile estimation where a classic reference is Koenker and Bassett (1978). On non-parametric quantile estimation, recent references are Samanta (1989), Lejeune and

Sarda (1988), and Xiang (1996) where a newer application of non-parametric quantile on financial data was made by Abberger (1997).

Let $Z = (Y, X')'$ be a stationary random vector with the joint density $f(y, x)$ and the cumulative distribution function (CDF) $F(y, x)$. Then the conditional CDF is

$$F(y|x) = \frac{\int_{-\infty}^{y} f(\eta, x)\,\mathrm{d}\eta}{F_X(x)}, \tag{9}$$

where $F_X(x)$ is the marginal CDF of X. Hence, the quantile function can be found as

$$q_\alpha = \inf\{y \in \mathbb{R} | F(y|x) \geq \alpha\}, \tag{10}$$

where $F(y|x)$ can be estimated from

$$F(y|x) = \frac{\sum_{t=1}^{n} K((y_t - y)/h) F_1((x_t - x)/h_1)}{\sum_{t=1}^{n} K_1((x_t - x)/h_1)} \tag{11}$$

where $K(\cdot)$, $K_1(\cdot)$, h, and h_1 are the kernel functions and the bandwidths corresponding to y and x, respectively, and $F_1(u) = \int_{-\infty}^{u} K_1(z)\,\mathrm{d}z$. The optimal bandwidth selection can be made by modifying the standard cross-validation method. This can be represented as

$$h_{\mathrm{CV}} = \arg\min_h \sum_{t=1}^{n} \rho_\alpha(y_t - q_\alpha^{-t}(h)) \tag{12}$$

where $\rho_\alpha(\cdot)$ is the loss function defined by Koenker and Bassett (1978), and $q_\alpha^{-t}(h) = \inf\{y \in R | F^{-t}(y|x) \geq \alpha\}$ and $F^{-t}(y|x)$ is a leave-one-out estimator of the conditional CDF.

Estimating and forecasting VaR via non-parametric quantile regression is straightforward. For instance, to estimate $\mathrm{VaR}_t(\alpha) \equiv q_\alpha$, one needs to estimate the conditional CDF by using the kernel method and then estimate corresponding quantiles. We have both used the Gaussian and Epanecknikov kernels. In our empirical section, this method will be denoted as $\mathrm{NPQ}_i(m)$, where $i = 1$ when Gaussian kernel $K(u) = (1/\sqrt{2\pi})\exp(-(1/2)u^2)$ is used and $i = 2$ when Epanecknikov kernel $K(u) = (3/4)(1 - u^2)1(|u| < 1)$ is used, and m is number of most recent historical observations used in estimating the conditional CDF. Optimal bandwidths are chosen by the robustified cross-validation process explained before. The difference between this approach and the one Butler and Schachter (1998) is that this method directly estimates the quantiles. In Section 4, we use $x_t = y_{t-1}$ and $m = 200$. For each out-of-sample forecasting point, we have dropped one distant observation and add a recent one. The optimal bandwidth is chosen for each out-of-sample data point.

2.3. Monte Carlo simulation methods

In this method, an underlying stochastic process which is assumed to govern the dynamics of the asset prices is used to simulate the future values of the asset, see Hull (1997) or Jorion (2000). One of the most popular stochastic process in the asset pricing context is the geometric Brownian motion given as

$$\frac{\mathrm{d}S_t}{S_t} = \mu_t\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t \tag{13}$$

where W_t is a standard Wiener process, and μ_t and σ_t are the drift and the volatility parameters, respectively. This is a rare example of an explicitly solvable stochastic differential equation with its solution being

$$S_t = S_0 \exp([\mu_t - \tfrac{1}{2}\sigma_t^2]t + \sigma_t W_t). \tag{14}$$

See Brodie and Glasserman (1998) for more discussion. Thus, simulating values of S_t reduces to simulating values of W_t. The Monte Carlo method based on Eq. (14) will be denoted as MC_1.

Alternatively, $\{S_t\}$ can be simulated recursively from the simple discretization of Eq. (13), i.e.

$$\Delta S_t = S_{t-1}(\mu_{t-1}\Delta t + \sigma_{t-1}\varepsilon_t\sqrt{\Delta t}) \tag{15}$$

where ε_t is zero mean, unit variance random error term, and μ_t and σ_t are drift and volatility parameters in discrete time. In practice, the above stochastic process with an infinitesimally small increment dt is approximated by small discrete moves of Δt. Then one can simulate the sample paths of the above process with estimated drift and volatility parameters. See Jorion (2000, p. 293) for an illustration. This Monte Carlo method based on Eq. (15) will be denoted as MC_2.

One possible extension to the above model is to integrate a time varying volatility structure to the above equations. In our empirical analysis, we consider three such extensions of MC_2 with the volatility coefficients σ_t being estimated by $GARCH_N$, $EGARCH_{GED}$, and RM(0.94).

The number of Monte Carlo replications used in Section 4 is 5000. Once the sample paths of financial asset is obtained, VaR measures can be estimated by computing the empirical quantile of the return distribution. The drift parameter is estimated by the sample average of the asset return values.

2.4. *Extreme value theory*

Risk management is primarily concerned with the risk of low-probability events that could lead to catastrophic losses. However, all the VaR methodology we have reviewed so far ignore extreme events and directly focus on risk measures that accommodate the whole return distribution. In risk management, these extreme observations are used to model the tails of return distributions. The focus of EVT is, unlike the other parametric and non-parametric methods used in VaR methodology, to model the tails rather than the entire return distributions. There has been various theoretical and empirical studies in this field. For instance, Embrechts et al. (1997) give an excellent review of EVT. Longin (1996, 2000) use GEV to estimate the tail index via maximum likelihood estimation. Neftçi (2000), on the other hand uses GPD. We consider the three EVT models, namely, GEV, GPD, and Hill's tail index estimation.

2.4.1. *Generalized extreme value distribution*

Consider the return series $\{y_t\}_{t=1}^T$ of a financial asset and the ordered return series $\{y_{(t)}\}_{t=1}^T$ in increasing order $y_{(t)} \leq y_{(t+1)}$ for all t. The sample minimum is $y_{(1)}$ over T sample period. If the returns are i.i.d. with the CDF $F_Y(y)$, then the CDF of the minimal return, denoted by $G_Y(y)$, is given by Longin (1996, 2000)

$$G_Y(y) = \Pr(y_{(1)} \le y) = 1 - \Pr(y_{(1)} > y) = 1 - \prod_{t=1}^{T} \Pr(y_t > y)$$
$$= 1 - \prod_{t=1}^{T} [1 - \Pr(y_t \le y)] = 1 - [1 - F_Y(y)]^T. \quad (16)$$

Thus, $G(y)$ is degenerated as $T \to \infty$. Hence, we seek a limit law $H_X(x)$ with which a normalization $x_T = (y_{(1)} - \beta_T)/\delta_T$ does not degenerate as $T \to \infty$ for suitable normalizing constants β_T and $\delta_T > 0$. The limiting distribution of x_T is the generalized extreme value (GEV) distribution of Mises (1936) and Jenkinson (1955) of the form

$$H_X(x) = 1 - \exp[-(1 + \tau x)^{1/\tau}] \quad (17)$$

for $1 + \tau x > 0$. The corresponding limiting density function of $\{x_T\}$ as $T \to \infty$, obtained by differentiating $H_X(x)$, is given by

$$H_X(x) = (1 + \tau x)^{(1/\tau)-1} \exp[-(1 + \tau x)^{1/\tau}]. \quad (18)$$

Hence, the approximate density of $y_{(1)}$ for given T may be obtained by change of variables which is

$$H_Y(x_T) = \frac{1}{\delta_T}(1 + \tau x_T)^{(1/\tau)-1} \exp(-(1 + \tau x_T)^{1/\tau}), \quad (19)$$

where $1/\delta_T$ is the Jacobian of the transformation.

Hence, the three parameters $\theta_T = (\tau, \beta_T, \delta_T)'$ may be estimated by maximum likelihood method. To implement it, Longin (1996, 2000) partition T samples into m non-overlapping subsamples each with n observations. In other words, if $T = mn$, the ith subsample of the return series is $\{y_{(i-1)n+j}\}_{j=1}^{n}$ for $i = 1, \ldots, m$. If $T < mn$, we drop some observation in the first subsample so that it has less than n observations. The collection of the subperiod minima is then $\{y_{n,i}\}$, where $y_{n,i} = \min_{1 \le j \le n}\{y_{(i-1)n+j}\}$, $i = 1, \ldots, m$. The likelihood function of the subperiod minima is

$$\prod_{i=1}^{m} H_Y(x_n) = \prod_{i=1}^{m} H_Y\left(\frac{y_{n,i} - \beta_n^i}{\delta_n^i}\right). \quad (20)$$

Assuming $\theta_n^i = \theta_n$ for all subperiods $i = 1, \ldots, m$, θ_n can be estimated for a numerical optimization of the log likelihood.

Consider the probability that the subperiod minimum $y_{n,i}$ is less than y_n^* under the limit law (Eq. (17)). Denoting $x_n^* = (y_n^* - \beta_n)/\delta_n$, it is

$$H_X(x_n^*) = H_X\left(\frac{y_n^* - \beta_n}{\delta_n}\right) = \Pr\left(\frac{y_{n,i} - \beta_n}{\delta_n} \le \frac{y_n^* - \beta_n}{\delta_n}\right) = \Pr(y_{n,i} \le y_n^*), \quad (21)$$

which is therefore, equal to

$$G_Y(y_n^*) = 1 - [1 - F_Y(y_n^*)]^n = 1 - (1 - \alpha)^n, \quad (22)$$

where the second equality holds if $y_n^* = \mathrm{VaR}(\alpha)$. Hence, equating Eqs. (21) and (22), we get

$$H_X(x_n^*) = 1 - \exp(-(1 + \tau x_n^*)^{1/\tau}) = 1 - (1 - \alpha)^n \tag{23}$$

which yields the VaR forecasts

$$y_n^* = \mathrm{VaR}(\alpha) = \beta_n - \frac{\delta_n}{\tau}\{1 - [-\ln(1 - \alpha)^n]^\tau\}. \tag{24}$$

We denote this method as Longin(n), where n is the size of the subperiod. In Section 4, we use $n = 10$ and 20. Tsay (2000) provides an excellent exposition of this method and other VaR models.

2.4.2. Generalized pareto distribution

An alternative approach to GEV method is based on *exceedances over threshold* (Smith, 1989; Davison and Smith, 1990). According to this approach, we fix some high threshold u and look at all exceedances e over u. The distribution of excess values is given by Neftçi (2000)

$$F_u(e) = \Pr(X < u + e | X > u) = \frac{F(u + e) - F(u)}{1 - F(u)}, \quad e > 0. \tag{25}$$

Pickands (1975) shows that the asymptotic form of $F_u(e)$ is

$$H(e) = 1 - \left(1 - \frac{\tau e}{\delta}\right)^{1/\tau}, \tag{26}$$

where $\delta > 0$ and $1 - (\tau e/\delta) > 0$. This is known as the generalized Pareto distribution (GPD) with its density

$$H(e) = \frac{1}{\delta}\left(1 - \frac{\tau e}{\delta}\right)^{(1/\tau)-1}. \tag{27}$$

Let $\{e_i\}_{i=1}^n$ be the sample of exceedances over threshold with its sample size n. The likelihood $\prod_{i=1}^n h(e_i)$ may be maximized to estimate $\theta = (\tau\,\delta)'$. Once $\hat{\theta} = (\hat{\tau}\,\hat{\delta})'$ is estimated VaR(α) can be estimated as follows. From Eqs. (25) and (26), we get

$$[1 - F(u + e)] = [1 - F(u)][1 - H(e)]. \tag{28}$$

From this, by letting $[1 - F(u + e)] = \alpha$, $[1 - F(u)] = n/T$, and using the GPD distribution $H(\mathrm{VaR}_t(\alpha))$ in Eq. (26), we obtain the VaR estimate

$$\mathrm{VaR}_t(\alpha) = -\frac{\hat{\delta}}{\hat{\tau}}\left(1 - \left(\frac{T\alpha}{n}\right)^{\hat{\tau}}\right), \tag{29}$$

where T is the total observations and n is the number of exceedances.

The above discussion is for the upper tail (short position). In this paper, we focus on lower tails. However, if one uses negative return series $\{-y_t\}_{t=1}^T$ for the variable X in Eq. (25), the above discussion continues to apply to the lower tails (long position). Let $x_t = -y_t$. In choosing the threshold value u, we follow Neftçi (2000): $u = 1.176 \times \hat{\sigma}_T$, where $\hat{\sigma}_T$ is the standard deviation of $\{x_t\}_{t=1}^T$ from the whole sample and $1.176 = \Phi_t^{-1}(0.10) = 1.440\sqrt{(v-2)/v}$ with $t(6)$ distribution being assumed. Therefore, the extreme observations (exceedances over the thresholds) would belong to 10% tails if its true distribution is indeed $t(6)$. The number of $\{x_t\}_{t=1}^T$ that exceeds u is n.

2.4.3. Hill estimator

As before, we assume the return series $\{y_t\}_{t=1}^T$ are i.i.d. and denote the ordered return series as $\{y_{(t)}\}_{t=1}^T$ in increasing order. Suppose $y_{(n)} < 0$ and $y_{(n+1)} > 0$ so that n is the number of negative returns in the T observations. The GEV distribution in Eq. (17) with $\tau < 0$ is known as the Fréchet distribution with the CDF $F_Y(y) = \exp(-y^{1/\tau})$, $y < 0$. As shown in Embrechts et al. (1997, p. 321), it reduces to

$$F_Y(y) = 1 - Cy^{1/\tau}, \quad |y| \geq u \geq 0 \tag{30}$$

where $C = u^{-1/\tau}$ is a slowly varying function with u being the known threshold. A popular estimator of τ is due to Hill (1975) who shows that its maximum likelihood estimator is

$$\hat{\tau} = -\frac{1}{k}\sum_{t=1}^{k} \ln|y_{(t)}| + \ln|y_{(k+1)}|, \tag{31}$$

where $k \equiv k(n) \to \infty$ and $k(n)/n \to 0$. It is known that $\hat{\tau} \to_p \tau$ as $n \to \infty$ (Mason, 1982). We choose the sample fraction k using a bootstrap method of Danielsson et al. (2001). Once τ is estimated, the VaR estimate can be found from (see Embrechts et al., 1997, p. 347):

$$\mathrm{VaR}_t(\alpha) = \left[\frac{n}{k}(1-\alpha)\right]^{\hat{\tau}} y_{(k+1)}. \tag{32}$$

3. Evaluating VaR models

We consider tests of Christoffersen (1998) and White (2000) for VaR forecast evaluation. Our evaluation of out-of-sample forecasts proceeds as follows. There are P predictions in all for each model. The first prediction is based on the model with parameters estimated using data from 1 to R, the second on the model with parameters estimated using data from 2 to $R+1, \ldots,$ and the last on the model with parameters estimated using data $P-1$ to $R+P-1 = T$. Based on the estimated models using a series of rolling samples, each of size R, one-step ahead forecasts are generated for P post-samples, resulting in P forecasts to evaluate each model.

3.1. Christoffersen tests for coverage probability

We begin with three likelihood ratio tests of Christoffersen (1998). Let l be the number of models ($k = 1, \ldots, l$) to be compared with the benchmark model ($k = 0$). We consider $l = 26$ models plus a benchmark model with total 27 models in Section 4. Let the indicator $d_t^k \equiv 1(y_t < \mathrm{VaR}_t^k(\alpha))$, $t = R, \ldots, T$, denote for the case when return falls beyond the VaR forecast estimated from model k. Let the probability of the unconditional coverage failure be denoted as $p_k^\alpha = \Pr[y_t < \mathrm{VaR}_t^k(\alpha)] = \Pr(d_t^k = 1)$. As the indicator $\{d_t^k\}$ has a binomial distribution, the likelihood is $L(p_k^\alpha) = (1-p_k^\alpha)^{n_0}(p_k^\alpha)^{n_1}$, where $n_0 = \sum_{t=R}^T (1-d_t^k)$ and $n_1 = \sum_{t=R}^T d_t^k$ are the number of 0's and 1's in the indicator sequence $\{d_t^k\}_{t=1}^T$. Note that

$n_0^k + n_1^k = P$. The indices α and k in d_t^k, p_k^α, $\mathrm{VaR}_t^k(\alpha)$, n_0^k, n_1^k will often be suppressed in the following sections.

First, we test whether the probability of the unconditional coverage failure, $p = \Pr[y_t < \mathrm{VaR}_t(\alpha)]$, is equal to α. That is to test $H_0: p = \alpha$ against $H_1: p \neq \alpha$. As the indicator d_t has a binomial distribution the likelihood is $L(p) = (1-p)^{n_0} p^{n_1}$. Under the null, it is $L(a) = (1-\alpha)^{n_0}\alpha^{n_1}$, and, thus, the likelihood ratio test statistic is

$$\mathrm{LR}_1 = -2 \ln \left(\frac{L(\alpha)}{L(\hat{p})}\right) \xrightarrow{d} \chi(1), \tag{33}$$

where $\hat{p} = n_1/(n_0 + n_1)$ is the maximum likelihood estimator of p.

The second test is to check whether the process $\{d_t\}$ is serially independent. If the transition probability of the first-order Markov chain is denoted as $\pi_{ij} = \Pr(d_t = j | d_{t-1} = i)$, then the likelihood ratio of independence can be tested by

$$\mathrm{LR}_2 = -2 \ln \left[\frac{L(\hat{p})}{L(\hat{\pi}_{01}, \hat{\pi}_{11})}\right] \xrightarrow{d} \chi(1), \tag{34}$$

where

$$L(\hat{\pi}_{01}, \hat{\pi}_{11}) = (1-\hat{\pi}_{01})^{n_{00}} \hat{\pi}_{01}^{n_{01}} (1-\hat{\pi}_{11})^{n_{10}} \hat{\pi}_{11}^{n_{11}}, \tag{35}$$

where n_{ij} is the number of observations with value i followed by j, $\hat{\pi}_{01} = n_{01}/(n_{00} + n_{01})$, and $\hat{\pi}_{11} = n_{11}/(n_{10} + n_{11})$. By combining the two tests, the third test for conditional coverage can be constructed,

$$\mathrm{LR}_3 = \mathrm{LR}_1 + \mathrm{LR}_2 \xrightarrow{d} \chi(2). \tag{36}$$

3.2. Forecast evaluation criteria

Our primary objective is to compare the various VaR models with the most popular model (the benchmark). We use the RM(0.94) as a benchmark model and call it model 0. When several models using the same data are compared in predictive ability, it is crucial to take into account the dependence among the models. Failing to do so will result in the data-snooping problem which occurs when a model is searched extensively until a match with the given data is found. Conducting inference without taking into account specification search is commonly referred to as "data-snooping" and can be extremely misleading (cf. Lo and MacKinlay, 1999, Chapter 8). White (2000) develops a noble test to compare multiple models in predictive ability accounting for specification search, built on West (1996) and Diebold and Mariano (1995).

As will be discussed shortly in the next subsection, comparison of l models via given forecast criteria can be formulated as hypothesis testing of some suitable moment conditions of the loss-differential f. Consider an $l \times 1$ vector of moments, $E(f^*)$, where $f^* = f(Z, \beta^*)$ is an $l \times 1$ vector with elements $f_k^* \equiv f_k(Z, \beta^*)$ for a random vector $Z = (Y, X')'$ and $\beta^* \equiv \mathrm{plim}\, \hat{\beta}_n$. Hypothesis testing for $E(f^*)$ can be conducted whenever the $l \times 1$ sample moment vector

$$\bar{f} = P^{-1} \sum_{t=R}^{T} f(Z_{t+1}, \hat{\beta}_t) \tag{37}$$

has a continuous limiting distribution. For example, when we compare model k with the benchmark model ($k = 0$) using a loss function Loss, the kth element of the $l \times 1$ sample moment vector $\bar{f}$ is the loss differential of models k and 0, i.e.

$$\bar{f}_k = \text{Loss}_k - \text{Loss}_0 \quad (k = 1, \ldots, l). \tag{38}$$

We now define the three loss functions for Loss.

The first loss function A_k^α is based on the negative quasi-log likelihood of Bertail et al. (2000), i.e.

$$A_k^\alpha = P^{-1} \sum_{t=R}^{T} |y_t - \text{VaR}_t^k(\alpha)| \times [\alpha d_t^k + (1-\alpha)(1 - d_t^k)], \tag{39}$$

which weights the observed deviation from the VaR with the probability with which it is supposed to occur. Smaller A_k^α indicates a better goodness-of-fit.

The second loss function B_k^α is based on the negative tail mean return, which is defined as

$$B_k^\alpha = -P^{-1} \sum_{t=R}^{T} \frac{y_t d_t^k}{n_1^k}, \tag{40}$$

where $n_1^k = \sum_{t=R}^{T} d_t^k$ is the number of tail returns to be used in computing B_k^α. The model with smaller B_k^α is a better one.

The third loss function C_k^α is the average of likelihood ratio statistic LR_1 in Eq. (33), i.e.

$$\begin{aligned} C_k^\alpha \equiv P^{-1}\text{LR}_1 &= P^{-1}[-2\ \ln L(a) + 2\ \ln L(\hat{p}_k^\alpha)] \\ &= P^{-1} \sum_{t=R}^{T} -2[d_t^k\ \ln(\alpha) + (1 - d_t^k) \ln(1-\alpha)] \\ &\quad + 2[d_t^k\ \ln(\hat{p}_k^\alpha) + (1 - d_t^k)\ \ln(1 - \hat{p}_k^\alpha)]. \end{aligned} \tag{41}$$

As the smaller LR_1 indicates that the coverage probability p is closer to α, the model with lower C_k^α generates the VaR forecasts with a better coverage probability and, thus, C_k^α is to be minimized. $C_k^\alpha = 0$ if $\hat{p}_k^\alpha = \alpha$ and $C_k^\alpha > 0$ if $\hat{p}_k^\alpha \neq \alpha$. It should be noted that the Christoffersen test LR_1 is to test the null hypothesis $H_0 : p_k^\alpha = \alpha$ for a given model k and it is not to compare models, while the White's reality check using C_k^α is for model selection and to compare models. When $n_1^k = 0$, $\hat{p}_k^\alpha = 0$ and, thus, $\ln(\hat{p}_k^\alpha)$ and C_k^α are not defined.

All of the three objective functions A_k^α, B_k^α, and C_k^α for our VaR reality check will be minimized. Note that all three loss functions are expressed as sample moments of the form $P^{-1} \sum_{t=R}^{T} x_t$ with x_t being the respective summands expressed in Eqs. (39)–(41), so that test statistics can be formulated based on the sample moment vector of loss differentials as in Eq. (37).

These loss functions may be extended to incorporate the penalty of the BIS's capital adequacy coefficient θ, which depends on v, the number of coverage failure out of 250 days. Thus, $v = [n_1 250/P]$, where $[a]$ is the nearest integer to a. In our case $P = 522$ or 505. The following table (cf. Crouhy et al., 1998, p. 15) is used by the BIS for $\alpha = 0.01$. The BIS penalized loss functions, $\theta_k A_k^\alpha$, $\theta_k B_k^\alpha$, and $\theta_k C_k^\alpha$ may be considered. However, we leave them

v	4 or less	5	6	7	8	9	10 or more
θ	3	3.4	3.5	3.65	3.75	3.85	4

for future research and focus on A_k^α, B_k^α, and C_k^α in this paper. Each loss function has its advantages and disadvantages depending on the need and perspective. We have used three different loss functions to reflect the notion that different agents may have different objectives to evaluate risk measurement forecasts. Hence, we use all three objectives in our empirical section.

3.3. Reality check for predictive ability

Suppose, one-step predictions are to be made for P prediction periods, indexed from R to T, so that $T = R + P - 1$. Here, P and R may increase as the sample size T increases. The first forecast is based on the model parameter estimator $\hat{\beta}_R$, formed using observations 1 through R, the next based on the model parameter estimator $\hat{\beta}_{R+1}$, formed using observations 2 through $R + 1$, and so forth, with the final forecast based on the model parameter estimator $\hat{\beta}_T$. Often, model comparison via forecast criteria can be conveniently formulated as hypothesis testing of some suitable moment conditions. Consider an $l \times 1$ vector of moments, $E(f^*)$, where $f^* = f(Z, \beta^*)$ is an $l \times 1$ vector with elements $f_k^* \equiv f_k(Z, \beta^*)$ for a random vector $Z = (Y, X')'$ and $\beta^* \equiv \text{plim}\,\hat{\beta}_T$. As discussed earlier, hypothesis testing for $E(f^*)$ can be conducted whenever the $l \times 1$ sample moment vector $\bar{f} = P^{-1}\sum_{t=R}^{T} f(Z_{t+1}, \hat{\beta}_t)$ has a continuous limiting distribution.

West (1996, Theorem 4.1) shows that under proper regularity conditions:

$$\sqrt{P}(\bar{f} - E(f^*)) \to N(0, \Omega) \text{ in distribution} \tag{42}$$

as $P \equiv P(T) \to \infty$ when $T \to \infty$, where Ω, is a $l \times l$ matrix

$$\Omega = \lim_{T\to\infty} \text{var}[P^{-1/2} \sum_{t=R}^{T} f(Z_{t+1}, \hat{\beta}_t)], \tag{43}$$

which is a complicated expression as Ω depends on the estimated parameter $\hat{\beta}_t$. When either

$$F \equiv E\left(\frac{\partial f^*}{\partial \beta}\right) = 0 \tag{44}$$

or $P/R \to 0$ as $T \to \infty$, Ω can be substantially simplified because then Ω does not depend on the estimated parameter $\hat{\beta}_t$ and

$$\Omega = \lim_{T\to\infty} \text{var}[P^{-1/2} \sum_{t=R}^{T} f(Z_{t+1}, \beta^*)], \tag{45}$$

which corresponds to West (1996, Theorem 4.1(a)) and the result of Diebold and Mariano (1995). Here, the effect of using $\hat{\beta}_t$ rather than β^* is asymptotically negligible. One can proceed as if β^* were known and were equal to $\hat{\beta}_t$ in period t. However, when $F \neq 0$, as is the case in this paper due to the use of the indicator d_t in our three loss functions, Ω is unknown and depends on $\hat{\beta}_t$. In this case, although Ω is not feasible to derive even asymptotically, a bootstrap procedure may be used to obtain the null distribution of the statistic.

When we compare a single model $(l = 1)$ with a benchmark we can use the tests by Diebold and Mariano (1995) and West (1996) with an appropriate estimator of Ω. When we compare multiple forecasting models $(l > 1)$ against a given benchmark model, however, sequential use of Diebold and Mariano (1995) and West (1996) tests may result in a data-snooping bias since the test statistics are mutually dependent due to the use of the same data. See Lo and MacKinlay (1999, Chapter 8) and White (2000) for more discussions on the biases due to data snooping. To account for possible bias due to data snooping, we use White (2000) procedure. The appropriate null hypothesis is that the best model is no better than a benchmark, expressed formally as

$$H_0 : \max_{1 \leq k \leq l} E(f_k^*) \leq 0. \tag{46}$$

This is a multiple hypothesis, the intersection of the one-sided individual hypotheses $E(f_k^*) \leq 0$, $k = 1, \ldots, l$. The alternative is that H_0 is false, i.e. that the best model is superior to the benchmark. White (2000) test statistic for H_0 in Eq. (46) is formed as follows:

$$\overline{V} \equiv \max_{1 \leq k \leq l} \sqrt{P}\bar{f}_k, \tag{47}$$

which converges in distribution to $\max_{1 \leq k \leq l} Z_k$ under H_0, where the limit random vector $Z = (Z_1, \ldots, Z_l)'$ is $N(0, \Omega)$. This null limit distribution is, however, unknown (due to unknown Ω) and not feasible to derive even asymptotically. White (2000) suggests to use the stationary bootstrap of Politis and Romano (1994) to obtain the null distribution of $\overline{V}$. This gives appropriate P-values for testing the null hypothesis that the best model has no predictive superiority relative to the benchmark (White, 2000, Corollary 2.4). The P-value is called the "reality check P-value" for data snooping. White (2000, Proposition 2.5) also shows that the test's level can be driven to zero at the same time the power approaches to one as $\overline{V}$ diverges at rate $P^{1/2}$ under the alternative. Implementation of the reality check bootstrap and an illustrative example can be found in White (2000). See also Sullivan et al. (1998, 1999) for applications to the studies of technical trading rules and calendar effects in asset markets.

4. Results

We now evaluate the out-of-sample predictive ability of the models described in Section 2 using the evaluation methods in Section 3. Daily stock market index data for Japan (Nikkei 225) are obtained from Datastream. Logarithmic returns for Nikkei 225 index are analyzed from 2 January 1984 to 9 December 2000 (the date when we collected the data) with the total 4420 observations. Nikkei 225 is the stock market index of 225 companies listed in the Tokyo Stock Exchange. Those 225 companies are occasionally reviewed. A total of 30 of these companies are replaced in April 2000. This discontinuity may constitute less of a problem for our study since this period corresponds only to a minor portion of time span of the present paper. Besides, our major focus is on the crisis period including the data between 1995 and 1998. However, we note that TOPIX index is an alternative index of Japanese stock market and more comprehensive than Nikkei 225, while we leave it for future studies focusing on more recent observations.

The VaR models are estimated using $R = 2871$. The out-of-sample forecast evaluation is conducted over three subperiods, period 1 (2 January 1995–31 December 1996, $P = 522$), period 2 (1 January 1997–31 December 1998, $P = 522$), and period 3 (1 January 1999–9 December 2000, $P = 505$). Period 2 includes the financial turmoil which started in late 1997. The empirical analysis conducted in this paper have a number of implications for predictive performance of various VaR models in various dimensions, for three out-of-sample periods, for two tail probabilities ($\alpha = 0.01, 0.05$), and for three loss functions (A, B, and C).

In Table 1, the results from Christoffersen (1998) tests for coverage probabilities corresponding to various VaR models are presented. One of the most important findings is that the risk forecasting performance of EVT models turn out to be much worse than that of conventional VaR models such as TGARCH and GARCH volatility models. For instance, in period 1 (pre-crisis period), for $\alpha = 0.05$, none of the EVT models produce satisfactory coverage probabilities. However, for the same period and the same α, many conventional VaR models generate reasonable conditional and unconditional coverage probabilities. The coverage estimates obtained by the TGARCH model, except for the crisis period with $\alpha = 0.05$, are close to the true coverages. For the crisis period, risk forecast performance of many models were less than satisfactory. For period 2 with $\alpha = 0.05$, only four models could produce favorable coverage probabilities namely, MA(200), RM(0.97), $\mathrm{MC}_2 + \mathrm{RM}(0.94)$, and $\mathrm{MC}_2 + \mathrm{GARCH}_N$. For period 2 with $\alpha = 0.01$, relatively more successful risk forecasts are obtained. In this case, 12 models out of 27 had favorable risk forecasts. HS has shown some satisfactory performance for this case. For the post-crisis period more optimistic picture can be drawn. In this period, many of the risk models could produce favorable coverages. For $\alpha = 0.05$, except for all EVT models and the two non-parametric methods (namely, hybrid HS and NPQ), most methods successfully capture the required coverages. The performance of most risk forecasts appear to be more successful for the post crisis period (especially with $\alpha = 0.01$) which can indicate that severe effects of crisis has died out.

As a general conclusion drawn from this analysis is that the EVT models do not produce superior risk forecasts than that of more conventional VaR models. These findings are in contrast with the conclusions drawn by Ho et al. (2000) and Danielsson and Morimoto (2000) where EVT models are claimed to be more successful than the traditional VaR methods.

We now turn to the results obtained from the reality check which are presented in Tables 2–4, where RC_1 and RC_2 denote P-values of the test by White (2000) computed using the stationary bootstrap of Politis and Romano (1994). The bootstrap reality check P-values are computed with 1000 bootstrap resamples and the bootstrap smoothing parameter $q = 0.75$. See Politis and Romano (1994) and White (2000) for the details. The P-values for $q = 0.25$ and 0.50 are similar and are not reported. RC_1 is to compare each model k ($k = 1, \dots, 26$) with the benchmark model RM(0.94) ($k = 0$). RC_2 is to compare the best of the first l ($l = 1, \dots, 26$) models with the benchmark. RC_1 may be considered as a bootstrap version of Diebold and Mariano (1995) and West (1996) with taking into account the fact that $F \equiv E(\partial f^* / \partial \beta) \neq 0$.

Table 1
Christoffersen tests[a]

k	Model	Period 1						Period 2						Period 3					
		$\hat{p}_k^{0.05}$	LR_1	LR_2	$\hat{p}_k^{0.01}$	LR_1	LR_2	$\hat{p}_k^{0.05}$	LR_1	LR_2	$\hat{p}_k^{0.01}$	LR_1	LR_2	$\hat{p}_k^{0.05}$	LR_1	LR_2	$\hat{p}_k^{0.01}$	LR_1	LR_2
0	RM(0.94)	0.069	3.59	1.29	0.013	0.56	0.19	0.073	5.09	0.27	0.015	1.30	2.64	0.048	0.06	0.51	0.018	2.55	0.33
1	RM(0.97)	0.061	1.34	0.64	0.013	0.56	0.19	0.067	2.94	0.06	0.015	1.30	2.64	0.048	0.61	0.03	0.010	0.00	0.10
2	RM(0.90)	0.075	5.92	0.37	0.015	1.30	0.25	0.075	5.92	0.00	0.015	1.30	2.64	0.052	0.03	0.26	0.018	2.55	0.33
3	MA(200)	0.069	3.59	2.36	0.017	2.29	0.32	0.065	2.34	0.29	0.023	6.53	1.21	0.040	1.21	0.03	0.010	0.00	0.10
4	$Garch_N$	0.075	5.92	0.43	0.010	0.01	0.10	0.079	7.75	0.02	0.010	0.01	0.10	0.056	0.32	0.10	0.012	0.17	0.17
5	$Garch_t$	0.065	2.34	0.02	0.012	0.11	0.14	0.071	4.31	0.19	0.012	0.11	0.14	0.048	0.06	0.51	0.014	0.69	0.23
6	$Garch_{GED}$	0.063	1.81	0.00	0.012	0.11	0.14	0.073	5.09	0.02	0.008	0.31	0.06	0.005	0.00	0.38	0.014	0.69	0.23
7	$Egarch_N$	0.061	1.34	0.00	0.015	1.30	0.25	0.113	32.81	0.09	0.027	10.25	0.77	0.071	4.33	0.20	0.016	1.49	0.26
8	$Egarch_t$	0.071	4.31	0.18	0.013	0.56	0.19	0.117	36.43	0.24	0.038	24.65	0.07	0.075	5.96	0.00	0.020	3.83	0.40
9	$Egarch_{GED}$	0.060	0.94	0.01	0.013	0.56	0.19	0.107	27.67	0.00	0.023	6.53	0.57	0.071	4.33	0.05	0.016	1.49	0.26
10	$Tgarch_N$	0.067	2.94	1.11	0.006	1.12	0.03	0.083	9.79	0.93	0.013	0.56	0.20	0.056	0.32	0.30	0.018	2.55	0.37
11	$Tgarch_t$	0.063	1.81	0.00	0.010	0.01	0.10	0.071	4.31	0.19	0.008	0.31	0.06	0.052	0.03	0.26	0.016	1.49	0.29
12	$Tgarch_{GED}$	0.063	1.81	0.00	0.008	0.31	0.06	0.079	7.75	0.62	0.012	0.11	0.14	0.056	0.32	0.10	0.016	1.49	0.29
13	HS	0.046	0.17	0.66	0.006	1.12	0.04	0.106	26.03	0.77	0.017	2.29	0.32	0.040	1.21	0.03	0.006	0.98	0.04
14	HS(200, 0.99)	0.192	132.77	2.29	0.063	67.78	1.62	0.255	244.75	0.47	0.141	259.17	0.08	0.188	123.24	0.69	0.079	98.30	0.02
15	HS(200, 0.97)	0.180	115.14	0.80	0.090	117.62	2.51	0.255	244.75	0.00	0.171	352.01	0.72	0.185	117.36	1.12	0.097	138.96	0.31
16	MC_1	0.042	0.70	1.04	0.008	0.31	0.06	0.098	19.90	0.26	0.038	24.65	4.26	0.040	1.21	0.03	0.014	0.69	0.23
17	MC_2	0.042	0.70	0.01	0.008	0.31	0.06	0.098	19.90	0.26	0.035	19.37	2.14	0.038	1.75	0.08	0.012	0.17	0.17
18	$MC_2 + Garch_N$	0.060	0.94	0.70	0.015	1.30	0.25	0.067	2.94	0.06	0.019	3.50	0.40	0.046	0.21	0.68	0.018	2.55	0.36
19	$MC_2 + Egarch_{GED}$	0.052	0.04	0.26	0.017	2.29	0.31	0.104	24.44	0.03	0.046	36.43	0.65	0.062	1.31	0.54	0.022	5.32	1.30
20	MC_2 + RM(0.94)	0.061	1.34	0.64	0.015	1.30	0.25	0.067	2.94	0.06	0.025	8.31	4.39	0.044	0.45	0.86	0.020	3.83	0.45
21	NPQ_1	0.205	154.47	0.65	0.073	87.56	3.44	0.263	259.90	0.21	0.148	281.67	0.02	0.208	154.20	0.60	0.099	143.72	0.21
22	NPQ_2	0.242	219.02	0.02	0.144	270.35	0.18	0.290	315.39	0.14	0.225	530.73	1.19	0.244	215.89	0.17	0.157	298.45	2.09
23	Neftci	0.098	19.90	0.90	0.002	5.15	0.00	0.171	101.16	1.87	0.008	0.31	0.06	0.087	12.19	0.33	0.002	4.88	0.00
24	Longin(10)	0.090	14.47	1.88	0.006	1.12	0.03	0.163	90.46	0.37	0.031	14.52	2.90	0.081	8.84	0.06	0.006	0.98	0.04
25	Longin(20)	0.115	34.60	0.75	0.008	0.31	0.06	0.184	120.92	1.21	0.036	21.96	4.80	0.107	26.48	0.25	0.010	0.00	0.10
26	Hill	0.008	30.08	0.06	0.002	5.15	0.00	0.159	85.28	0.16	0.044	33.47	2.91	0.028	6.20	0.86	0.004	2.40	0.02

[a] The sample period of the data is from 2 January 1984 to 9 December 2000 (the date when we collected the data) with the total 4420 observations. The models are estimated using $R = 2871$ observations. The out-of-sample forcast evaluation is conducted over three subperiods, period 1 (2 January 1995–31 December 1996, $P = 522$), period 2 (1 January 1997–31 December 1998, $P = 522$), and period 3 (1 January 1999–9 December 2000, $P = 505$). LR_1 and LR_2 are the estimated statistics of Christoffersen (1998). LR_3 is not reported for space, which is the sum of LR_1 and LR_2. LR_1 and LR_2 follow asymptotically $\chi(1)$ with the 95% critical value 3.84. LR_3 is asymptotically χ^2-distributed and its 95% critical value is 5.99.

Table 2
Reality check using the loss function A[a]

k	Model	Period 1						Period 2						Period 3					
		$A_k^{0.05}$	RC_1	RC_2	$A_k^{0.01}$	RC_1	RC_2	$A_k^{0.05}$	RC_1	RC_2	$A_k^{0.01}$	RC_1	RC_2	$A_k^{0.05}$	RC_1	RC_2	$A_k^{0.01}$	RC_1	RC_2
0	RM(0.94)	1.76			2.92			2.42			4.02			1.98			3.28		
1	RM(0.97)	1.78	0.998	0.999	2.96	1.000	1.000	2.43	0.805	0.795	4.04	0.799	0.785	2.00	0.984	0.981	3.31	0.978	0.982
2	RM(0.90)	1.74	0.015	0.054	2.89	0.007	0.009	2.40	0.010	0.047	3.98	0.012	0.060	1.96	0.003	0.018	3.24	0.005	0.020
3	MA(200)	1.74	0.192	0.188	2.87	0.151	0.133	2.40	0.282	0.290	3.93	0.106	0.097	2.14	1.000	0.180	3.56	1.000	0.191
4	$Garch_N$	1.77	0.702	0.284	2.97	0.973	0.181	2.40	0.192	0.388	4.04	0.610	0.101	1.98	0.521	0.264	3.32	0.977	0.262
5	$Garch_t$	1.77	0.863	0.285	2.97	0.997	0.181	2.43	1.000	0.388	4.09	1.000	0.101	2.02	1.000	0.264	3.37	1.000	0.262
6	$Garch_{GED}$	1.79	0.995	0.287	2.99	1.000	0.000	2.42	0.649	0.388	4.05	0.859	0.101	2.02	1.000	0.264	3.35	1.000	0.262
7	$Egarch_N$	1.65	0.000	0.000	2.72	0.000	0.000	1.94	0.000	0.000	3.16	0.000	0.000	1.84	0.000	0.000	3.02	0.000	0.000
8	$Egarch_t$	1.56	0.000	0.000	2.58	0.000	0.000	1.89	0.000	0.000	3.07	0.000	0.000	1.78	0.000	0.000	2.90	0.000	0.000
9	$Egarch_{GED}$	1.65	0.000	0.000	2.73	0.000	0.000	1.97	0.000	0.000	3.21	0.000	0.000	1.86	0.000	0.000	3.05	0.000	0.000
10	$Tgarch_N$	1.79	0.906	0.000	3.00	0.978	0.000	2.48	0.994	0.000	4.16	0.996	0.000	2.03	0.969	0.000	3.37	0.990	0.000
11	$Tgarch_t$	1.76	0.497	0.000	2.94	0.724	0.000	2.51	1.000	0.000	4.21	1.000	0.000	2.00	0.830	0.000	3.33	0.962	0.000
12	$Tgarch_{GED}$	1.78	0.876	0.000	2.97	0.921	0.000	2.45	0.925	0.000	4.10	0.952	0.000	2.02	0.987	0.000	3.35	0.989	0.000
13	HS	1.96	1.000	0.000	3.62	1.000	0.000	2.02	0.000	0.000	3.57	0.000	0.000	2.11	1.000	0.000	3.63	1.000	0.000
14	HS(200, 0.99)	1.04	0.000	0.000	1.93	0.000	0.000	1.16	0.000	0.000	1.96	0.000	0.000	1.10	0.000	0.000	1.97	0.000	0.000
15	HS(200, 0.97)	1.08	0.000	0.000	1.71	0.000	0.000	1.21	0.000	0.000	1.79	0.000	0.000	1.15	0.000	0.000	1.77	0.000	0.000
16	MC_1	2.04	1.000	0.000	2.98	0.936	0.000	2.08	0.000	0.000	2.99	0.000	0.000	2.14	1.000	0.000	3.12	0.000	0.000
17	MC_2	2.04	1.000	0.000	2.98	0.942	0.000	2.07	0.000	0.000	2.99	0.000	0.000	2.14	1.000	0.000	3.12	0.000	0.000
18	$MC_2 + Garch_N$	1.91	1.000	0.000	2.78	0.000	0.000	2.56	1.000	0.000	2.75	0.000	0.000	2.13	1.000	0.000	3.10	0.000	0.000
19	$MC_2 + Egarch_{GED}$	1.70	0.020	0.000	2.48	0.000	0.000	2.03	0.000	0.000	2.92	0.000	0.000	1.92	0.000	0.000	2.77	0.000	0.000
20	MC_2 + RM(0.94)	1.83	1.000	0.000	2.66	0.000	0.000	2.52	1.000	0.000	3.66	0.000	0.000	2.06	1.000	0.000	2.99	0.000	0.000
21	NPQ_1	1.00	0.000	0.000	1.75	0.000	0.000	1.11	0.000	0.000	1.78	0.000	0.000	1.05	0.000	0.000	1.79	0.000	0.000
22	NPQ_2	0.89	0.000	0.000	1.23	0.000	0.000	1.02	0.000	0.000	1.32	0.000	0.000	0.95	0.000	0.000	1.28	0.000	0.000
23	Neftci	1.39	0.000	0.000	5.14	1.000	0.000	1.52	0.000	0.000	5.11	1.000	0.000	1.59	0.000	0.000	5.29	1.000	0.000
24	Longin(10)	1.44	0.000	0.000	3.21	1.000	0.000	1.55	0.000	0.000	3.16	0.000	0.000	1.67	0.000	0.000	3.48	1.000	0.000
25	Longin(20)	1.31	0.000	0.000	2.94	0.716	0.000	1.43	0.000	0.000	3.00	0.000	0.000	1.50	0.000	0.000	3.25	0.255	0.000
26	Hill	2.59	1.000	0.000	4.82	1.000	0.000	1.58	0.000	0.000	2.80	0.000	0.000	2.49	1.000	0.000	3.97	1.000	0.000

[a] The sample period of the data is from 2 January 1984 to 9 December 2000 (the date when we collected the data) with the total 4420 observations. The models are estimated using $R = 2871$ observations. The out-of-sample forcast evaluation is conducted over three subperiods, period 1 (2 January 1995–31 December 1996, $P = 522$), period 2 (1 January 1997–31 December 1998, $P = 522$), and period 3 (1 January 1999–9 December 2000, $P = 505$). A_k^{α} is the negative QLL for model k with α quantile. Smaller A_k^{α} indicates a better goodness of fit. RC_1 and RC_2 denote reality check P-values of White (2000) test computed using the stationary bootstrap of Politis and Romano (1994, PR). The bootstrap reality check P-values are computed with 1000 bootstrap resamples and the bootstrap smoothing parameter $q = 0.75$. See PR or White (1998) for the details. The P-values for $q = 0.25$ and 0.50 are similar and are not reported. RC_1 is to compare each model k ($k = 1, \ldots, 26$) with the benchmark model RM(0.94) ($k = 0$). RC_2 is to compare the best of the first l ($l = 1, \ldots, 26$) models with the benchmark model. RC_1 may be considered as a bootstrap version of Diebold and Mariano (1995) and West (1996).

Table 3
Reality check using the loss function B[a]

k	Model	Period 1						Period 2						Period 3					
		$B_k^{0.05}$	RC_1	RC_2	$B_k^{0.01}$	RC_1	RC_2	$B_k^{0.05}$	RC_1	RC_2	$B_k^{0.01}$	RC_1	RC_2	$B_k^{0.05}$	RC_1	RC_2	$B_k^{0.01}$	RC_1	RC_2
0	RM(0.94)	4.34			6.79			6.13			7.19			5.76			7.18		
1	RM(0.97)	4.48	0.765	0.751	6.79	0.000	0.000	6.34	0.676	0.692	8.19	0.689	0.717	5.85	0.585	0.588	8.22	0.681	0.682
2	RM(0.90)	4.08	0.150	0.257	6.28	0.120	0.135	6.07	0.376	0.629	7.19	0.000	0.528	5.46	0.181	0.327	7.18	0.000	0.490
3	MA(200)	4.46	0.648	0.412	6.47	0.395	0.399	6.54	0.739	0.730	8.61	0.734	0.670	6.17	0.745	0.487	8.75	0.699	0.602
4	$Garch_N$	4.23	0.387	0.456	6.68	0.473	0.606	5.94	0.212	0.589	5.70	0.249	0.530	5.24	0.136	0.336	7.39	0.556	0.729
5	$Garch_t$	4.33	0.511	0.477	7.11	0.627	0.690	6.19	0.651	0.591	7.62	0.626	0.585	5.63	0.319	0.344	7.32	0.533	0.750
6	$Garch_{GED}$	4.40	0.591	0.483	7.11	0.620	0.690	6.05	0.354	0.638	6.12	0.306	0.618	5.55	0.236	0.345	7.23	0.498	0.792
7	$Egarch_N$	4.68	0.779	0.553	6.65	0.455	0.756	5.56	0.145	0.326	8.12	0.631	0.623	5.17	0.162	0.362	7.55	0.596	0.876
8	$Egarch_t$	4.49	0.698	0.574	6.94	0.532	0.774	5.50	0.118	0.295	7.09	0.463	0.631	5.05	0.105	0.280	7.16	0.470	0.902
9	$Egarch_{GED}$	4.71	0.778	0.589	6.94	0.578	0.774	5.62	0.145	0.305	8.44	0.725	0.640	5.13	0.165	0.286	7.55	0.581	0.902
10	$Tgarch_N$	4.21	0.389	0.600	7.26	0.583	0.832	5.89	0.231	0.307	6.18	0.273	0.651	5.08	0.167	0.330	7.00	0.425	0.900
11	$Tgarch_t$	4.34	0.509	0.609	7.69	0.742	0.851	6.14	0.487	0.321	7.25	0.513	0.688	5.35	0.234	0.348	7.18	0.504	0.900
12	$Tgarch_{GED}$	4.34	0.484	0.609	7.75	0.686	0.880	5.91	0.219	0.321	6.29	0.333	0.692	5.25	0.185	0.351	7.18	0.531	0.900
13	HS	5.23	0.938	0.680	9.23	0.763	0.929	5.79	0.287	0.340	9.44	0.766	0.703	6.22	0.758	0.379	10.41	0.753	0.910
14	HS(200, 0.99)	3.07	0.014	0.021	4.39	0.139	0.590	4.01	0.004	0.008	5.09	0.226	0.582	3.58	0.019	0.018	4.81	0.138	0.536
15	HS(200, 0.97)	3.10	0.016	0.021	3.87	0.115	0.496	3.97	0.004	0.008	4.70	0.148	0.506	3.51	0.012	0.016	4.42	0.107	0.456
16	MC_1	5.34	0.929	0.028	8.47	0.766	0.505	5.92	0.370	0.008	7.70	0.593	0.511	6.22	0.780	0.016	7.93	0.655	0.464
17	MC_2	5.34	0.934	0.032	8.47	0.797	0.505	5.92	0.363	0.008	7.92	0.624	0.513	6.31	0.805	0.016	8.20	0.637	0.471
18	$MC_2 + Garch_N$	4.47	0.638	0.032	6.19	0.256	0.505	5.83	0.244	0.008	6.62	0.386	0.521	5.48	0.284	0.016	7.08	0.474	0.471
19	$MC_2 + Egarch_{GED}$	4.91	0.855	0.035	6.42	0.397	0.508	5.66	0.145	0.008	6.94	0.426	0.523	5.36	0.254	0.016	6.82	0.386	0.471
20	$MC_2 + RM(0.94)$	4.41	0.582	0.035	6.58	0.419	0.508	6.28	0.748	0.008	7.36	0.561	0.525	5.83	0.555	0.016	6.93	0.407	0.471
21	NPQ_1	3.00	0.009	0.031	4.42	0.155	0.513	3.96	0.007	0.008	5.03	0.194	0.525	3.44	0.013	0.012	4.61	0.120	0.479
22	NPQ_2	2.77	0.006	0.013	3.45	0.082	0.419	3.74	0.003	0.006	4.28	0.134	0.442	3.17	0.009	0.007	3.90	0.085	0.396
23	Neftci	4.09	0.262	0.013	11.04	0.734	0.541	4.85	0.029	0.006	10.54	0.717	0.528	4.93	0.115	0.007	14.33	0.734	0.562
24	Longin(10)	4.22	0.360	0.013	9.23	0.779	0.541	4.96	0.051	0.006	8.16	0.676	0.529	5.05	0.127	0.007	10.41	0.768	0.562
25	Longin(20)	3.87	0.115	0.013	8.47	0.778	0.541	4.71	0.019	0.006	7.82	0.613	0.529	4.58	0.078	0.007	8.97	0.755	0.565
26	Hill	8.47	0.855	0.387	11.04	0.747	0.541	5.01	0.048	0.006	7.44	0.571	0.529	6.86	0.850	0.015	11.85	0.784	0.579

[a] See the footnote for Table 2. B_k^{α} is the rescaled negative left tail mean return for model k with α quantile. A smaller B_k^{α} indicates a better model.

Table 4
Reality check using the loss function C[a]

k	Model	Period 1				Period 2				Period 3			
		$\alpha = 0.05$		$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.01$	
		RC_1	RC_2	RC_1	RC_2	RC_1	RC_2	RC_1	RC_2	RC_1	RC_2	RC_1	RC_2
1	RM(0.97)	0.121	0.116	0.000	0.000	0.183	0.191	0.409	0.402	0.393	0.398	0.184	0.207
2	RM(0.90)	0.821	0.290	0.739	0.558	0.726	0.237	0.000	0.402	0.497	0.670	0.000	0.207
3	MA(200)	0.441	0.400	0.772	0.752	0.231	0.267	0.890	0.658	0.721	0.891	0.195	0.207
4	Garch_N	0.817	0.476	0.348	0.827	0.888	0.343	0.283	0.669	0.583	0.912	0.196	0.220
5	Garch_t	0.243	0.480	0.304	0.835	0.198	0.344	0.221	0.672	0.393	0.946	0.183	0.226
6	$\text{Garch}_{\text{GED}}$	0.186	0.483	0.285	0.835	0.368	0.354	0.390	0.696	0.477	0.946	0.182	0.230
7	Egarch_N	0.196	0.521	0.648	0.897	0.997	0.604	0.913	0.769	0.818	0.967	0.339	0.273
8	Egarch_t	0.603	0.571	0.391	0.904	0.999	0.647	0.983	0.833	0.850	0.972	0.729	0.333
9	$\text{Egarch}_{\text{GED}}$	0.188	0.515	0.396	0.904	0.993	0.657	0.900	0.843	0.797	0.974	0.327	0.333
10	Tgarch_N	0.385	0.534	0.536	0.908	0.909	0.679	0.351	0.865	0.601	0.974	0.406	0.354
11	Tgarch_t	0.232	0.537	0.357	0.909	0.358	0.686	0.368	0.870	0.507	0.974	0.296	0.354
12	$\text{Tgarch}_{\text{GED}}$	0.210	0.537	0.405	0.909	0.860	0.693	0.275	0.870	0.561	0.974	0.277	0.354
13	HS	0.207	0.450	0.524	0.922	0.988	0.721	0.628	0.871	0.752	0.985	0.321	0.423
14	HS(200, 0.99)	1.000	0.736	0.999	0.957	1.000	0.800	1.000	0.923	1.000	0.989	1.000	0.728
15	HS(200, 0.97)	1.000	0.777	1.000	0.966	1.000	0.817	1.000	0.939	1.000	0.991	0.999	0.771
16	MC_1	0.285	0.791	0.425	0.966	0.974	0.822	0.980	0.953	0.734	0.991	0.262	0.776
17	MC_2	0.294	0.793	0.432	0.966	0.987	0.822	0.955	0.954	0.786	0.992	0.195	0.779
18	$MC_2 + \text{Garch}_N$	0.136	0.793	0.712	0.971	0.180	0.835	0.733	0.964	0.590	0.998	0.000	0.779
19	$MC_2 + \text{Egarch}_{\text{GED}}$	0.160	0.775	0.717	0.976	0.989	0.838	0.989	0.965	0.675	0.998	0.802	0.795
20	$MC_2 + \text{RM}(0.94)$	0.129	0.775	0.754	0.984	0.103	0.840	0.911	0.972	0.684	0.999	0.815	0.796
21	NPQ_1	1.000	0.795	1.000	0.986	1.000	0.854	1.000	0.975	1.000	0.999	1.000	0.829
22	NPQ_2	1.000	0.812	1.000	0.986	1.000	0.861	1.000	0.975	1.000	0.999	1.000	0.862
23	Neftci	0.969	0.838	0.913	0.988	1.000	0.870	0.336	0.977	0.931	0.999	0.622	0.884
24	Longin(10)	0.970	0.842	0.541	0.988	1.000	0.875	0.940	0.978	0.888	0.999	0.323	0.884
25	Longin(20)	0.998	0.855	0.426	0.988	1.000	0.878	0.968	0.978	0.979	0.999	0.186	0.884
26	Hill	1.000	0.898	0.903	0.988	1.000	0.880	0.985	0.981	0.943	0.999	0.447	0.892

[a] See the footnote for Table 2. As $C_k^\alpha \times P = \text{LR}_1$ is reported in Table 1, we do not report C_k^α here again for space. $C_k^\alpha = P^{-1} \times \text{LR}_1$, where $P = 522$ for periods 1, 2, and $P = 505$ for period 3. A smaller C_k^α indicates a better model.

Table 2 presents the reality check results using the loss function A, the negative quasi-log likelihood. Results obtained here suggest that the predictive power of many VaR models such as EVT methods, ARCH models, and MC methods, perform better than the benchmark model RM(0.94). Many of the P-values corresponding to RC_1 and RC_2 are generally small, indicating that many models generate better forecasts. Furthermore, some models which do not produce significant forecast gain during pre-crisis period, appear to have superior forecasts during crisis. For instance, EVT models and Monte Carlo models show high P-values close to 1 for RC_1 with $\alpha = 0.01$ during pre- and post-crisis periods. However, these models produce the P-values equal to 0 for RC_1 during the crisis. This implies that the benchmark model works relatively better during the normal times than during the crisis period.

Table 3 presents the reality check results for the loss function B. We obtain less significant predictive gains over the benchmark model under the loss function B than under the loss A. In terms of the loss B, predictive gains can mostly be obtained with $\alpha = 0.05$. For instance, for $\alpha = 0.05$,the predictive power of many models improves over RM. On the other hand, for $\alpha = 0.01$, none of these models could attain predictive gains over RM. Although some models such as NPQ, HS, $MC_2 + RM(0.94)$, and some EVT methods produce some favorable predictive performance, consistent and uniform prediction improvements may not be observed with respect to the loss function B.

Our final reality check analysis is based on the loss function C which is the likelihood ratio of the unconditional coverage probabilities. The results are presented in the Table 4. It is clear that none of the alternative VaR models produce predictive performance superior to RM(0.94) in terms of the loss C.

Therefore, for different periods, different tail probability levels, and for different loss criteria, obtained are different performances for risk forecasting. Unlike the previous findings in the literature, e.g. Danielsson and Morimoto (2000) for Japan and Ho et al. (2000) for emerging markets, EVT models have not shown better predictive ability than conventional models. The difference between our results and the results obtained by Danielsson and Morimoto (2000) may stem from the following facts: (i) in our study a wider variety of VaR models are used than that of Danielsson and Morimoto; (ii) we have applied three different loss functions which account for various aspects of risk forecasts; (iii) we use Nikkei 225 index while they use TOPIX index.

Interestingly, some traditional methods such as TGARCH and Monte Carlo methods have proved to be more successful than the EVT models even during the crisis period. RM methods work relatively well in normal periods. Moreover, better risk forecasts can be obtained from the post-crisis period which can be interpreted that the severe impact of financial turmoil has died out. To conclude, forecast evaluation of VaR models has not revealed a systematic ranking among the models. The policy implications of these findings highlights the challenges faced with the policy analysts and academics who wish to use risk based capital adequacy requirements in banking sector to prevent potential financial crisis.

5. Conclusions

In this paper, a comprehensive predictive assessment of various VaR models is studied for the Japanese stock market. We have made the assessment of on 27 VaR models using three different loss functions via White (2000) reality check. The Riskmetrics model, RM(0.94), is used as a benchmark model for the comparison. The reality check with the various loss functions indicates that none of the methods studied in this paper exhibits consistently superior predictive ability for all periods. Unlike the recent findings in the literature, we could not confirm that the EVT models produce superior risk forecasts for crisis period. This finding highlights the difficulties in risk modelling and attempts to unify risk measurement practice to avoid potential financial crisis. To further investigate this important topic, more work on both theoretical and applied aspects of risk modelling is necessary.

References

Abberger, K., 1997. Quantile smoothing in financial time series. Statistical Papers 38, 125–148.

Alexander, C., 1998. Volatility and correlation: measurement, methods and applications. In: Alexander, C. (Ed.), Risk Management and Analysis, Vol. 1. Wiley, New York, Chapter 4, pp. 125–168.

Beim, D.O., Calomiris, C.W., 2001. Emerging Financial Markets. McGraw-Hill, New York.

Bertail, P., Haefke, C., Politis, D.N., White, H., 2000. A Subsampling Approach to Estimating the Distribution of Diverging Statistics with Applications to Assessing Financial Market Risks. UCSD Discussion Paper 2000/2001.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Boudoukh, J., Richardson, M., Whitelaw, R., 1998. The best of both worlds. Risk 11, 64–67.

Brodie, M., Glasserman, P., 1998. Simulation for option pricing and risk management. In: Alexander, C. (Ed.), Risk Management and Analysis, Vol. 1. Wiley, New York, Chapter 5, pp. 173–207.

Butler, J.S., Schachter, B., 1998. Estimating value-at-risk with a precision measure by combining kernel estimation with historical simulation. Review of Derivatives Research 1, 371–390.

Christoffersen, P.F., 1998. Evaluating interval forecasts. International Economic Review 39 (4), 841–864.

Christoffersen, P.F., Errunza, V.R., 2000. Towards a global financial architecture: capital mobility and risk management. Emerging Market Review 1, 2–19.

Crouhy, M., Galai, D., Mark, R., 1998. The New 1998 regulatory framework for capital adequacy. In: Alexander, C. (Ed.), Risk Management and Analysis, Vol. 1. Wiley, New York, Chapter 1, pp. 1–37.

Corsetti, S., Pesenti, P., Roubini, N., 1999. What caused the asian currency and financial crisis. Japan and World Economy 11 (3), 305–373.

Danielsson, J., 2000. The Emperor Has No Clothes: Limits to Risk Modelling. London School of Economics, London.

Danielsson, J., deHaan, L., Peng, L., deVries, C.G., 2001. Using a bootstrap method to choose the sample fraction in tail index estimation. Journal of Multivariate Analysis 76 (2), 226–248.

Danielsson, J., deVries, C.G., 1997. Beyond the Sample: Extreme Quantile and Probability Estimation. LSE and Tinbergen Institute.

Danielsson, J., Morimoto, J., 2000. Forecasting Extreme Financial Risk: A Critical Analysis of Practical Methods for the Japanese Market. IMES Discussion Paper Series-2000-E-8.

Davison, A.C., Smith, R.L. 1990. Models for exceedances over high thresholds. Journal of Royal Statistical Society B 52, 393–442.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.

Diebold, F.X., Santomera, A. 1999. Financial risk management in a volatile global environment. Asia Risk December, 35–36.

Dornbusch, R., 1998a. After Asia: New Directions for the International Financial System. MIT Press, New York.

Dornbusch, R., 1998b. Asian Crisis Themes. MIT Press, New York.

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

Embrechts P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer Verlag, New York.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. Econometrica 50, 987–1008.

FRBNY, 2000., Economic Policy Review, Special Issue: Lessons from Recent Crises in Asian and Other Emerging Markets, Vol. 6, No. 3. Federal Reserve Bank of New York, New York.

Fischer, S., 1998. The Asian Crisis, the IMF, and the Japanese Economy, www.imf.org.

Glosten, L.R., Jaganathan, R., Runkle, D., 1993. On the relationship between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779–1801.

Goldstein M., Kaminsky, G.L., Reinhart, C.M., 2000. Assessing Financial Vulnerability: An Early Warning System for Emerging Markets. Institute for International Economics, Washington, DC.

Haggard, S., 2000. The Political Economy of the Asian Financial Crisis. Institute for International Economics, Washington, DC.

Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. Annals of Statistics 19, 1547–1569.

Ho, L.C., Burridge, P., Cadle, J., Theobald, M., 2000. Value-at-risk: applying the extreme value approach to asian markets in recent financial turmoil. Pacific–Basin Finance Journal 8, 249–275.

Hull, J., 1997. Options Futures and Other Derivatives. Prentice-Hall, New York.

Jenkinson, A.F., 1955. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society 81, 145–158.

Morgan, J.P., 1995. Riskmetrics Technical Manual, 3rd Edition.

Jorion, P., 2000. Value-at-Risk, 2nd Edition. McGraw Hill, New York.

Krugman, P., 1998. Japan's Bank Bailout: Some Simple Arithmetic. MIT Press, New York.

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46 (1), 33–50.

Lejeune, M., Sarda, P., 1988. Quantile regression: a non-parametric approach. Computational Statistics and Data Analysis 6, 229–239.

Lo, A.W., MacKinlay, A.C., 1999. A Non-Random Walk Down Wall Street. Princeton University Press, Princeton.

Longin, F., 1996. The asymptotic distribution of extreme stock market returns. Journal of Business 69 (3), 383–408.

Longin, F., 2000. From value-at-risk to stress testing: the extreme value approach. Journal of Money Banking and Finance 24, 1097–1130.

Mason, D.M., 1982. Law of large numbers for sum of extreme values. Annals of Probabilty 10, 754–764.

Mikitani, R., Posen, A.S., 2000. Japan's Financial Crisis and its Parellels to US Experience. Special Report 13, Institute for International Economics.

Mises, R.V., 1936. La Distribution de la plus grande de *n* valeurs, Reprinted in Selected Papers II. American Mathematical Society, Providence, RI, 1954, pp. 271–294.

Mishkin, F.S., 1999. Lessons from the Asian Crisis. NBER Working Paper No. 7102.

Mishkin, F.S., 2000. Financial Policies and the Prevention of Financial Crises in Emerging Market Countries. Graduate School of Business, Columbia University, Columbia.

Mittnik, S., Paolella, M.S., Rachev, S.T., 1998. Unconditional and conditional distributional model for the nikkei index. Asia Pacific Financial Markets 5, 99–128.

Neftçi, S., 2000. Value-at-risk calculations, extreme events, and tail estimation. Journal of Derivatives Spring, 23–38.

Nelson, D.B., 1991. Conditional heteroscedasticity in asset returns: a new approach. Econometrica 59 (2), 347–370.

Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. Journal of American Statistical Association 89, 1303–1313.

Samanta, M., 1989. Non-parametric estimation of conditional quantiles. Statistics and Probability Letters 7, 407–412.

Smith, R.L., 1989. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. Statistical Science 4, 367–393.

Stulz, R., 2000. Why Risk Management is not a Rocket Science? Financial Times, Mastering Risk Series.

Sullivan, R., Timmermann, A., White, H., 1998. Dangers of Data-Driven Inference: The Case of Calendar Effects in Stock Returns, UCSD.

Sullivan, R., Timmermann, A., White, H., 1999. Data-snooping, technical trading rule performance, and the bootstrap. Journal of Finance 54, 1647–1692.

Tsay, R., 2000. Extreme Values, Quantile Estimation and Value-at-Risk. Graduate School of Business, University of Chicago, Chicago.

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

White, H., 2000. A reality check for data snooping. Econometrica 68 (5), 1097–1126.

Xiang, X., 1996. A kernel estimator of a conditional quantiles. Journal of Multivariate Analysis 59, 206–216.