Evaluating Predictive Performance of Value-at-Risk Models in Emerging Markets: A Reality Check

YONG BAO,[1] TAE-HWY LEE[2]* AND BURAK SALTOĞLU[3]
[1] *University of Texas, San Antonio, USA*
[2] *University of California, Riverside, USA*
[3] *Marmara University, Turkey*

ABSTRACT

We investigate the predictive performance of various classes of value-at-risk (VaR) models in several dimensions—unfiltered versus filtered VaR models, parametric versus nonparametric distributions, conventional versus extreme value distributions, and quantile regression versus inverting the conditional distribution function. By using the reality check test of White (2000), we compare the predictive power of alternative VaR models in terms of the empirical coverage probability and the predictive quantile loss for the stock markets of five Asian economies that suffered from the 1997–1998 financial crisis. The results based on these two criteria are largely compatible and indicate some empirical regularities of risk forecasts. The Riskmetrics model behaves reasonably well in tranquil periods, while some extreme value theory (EVT)-based models do better in the crisis period. Filtering often appears to be useful for some models, particularly for the EVT models, though it could be harmful for some other models. The CaViaR quantile regression models of Engle and Manganelli (2004) have shown some success in predicting the VaR risk measure for various periods, generally more stable than those that invert a distribution function. Overall, the forecasting performance of the VaR models considered varies over the three periods before, during and after the crisis. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS CaViaR; coverage probability; filtering; quantile loss; reality check; stress testing; VaR

INTRODUCTION

Increasing financial fragility in emerging markets and the extensive use of derivative products in developed countries can be characterized as two distinct features of the financial world over the last decade. Consequently, effective use of risk measurement tools has been suggested as a main panacea for mitigating growing financial risks. A uniform risk measurement methodology called value-at-risk (VaR) has received a great deal of attention from both regulatory and academic fronts.[1] During

---

*Correspondence to: Tae-Hwy Lee, Department of Economics, University of California, Riverside, CA 92521, USA.
E-mail: tae.lee@ucr.edu
[1] Other measures of risk have also been proposed in the literature. See Granger (2002) for an excellent discussion.

a short span of time, numerous papers have studied various aspects of VaR methodology. The recent research in this field has progressed so rapidly that comparing the relative predictive performance of different VaR models has not yet been matched. This comparison will provide valuable information, since precise risk forecasts are vital for risk practitioners and regulators.

Consider a financial return series $\{r_t\}_{t=1}^T$, generated by the probability law $\Pr(r_t \leq r | \mathcal{F}_{t-1}) \equiv F_t(r)$ conditional on the information set $\mathcal{F}_{t-1}$ (σ-field) at time $t-1$. Suppose $\{r_t\}$ follows the stochastic process

$$r_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t \tag{1}$$

where $\mu_t = E(r_t | \mathcal{F}_{t-1})$, $\sigma_t^2 = E(\varepsilon_t^2 | \mathcal{F}_{t-1})$ and $\{z_t\} \equiv \{\varepsilon_t / \sigma_t\}$ has the conditional distribution function $G_t(z) \equiv \Pr(z_t \leq z | \mathcal{F}_{t-1})$. The VaR with a given tail probability $\alpha \in (0, 1)$, denoted by $q_t(\alpha)$, is defined as the conditional quantile

$$F_t(q_t(\alpha)) = \alpha \tag{2}$$

which can be estimated by inverting the distribution function

$$q_t(\alpha) = F_t^{-1}(\alpha) = \mu_t + \sigma_t G_t^{-1}(\alpha) \tag{3}$$

Hence a VaR model involves the specification of $F_t(\cdot)$, or μ_t, σ_t^2, $G_t(\cdot)$. For a given model with the conditional mean μ_t, this paper considers different approaches to modelling the conditional variance σ_t^2 and conditional distribution $F_t(\cdot)$ or $G_t(\cdot)$. See Table I, where the various models are classified as 'unfiltered' if a VaR model involves the specification of $F_t(\cdot)$, or 'filtered' if a VaR model involves the specification of μ_t, σ_t^2, $G_t(\cdot)$. The filtered VaR models are computed using the standardized return series $z_t = (r_t - \mu_t)/\sigma_t$ with σ_t^2 estimated by a GARCH(1, 1) model.

If the dependence structure of $\{r_t\}$ can be fully described by the first two conditional moments, that is, $F_t(\cdot)$ belongs to a location-scale family, $\{z_t\}$ may be independently and identically distributed (IID) so that $G_t(\cdot) = G(\cdot)$, for which a parametric distribution may be used: e.g., the normal distribution, the Student-t distribution, the generalized error distribution (Nelson, 1991), the exponential generalized beta distribution (Wang *et al*., 2001), the stable Paretian distribution (Mittnik *et al*., 2002) and the mixture of normal distributions (Venkataraman, 1997). We can also estimate the distribution nonparametrically: e.g., the semi-parametric model of Engle and González-Rivera (1991), the historically simulated density and the nonparametric density. When $\{z_t\}$ is not IID, its time-varying conditional distribution $G_t(\cdot)$ may be modelled parametrically (e.g., Hansen, 1994; Harvey and Siddique, 1999, 2000), nonparametrically (e.g., Gallant *et al*., 1991; Hall *et al*., 1999; Cai, 2002), or via a time-varying mixture of some distributions.

The popularity of the parametric models stems mainly from their intuitive appeal and simplicity. However, most conventional parametric specifications and some nonparametric distributions have failed in capturing some rare events that took place in emerging financial markets over the last decade. This inadequacy has led researchers to model directly the tail behaviour of a distribution parametrically rather than the whole distribution. To fill this gap, recent studies on risk modelling have found an interesting avenue in this direction, leading the extreme value theory (EVT) distributions to become popular. The most commonly used EVT distributions in the literature include the generalized extreme value (GEV) distribution of von Mises (1936) and Jenkinson (1955), the generalized Pareto (GP) distribution of Balkema and de Haan (1974) and Pickands (1975), and the Hill

Table I. VaR models and mnemonics

	Unfiltered	Filtered
Normal distribution		Normal*
Historical distribution	HS	HS*
Monte Carlo distribution	MC	MC*
NP distribution	NP	NP*
EVT distributions*	GP	GP*
	GEV	GEV*
	HILL	HILL*
No distribution	CaViaR$_S$	
	CaViaR$_A$	

Notes:

(1) This table defines mnemonics for the models used in this paper.

(2) The filtered VaR models (using z) are computed using the standardized returns $z_t = r_t/\sigma_t$ where σ_t^2 is estimated by a GARCH(1, 1) model. The filtered models are denoted with *.

(3) The acronyms stand for the following methods: HS = historical simulation; MC = Monte Carlo; NP = nonparametrically estimated distribution of Hall *et al*. (1999) and Cai (2002); GP = generalized Pareto distribution; GEV = generalized extreme value distribution; Hill = method based on Hill (1975); CaViaR$_S$ = symmetric CaViaR model of Engle and Manganelli (2004); CaViaR$_A$ = asymmetric CaViaR model of Engle and Manganelli (2004).

(4) CaViaR models do not need a distribution as they are not based on inversion of the distribution to estimate quantiles. We do not consider the filtered version of CaViaR models for z_t because $\{z_t\}$ is nearly IID and its quantiles may not exhibit any dependence.

(5) In addition to the above models, we also include the popular Riskmetrics model in Tables III and IV as a benchmark. The Riskmetrics model is similar to Normal* as it uses the normal distribution, but different from Normal* as its σ_t^2 is updated by EWMA instead of σ_t^2 being estimated by GARCH(1, 1) as we do for Normal*.

(1975) estimator. While EVT was developed under the IID assumption on the series in question, the theory has been extended to serially dependent observations provided that the dependence is weak. See Berman (1964) and Leadbetter *et al*. (1983). Hence EVT distributions could be directly applicable to the return series which has long as well as short memory. Recently, serious research has been conducted in this field. See Longin (1996), Danielsson and de Vries (1997), Pownall and Koedijk (1999) and Neftçi (2000). Certain problems with the EVT methodology have also been documented in Diebold *et al*. (2000).

Another question is how to model σ_t^2. The conditional variance σ_t^2 can be estimated with various volatility models. We can estimate it nonparametrically (e.g., Bühlman and McNeil, 2001), parametrically (e.g., Engle, 1982; Bollerslev, 1986; Taylor, 1986). See Poon and Granger (2003) for an excellent survey and references therein. In light of the fact that VaR is essentially a quantile of some distribution, we do not include various volatility models in this paper and focus instead on the distribution $G_t(\cdot)$ or $F_t(\cdot)$. González-Rivera *et al*. (2004) and Hansen and Lunde (2004) found that in terms of out-of-sample predictive ability some simple volatility models often perform as well as more complex models, while their relative performance varies with users' evaluation criteria. Therefore, we consider only a simple GARCH(1,1) model for σ_t^2.

Table II. Three out-of-sample evaluation periods

	Period 1 (before crisis)	Period 2 (during crisis)	Period 3 (after crisis)
Out-of-sample period	1/1/1996–12/31/1996 $P = 261$	7/1/1997–6/30/1998 $P = 261$	1/1/1999–12/31/1999 $P = 261$
In-sample period	1/1/1988–12/31/1995 $R = 2086$	1/1/1988–6/30/1997 $R = 2476$	1/1/1988–12/31/1998 $R = 2869$

The essential problem is that we usually do not know the true data generating process (DGP). While the conventional VaR models (assuming normality or its extensions) have been criticized for their inadequacy during the recent Asian financial turmoil, models based on the EVT distributions are claimed to perform better during the crisis period. On the other hand, the nonparametric model of Hall *et al.* (1999) and Cai (2002) requires very weak assumptions and has a generic advantage compared with the EVT or parametric distributions to capture higher-order dependency beyond that specified in some particular parametric models or the simplest IID assumption, and hence avoid potential misspecifications. However, this generality does not necessarily guarantee the superior predictive performance of the nonparametric model, especially in small samples.

The aim of our paper is to compare various VaR models in the following dimensions: (i) Whether to filter or not?[2] (ii) Whether to model the distribution $G_t(\cdot)$ or $F_t(\cdot)$ parametrically or nonparametrically? (iii) Whether to model the whole distribution or the tails only? (iv) Whether to model quantiles directly or to estimate the quantiles via the inverse of $F_t(\cdot)$ or $G_t(\cdot)$ as in (3)? A direct way to estimate the conditional quantile is the CaViaR model of Engle and Manganelli (2004), where an autoregressive quantile model is used to estimate the conditional quantile of unfiltered data.

We apply these different approaches to VaR modelling to the stock markets of five Asian economies (Indonesia, Korea, Malaysia, Taiwan and Thailand), that suffered the 1997–1998 financial crisis. Thus our exercise can be regarded as a 'stress testing' under different market scenarios. See Table II. Various VaR models are compared in terms of the predictive likelihood function for quantile forecasts using the reality check tests of White (2000) and Hansen (2001), and also in terms of the tail interval forecast (empirical tail coverage probability), for the one-step-ahead VaR predictions at $\alpha = 0.01$ and 0.05.

Our results indicate some empirical regularities of risk forecasts. The Riskmetrics model behaves reasonably well in tranquil periods, while some EVT models do better in the crisis period. Filtering often appears to be useful for some models (particularly the EVT models), though it could be harmful for some other models. The CaViaR models have shown some success in predicting the VaR risk measure for various periods, generally comparable to the VaR models that invert a distribution function. Overall, the forecasting performance of the VaR models considered varies for different tails, over the three different periods, and for the five different economies.

---

[2] Following Hull and White (1998) and Barone-Adesi *et al.* (2002), a VaR model in this paper is said to be 'filtered' if it is applied to $z_t = (r_t - \mu_t)/\sigma_t$, the standardized demeaned return series using a time-varying volatility model. A VaR model is said to be 'unfiltered' if it is applied to r_t or if the volatility is assumed to be a constant ($\sigma_t^2 = \sigma^2$). One exception to this terminology is for the Monte Carlo distribution.

The organization of this paper is as follows. In the next section we discuss various VaR models. In the third section forecast evaluation criteria and the predictive ability tests are discussed. The fourth section presents the empirical results, and a final section concludes.

VAR MODELS

As defined in equations (1)–(3), the computation of $q_t(\alpha)$ amounts to computing the quantile of the distribution of $\{z_t\}$ or $\{r_t\}$. Given the specifications of μ_t and σ_t^2, it can be seen from (3) that VaR models will be determined by the choice of $G_t(\cdot)$ or $F_t(\cdot)$. We consider five types of distribution: namely, normal distribution, historically simulated distribution (HS), Monte Carlo simulated distribution (MC), nonparametrically estimated distribution (NP) and EVT-based distribution. We also consider the CaViaR model of Engle and Manganelli (2004), that is to estimate the VaR from a quantile regression rather than from inverting $G_t(\cdot)$ or $F_t(\cdot)$.

The various models described in this section are summarized in Table I. We use an asterisk * to denote a filtered model. For example, NP* denotes that VaR is estimated by applying the nonparametric distribution to the filtered series $\{z_t\}$.

Throughout, for the purpose of comparison, we include the Riskmetrics model of J.P. Morgan (1995) as a benchmark model

$$\hat{q}_t(\alpha) = \hat{\mu}_t + \hat{\sigma}_t \Phi^{-1}(\alpha) \tag{4}$$

where $\Phi(\cdot)$ is the standard normal distribution function so that $\Phi^{-1}(0.01) = -2.326$ and $\Phi^{-1}(0.05) = -1.645$, and $\hat{\sigma}_t$ is given recursively by the exponentially weighted moving average (EWMA)

$$\hat{\sigma}_t^2 = 0.94\hat{\sigma}_{t-1}^2 + 0.06(r_{t-1} - \hat{\mu}_t)^2 \tag{5}$$

with $\hat{\mu}_t = \dfrac{1}{t-1}\sum_{j=1}^{t-1} r_j$.[3]

Normal distribution
We consider the standard normal distribution $\Phi(\cdot)$ for $G_t(\cdot)$ so that $q_t(\alpha) = \mu_t + \sigma_t\Phi^{-1}(\alpha)$ with σ_t^2 estimated by the GARCH(1, 1) model $\sigma_t^2 = a_0 + a_1 r_{t-1}^2 + a_2\sigma_{t-1}^2$ with $\mu_t = 0$. This model is denoted as Normal* in Table I. Normal* and Riskmetrics differ in σ_t^2 and μ_t, with the same $\Phi(\cdot)$ for $G_t(\cdot)$.

Historical distribution
One approach to VaR modelling is to estimate the quantile nonparametrically. A conventional way is to use the historically simulated distribution (HS). The idea behind historical simulation is to assume that the distribution of returns $\{r_t\}$ will remain the same in the past and in the future, and hence the empirical distribution of historical returns will be used in forecasting VaR. See Jorion (2000, p. 221) for more details.

---

[3] We use $\hat{\mu}_t = \dfrac{1}{t-1}\sum_{j=1}^{t-1} r_j$ for the Riskmetrics model and for MC models following the convention for these two models. For the other models in this paper, we set $\mu_t = 0$.

The key assumption of HS is that the series under consideration is IID. For periods of greater turmoil, it can turn out to be a very bad measure of risk since risk can change significantly. Therefore, a more appropriate path to pursue is to put different weights on historical observations of $\{r_t\}$. The ideas of volatility updating (Hull and White, 1998) and filtering (Barone-Adesi *et al.*, 2002) are in the direction of using the filtered series $\{z_t\}$ instead of $\{r_t\}$. The filtered HS model will be denoted as HS* when σ_t^2 is estimated from a GARCH(1, 1) model. See Table I.

Monte Carlo distribution

The underlying stochastic process that governs the dynamics of asset prices may be calibrated for the asset's future values. A popular, simple stochastic process is the geometric Brownian motion given by

$$dS_t/S_t = \mu_t dt + \sigma_t dW_t \tag{6}$$

where S_t is the asset price at time t, W_t is a standard Wiener process, and μ_t and σ_t are the drift and volatility parameters, respectively. The solution to this stochastic differential equation is $S_t = S_0 \exp([\mu_t - \frac{1}{2}\sigma_t^2]t + \sigma_t W_t)$. See Broadie and Glasserman (1998). Thus, simulating S_t amounts to simulating W_t. Since we are predicting one-step-ahead VaR, it can be written as

$$S_t = S_{t-1} \exp\left(\left[\mu_t - \frac{1}{2}\sigma_t^2\right] + \sigma_t z_t\right)$$

where z_t is simulated from a standard normal distribution. We do it N times, from which the empirical αth quantile of $r_t \equiv \log(S_t/S_{t-1})$ is estimated. Later, we set $N = 1000$.

When σ_t^2 is estimated from the unconditional variance $\frac{1}{t-2}\sum_{j=1}^{t-1}(r_j - \hat{\mu}_j)^2$ with $\hat{\mu}_t = \frac{1}{t-1}\sum_{j=1}^{t-1}r_j$, the VaR model will be denoted as MC. When σ_t^2 is estimated by the conditional variance model of GARCH(1, 1) it will be denoted as MC*. See Table I.

Nonparametrically estimated distribution

We also use a nonparametrically estimated conditional distribution following Hall *et al.* (1999) and Cai (2002). If Y and X are stationary, the conditional distribution function of Y given $X = x_t$ can be estimated through the 'weighted' Nadaraya–Watson estimator

$$F(y|x_t) = \frac{\sum_{i=1}^{n} p_i k_h(x_i - x_t)\mathbf{1}(Y_i \leq y)}{\sum_{i=1}^{n} p_i k_h(x_i - x_t)} \tag{7}$$

where $K_h(\cdot)$ is a kernel function with bandwidth parameter h, $\mathbf{1}(\cdot)$ is an indicator function, and the weights $p_i \equiv p_i(x_t)$ are obtained from a constrained maximization problem

$$\max \sum_{i=1}^{n} \log p_i$$

subject to

$$\sum_{i=1}^{n} p_i g_i = 0, \quad \sum_{i=1}^{n} p_i = 1 \quad \text{and} \quad p_i \geq 0$$

where $g_i \equiv (x_i - x_t)K_h(x_i - x_t)$.[4] The weights p_i in (7) can be regarded as the local empirical likelihoods. See Owen (2001). The solution to the above constrained maximization problem is $p_i = n^{-1}(1 + \lambda g_i)^{-1}$, where the Lagrangian multiplier λ is chosen to maximize $L_n(\lambda) = (nh)^{-1}\Sigma_{i=1}^{n}\log(1 + \lambda g_i)$

The quantile function can then be found as

$$\hat{q}_t(\alpha) = \inf\{y \in \mathbb{R}|\hat{F}(y|x_t) \geq \alpha\} \tag{8}$$

where $\hat{F}(y|x_t)$ is estimated using the optimal bandwidth h selected by the standard cross-validation based on the quantile estimation loss function of Koenker and Bassett (1978). See equation (29) later. We report the results using the Gaussian kernel $K_h(u) = (2\pi)^{-1/2}\exp(-(u/h)^2/2)$, while other kernels give similar results.

The NP distribution in (7) is the 'conditional' distribution. Of course, when we have an IID series, the conditional distribution should be the same as the unconditional distribution; otherwise, the NP method may capture some unknown higher-order dependency structure of a non-IID series. Therefore, we employ the NP method to both unfiltered and filtered data, and we denote them as NP and NP*, respectively, as summarized in Table I. Later, we set $(y_t \ x_t) = (r_t \ r_{t-1})$ for NP, and $(y_t \ x_t) = (z_t \ z_{t-1})$ for NP*.[5]

Extreme value distributions

All the previous methods estimate the quantiles using information from the whole distribution. Alternatively, since the quantiles at 1% or 5% are 'extreme' values for a distribution, we can focus on modelling the tails directly. That brings us to the extreme value theory. Embrechts *et al.* (1997) provided a treatise on various aspects of EVT; Longin (1996, 2000) used the generalized extreme value distribution to estimate the tail index; McNeil and Frey (2000) and Neftçi (2000) used the generalized Pareto distribution. In the following discussion, we assume that the series in question, $\{y_t\}_{t=1}^{n}$, is IID. Nevertheless, as mentioned before, we can apply the EVT models to weakly dependent series (unfiltered data) too.

Generalized extreme value distribution

For a series $\{y_t\}$, consider the ordered series $\{y_{(t)}\}$ in increasing order $y_{(t)} \leq y_{(t+1)}$ for all t. The sample minimum is $y_{(1)}$ over an n-sample period. If $\{y_t\}$ is IID with the CDF $F_Y(y)$, then the CDF of the minimum, denoted by $G_Y(y)$, is given by

---

[4] The first constraint $\Sigma_{i=1}^{n}p_i(x_i - x_t)K_h(x_i - x_t) = 0$ is the so-called 'discrete moment condition', which is not satisfied by the Nadaraya–Watson estimator. This equation makes an extra term in the asymptotic bias of the Nadaraya–Watson estimator in comparison to the local linear estimator. That is why this constraint is imposed. See Fan and Gijbels (1996, p. 63) and Cai (2002, p. 172). We thank Zongwu Cai for pointing this out to us.

[5] This choice is not in favour of the NP* method because $\{z_t\}$ is likely to exhibit the near-IID property. Nevertheless, we choose $x_t = z_{t-1}$ and consider only the univariate case in this paper. We expect the NP and NP* models would perform better when x_t is chosen from some relevant variables that explain the returns. For example, Fama and French (1993) and Lakonishok *et al.* (1994) identify factors like firm size, book to market ratio and earnings–growth extrapolation for the variable x_t. We do not consider these factors in the present study.

$$G_Y(y) = \Pr(y_{(1)} \leq y) = 1 - \Pr(y_{(1)} > y) = 1 - \prod_{t=1}^{n} \Pr(y_t > y)$$

$$= 1 - \prod_{t=1}^{n}[1 - \Pr(y_t \leq y)] = 1 - [1 - F_Y(y)]^n \tag{9}$$

Since $G_Y(y)$ degenerates as $n \to \infty$, we seek a limit law $H_X(x)$ with which a normalization $x_n = (y_{(1)} - \beta_n)/\delta_n$ does not degenerate as $n \to \infty$ for some suitable normalizing constants β_n and $\delta_n > 0$. The limiting distribution of x_n is the generalized extreme value distribution of von Mises (1936) and Jenkinson (1955) of the form

$$H_X(x) = 1 - \exp\left(-(1 + \tau x)^{\frac{1}{\tau}}\right) \tag{10}$$

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

$$h_X(x) = (1 + \tau x)^{\frac{1}{\tau}-1} \exp\left(-(1 + \tau x)^{\frac{1}{\tau}}\right) \tag{11}$$

so that the approximate density of $y_{(1)}$ for given n, by change of variables, is

$$h_X(x_n) = \frac{1}{\delta_n}(1 + \tau x_n)^{\frac{1}{\tau}-1} \exp\left(-(1 + \tau x_n)^{\frac{1}{\tau}}\right) \tag{12}$$

Hence the three parameters $\theta_n = (\tau\ \beta_n\ \delta_n)'$ may be estimated by maximum likelihood. To implement it, Longin (1996, 2000) partitioned the sample into g non-overlapping subsamples each with m observations. In other words, if $n = gm$, the ith subsample of the series is $\{y_{(i-1)m+j}\}_{j=1}^{m}$ for $i = 1, \ldots, g$. If $n < gm$, we drop some observations in the first subsample so that it has fewer than m observations. The collection of subperiod minima is then $\{y_{m,i}\}$, where $y_{m,i} = \min_{1 \leq j \leq m}\{y_{(i-1)m+j}\}$, $i = 1, \ldots, g$. The likelihood function of $\{y_{m,i}\}$ is

$$\prod_{i=1}^{g} h_Y(x_m) = \prod_{i=1}^{g} h_Y\left(\frac{y_{m,i} - \beta_m^i}{\delta_m^i}\right) \tag{13}$$

Assuming $\theta_m^i = \theta_m$ for all subperiods, $i = 1, \ldots, g$, we can estimate θ_m from a numerical optimization of the (log) likelihood.

Next, consider the probability that the subperiod minimum $y_{m,i}$ is less than y_m^* under the limit law (10). Denoting $x_m^* = \frac{y_m^* - \beta_m}{\delta_m}$, we have

$$H_X(x_m^*) = H_X\left(\frac{y_m^* - \beta_m}{\delta_m}\right) = \Pr\left(\frac{y_{m,i} - \beta_m}{\delta_m} \leq \frac{y_m^* - \beta_m}{\delta_m}\right) = \Pr(y_{m,i} \leq y_m^*) \tag{14}$$

which is therefore equal to

$$G_Y(y_m^*) = 1 - \left[1 - F_Y(y_m^*)\right]^m = 1 - (1-\alpha)^m \tag{15}$$

where the second equality holds if $y_m^* = q(\alpha)$. Hence, equating (14) and (15), we get

$$H_X(x_m^*) = 1 - \exp\left(-(1 + \tau y_m^*)^{\frac{1}{\tau}}\right) = 1 - (1-\alpha)^m \tag{16}$$

which yields

$$q(\alpha) = \beta_m - \frac{\delta_m}{\tau}\left\{1 - \left[-\ln(1-\alpha)^m\right]^\tau\right\} \tag{17}$$

Later, we use $m = 10$. We employ the GEV distribution to both unfiltered ($y_t = r_t$) and filtered data ($y_t = z_t$).

Generalized Pareto distribution
An alternative EVT approach is based on *exceedances over threshold* (Smith, 1989; Davison and Smith, 1990). According to this approach, we fix some low threshold u and look at all exceedances e over u. To be consistent with the other part of this paper, we discuss here the exceedance distribution for the left tail. The distribution of excess values is given by

$$\Pr(Y > u - e | Y < u) = \frac{F(u) - F(u - e)}{F(u)}, \quad e > 0 \tag{18}$$

Balkema and de Haan (1974) and Pickands (1975) showed that the asymptotic form of $\Pr(Y > u - e | Y < u)$ is

$$H(e) = 1 - \left(1 - \frac{\tau e}{\delta}\right)^{1/\tau} \tag{19}$$

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

$$h(e) = \frac{1}{\delta}\left(1 - \frac{\tau e}{\delta}\right)^{1/\tau - 1}$$

For $\{y_t\}_{t=1}^n$, we can estimate $(\tau, \delta)'$ by maximizing $\Pi_{i=1}^m h(e_i)$, where $\{e_i\}_{i=1}^m$ is the sample of exceedances over the threshold u. Denote the MLE of $(\tau, \delta)'$ by $(\hat{\tau}, \hat{\delta})'$. Then from (18) and (19),

$$\frac{F(u - e)}{F(u)} = \left(1 - \frac{\tau e}{\delta}\right)^{1/\tau}$$

which gives, if we estimate $F(u)$ by m/n,

$$\hat{F}(u-e) = \frac{m}{n}\left(1 - \frac{\hat{\tau}e}{\hat{\delta}}\right)^{1/\hat{\tau}}$$

or equivalently for $y < u$,

$$\hat{F}(y) = \frac{m}{n}\left(1 - \frac{\hat{\tau}(u-y)}{\hat{\delta}}\right)^{1/\hat{\tau}} \tag{20}$$

Immediately, the αth quantile can be estimated by setting $\hat{F}(y) = \alpha$ and hence

$$\hat{q}(\alpha) = u - \frac{\hat{\delta}}{\hat{\tau}}\left(1 - \left(\frac{na}{m}\right)^{\hat{\tau}}\right) \tag{21}$$

An important issue in implementing the GP approach is how to choose the threshold u. For example, if we are interested in the 5% quantile, then the chosen u must be larger (enough) than $q(0.05)$. We follow Neftçi (2000) and use the empirical 10% quantile. One possible extension is that we may estimate the threshold value u to decide which extremes are really extremes, see Gonzalo and Olmo (2004).

Hill estimator
Denote the ordered series as $\{y_{(t)}\}_{t=1}^{n}$ in increasing order. Suppose $y_{(m)} < 0$ and $y_{(m+1)} > 0$ so that m is the number of negative observations in the sample. The GEV distribution (10) with $\tau < 0$ is known as the Fréchet distribution with the CDF $F(y) = \exp(-|y|^{\frac{1}{\tau}})$, $y < 0$. As shown in Embrechts *et al*. (1997, p. 325), it reduces to

$$F(y) = 1 - C|y|^{\frac{1}{\tau}}, \quad |y| \geq u \geq 0 \tag{22}$$

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 showed that its maximum likelihood estimator is

$$\hat{\tau} = -\frac{1}{k}\sum_{t=1}^{k}(\ln|y_{(m-k-t)}| - \ln|y_{(m-k)}|) \tag{23}$$

where $k \equiv k(m) \rightarrow \infty$ and $k(m)/m \rightarrow 0$. It is known that $\hat{\tau} \xrightarrow{p} \tau$ as $m \rightarrow \infty$ (Mason, 1982). We can 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

$$\hat{q}(\alpha) = \left[\frac{m}{k}(1-\alpha)\right]^{\hat{\tau}} y_{(k+1)} \tag{24}$$

See Embrechts *et al*. (1997, p. 347).

Conditional autoregressive VaR
Engle and Manganelli (2004) suggested that VaR can be estimated by modelling the quantiles directly rather than inverting a distribution. The idea is similar to the GARCH modelling and VaR is modelled autoregressively as

$$q_t(\alpha) = a_0 + a_1 q_{t-1}(\alpha) + h(x_t|\theta) \qquad (25)$$

where $x_t \in \mathcal{F}_{t-1}$, θ is a parameter vector and $h(\cdot)$ is a function to explain the VaR model. This model is called the CaViaR model. In our study two specifications of the CaViaR model are chosen: Symmetric CaViaR (CaViaR$_S$)

$$q_t(\alpha) = a_0 + a_1 q_{t-1}(\alpha) + a_2 |r_{t-1}| \qquad (26)$$

Asymmetric CaViaR (CaViaR$_A$)

$$q_t(\alpha) = a_0 + a_1 q_{t-1}(\alpha) + a_2 |r_{t-1}| + a_3 |r_{t-1}| \cdot \mathbf{1}(r_{t-1} < 0) \qquad (27)$$

The estimation can be made via quantile regression. Due to the nondifferentiable absolute function the estimation can be achieved by a genetic algorithm. See Price and Storn (1997). Following Engle and Manganelli (2004), we apply the CaViaR model directly to the return series $\{r_t\}$ (not to $\{z_t\}$).

COMPARING VAR MODELS

We compare VaR models through their performance in terms of out-of-sample one-step-ahead predictive ability. Suppose we have a sample of total T observations and we split it into an in-sample part of size R and an out-of-sample part of size P so that $T = P + R$. We use a rolling window scheme. That is, the $(t - R)$th prediction is based on observations $t - R$ through $t - 1$, $t = R + 1, \ldots, T$. Let a benchmark model be indexed by $k = 0$ and the l competing models by $k = 1, \ldots, l$. Let $q_{k,t}(\alpha, \beta_k)$ be the VaR forecast using Model k ($k = 0, \ldots, l$), for which a loss function $L(q_{k,t}(\alpha, \beta_k))$ will be defined. Then the loss differential between Model 0 (benchmark) and Model k is

$$\hat{f}_{k,t} \equiv f(\hat{q}_{0,t}(\alpha), \hat{q}_{k,t}(\alpha); \hat{\beta}_{t-1}) = L(\hat{q}_{0,t}(\alpha; \hat{\beta}_{0,t-1})) - L(\hat{q}_{k,t}(\alpha; \hat{\beta}_{k,t-1})), \quad k = 1, \ldots, l$$

where $\hat{\beta}_{t-1}$ collects the estimated parameters using the information up to time $t - 1$ from both models. When the pseudo-true parameter vector $\beta^\dagger$ (the probability limit of the estimator $\hat{\beta}_{t-1}$) and the corresponding pseudo-true quantiles are used, we may define analogously $f_{k,t}^\dagger$. Stacking $\hat{f}_{k,t}$ and $f_{k,t}^\dagger$ for $k = 1, \ldots, l$ gives the $l \times 1$ vectors $\hat{f}_t$ and $f_t^\dagger$. Testing for the unconditional predictive ability hypothesis can be conducted in terms of $E(f_t^\dagger)$ (or $E(f^\dagger)$ assuming stationarity) as in West (1996), while the conditional predictive ability testing in terms of $E(f_t^\dagger|\mathcal{F}_{t-1})$ as in Giacomini and White (2003).

Evaluation criteria

We evaluate and compare various VaR forecast models in terms of the predictive quantile loss and the empirical coverage probability.

We use the 'check' function of Koenker and Bassett (1978). As we conduct an out-of-sample analysis to compare predictive ability of the various VaR models, the check function may be regarded as a 'predictive' quasi-likelihood, as discussed in Bertail *et al.* (2004) and Komunjer (2004). In what follows, the index k may be suppressed to simplify notation. Hence, the expected loss of $q_t(\alpha)$ for a given α is:

$$Q(\alpha) = E[\alpha - \mathbf{1}(r_t < q_t(\alpha))][r_t - q_t(\alpha)] \tag{28}$$

The loss $Q(\alpha)$ can provide a measure of the lack-of-fit of a quantile model. The expected check function $Q(\alpha)$ can be evaluated from the out-of-sample VaR forecasts

$$\hat{Q}_P(\alpha) = P^{-1} \sum_{t=R+1}^{T} [\alpha - \mathbf{1}(r_t < \hat{q}_t(\alpha))][r_t - \hat{q}_t(\alpha)] \tag{29}$$

where $\hat{q}_t(\alpha) = \hat{F}_t^{-1}(\alpha) = \hat{\mu}_t + \hat{\sigma}_t \hat{G}_t^{-1}(\alpha)$ with $\hat{F}_t^{-1}(\cdot)$, $\hat{\mu}_t$, $\hat{\sigma}_t$, $\hat{G}_t^{-1}(\cdot)$ estimated using the information $\mathcal{F}_{t-1}$. A model that gives the VaR forecast $\hat{q}_t(\alpha)$ with the minimum value of $\hat{Q}_P(\alpha)$ is the preferred model.

If $F(y_t)$ is continuous in a neighbourhood of $q_t(\alpha)$, $q_t(\alpha)$ minimizes $Q(\alpha)$ and makes a condition for the correct conditional coverage probability

$$\alpha = E[\mathbf{1}(r_t < q_t(\alpha))|\mathcal{F}_{t-1}] \tag{30}$$

i.e., $\{\alpha - \mathbf{1}(r_t < q_t(\alpha))\}$ is a martingale difference sequence (MDS).[6] Given the nominal conditional coverage probability $\alpha = E[\mathbf{1}(r_t < q_t(\alpha))|\mathcal{F}_{t-1}]$, the empirical conditional coverage probability constructed for the VaR forecasts $\{\hat{q}_t(\alpha)\}_{t=R+1}^{T}$ can be computed from

$$\hat{\alpha}_P = \frac{1}{P} \sum_{t=R+1}^{T} \mathbf{1}(r_t < \hat{q}_t(\alpha)) \tag{31}$$

A model that gives the VaR forecast $\hat{q}_t(\alpha)$ with $\hat{\alpha}_P$ closest to its nominal value α is the preferred model.

Reality check

When several models using the same data are compared in terms of 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 specification search into account can be extremely misleading (see Lo and MacKinlay, 1999, chapter 8). White (2000) developed a noble test of superior unconditional predictive ability among multiple models accounting for specification search, built on Diebold and Mariano (1995) and West (1996).

Our interest is to compare all the models with a benchmark. An appropriate null hypothesis is that all the models are no better than a benchmark, i.e., $H_0 : \max_{1 \le k \le l} E(f_k^\dagger) \le 0$. This is a multiple hypothesis, the intersection of the one-sided individual hypotheses $E(f_k^\dagger) \le 0$, $k = 1, \ldots, l$. The alternative is that H_0 is false, that is, the best model is superior to the benchmark. If the null hypothesis is rejected, there must be at least one model for which $E(f_k^\dagger)$ is positive. Suppose that $\sqrt{P}(\bar{f} - E(\mathbf{f}^\dagger)) \xrightarrow{d} N(0, \Omega)$ as $P(T) \to \infty$ when $T \to \infty$, for Ω positive semi-definite. White's (2000) test statistic for H_0 is formed as $\bar{V} \equiv \max_{1 \le k \le l} \sqrt{P} \bar{f}_k$, where $\bar{f}_k = P^{-1/2} \Sigma_{t=R+1}^{T} \hat{f}_{k,t}$. However, as the null limiting distribution of $\bar{V}$ is unknown, White (2000, theorem 2.3) showed that the distribution of $\sqrt{P}(\bar{\mathbf{f}}^* - \bar{\mathbf{f}})$ converges to that of $\sqrt{P}(\bar{\mathbf{f}} - E(\mathbf{f}^\dagger))$, where $\bar{\mathbf{f}}^*$ is obtained from the stationary bootstrap of Politis and Romano

---

[6] See, e.g., Giacomini and Komunjer (2002). The MDS property has been used in forming a conditional moment test for the quantile models. See, e.g., Zheng (1998), Engle and Manganelli (2004), Bierens and Ginther (2001).

(1994). By the continuous mapping theorem this result extends to the maximal element of the vector $\sqrt{P}(\bar{\mathbf{f}}^* - \bar{\mathbf{f}})$, so that the empirical distribution of

$$\bar{V}^* = \max_{1 \le k \le l} \sqrt{P}(\bar{f}_k^* - \bar{f}_k) \tag{32}$$

may be used to compute the *p*-value of $\bar{V}$ (White, 2000, corollary 2.4). This *p*-value is called the 'reality check *p*-value'.

We note that White's reality check is conservative when a poor model is included in the set of l competing models. The inclusion of $\bar{\mathbf{f}}_k$ in (32) guarantees that the statistic satisfies the null hypothesis $E(\bar{\mathbf{f}}_k^* - \bar{\mathbf{f}}_k) = 0$ for all *k*. This setting makes the null hypothesis the least favourable to the alternative and consequently, it renders a very conservative test. When a poor model is introduced, the reality check *p*-value for Model *k* becomes very large and, depending on the variance of $\bar{\mathbf{f}}_k$, it may remain large even after the inclusion of better models. Hansen (2001) considered the following modification to (32):

$$\bar{V}^* = \max_{1 \le k \le l} \sqrt{P}(\bar{f}_k^* - g(\bar{f}_k)) \tag{33}$$

where different $g(\cdot)$ functions will produce different bootstrap distributions that are compatible with the null hypothesis. Hansen (2001) recommended setting $g(\cdot)$ as a function of the variance of $\bar{\mathbf{f}}_k$, i.e.

$$g(\bar{f}_k) = \begin{cases} 0 & \text{if } \bar{f}_k \le -A_k \\ \bar{f}_k & \text{if } \bar{f}_k > -A_k \end{cases} \tag{34}$$

where $A_k = \frac{1}{4} P^{-1/4} \sqrt{var(P^{1/2} \bar{f}_k)}$ with the variance estimated from the bootstrap samples. In our empirical section, we report two reality check *p*-values: with $g(\bar{f}_k) = \bar{f}_k$ as in (32) (denoted as *White*) and with $g(\bar{f}_k)$ determined from (34) (denoted as *Hansen*). When $E(f_k^\dagger) = 0$ for all $1 \le k \le l$, then the reality check *p*-value of White (2000) will provide an asymptotically correct size. However, when some models are dominated by the benchmark model, i.e., $E(f_k^\dagger) < 0$ for some $1 \le k \le l$, then the reality check *p*-value of White (2000) will make a conservative test. So, when bad models are included in the set of competing models, White's test tends to behave conservatively. Hansen's (2001) modification is basically to remove those (very) bad models in the comparison. Hansen (2001) contains two modifications. One is what we employ here. The other is to take a maximum over the standardized statistics. There might be additional gains in power from the standardization, as noted in Hansen (2001, 2003).

Remarks: White's theorem 2.3 is obtained under the assumption of differentiability of the loss function (as in West, 1996, assumption 1). Also, White's theorem 2.3 is obtained under the assumption that either (a) the same loss function is used for estimation and prediction, or (b) (*P/R*) log log*R* $\to$ 0 as $T \to \infty$ so that the effect of parameter estimation vanishes asymptotically (as in West, 1996, theorem 4.1(a)). Thus White's theorem 2.3 does not immediately apply to the nonsmooth functions and the presence of estimated parameters. Nevertheless, it is noted in White (2000, p. 1113) that the results analogous to theorem 2.3 can be established under similar conditions used in deriving the asymptotic normality of the least absolute deviations estimator. When no parameter estimation is involved, White's (2000) procedure is applicable to nondifferentiable *f*. We expect that the approach

of Randles (1982) and McCracken (2000, assumption 4) may be useful here, where the condition $E(\partial \mathbf{f}^\dagger/\partial\beta) = 0$ is modified to $\partial(E\mathbf{f}^\dagger)/\partial\beta = 0$ to exploit the fact that the expected loss function may be differentiable even when the loss function is not. We conjecture that when parameter estimation is involved, White's (2000) procedure continues to hold either when $\partial(E\mathbf{f}^\dagger)/\partial\beta = 0$ or when P grows at a suitably slower rate than R. Since we are using different criteria for in-sample estimation and forecast evaluation, there is no reason to expect that $\partial(E\mathbf{f}^\dagger)/\partial\beta = 0$. Hence it is important to have very large R compared to P. In our empirical section, we thus have $R = 2086$, 2476 or 2869, which are much larger than $P = 261$.

EMPIRICAL RESULTS

From Datastream, we retrieve the Indonesia Jakarta Stock Exchange Composite Price Index, Korea Stock Exchange Composite Price Index, Malaysia Kuala Lumpur Stock Exchange Composite, Taiwan Weight Index and Thailand S.E.T. Price Index. The return series is given by the log difference of price indices, then multiplied by 100. To investigate the performance of VaR models under different circumstances, we use three out-of-sample evaluation periods, which we denote as the before-crisis period, the crisis period and the after-crisis period, respectively. Thus our exercise can be regarded as a 'stress testing' under different scenarios. The three periods are summarized in Table II.

For all three periods the estimation sample (in-sample) starts from January 1, 1988. For each period, we split the whole sample into an in-sample period and an out-of-sample period for one year ($P = 261$). The first period ends at December 31, 1996 with total $T(= R + P) = 2347$ observations; the second period, which covers the 1997–1998 Asian financial crisis, ends at June 30, 1998 with total $T = 2737$ observations; the third period, after the crisis, ends at December 31, 1999 with total $T = 3130$ observations.

Tail coverage probability

In Tables IIIA–C, we present the empirical coverage probability $\hat{\alpha}_P$ (the relative frequency of the violations) for Periods 1, 2 and 3, respectively.[7] We investigate the three periods in order to compare the potential changes in the risk forecast precision across the three different periods.

As shown in Table IIIA, for the pre-crisis period (Period 1) with $\alpha = 0.05$, the conventional Riskmetrics model has a satisfactory forecast. For most countries the predicted coverage probabilities are very close to the nominal coverage. Both the symmetric and asymmetric CaViaR models do well. Unfiltered EVT models behave rather poorly for this tail probability level $\alpha = 0.05$. Most of the EVT models perform rather poorly. The result marginally improves when filtering is applied, but is still far from being satisfactory. The same finding is also applicable for HS and NP, even though these nonparametric models have better forecasting performance than that of the EVT-based models. As in the case of EVT models, filtering also improves the prediction quality of the HS, MC and NP models. Both symmetric and asymmetric CaViaR models are quite satisfactory.

During the pre-crisis period with $\alpha = 0.01$, the risk forecasts of Riskmetrics are fine but there are many other alternative models which do better. For instance, filtered nonparametric (both HS* and NP*), filtered Monte Carlo (MC*) and filtered GP (GP*) models provide good empirical coverage

---

[7] When $1\,(y_t < \hat{q}_t(\alpha)) = 1$, it will be said that a violation occurs.

Table IIIA. Empirical coverage probability, Period 1

		Indonesia	Korea	Malaysia	Taiwan	Thailand
5%	Riskmetrics	0.0460	0.0651	0.0421	0.0307	0.0728
	Normal*	0.0383	0.0575	0.0192	0.0192	0.0498
	HS	0.0383	0.0460	0.0153	0.0115	0.0383
	HS*	0.0498	0.0766	0.0268	0.0192	0.0536
	MC	0.0077	0.0192	0.0153	0.0115	0.0192
	MC*	0.0383	0.0307	0.0192	0.0192	0.0421
	NP	0.0383	0.0498	0.0153	0.0115	0.0613
	NP*	0.0498	0.0766	0.0268	0.0230	0.0536
	GP	0.0920	0.0766	0.0268	0.0192	0.0805
	GP*	0.0805	0.0766	0.0345	0.0192	0.0690
	GEV	0.0000	0.0000	0.0000	0.0000	0.0038
	GEV*	0.0077	0.0077	0.0038	0.0077	0.0077
	HILL	0.0383	0.0115	0.0153	0.0038	0.0077
	HILL*	0.0421	0.0230	0.0153	0.0153	0.0153
	CaViaR$_S$	0.0536	0.0421	0.0307	0.0192	0.0498
	CaViaR$_A$	0.0766	0.0421	0.0307	0.0192	0.0498
1%	Riskmetrics	0.0268	0.0115	0.0230	0.0153	0.0153
	Normal*	0.0230	0.0115	0.0153	0.0153	0.0153
	HS	0.0077	0.0000	0.0000	0.0038	0.0077
	HS*	0.0230	0.0077	0.0115	0.0115	0.0115
	MC	0.0038	0.0038	0.0153	0.0038	0.0077
	MC*	0.0230	0.0115	0.0153	0.0153	0.0153
	NP	0.0077	0.0000	0.0000	0.0038	0.0077
	NP*	0.0230	0.0077	0.0115	0.0115	0.0115
	GP	0.0077	0.0038	0.0038	0.0038	0.0077
	GP*	0.0230	0.0115	0.0115	0.0115	0.0115
	GEV	0.0000	0.0000	0.0000	0.0000	0.0000
	GEV*	0.0000	0.0000	0.0000	0.0038	0.0000
	HILL	0.0000	0.0000	0.0000	0.0000	0.0000
	HILL*	0.0077	0.0077	0.0038	0.0077	0.0077
	CaViaR$_S$	0.0077	0.0077	0.0115	0.0115	0.0077
	CaViaR$_A$	0.0192	0.0077	0.0115	0.0115	0.0115

Note: The number in each cell refers to the frequency at which the actual return falls short of the VaR forecast in the out-of-sample period (1/1/1996–12/31/1996).

probabilities. Unfiltered methods, on the other hand, mostly overstate the VaR forecasts (i.e., understate the coverage probability). Both CaViaR models are quite satisfactory. Overall, no model appears to be particularly superior to Riskmetrics, which works reasonably well for both $\alpha = 0.05$ and 0.01 in Period 1.

In Table IIIB, we observe different findings for the crisis period (Period 2). Most models fail to generate correct coverage probabilities. Unlike the findings obtained in Period 1, most models understate the VaR forecasts (i.e., overstate the coverage probability). The Riskmetrics, MC, HS and other conventional models here produce rather poor coverage for $\alpha = 0.05$. For this quantile level and period, a consistent risk forecasting across the five countries is not a trivial task. Among all the models, the filtered Hill is relatively better but not satisfactory. CaViaR's coverage performance here is also less satisfactory.

Table IIIB. Empirical coverage probability, Period 2

		Indonesia	Korea	Malaysia	Taiwan	Thailand
5%	Riskmetrics	0.0920	0.1111	0.0805	0.0843	0.0690
	Normal*	0.0766	0.1073	0.0958	0.0651	0.0881
	HS	0.2337	0.2644	0.2682	0.0192	0.1686
	HS*	0.1149	0.1111	0.1379	0.0613	0.0881
	MC	0.1379	0.2222	0.2490	0.0192	0.1494
	MC*	0.0766	0.1034	0.0958	0.0613	0.0766
	NP	0.2337	0.2644	0.2682	0.0192	0.1686
	NP*	0.2759	0.2989	0.3142	0.1456	0.2989
	GP	0.2950	0.2720	0.3142	0.0421	0.2261
	GP*	0.1609	0.1226	0.1533	0.0690	0.0958
	GEV	0.0421	0.0805	0.0345	0.0000	0.0077
	GEV*	0.0077	0.0153	0.0153	0.0077	0.0038
	HILL	0.1648	0.1571	0.1686	0.0038	0.0230
	HILL*	0.1034	0.0498	0.0651	0.0192	0.0383
	CaViaR$_S$	0.1418	0.0920	0.1303	0.0613	0.0996
	CaViaR$_A$	0.1992	0.0843	0.1264	0.0613	0.1034
1%	Riskmetrics	0.0421	0.0268	0.0230	0.0230	0.0115
	Normal*	0.0383	0.0421	0.0460	0.0192	0.0307
	HS	0.0843	0.1034	0.0843	0.0038	0.0192
	HS*	0.0383	0.0307	0.0153	0.0115	0.0077
	MC	0.0805	0.1456	0.1303	0.0115	0.0498
	MC*	0.0421	0.0383	0.0498	0.0192	0.0192
	NP	0.0843	0.1034	0.0881	0.0038	0.0192
	NP*	0.1648	0.2261	0.1686	0.0421	0.1571
	GP	0.1073	0.1379	0.1149	0.0077	0.0307
	GP*	0.0307	0.0307	0.0192	0.0115	0.0077
	GEV	0.0077	0.0153	0.0038	0.0000	0.0038
	GEV*	0.0077	0.0038	0.0000	0.0000	0.0000
	HILL	0.0421	0.0805	0.0345	0.0000	0.0077
	HILL*	0.0077	0.0153	0.0153	0.0077	0.0038
	CaViaR$_S$	0.0690	0.0383	0.0230	0.0192	0.0077
	CaViaR$_A$	0.0575	0.0307	0.0153	0.0192	0.0115

Note: The number in each cell refers to the frequency at which the actual return falls short of the VaR forecast in the out-of-sample period (7/1/1997–6/30/1998).

For $\alpha = 0.01$, some EVT-based methods, filtered GP (GP*) and filtered Hill (HILL*), perform better than the other filtered VaR methods and Riskmetrics. The filtered NP (NP*) method produces very poor coverage, much worse than the unfiltered NP. HS* is relatively fine too. During Period 2 and for $\alpha = 0.01$, Taiwan appears to produce the best risk forecast precision among the five economies, perhaps because Taiwan suffered least from the crisis. The CaViaR models show an improved coverage performance for $\alpha = 0.01$ compared to their performance for $\alpha = 0.05$ (although not satisfactory either).

Table IIIC shows that the post-crisis period (Period 3) is somewhat similar to the pre-crisis period (Period 1). For $\alpha = 0.05$ filtering helps HS and NP. HS* and NP* produce better forecasting accuracy than the filtered EVT-based models. Both CaViaR specifications do well, as in the case of the pre-crisis period. For the $\alpha = 0.01$ probability level, as in the case of Period 1, many models produce good coverage probability.

Table IIIC. Empirical coverage probability, Period 3

		Indonesia	Korea	Malaysia	Taiwan	Thailand
5%	Riskmetrics	0.0268	0.0613	0.0421	0.0498	0.0230
	Normal*	0.0268	0.0690	0.0460	0.0421	0.0268
	HS	0.1073	0.1303	0.0460	0.0230	0.0498
	HS*	0.0421	0.0766	0.0575	0.0421	0.0268
	MC	0.0383	0.1111	0.0345	0.0230	0.0460
	MC*	0.0268	0.0575	0.0421	0.0421	0.0230
	NP	0.1111	0.1303	0.0460	0.0307	0.0958
	NP*	0.0421	0.0766	0.0575	0.0421	0.0383
	GP	0.1609	0.1648	0.1111	0.0421	0.0958
	GP*	0.0843	0.0805	0.0728	0.0460	0.0421
	GEV	0.0000	0.0077	0.0038	0.0000	0.0000
	GEV*	0.0077	0.0038	0.0038	0.0038	0.0000
	HILL	0.0575	0.0230	0.0345	0.0038	0.0077
	HILL*	0.0345	0.0192	0.0268	0.0153	0.0077
	CaViaR$_S$	0.0728	0.0421	0.0575	0.0460	0.0307
	CaViaR$_A$	0.0843	0.0536	0.0651	0.0460	0.0307
1%	Riskmetrics	0.0038	0.0038	0.0115	0.0153	0.0000
	Normal*	0.0115	0.0115	0.0192	0.0230	0.0000
	HS	0.0077	0.0153	0.0115	0.0038	0.0000
	HS*	0.0115	0.0038	0.0115	0.0077	0.0000
	MC	0.0115	0.0230	0.0153	0.0038	0.0077
	MC*	0.0115	0.0115	0.0192	0.0192	0.0000
	NP	0.0077	0.0192	0.0115	0.0038	0.0000
	NP*	0.0115	0.0038	0.0115	0.0038	0.0000
	GP	0.0268	0.0230	0.0192	0.0038	0.0077
	GP*	0.0077	0.0115	0.0115	0.0115	0.0000
	GEV	0.0000	0.0000	0.0000	0.0000	0.0000
	GEV*	0.0000	0.0000	0.0000	0.0000	0.0000
	HILL	0.0000	0.0077	0.0038	0.0000	0.0000
	HILL*	0.0077	0.0038	0.0038	0.0038	0.0000
	CaViaR$_S$	0.0115	0.0000	0.0115	0.0115	0.0000
	CaViaR$_A$	0.0153	0.0038	0.0077	0.0153	0.0000

Note: The number in each cell refers to the frequency at which the actual return falls short of the VaR forecast in the out-of-sample period (1/1/1999–12/31/1999).

Overall, filtering is often useful, sometimes marginally, but sometimes it can make matters worse. The Riskmetrics model is not much worse than the best of the other competing models in all three periods including Period 2. Most of the models fail for Period 2, while HILL* works well for Period 2, particularly for $\alpha = 0.01$. The CaViaR models work well in Periods 1 and 3.

Quantile loss

In Tables IVA–C, for each of Periods 1, 2 and 3, we present the out-of-sample average quantile loss values $\hat{Q}_P(0.05)$ and $\hat{Q}_P(0.01)$ as defined in (29). The reality check p-values of White (2000) and Hansen (2001) are also reported, with the Riskmetrics model as our benchmark model since it is the most widely used model for practitioners. The null hypothesis is that none of the 15 competing models can beat the Riskmetrics model. A significant (small) reality check p-value is in favour of the alternative hypothesis that there is a model that beats the Riskmetrics benchmark in terms of the

Table IVA. Predictive quantile loss, Period 1

		Indonesia	Korea	Malaysia	Taiwan	Thailand
5%	Riskmetrics	0.1268	0.1152	0.0998	**0.1540**	0.1459
	Normal*	0.1295	0.1181	0.1065	0.1663	0.1440
	HS	0.1189	0.1152	0.1093	0.2231	0.1486
	HS*	0.1253	0.1204	0.1011	0.1684	0.1437
	MC	0.1585	0.1220	0.1194	0.2198	0.1575
	MC*	0.1305	0.1189	0.1075	0.1705	0.1446
	NP	0.1188	0.1165	0.1091	0.1969	0.1407
	NP*	0.1253	0.1205	0.1009	0.1629	**0.1404**
	GP	0.1246	**0.1149**	0.0993	0.1861	0.1490
	GP*	0.1269	0.1203	**0.0987**	0.1637	0.1442
	GEV	0.2311	0.1988	0.2218	0.3770	0.2921
	GEV*	0.1805	0.1692	0.1681	0.2497	0.2160
	HILL	0.1231	0.1356	0.1201	0.3159	0.2270
	HILL*	0.1265	0.1271	0.1228	0.2002	0.1660
	CaViaR$_S$	**0.1160**	0.1222	0.1049	0.1626	0.1416
	CaViaR$_A$	0.1519	0.1227	0.1050	0.1626	0.1420
	White	0.362	0.891	0.597	1.000	0.703
	Hansen	0.194	0.601	0.426	0.595	0.582
1%	Riskmetrics	0.0417	**0.0311**	0.0374	**0.0580**	0.0494
	Normal*	0.0377	0.0329	0.0348	0.0597	0.0491
	HS	0.0392	0.0332	0.0329	0.0683	0.0573
	HS*	0.0399	0.0332	0.0332	0.0588	0.0470
	MC	0.0422	0.0315	0.0317	0.0615	0.0535
	MC*	0.0378	0.0330	0.0347	0.0592	0.0486
	NP	0.0391	0.0331	0.0328	0.0682	0.0465
	NP*	0.0399	0.0332	0.0332	0.0588	0.0471
	GP	0.0389	0.0314	**0.0313**	0.0645	0.0543
	GP*	0.0380	0.0331	0.0333	0.0590	0.0470
	GEV	0.0915	0.0583	0.0748	0.0892	0.1020
	GEV*	0.0605	0.0464	0.0525	0.0696	0.0636
	HILL	0.0451	0.0399	0.0427	0.0718	0.0627
	HILL*	0.0377	0.0344	0.0332	0.0598	0.0484
	CaViaR$_S$	**0.0300**	0.0339	0.0329	0.0583	**0.0437**
	CaViaR$_A$	0.0552	0.0337	0.0319	0.0599	0.0440
	White	0.095	0.888	0.226	0.892	0.433
	Hansen	0.090	0.863	0.222	0.837	0.363

Note: The number in each cell refers to the out-of-sample (1/1/1996–12/31/1996) average quantile loss; 'White' refers to the bootstrap reality check *p*-value of White (2000); 'Hansen' refers to the bootstrap reality check *p*-value of Hansen (2001). We use 1000 bootstrap samples and the stationary bootstrap smoothing parameter *q* = 0.25. Riskmetrics is the benchmark in the reality check. The best model for each country with the smallest out-of-sample average quantile loss is in bold font.

out-of-sample average loss values $\hat{Q}_P(\alpha)$. The best model in $\hat{Q}_P(\alpha)$ is indicated in bold font in the tables for each country and for each α. Riskmetrics is selected several times as the best model (three times out of 10 in Period 1, once out of 10 in Period 2, and twice out of 10 in Period 3), while there are models with lower loss values. However, the reality check *p*-values are generally quite large and insignificant, indicating that Riskmetrics is not dominated by any other models for all three periods.

Table IVB. Predictive quantile loss, Period 2

		Indonesia	Korea	Malaysia	Taiwan	Thailand
5%	Riskmetrics	0.3885	0.3461	0.3123	0.2071	0.2689
	Normal*	0.3713	0.3499	0.3166	0.1996	0.2745
	HS	0.5743	0.6358	0.4979	0.2165	0.3072
	HS*	0.3959	0.3589	0.3303	0.1992	0.2750
	MC	0.4191	0.5712	0.4491	0.2087	0.2846
	MC*	0.3732	0.3475	0.3186	0.1988	0.2711
	NP	0.5746	0.6361	0.4984	0.2164	0.3074
	NP*	0.6395	0.7765	0.5954	0.2346	0.4708
	GP	0.6664	0.6897	0.5997	**0.1928**	0.3750
	GP*	0.4291	0.3666	0.3474	0.2007	0.2788
	GEV	**0.3556**	0.3568	**0.3034**	0.3549	0.3033
	GEV*	0.4868	0.3861	0.4027	0.2672	0.3950
	HILL	0.4680	0.4700	0.3743	0.3038	**0.2559**
	HILL*	0.3799	**0.3326**	0.3128	0.2196	0.2889
	CaViaR$_S$	0.4644	0.3480	0.3242	0.1982	0.2746
	CaViaR$_A$	0.5438	0.3422	0.3217	0.1979	0.2815
	White	0.599	0.749	0.862	0.498	0.786
	Hansen	0.267	0.510	0.662	0.268	0.581
1%	Riskmetrics	0.1284	**0.0935**	0.0926	0.0577	0.0721
	Normal*	0.1352	0.1118	0.1072	0.0620	0.0717
	HS	0.2245	0.2432	0.1656	0.0655	0.0773
	HS*	0.1322	0.1042	0.0982	0.0577	0.0724
	MC	0.2030	0.3164	0.2182	0.0611	0.1021
	MC*	0.1373	0.1112	0.1087	0.0606	**0.0704**
	NP	0.2258	0.2443	0.1664	0.0655	0.0775
	NP*	0.3795	0.4937	0.2602	0.0834	0.1911
	GP	0.2702	0.3046	0.1924	0.0621	0.0855
	GP*	0.1290	0.1052	0.0973	0.0580	0.0720
	GEV	**0.1115**	0.1040	0.0916	0.0858	0.0943
	GEV*	0.1679	0.1046	0.1227	0.0691	0.1131
	HILL	0.1502	0.1724	0.1209	0.0692	0.0755
	HILL*	0.1284	0.0982	0.0916	**0.0576**	0.0821
	CaViaR$_S$	0.1913	0.1058	0.0937	0.0596	0.0744
	CaViaR$_A$	0.1669	0.0958	**0.0858**	0.0578	0.0748
	White	0.739	0.996	0.886	0.935	0.900
	Hansen	0.514	0.913	0.606	0.878	0.844

Note: The number in each cell refers to the out-of-sample (7/1/1997–6/30/1998) average quantile loss; 'White' refers to the bootstrap reality check *p*-value of White (2000); 'Hansen' refers to the bootstrap reality check *p*-value of Hansen (2001). We use 1000 bootstrap samples and the stationary bootstrap smoothing parameter $q = 0.25$. Riskmetrics is the benchmark in the reality check. The best model for each country with the smallest out-of-sample average quantile loss is in bold font.

(Only one exception is for Period 3, Thailand, $\alpha = 0.01$, for which case MC is the best and significantly better than Riskmetrics.)

In Table IVA, CaViaR$_S$, GP, GP* and NP* generate (marginally but not significantly) better quantile loss than Riskmetrics, while many models (particularly GEV and GEV*) are much worse. None of the models can produce a significantly better risk forecast than Riskmetrics. Even though there

Table IVC. Predictive quantile loss, Period 3

		Indonesia	Korea	Malaysia	Taiwan	Thailand
5%	Riskmetrics	**0.2079**	**0.2602**	0.1956	0.1952	0.1916
	Normal*	0.2200	0.2674	0.1919	0.1953	0.1949
	HS	0.2353	0.3142	0.1918	0.2058	0.1939
	HS*	0.2088	0.2707	0.1917	0.1965	0.1934
	MC	0.2122	0.2807	0.1994	0.2057	0.1943
	MC*	0.2251	0.2659	0.1928	0.1974	0.2008
	NP	0.2340	0.3145	0.1918	**0.1792**	0.1990
	NP*	0.2085	0.2708	0.1917	0.1846	**0.1832**
	GP	0.2831	0.3601	0.2089	0.1870	0.2041
	GP*	0.2116	0.2723	0.1942	0.1944	0.1897
	GEV	0.3311	0.3187	0.3101	0.3531	0.3107
	GEV*	0.3644	0.3484	0.2800	0.2784	0.3371
	HILL	0.2103	0.2634	0.2033	0.3056	0.2348
	HILL*	0.2166	0.2791	0.2026	0.2259	0.2249
	CaViaR$_S$	0.2448	0.2648	**0.1856**	0.1918	0.1857
	CaViaR$_A$	0.2741	0.2668	0.1893	0.1912	0.1871
	White	0.991	0.994	0.677	0.284	0.614
	Hansen	0.840	0.895	0.553	0.191	0.454
1%	Riskmetrics	0.0588	0.0708	0.0554	0.0577	0.0560
	Normal*	0.0641	0.0666	0.0558	0.0569	0.0501
	HS	0.0540	0.0705	0.0548	0.0651	0.0531
	HS*	0.0650	0.0673	0.0540	0.0541	0.0566
	MC	**0.0532**	0.0786	0.0559	0.0574	**0.0480**
	MC*	0.0645	**0.0664**	0.0550	0.0558	0.0519
	NP	0.0540	0.0808	0.0550	0.0651	0.0527
	NP*	0.0649	0.0673	0.0542	**0.0531**	0.0531
	GP	0.0572	0.0787	0.0578	0.0596	0.0489
	GP*	0.0659	0.0667	0.0535	0.0542	0.0570
	GEV	0.1237	0.0909	0.1004	0.0862	0.0926
	GEV*	0.1197	0.0900	0.0830	0.0714	0.0941
	HILL	0.0635	0.0691	0.0616	0.0697	0.0650
	HILL*	0.0716	0.0720	0.0548	0.0576	0.0664
	CaViaR$_S$	0.0807	0.0694	0.0532	0.0539	0.0625
	CaViaR$_A$	0.0901	0.0691	**0.0526**	0.0539	0.0601
	White	0.568	0.654	0.720	0.314	0.000
	Hansen	0.287	0.573	0.704	0.283	0.000

Note: The number in each cell refers to the out-of-sample (1/1/1999–12/31/1999) average quantile loss; 'White' refers to the bootstrap reality check *p*-value of White (2000); 'Hansen' refers to the bootstrap reality check *p*-value of Hansen (2001). We use 1000 bootstrap samples and the stationary bootstrap smoothing parameter $q = 0.25$. Riskmetrics is the benchmark in the reality check. The best model for each country with the smallest out-of-sample average quantile loss is in bold font.

are some models which produce smaller quantile loss values than Riskmetrics, this has not been observed uniformly across countries. Filtering generally helps for Taiwan and Thailand, while it can be harmful (e.g., for Malaysia).

The result for Period 2 is reported in Table IVB. In the crisis period, more EVT models are selected as the best models. For $\alpha = 0.05$, the best models for all five countries are the EVT models—GP,

GEV, HILL or HILL*. For $\alpha = 0.01$, the best models are Riskmetrics, MC*, GEV, HILL* or CaViaR$_A$. However, none of these models are significantly better than the benchmark Riskmetrics model. During this period, filtering works for HS*, MC*, GP* and HILL* in terms of the quantile loss. As an unfiltered model assumes unconditional variance, it may be unlikely that an unfiltered model is better than the filtered one. However, this could happen. For example, NP* is worse than NP in terms of both the empirical tail coverage $\hat{\alpha}_P$ and the predictive quantile loss $\hat{Q}_P(\alpha)$, during the crisis period, for both $\alpha = 0.05, 0.01$, for all five countries. For NP*, this may be due to the issue mentioned earlier in footnote 5.

The post-crisis period result is presented in Table IVC, where the best models are often selected from MC, MC*, NP or NP*. There are some forecasting improvements in terms of $\hat{Q}_P(\alpha)$ and more so with $\alpha = 0.01$. For Thailand with $\alpha = 0.01$, MC makes a significantly better VaR forecasting than the benchmark. For the other cases, Riskmetrics is not statistically dominated by any other models.

Figures 1–3 plot the out-of-sample loss values with $\alpha = 0.05$

$$[\alpha - 1(r_t < \hat{q}_t(\alpha))][r_t - \hat{q}_t(\alpha)] = \begin{cases} 0.95 \cdot |r_t - \hat{q}_t(\alpha)| & \text{if } r_t < \hat{q}_t(\alpha) \\ 0.05 \cdot |r_t - \hat{q}_t(\alpha)| & \text{if } \hat{q}_t(\alpha) < r_t \end{cases}$$

for each time of $t = R + 1, \ldots, T$. This is the summand of the out-of-sample average loss $\hat{Q}_P(\alpha)$ in equation (29). For the sake of space, we present only the figures for Korea with $\alpha = 0.05$, while all other figures for the five economies with $\alpha = 0.05, 0.01$ are available upon request and deliver largely the same features. Note that the larger weight 0.95 is given to the loss when a violation occurs and thus once there is a violation the loss value increases at that time (generating a spike in plots).

Reading Figure 1 for Period 1, we note that when two models have the same number of violations and thus the same value for $\hat{\alpha}_P$, a model which produces a larger $\hat{Q}_P(\alpha)$ loss value is a worse one. For instance, for Period 1 for Korea with $\alpha = 0.05$, both GP and GP* produce the same number of violations with $\hat{\alpha}_P = 0.0766$, while GP is better than GP* because they have $\hat{Q}_P(\alpha) = 0.1149$ and 0.1203, respectively. This can be observed from Figure 1 where both GP and GP* have spikes at the same points of time but the spikes for GP* are higher. When two models have the same quantile loss $\hat{Q}_P(\alpha)$, a model which produces $\hat{\alpha}_P$ closer to α is a better one. For instance, both Riskmetrics and HS produce the same quantile loss $\hat{Q}_P(0.05) = 0.1152$. But HS is better than Riskmetrics because HS has $\hat{\alpha}_P = 0.0460$ and Riskmetrics has $\hat{\alpha}_P = 0.0651$. On the other hand, some models can be bad (or good) in terms of both $\hat{\alpha}_P$ and $\hat{Q}_P(\alpha)$. For example, GEV, GEV* and HILL have shown no spikes in the plots, i.e., none or few violations (and thus they produce bad empirical coverage probability), but they also have the largest quantile loss values. They are less adequate models in both criteria of $\hat{\alpha}_P$ and $\hat{Q}_P(\alpha)$.

Similar interpretations of the plots can be drawn for the crisis period (Figure 2) and the post-crisis period (Figure 3). It should be noted that the vertical scales of the graphs are different for Figures 1–3. If we look at the plots for HS, MC, NP, NP* and GP in the crisis period, there are a large number of violations (which make $\hat{\alpha}_P$ too large) and many large spikes in the loss (which make the average loss $\hat{Q}_P(\alpha)$ quite large). The worst model in Period 2 is NP*, which has produced a very volatile loss plot indicating a poor predictive performance, which can easily be verified by Table IIIB ($\hat{\alpha}_P = 0.2989$) and Table IVB ($\hat{Q}_P(0.05) = 0.7765$). The best model in Period 2 is filtered Hill (HILL*) that produces both very good coverage probability ($\hat{\alpha}_P = 0.0498$) and the smallest average quantile loss value ($\hat{Q}_P(0.05) = 0.3326$). In Period 3, while the plots for HS, MC, NP, NP* and GP remain very volatile, Riskmetrics becomes a good model again (as was the case for Period 1), with a reasonable empirical coverage probability $\hat{\alpha}_P = 0.0613$ as shown in Table IIIC and with the small-
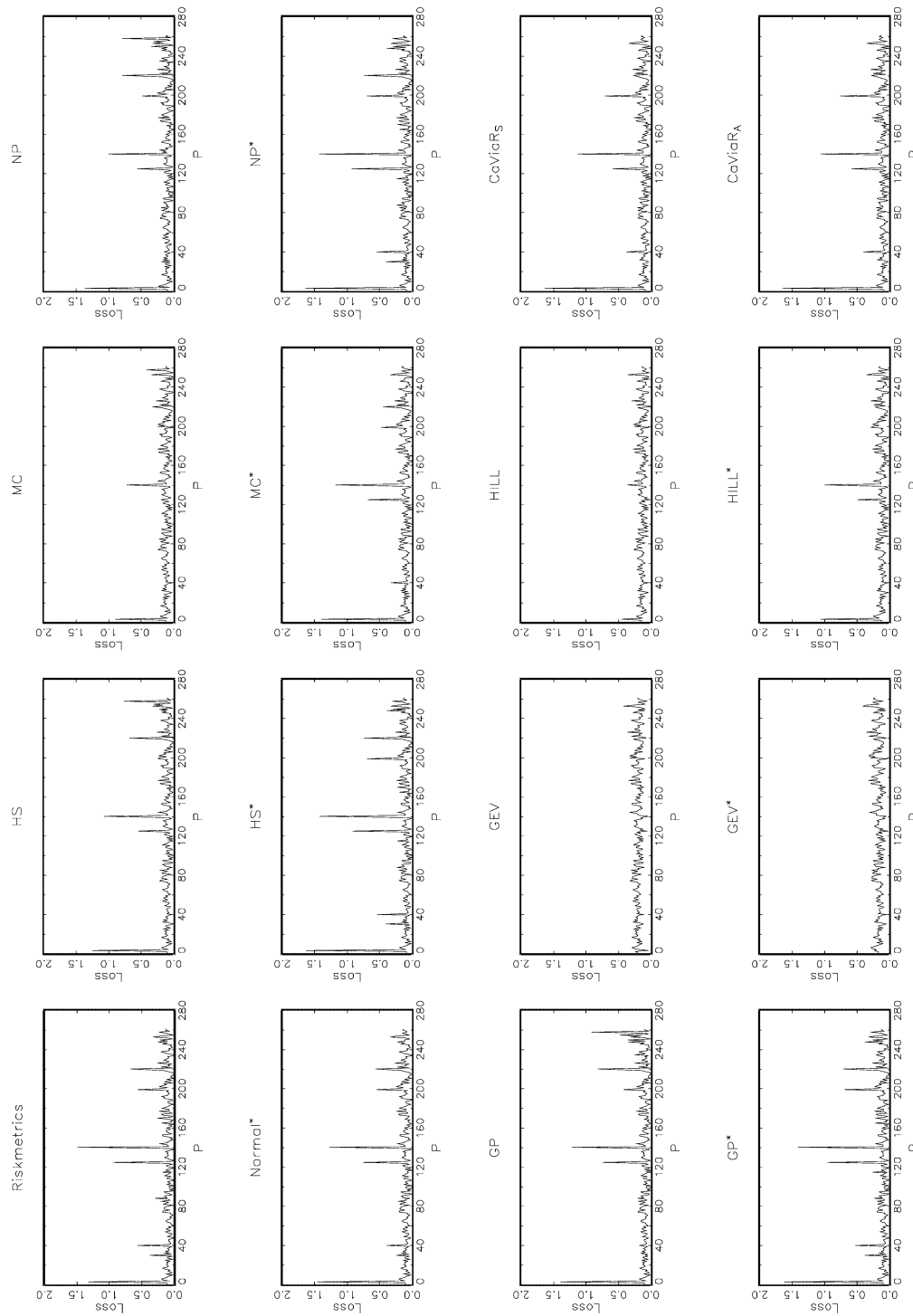
Figure 1. Out-of sample loss for korea with $\alpha = 5\%$ [Period 1]
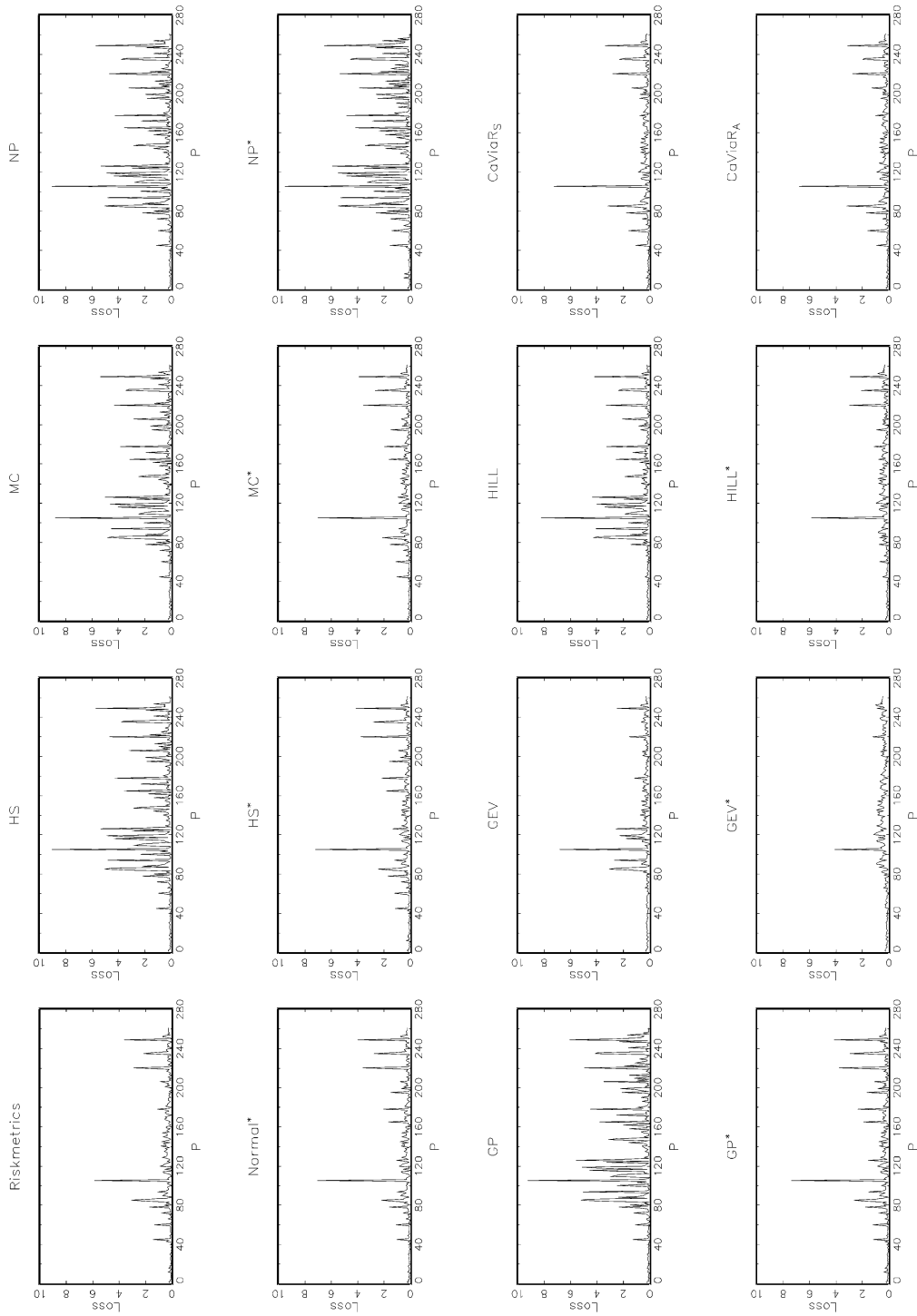
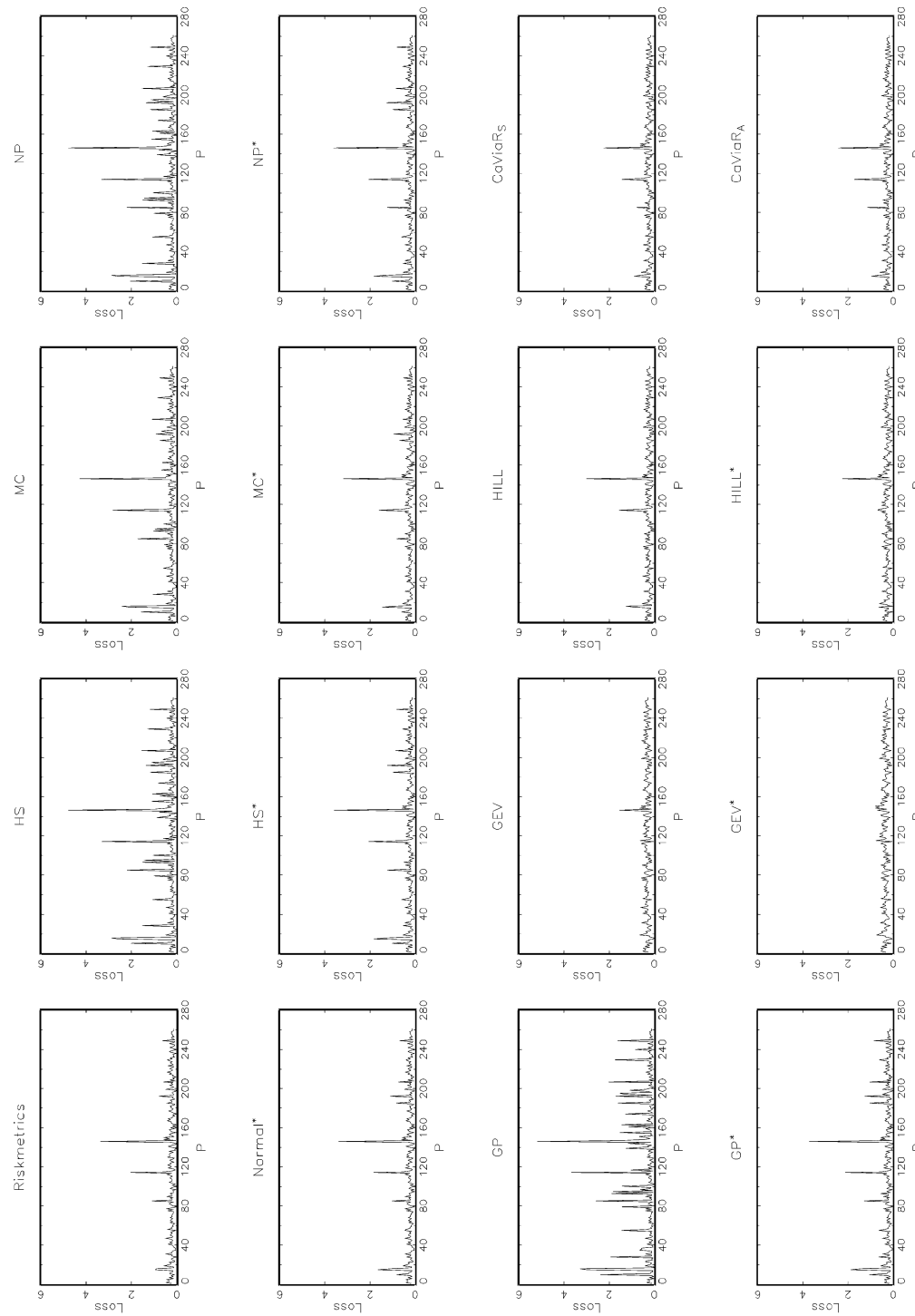Figure 2. Out-of sample loss for korea with $\alpha = 5\%$ [Period 2]

Figure 3. Out-of sample loss for korea with $\alpha = 5\%$ [Period 3]

est loss $\hat{Q}_P(0.05)$ as shown in Table IVC. A final note on the figures for most of the comparable models is that the dates in which the violations took place are similar. In other words, most models capture the violations at the same time as they occur.

CONCLUSIONS

In this paper we have studied a comparative risk forecast experiment for five emerging markets. Our findings are summarized as follows. (i) Based on $\hat{\alpha}_P$ (the coverage probability or the number of violations), the Riskmetrics model behaves reasonably well before and after the crisis, while some EVT models do better in the crisis period. Filtering appears to be useful for some models (particularly the EVT models), though it may be harmful for other models. The forecasting performance of different models varies with $\alpha = 0.05$ and 0.01. (ii) The results based on $\hat{Q}_P(\alpha)$ (the predictive quantile loss) are largely compatible with those based on $\hat{\alpha}_P$. While the Riskmetrics and other conventional models work reasonably well before and after the crisis, the EVT models work better in the crisis period. However, we cannot reject that the Riskmetrics model cannot be beaten even during the crisis period. The CaViaR models have shown some success in predicting the VaR risk measure across various periods.

Our experiment demonstrates that risk forecasting during the crisis period is more diffcult and yields poorer results than during tranquil periods, and most VaR models generally behave similarly before and after the crisis, but differently in the crisis period. Hence, it may be promising to consider the regime-switching VaR models as in Guidolin and Timmermann (2003) and Li and Lin (2004).

ACKNOWLEDGEMENTS

REFERENCES

Balkema AA, de Haan L. 1974. Residual lifetime at great age. *Annals of Probability* **2**: 792–804.
Barone-Adesi G, Giannopoulos K, Vosper L. 2002. Backtesting derivative portfolios with filtered historical simulation (FHS). *European Financial Management* **8**: 31–58.
Berman SM. 1964. Limiting theorems for the maximum term in stationary sequences. *Annals of Mathematical Statistics* **35**: 502–516.
Bertail P, Haefke C, Politis DN, White H. 2004. A subsampling approach to estimating the distribution of diverging statistics with applications to assessing financial market risks. *Journal of Econometrics* **120**: 295–326.
Bierens HJ, Ginther D. 2001. Integrated conditional moment testing of quantile regression models. *Empirical Economics* **26**: 307–324.
Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**: 307–327.

Brodie M, Glasserman P. 1998. Simulation for option pricing and risk management. In *Risk Management and Analysis*, Vol. 1, Alexander C (ed.). John Wiley & Sons: New York.

Bühlmann P, McNeil AJ. 2001. An algorithm for nonparametric GARCH modeling. Working Paper, Seminar für Statistik, Federal Institute of Technology, Zurich, Germany.

Cai Z. 2002. Regression quantiles for time series. *Econometric Theory* **18**: 169–192.

Danielsson J, de Vries CG. 1997. Beyond the sample: extreme quantile and probability estimation. Working Paper, London School of Economics and Tinbergen Institute.

Danielsson J, de Haan L, Peng L, de Vries CG. 2001. Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis* **76**: 226–248.

Davison AC, Smith RL. 1990. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society, Series B* **52**: 393–442.

Diebold FX, Mariano RS. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253–263.

Diebold FX, Schuermann T, Stroughair J. 2000. Pitfalls and opportunities in the use of extreme value theory in risk management. *Journal of Risk Finance* **1**: 30–36.

Embrechts P, Klüppelberg C, Mikosch T. 1997. *Modeling Extremal Events for Insurance and Finance*. Springer-Verlag: New York.

Engle RF. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica* **50**: 987–1008.

Engle RF, González-Rivera G. 1991. Semiparametric ARCH models. *Journal of Business and Economic Statistics* **9**: 345–360.

Engle RF, Manganelli S. 2004. CaViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics* **22**(4): 367–381.

Fama E, French KR. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* **33**: 3–56.

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

Gallant AR, Hsieh DA, Tauchen GE. 1991. On fitting a recalcitrant series: the pound/dollar exchange rate, 1974–1983. In *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics,* Barnett WA (ed.). Cambridge University Press: Cambridge.

Giacomini R, Komunjer I. 2002. Evaluation and combinations of conditional quantile forecasts. Working Paper, UCSD and Caltech.

Giacomini R, White H. 2003. Tests of conditional predictive ability. Working Paper, Department of Economics, University of California, San Diego.

González-Rivera G, Lee T-H, Mishra S. 2004. Forecasting volatility: a reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting* **20**(4): 629–645.

Gonzalo J, Olmo J. 2004. Which extreme values are really extremes. *Journal of Financial Econometrics* **2**(3): 349–369.

Granger CWJ. 2002. Some comments on risk. *Journal of Applied Econometrics* **17**: 447–456.

Guidolin M, Timmermann A. 2003. Value at risk and expected shortfall under regime switching. UCSD.

Hall P, Wolff RCL, Yao Q. 1999. Methods for estimating a conditional distribution function. *Journal of the American Statistical Association* **94**: 154–163.

Hansen BE. 1994. Autoregressive conditional density estimation. *International Economic Review* **35**: 705–730.

Hansen PR. 2001. An unbiased and powerful test for superior predictive ability. Working Paper, Department of Economics, Brown University.

Hansen PR. 2003. Asymptotic tests of composite hypotheses. Working Paper, Stanford University.

Hansen PR, Lunde A. 2004. A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics*, in press.

Harvey C, Siddique A. 1999. Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis* **34**: 465–488.

Harvey C, Siddique A. 2000. Conditional skewness in asset pricing tests. *Journal of Finance* **55**: 1263–1295.

Hill BM. 1975. A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **3**: 1163–1174.

Hull J, White A. 1998. Incorporating volatility updating into historical simulation method for value at risk. *Journal of Risk* **1**: 5–19.

Jenkinson AF. 1955. Distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society* **81**: 145–158.

Morgan JP. 1995. *Riskmetrics Technical Manual*, 3rd edn.

Jorion P. 2000. *Value at Risk*, 2nd edn. McGraw Hill: New York.

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

Komunjer I. 2004. Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics*, in press.

Lakonishok J, Shleifer A, Vishny RW. 1994. Contrarian investment, extrapolation, and risk. *Journal of Finance* **49**: 1541–1578.

Leadbetter MR, Lindgren G, Rootzén H. 1983. *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag: New York.

Li M-YL, Lin H-WW. 2004. Estimating value-at-risk via Markov switching ARCH models—an empirical study on stock index returns. *Applied Economics Letters* **11**(11): 679–691.

Lo AW, MacKinlay AC. 1999. *A Non-Random Walk Down Wall Street*. Princeton University Press: Princeton, NJ.

Longin F. 1996. The asymptotic distribution of extreme stock market returns. *Journal of Business* **69**: 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 DM. 1982. Law of large numbers for sum of extreme values. *Annals of Probability* **10**: 754–764.

McCracken MW. 2000. Robust out-of-sample inference. *Journal of Econometrics* **99**: 195–223.

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

Mittnik S, Paolella MS, Rachev ST. 2002. Stationarity of stable power-GARCH processes. *Journal of Econometrics* **106**: 97–107.

Neftçi S. 2000. Value at risk calculations, extreme events, and tail estimation. *Journal of Derivatives* **7**: 23–38.

Nelson DB. 1991. Conditional heteroscedasticity in asset returns: a new approach. *Econometrica* **59**: 347–370.

Owen AB. 2001. *Empirical Likelihood*. Chapman and Hall: London.

Pickands III J. 1975. Statistical inference using extreme order statistics. *Annals of Statistics* **3**: 119–131.

Politis DN, Romano JP. 1994. The stationary bootstrap. *Journal of the American Statistical Association* **89**: 1303–1313.

Poon S, Granger C. 2003. Forecasting volatility in financial markets: a review. *Journal of Economic Literature* **41**: 478–539.

Pownall RAJ, Koedijk KG. 1999. Capturing downside risk in financial markets: the case of the Asian crisis. *Journal of International Money and Finance* **18**: 853–870.

Price K, Storn R. 1997. Differential evolution. *Dr. Dobb's Journal* 2**64**: 18–24.

Randles RH. 1982. On the asymptotic normality of statistics with estimated parameters. *Annals of Statistics* **10**: 462–474.

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

Taylor SJ. 1986. *Modeling Financial Time Series*. John Wiley & Sons: New York.

Venkataraman S. 1997. Value at risk for a mixture of normal distributions: the use of quasi-Bayesian estimation techniques. *Economic Perspectives*, March 2–13, Federal Reserve Bank of Chicago.

von Mises R. 1936. La distribution de la plus grande de n valeurs. *Rev Math Union Interbalcanique* **1**: 141–160. [reproduced: *Selected Papers of Richard von Mises*, American Mathematical Society, 1964, vol. **2**, 271–294].

Wang K, Fawson C, Barrett C, McDonald JB. 2001. A flexible parametric GARCH model with an application to exchange rates. *Journal of Applied Econometrics* **16**: 521–536.

West KD. 1996. Asymptotic inference about predictive ability. *Econometrica* **64**: 1067–1084.

White H. 2000. A reality check for data snooping. *Econometrica* **68**: 1097–1126.

Zheng JX. 1998. A consistent nonparametric test of parametric regression models under conditional quantile restrictions. *Econometric Theory* **14**: 123–138.

Authors' biographies:

Yong Bao is Assistant Professor of Economics at the University of Texas, San Antonio. He earned his Ph.D. in Economics in 2004 from the University of California, Riverside. His major research fields are time series analysis and finite sample econometrics.

Tae-Hwy Lee is Professor of Economics at the University of California, Riverside. He got a Ph.D. in Economics (1990) from UCSD. His research interests are time series econometrics and financial econometrics. He is currently Visiting Professor of Economics at California Institute of Technology for 2005–2006.

Burak Saltoğlu is Professor of Economics at Marmara University, Turkey. He got a Ph.D. in Economics (1996) from University of Essex, U.K. His major research fields are financial econometrics and financial risk management.

Authors' addresses:

Yong bao, Department of Economics, University of Texas, San Antonio, TX 78249, USA.

Tae-Hwy Lee, Department of Economics, University of California, Riverside, CA 92521, USA.

Burak Saltoğlu, Department of Economics, Marmara University, Istanbul 81040, Turkey.