THE PHILIPPINE REVIEW OF ECONOMICS
VOL. XLVI NO. 2 DECEMBER 2009 PP. 91-121

Measuring market risk using extreme value theory

Jose Oliver Q. Suaiso and Dennis S. Mapa*
University of the Philippines School of Statistics

The adoption of Basel II standards by the Bangko Sentral ng Pilipinas initiates financial institutions to develop value-at-risk (*VaR*) models to measure market risk. In this paper, two *VaR* models are considered using the peaks-over-threshold (POT) approach of the extreme value theory (EVT): (1) static EVT model, which is the straightforward application of POT to the bond benchmark rates; and (2) dynamic EVT model, which applies POT to the residuals of the fitted AR-GARCH model. The results are compared with traditional *VaR* methods such as RiskMetrics and AR-GARCH-type models. The relative size, accuracy, and efficiency of the models are assessed using mean relative bias, backtesting, likelihood ratio tests, loss function, mean relative scaled bias, and computation of market risk charge. Findings show that the dynamic EVT model can capture market risk conservatively, accurately, and efficiently. It is also practical to use because it has the potential to lower a bank's capital requirements. Comparing the two EVT models, the dynamic model is better than static as the former can address some issues in risk measurement and effectively capture market risks.

JEL classification: C01, C22, G12
Keywords: extreme value theory, peaks-over-threshold, value-at-risk, market risk, risk management

1. Introduction

Everyday banks operate and conduct activities that have inherent risks. By exposing themselves more to such activities, banks expect to gain more revenue. In bond trading, for example, banks take the risk of losing or gaining since the

*The authors are grateful to Dr. Joselito C. Magadia and Dr. Reynaldo Rey and the participants at the 9th Faculty-Student National Conference on Statistical Sciences (Financial Econometrics Session) and at a seminar held at the School of Statistics, University of the Philippines Diliman. Errors and omissions are sole responsibilities of the authors.

movement of interest rates is uncertain. They can earn more if the volume of investment is increased, but this could result in big losses in the event of an unfavorable market scenario. Such type of risk is called market risk or the risk of losses in on- and off-balance sheet positions arising from movements in market prices [BSP 2006a].

So for banks to cover any unexpected losses arising from their risk-taking activities, the Bangko Sentral ng Pilipinas (BSP) requires them to maintain capital that is sufficient to act as buffer against insolvency. This requirement is set under BSP Circular 538 or the Revised Risk-Based Capital Adequacy Framework. This circular is essentially Basel II—a framework for international capital adequacy standards. Circular 538 states that local banks must maintain their minimum capital adequacy ratio (CAR) at 10 percent. In formula, CAR is computed as:

$$CAR = \frac{capital}{\begin{array}{l} sum\ of\ risk\ weighted\ assets\ for \\ credit\ risk + market\ risk + operational\ risk \end{array}} \geq 10\% \qquad (1)$$

To meet the CAR requirement, banks may either increase their capital or decrease the level of risk in their assets. Increasing capital would be impractical because it would be too costly for a bank. To minimize risks, the bank may implement different strategies such as asset reclassification, credit risk-mitigation techniques, or advanced risk management systems. Minimizing risks therefore would be a better option, as it may be less costly to implement.

To compute for CAR, the procedure of putting a value to the risks in assets is very crucial to banks. It is because their measure of risks will determine their required capital and apparently their compliance with BSP's regulations. So to guide banks on how to measure risks, Basel II suggests different approaches. In market risk, for example, banks may define and measure market risk using internal models approach (IMA). Under IMA, capital is computed as a factor of bank's portfolio, which is based on the internal model assumptions on the behavior asset prices and yields. Thus IMA requires an internal model to measure market risk. In fact, BSP will allow banks to use IMA in year 2010 as part of BSP's Basel II Implementation Plan. Allowing banks to use internal models such as value-at-risk will benefit them since the capital will be commensurate to the current risks measured by the model. And it is expected that the banks' required capital could be lowered if a *VaR* model is used. Therefore the implementation of Basel II in the Philippines motivates a study on market risk measurement models.

While several market risk models have been developed, researchers have noted some common issues in risk measurement. These issues typically arise from the disagreement between model assumptions and actual observations and/or empirical studies on financial data. Among these issues are (1) the non-normality and/or fat-tailedness of price distribution of assets, (2) the presence of serial correlation and heteroskedasticity in financial time series, and (3) the problem of whether to model the entire distribution or only the tails. To address these issues, this paper proposes to use the extreme value theory (EVT) approach in measuring market risk.

The paper is structured as follows: section 2 defines value-at-risk (*VaR*) and discusses existing *VaR* models such as RiskMetrics and autoregressive moving average-generalized autoregressive conditional heteroskedasticity (ARMA-GARCH)-type models; section 3 introduces extreme value theory, its types of approaches, and *VaR* models using EVT; section 4 discusses some procedures in assessing *VaR* model by using measures of relative size, accuracy, and efficiency; section 5 presents the empirical results; and section 6 concludes the paper.

2. Value-at-risk models

One popular measure of market risk is the value-at-risk. It is a measure of the maximum potential loss of a financial position during a given time period with a given probability. The calculation of *VaR* is aimed at making a statement of the following form: "We are X percent certain that we will not lose more than V pesos in the next N days." Here V is the *VaR* of the portfolio, X is the confidence level, and N is the time horizon [Jorion 2000].

Following Tsay [2005], we define *VaR* in statistical terms. Suppose at the time index t we are interested in the risk of a financial position for the next ℓ periods. Let $\Delta V(\ell)$ be the change in value of the assets in the financial position from time t to $t+\ell$. This quantity is measured in pesos and is a random variable at the time index t. Denote the cumulative distribution function (CDF) of $\Delta V(\ell)$ by $F_\ell(x)$. We define the *VaR* of a *long position*[1] over the time horizon ℓ with probability p as

$$p = \Pr\left[\Delta V(\ell) \leq VaR\right] = F_\ell(VaR) \tag{2}$$

Since the holder of a long financial position suffers a loss when $\Delta V(\ell) < 0$, the *VaR* value above is negative when p is small, signifying a loss. From the definition, the probability that the holder would encounter a loss greater than or equal to *VaR* over the time horizon ℓ is p. Alternatively, we can interpret *VaR* as follows: With probability $(1-p)$, the holder would encounter a loss less than or equal to *VaR* over the time horizon ℓ.

[1] In bond trading for example, long position happens when a trader buys a bond and then sells it later, thinking that the price of the bond is likely to go up in the future.

2.1. The RiskMetrics model

J.P. Morgan developed RiskMetrics method to calculate *VaR*. Let r_t denote the daily *log return*[2], F_{t-1} the information set available at time $t-1$, and μ_t and h_t the conditional mean and the conditional variance of r_t, respectively. RiskMetrics assumes that

$$r_t | F_{t-1} \sim N(\mu_t, h_t)$$

where $\mu_t = 0$, $h_t = \alpha h_{t-1} + (1-\alpha) r_{t-1}^2$ and $1 > \alpha > 0$. A typical value of α is 0.94 so the conditional variance can be written as $h_t = 0.94 h_{t-1} + 0.06\, r_{t-1}^2$. RiskMetrics defines 1-day 99 percent confidence level *VaR* as:

$$VaR = Amount\ of\ position \times F^{-1}(0.01) h_{t+1}^{1/2} \tag{3}$$

where $F^{-1}(0.01)$ is the 1 percent quantile of a standard normal distribution (i.e., $F^{-1}(0.01) = 2.326$) and h_{t+1} is the 1-step ahead forecast of the conditional variance given by $h_{t+1} = \alpha h_t (1-\alpha) r_t^2$.

2.2. ARMA-GARCH models

Econometric models such as ARMA-GARCH type can also be used to calculate *VaR*. These models, however, do not forecast the *VaR*; rather they forecast the return and conditional variance of the time series. Here the asset return is forecasted using ARMA models, and the volatility is forecasted using the volatility models such as GARCH, exponential GARCH (EGARCH), GARCH-in-Mean (GARCH-M), and integrated in variance GARCH (IGARCH). These volatility models can capture the observed *volatility clustering*[3] in financial time-series data.

An ARMA(p,q)-GARCH(u,v) process is represented by the time-series model below:

$$\begin{aligned} r_t &= \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + a_t - \sum_{j=1}^{q} \theta_j a_{t-j} \\ h_t &= \alpha_0 + \sum_{i=1}^{u} \alpha_i a_{t-i}^2 + \sum_{j=1}^{v} \beta_j h_{t-j} \end{aligned} \tag{4}$$

where the innovation $a_t = h_t^{1/2} \varepsilon_t$ and ε_t is the error term. Assuming that ε_t is normally distributed, *VaR* at time $t+1$ is given by the following equation:

$$VaR = \hat{r}_{t+1} - F^{-1}(p) \hat{h}_{t+1}^{1/2} \tag{5}$$

---

[2] A typical unit of observation used in analyzing financial data is the log return of the price of assets. If P_t is the price of an asset at time index t, then log return (r_t) is defined as $r_t = \ln(P_t / P_{t-1})$.

[3] Volatility clustering is commonly observed in financial time series. It is a feature of time series wherein large shocks tend to be followed by large shocks for certain time periods, and small shocks tend to be followed by small shocks for certain time periods.

where $F^{-1}(p)$ is the $(p)100\%$ quantile of a standard normal distribution. Assuming that ε_t is distributed as a standardized Student-t distribution with v degrees of freedom, the *VaR* at time $t+1$ is

$$VaR = r_{t+1} - t_v^* - t_v^*(p)h_{t+1}^{1/2} \tag{6}$$

where $t_v^*(p)$ is the pth quantile of a standardized Student-t distribution with v degrees of freedom.

Here the computed *VaR* is actually the *VaR* of the log return r_t. The final *VaR* figure is computed by multiplying ARMA-GARCH *VaR* estimates by the asset amount or the mark-to-market value of the financial position.

While these models can capture serial autocorrelation and volatility clustering, the observed non-normality and heavy-tails of the distributions of financial data still pose problems to a risk manager. Thus, skewness risk and kurtosis risk are still other issues in risk measurement. So to address these problems, this paper proposes to use extreme value theory in measuring market risk.

3. *VaR* models using extreme value theory

Extreme value theory extends the central limit theorem—which deals with the distribution of the average i.i.d. (independent and identically distributed) variables drawn from an unknown distribution—to the distribution of their tails [Jorion 2000]. Basically EVT models extreme risks or events that take their value from the tail of the distribution. It is a tool that attempts to provide the best possible estimate of the tail area of the distribution [McNeil 1999].

There are two general types of extreme value models: unconditional and conditional. The unconditional types—considered as the oldest group of EVT models—are called block maxima models. The conditional types—regarded as the modern EVT approaches—are called peaks-over-threshold (POT) models. These two models can be used in estimating *VaR*.

3.1. Block maxima

Following Tsay [2005], suppose there is a collection of n daily returns $\{r_1,\ldots,r_n\}$, where $r_{(1)}$ is the smallest order statistic and $r_{(n)}$ is the maximum order statistic. For the case of *long position*, calculation of *VaR* is concerned with the properties of $r_{(1)}$. Assume that the returns r_t are serially independent with a common CDF $F(x)$ and that the range of the return r_t is $(-\infty,+\infty)$. The CDF of $r_{(1)}$, denoted by $F_{n,1}(x)$, is given by[4]

$$F_{n,1}(x)=1-\left[1-F(x)\right]^n \tag{7}$$

Suppose we find two sequences $\{\beta_n\}$ and $\{\alpha_n\}$, where $\alpha_n > 0$, such that the distribution of $r_{(1^*)} \equiv (r_{(1)} - \beta_n)/\alpha_n$ converges to a non-degenerated distribution as n goes to infinity. The sequence $\{\beta_n\}$ is a *location* series and $\{\alpha_n\}$ is a series of *scaling* factors. Under the independence assumption, the limiting distribution of the normalized minimum $r_{(1^*)}$ is given by

$$F_*(x)=\begin{cases}1-\exp\left[-(1+kx)^{1/k}\right] & \text{if } k\neq 0,\\ 1-\exp\left[-\exp(x)\right] & \text{if } k=0,\end{cases} \tag{8}$$

for $x<-1/k$ if $k<0$ and for $x>-1/k$ if $k>0$, where the subscript * denotes the minimum. The parameter k is called the *shape* parameter, which governs the tail behavior of the limiting distribution. The limiting distribution of (8) is the generalized extreme value (GEV) distribution for the minimum. This distribution has three parameters—k_n, β_n, α_n—that can be estimated by maximum likelihood estimation (MLE), all of which are dependent on n. Now the estimation is done by partitioning first the sample into g non-overlapping subsamples of length n, such that the number of observations T in the sample is, for simplicity, equal to $n*g$. Let $r_{n,i}$ be the minimum of the ith subsample. When n is sufficiently large, $x_{n,i} = (r_{n,i} - \beta_n)/\alpha_n$ should follow an extreme value distribution and the collection of subsample minima $\left\{r_{n,i} \middle| i=1,\dots,g\right\}$ is considered as a sample of g observations from the extreme value distribution. Now given a small probability p, and plugging in the maximum likelihood estimates of GEV parameters, 1-day VaR of a long position in asset with return r_t is[5]

$$VaR=\begin{cases}\hat{\beta}_n-\dfrac{\hat{\alpha}_n}{k_n}\left\{1-\left[-n\ln(1-p)\right]^{\hat{k}_n}\right\} & \text{if } \hat{k}_n\neq 0\\ \hat{\beta}_n+\hat{\alpha}_n\ln\left[-n\ln(1-p)\right] & \text{if } \hat{k}_n=0\end{cases} \tag{9}$$

One of the difficulties in implementing block maxima models is that the procedure requires a lot of data. The collected sample data are even reduced as the block maximum/minimum points are only considered. The choice of subsample size n is also a problem as it is not clearly defined [Tsay 2005]. Block maxima is also an unconditional approach, and hence does not consider the effects of explanatory variables such as volatility [Tsay 2005].

---

[4] See Tsay [2005] for the derivation of equation (7).

[5] Details on the derivation of formula (9) are given by Tsay [2005].

3.2. Peaks-over-threshold

Another type of EVT model is the peaks-over-threshold (POT), which is generally considered to be the most useful for practical applications [McNeil 1999]. In contrast with block maxima model, POT can efficiently use extreme values even when data are limited. It is used to model large observations that exceed a high threshold. Following Tsay [2005] and McNeil [1999], the POT modeling procedures are as follows:[6]

Let $X_1, X_2, \ldots$ be identically distributed random variables with unknown underlying distribution function $F(x) = P\{X_i \leq x\}$. These can be daily returns r_t on financial asset or portfolio. Specify a high threshold u in the sample (e.g., 99th quantile). Suppose that the ith exceedance occurs at day t_i (i.e., $x_t \geq u$). The POT approach then considers the data $(t_i, x_{t_i} - u)$, where $x_{t_i} - u$ is the exceedance over the threshold u, and t_i is the time at which the ith exceedance occurs. Now let $y = x_t - u$. The conditional distribution of y given $x_t \geq u$ or the *excess distribution function* over the high threshold u, $F_u(y)$, is defined as

$$F_u(y) = P\{X - u \leq y | X > u\} = \frac{F(y+u) - F(u)}{1 - F(u)} \tag{10}$$

for $0 \leq y < x_0 - u$ where $x_0 \leq \infty$ is the right endpoint of F. The Pickands-Balkema-de Haan theorem (see Appendix) states that under certain conditions F_u converges to the generalized Pareto distribution (GPD) $G_{\xi,\beta}$,

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - \exp(-x/\beta) & \text{if } \xi = 0, \end{cases} \tag{11}$$

where $\beta > 0$, and where $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ when $\xi < 0$. The parameter ξ is the *shape* parameter that determines the rate at which the tail disappears, and β is the *scaling* parameter.

Setting $x = u + y$ and using (11), equation (10) can be written as

$$F(x) = (1 - F(u)) G_{\xi,\beta}(x - u) + F(u) \tag{12}$$

To estimate (12), $G_{\xi,\beta}(x - u)$ is estimated using MLE, whereas $F(u)$ is estimated by its empirical estimator. The empirical estimator of $F(u)$ is $(n - N_u)/n$, where n is the sample size and N_u is the number of exceedances of the threshold u. Plugging in the maximum likelihood estimates of the GPD parameters and $(n - N_u)/n$, equation (12) is estimated by

---

[6] For ease in presentation, the right-hand side of return distribution is considered in this section. A positive threshold will then be used.

$$\hat{F}(x) = 1 - \frac{N_u}{n}\left(1 + \hat{\xi}\frac{x-u}{\beta}\right)^{-1/\hat{\xi}} \tag{13}$$

Formula (13) is called the tail estimator. Finally, given a small probability p, and define $q = 1 - p$ such that $q > F(u)$, a (1-p)100% confidence level 1-day *VaR* estimate is calculated by inverting the tail estimator formula in (13):

$$VaR_q = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{n}{N_u}(1-q)\right)^{-\hat{\xi}} - 1\right) \tag{14}$$

This type of EVT model is termed as *static* model [McNeil and Frey 2000]. In this model, the POT is applied directly to the raw return data where the underlying distribution F is assumed to be stationary or unconditional. Hence, it does not factor in the volatility of market prices, serial correlation, and volatility clustering.

The other EVT *VaR* model based on POT is called *dynamic* model. In this model, the conditional distribution of F is considered and the volatility of returns is captured. It is dynamic in a sense that it is capable of reacting to fluctuations in market prices, hence enabling the model to capture the current risks. The dynamic EVT approach involves the following procedures [McNeil and Frey 2000]:

Let r_t the return at time t be defined by the following model:

$$r_t = u_t + h_t^{1/2} Z_t \tag{15}$$

where $h_t^{1/2}$ and μ_t are the volatility and expected return, respectively, at time t; Z_t are the noise variables of the process with an identical unknown distribution $F_z(z)$. Using this model and given a small probability p, a 1-day (1-p)100% confidence level *VaR* is estimated by

$$VaR_q^t = \hat{\mu}_{t+1} + \hat{h}_{t+1}^{1/2} VaR(Z)_q \tag{16}$$

(As defined in equation (14), $q = 1 - p$.) Computing VaR_q^t above entails the following two-stage approach:

Suppose at the close of day t consider a time window containing the last n returns $r_{t-n+1}, \ldots, r_t$.

(1) An AR model with GARCH errors is fitted to the historical data by MLE. The estimates of the conditional mean $(\hat{\mu}_{t-n+1}, \ldots, \hat{\mu}_t)$ and standard deviation

series $\hat{h}_{t-n+1}^{1/2},\ldots,\hat{h}_t^{1/2}$ are calculated recursively from the model. If the model is adequate, the residuals are extracted and calculated as

$$\left(z_{t-n+1},\ldots,z_t\right)=\left(\frac{r_{t-n+1}-\hat{\mu}_{t-n+1}}{\hat{h}_{t-n+1}^{1/2}},\ldots,\frac{r_t-\hat{\mu}_t}{\hat{h}_t^{1/2}}\right) \tag{17}$$

The residuals should be i.i.d. if the model is adequate. These residuals can be considered as realizations of the unobserved, independent noise variables $Z_{t-n+1},\ldots,Z_t$. The last step in stage 1 is to calculate the 1-step ahead forecasts of $\hat{\mu}_{t+1}$ and $\hat{h}_t^{1/2}$.

(2) Stage 2 begins by confirming that the residuals have fat-tails or leptokurtic. This can be done by constructing QQ-plot of the residuals against the normal distribution or by computing the *excess kurtosis*. EVT is then applied to the residuals if they have fat-tails. A fixed high threshold u is chosen, and it is assumed that residuals exceeding this threshold have a generalized Pareto distribution (GPD). The GPD parameters are estimated using MLE, and then $VaR(Z)_q$ is calculated using (14). From here, VaR_q^t can be calculated.

In the study of McNeil and Frey [2000], it is shown that the dynamic model is better than static because the former does not just adequately measure losses in heavy-tailed distributions, but also captures volatility clustering, serial autocorrelation, and heteroskedasticity attributes of financial data. Results of their study also show that dynamic *VaR* forecasts exhibit sensitivity to price changes. This means that the model reacts quickly to changing volatility, enabling it to capture the risks in a timely manner.

4. *VaR* model assessment

With the inclusion of *VaR* models within the capital adequacy framework (Basel II), it can be expected soon that Philippine banks will use their *VaR* models in calculating the required minimum regulatory capital to be held against market risks. That is why the choice and evaluation of a *VaR* model is very crucial to a bank and to supervisors like the BSP. So to assess a *VaR* model, banks may look into a model's conservativism, accuracy, and efficiency.

A *VaR* model is considered *conservative* if it systematically produces high estimates of risk relative to other models. A model is *accurate* if its estimates are large enough to cover the true underlying risks. Accuracy can be assessed by analyzing the number of times the *VaR* estimates are lower than the losses and the magnitude of those losses. An *efficient VaR* model provides sufficient conservatism (i.e., accurate and conservative) while at the same time minimizes the capital that must be held [Engel and Gizycki 1999].

4.1. Mean relative bias

One measure of model conservatism is the mean relative bias (MRB) statistic [Hendricks 1996]. It tests whether different *VaR* models produce risk estimates of similar average size. Given T time periods and N *VaR* models, the mean relative bias of model i is calculated as

$$MRB_i = \frac{1}{T}\sum_{t=1}^{T}\frac{VaR_{it} - \overline{VaR_t}}{VaR_t} \tag{18}$$

where

$$\overline{VaR_t} = \frac{1}{N}\sum_{i=1}^{N} VaR_{it}$$

The mean relative bias is measured in percentage terms, so that an MRB value of, say, 0.05 implies that a given *VaR* model is 5 percent higher, on average, than the average estimates of all N *VaR* models.

4.2. Backtesting

Backtesting is a comparison of actual trading results with model-generated risk measures. Its objective is to test the quality and accuracy of the *VaR* model. In backtesting, the bank periodically counts the number of times that the risk measures were smaller than the trading outcome/losses; i.e., the bank counts the number of *VaR exceptions*. In interpreting the results of backtesting, the Basel Committee introduced the three-zone approach found in Table 1.

If the model falls into the green zone, then there is no problem with the quality or accuracy of the model. The yellow zone is an ambiguous zone and the conclusion of model inaccuracy is not definite. If the model falls into the red zone, there is a high probability that the model is inaccurate.

4.3. Likelihood ratio tests

If there is at least one exception, likelihood ratio (LR) tests can be used to assess if the model is inaccurate [Christoffersen 1998]. These tests are the LR test of unconditional coverage (LR_{uc}), LR test of independence (LR_{ind}), and LR test of conditional coverage (LR_{cc}).

Table 1. Backtesting zones based on 260 days covered and 99 percent VaR confidence level

Zone	No. of exceptions	Multiplicative factor (k)
Green	0	3.00
	1	3.00
	2	3.00
	3	3.00
	4	3.00
Yellow	5	3.40
	6	3.50
	7	3.65
	8	3.75
	9	3.85
Red	10 or more	4.00

4.3.1. LR test of unconditional coverage

The LR test of unconditional coverage checks if the proportion of *VaR* exceptions (also known as empirical failure rate) is equal to the prespecified level p (e.g., 1 percent for 99 percent confidence level *VaR*). The empirical failure rate π_1 is estimated by the following statistic:

$$\hat{\pi}_1 = \frac{1}{T}\sum_{t=1}^{T} I\left(r_t < VaR_t\left(p\right)\right) = \frac{T_1}{T} \tag{19}$$

where T is the total number of out-of-sample observations; $I(\bullet)$ is the indicator variable, which is equal to 1 if there is exception and zero otherwise; T_1 is the number of exceptions.

Now to test if the empirical failure rate π_1 is equal to p, a test of hypothesis is employed: $H_o : \pi_1 = p$ vs. $H_a : \pi_1 \neq p$. The LR_{uc} test statistic [Jorion 2000] is given by

$$LR_{uc} = 2 * \log\left[\frac{\left(1-\frac{T_1}{T}\right)^{T-T_1}\left(\frac{T_1}{T}\right)^{T_1}}{\left(1-p\right)^{T-T_1} p^{T_1}}\right] \tag{20}$$

where LR_{uc} is asymptotically distributed as chi-square with 1 degree of freedom. Hence if $LR_{uc} > \chi^2_{p,1}$, we reject the null hypothesis and conclude that the empirical failure rate is not equal to p, otherwise we accept H_o. Accepting H_o, however, does not always imply that the model is correctly specified. The failure rate could be within the prespecified level p but the *VaR* exceptions are clustered or not independent. A model with clustered *VaR* exceptions is a sign of inaccurate model [Christoffersen and Pelletier 2003].

4.3.2. LR *test of independence*

If H_o is not rejected in the LR_{uc} test, the *VaR* model is further assessed using the LR test of independence (LR_{ind}). It checks if the proportion of the clustered *VaR* exceptions is equal to the proportion of the independent *VaR* exceptions. Now consider the following statistics:

$$\pi_0 = \frac{T_{01}}{T_{01}+T_{00}}, \pi_1 = \frac{T_{11}}{T_{11}+T_{10}}, \pi = \frac{T_{01}+T_{11}}{T_{01}+T_{00}+T_{10}+T_{11}}$$

where T_{00} is the number of two consecutive days without *VaR* exception;

T_{10} is the number of days without *VaR* exception that is preceded by a day with *VaR* exception;

T_{11} is the number of two consecutive days with *VaR* exceptions;

T_{01} is the number of days with *VaR* exception that is preceded by day without a *VaR* exception;

π_0 is the proportion of *VaR* exceptions preceded by non-*VaR* exception;

π_1 is the proportion of two consecutive *VaR* exceptions.

Now to use the LR_{ind} test, another test of hypothesis is employed: $H_o : \pi_0 = \pi_1$ vs. $H_a : \pi_0 \neq \pi_1$. The LR_{ind} test statistic [Jorion 2000] is

$$LR_{ind} = 2 * \log\left[\frac{(1-\hat{\pi}_0)^{T_{00}}\, \hat{\pi}_0^{T_{01}}\, (1-\hat{\pi}_1)^{T_{10}}\, \hat{\pi}_1^{T_{11}}}{(1-\hat{\pi}_0)^{T_{00}+T_{10}}\, \hat{\pi}^{T_{01}+T_{11}}}\right] \tag{21}$$

where LR_{ind} is asymptotically distributed as chi-square with 1 degree of freedom. Hence, if $LR_{ind} > \chi^2_{p,1}$, we reject the null hypothesis and conclude that the exceptions are not independent. Accepting H_o, on the other hand, would require to further assess the model if the proportion of independent exceptions (π_0) or clustered exceptions (π_1) is not significantly different from the prespecified failure rate p.

4.3.3. LR *test of conditional coverage*

If the result of the LR_{ind} test shows that the *VaR* exceptions are independent, then there is still a need to check if the proportion of *VaR* exceptions is equal to the failure rate p. This third test is the LR test of conditional coverage (LR_{cc}). It employs the following test of hypothesis: $H_o : \pi_0 = \pi_1 = p$ vs. H_a : at least one of π_0, π_1 is not equal to p. The test statistic is given by Christoffersen [1998]:

$$LR_{cc} = LR_{uc} + LR_{ind} \tag{22}$$

where LR_{cc} is asymptotically distributed as chi-square with 2 degrees of freedom. Therefore if $LR_{ind} > \chi^2_{p,2}$, we reject H_o and conclude that at least one of π's is not equal to p; otherwise, the model is accurate.

4.4. Quadratic loss function

The Basel Committee on Banking Supervision [1996] notes that not only the number of *VaR* exceptions is important, but also the size of exceptions is a matter of regulatory concern. So the quadratic loss function may be used in assessing *VaR* model accuracy. Quadratic loss function does not just include the magnitude of exceptions, but it also penalizes more severely those large exceptions. The formula is given by Lopez [1998]:[7]

$$L_{t+1} = \begin{cases} 1 + (loss_{t+1} - VaR_t)^2 & \text{if } loss_{t+1} > VaR_t \\ 0 & \text{if } loss_{t+1} \leq VaR_t \end{cases}, \tag{23}$$

where L_{t+1} and $loss_{t+1}$ are the loss function value and trading loss, respectively, at time $t+1$, and VaR_t is the *VaR* forecast at time t. Loss function values are averaged across the out-of-sample period. Low average loss function indicates accuracy of a *VaR* model.

4.5. Market risk capital

Market risk capital (MRC) or risk charge is one measure of *VaR* efficiency. It is one of the provisions of Basel II to compute for the MRC when a bank uses IMA for market risk. The general MRC required for any given day t is computed as

$$MRC = Max\left[k\frac{1}{60}\sum_{i=1}^{60} VaR_{t-i}, VaR_{t-1} \right] \tag{24}$$

where k is the supervisory-determined multiplicative factor (see Table 1). To get the total capital adequacy requirements, the credit risk charge connected to the issuer of the financial instrument is added to the market risk charge applied for trading positions. Here, an *accurate VaR* model with low MRC is considered efficient.

---

[7] Formula (23) assumes that the loss and VaR values are both positive.

4.6. Mean relative scaled bias

Another efficiency measure is the mean relative scaled bias (MRSB). Computation of MRSB involves three steps. First, the number of *VaR* exceptions in the out-of-sample period is counted, and the percentage of outcomes covered is determined per model. Second, the *VaR* measures are then multiplied with a constant in order to obtain the desired level of coverage. In this step, each *VaR* model is scaled so that it covers 99 percent of losses (for 99 percent confidence level *VaR*). Finally, the MRB of each scaled *VaR* model is computed as in Section 4.1. The computed MRB in this last step will be the mean relative scaled bias (MRSB) [Hendricks 1996].

The second step requires calculation, on an ex-post basis, of the *multiple to obtain coverage* or the scaling factor. The scaling factor, ψ_i , for each *VaR* model i, is computed so that[8]

$$F_i = T * p \text{ where, } F_i = \sum_{t=1}^{T} \begin{cases} 1 \text{ if } loss_{t+1} > \Psi_i VaR_{it} \\ 0 \text{ if } loss_{t+1} \leq \Psi_i VaR_{it} \end{cases} \quad (25)$$

where T is the sample size, F_i is the expected number of exceptions, and $(1-p)$ is the *VaR* confidence level. So for a sample size of 260 days and *VaR* confidence level of 99 percent, ψ_i is computed such that the total number of exceptions F_i is equal to 2.6.

5. Results

This research uses the secondary market interest rates (MART1) of Philippine sovereign papers from October 1998 to September 2008. These benchmark rates series belong in one of the 12 tenors: 1-month, 3-month, 6-month, 1-year, 2-year, 3-year, 4-year, 5-year, 7-year, 10-year, 20-year, and 25-year. In-sample data cover the period from October 1998 to August 2007 (maximum of 2,263 data points); out-of-sample data cover the period from September 2007 to September 2008 (260 data points). Ten *VaR* models are then used to measure market risk in each tenor: RiskMetrics, AR-GARCH (normal & Student-*t*), AR-EGARCH (normal), AR-GARCH-M (normal & Student-*t*), AR-IGARCH (normal & Student-*t*), static EVT, and dynamic EVT.

After generating the *VaR* forecasts for each model in each tenor, the number exceptions are summarized in Table 2 below. Among the ten models, RiskMetrics has the poorest performance due to relatively high number of exceptions in all tenors. On the other hand, the top-performing models are the AR-IGARCH models (normal & Student-*t*) with zero exception in all 12 tenors, followed

[8] Formula (25) assumes that the loss and VaR values are both positive.

by AR-GARCH (Student-*t*) and AR-GARCH-M (Student-*t*), both having zero exception in 10 out of 12 tenors, and lastly dynamic EVT with zero exception in 9 out of 12 tenors.

Table 3 below summarizes the average MRB of each model per tenor. Two *VaR* models have near zero or lowest *absolute* average MRB, namely, the AR-IGARCH (normal) model with lowest absolute average MRB in seven tenors and the dynamic EVT model with lowest absolute average MRB in four tenors. It can be said that these two models are conservative or have less bias compared with the other *VaR* models.

Summarized in Table 4 is the accuracy ranking of the *VaR* models based on the results of backtesting. With all the top-five models falling into the green zone in all 12 tenors, it can be said that their accuracy is not significantly different from one another.

Table 5 shows the results of likelihood ratio tests. Five models with at least one exception in a particular tenor—AR-GARCH (normal), AR-GARCH (Student-*t*), AR-GARCH-M (normal), AR-GARCH-M (Student-*t*), and dynamic EVT—have passed the LR_{uc} test. This indicates that these *VaR* models can accurately predict losses 99 percent of the time. However, three models—RiskMetrics, AR-EGARCH (normal), and static EVT—failed in at least one of the LR tests. Failure in the LR tests implies that the assumption of 99 percent confidence level is not valid and/or the model is not correctly specified.

The results of the quadratic loss function (see Table 6) are similar to backtesting results. RiskMetrics is still considered least accurate since it has the highest average loss function in 11 tenors. The AR-IGARCH (normal and Student-*t*) models are still considered most accurate because of zero exception, thus resulting in zero percent average loss function in all tenors. Other models considered accurate are the AR-GARCH (Student-*t*) and AR-GARCH-M (Student-*t*), both with zero percent average loss functions in ten tenors, and the dynamic EVT model with zero loss functions in nine tenors.

Table 2. Summary of VaR exceptions and corresponding p-values

	RiskMetrics		*AR-GARCH, normal*		*AR-GARCH, Student-t*		*AR-EGARCH, normal*		*AR-GARCH-M, normal*	
Tenor	exceptions	p-value	exceptions	p-value	exceptions	p-value	exceptions	p-value	exceptions	p-value
1M	7	**0.0166**	3	0.4823	1	0.9267	9	**0.0014**	3	0.4823
3M	8	**0.0051**	3	0.4823	1	0.9267	8	**0.0051**	4	0.2636
6M	7	**0.0166**	4	0.2636	0	1.0000	4	0.2636	4	0.2636
1Y	10	**0.0003**	1	0.9267	0	1.0000	2	0.7342	1	0.9267
2Y	8	**0.0051**	1	0.9267	0	1.0000	5	0.1216	1	0.9267
3Y	6	**0.0482**	3	0.4823	0	1.0000	2	0.7342	3	0.4823
4Y	4	0.2636	1	0.9267	0	1.0000	3	0.4823	1	0.9267
5Y	8	**0.0051**	1	0.9267	0	1.0000	2	0.7342	1	0.9267
7Y	7	**0.0166**	1	0.9267	0	1.0000	3	0.4823	1	0.9267
10Y	8	**0.0051**	1	0.9267	0	1.0000	2	0.7342	1	0.9267
20Y	6	**0.0482**	2	0.7342	0	1.0000	7	**0.0166**	2	0.7342
25Y	6	**0.0482**	6	**0.0482**	0	1.0000	8	**0.0051**	6	**0.0482**

Note: P-values in bold font indicate significance at 5 percent level.

Table 2. Summary of VaR exceptions and corresponding p-values (continued)

	AR-GARCH-M, Student-t		*AR-IGARCH, normal*		*AR-IGARCH, Student-t*		*Static EVT*		*Dynamic EVT*	
Tenor	exceptions	p-value	exceptions	p-value	exceptions	p-value	exceptions	p-value	exceptions	p-value
1M	1	0.9267	0	1.0000	0	1.0000	9	**0.0014**	1	0.9267
3M	1	0.9267	0	1.0000	0	1.0000	7	**0.0166**	1	0.9267
6M	0	1.0000	0	1.0000	0	1.0000	6	**0.0482**	2	0.7342
1Y	0	1.0000	0	1.0000	0	1.0000	2	0.7342	0	1.0000
2Y	0	1.0000	0	1.0000	0	1.0000	2	0.7342	0	1.0000
3Y	0	1.0000	0	1.0000	0	1.0000	2	0.7342	0	1.0000
4Y	0	1.0000	0	1.0000	0	1.0000	2	0.7342	0	1.0000
5Y	0	1.0000	0	1.0000	0	1.0000	1	0.9267	0	1.0000
7Y	0	1.0000	0	1.0000	0	1.0000	1	0.9267	0	1.0000
10Y	0	1.0000	0	1.0000	0	1.0000	2	0.7342	0	1.0000
20Y	0	1.0000	0	1.0000	0	1.0000	4	0.2636	0	1.0000
25Y	0	1.0000	0	1.0000	0	1.0000	4	0.2636	0	1.0000

Note: P-values in bold font indicate significance at 5 percent level.

Table 3. Summary of average MRB of VaR models.

Tenor	*Risk Metrics*	*AR-GARCH, normal*	*AR-GARCH, Student-t*	*AR-EGARCH, normal*	*AR-GARCH-M, normal*	*AR-GARCH-M, Student-t*	*AR-IGARCH, normal*	*AR-IGARCH, Student-t*	*Static EVT*	*Dynamic EVT*
1M	-58 *	-37	21	-54	-38	18	93	107 **	-54	**3**
3M	-63 *	-36	10	-57	-37	9	126 **	98	-53	**3**
6M	-45 *	-18	11	-14	-18	11	21	96 **	-42	**-3**
1Y	-80 *	-40	71	-64	-41	71	**-21**	124 **	-62	42
2Y	-77 *	-49	69	-66	-49	70	**0**	98 **	-58	61
3Y	-69 *	-61	64	-58	-61	55	**17**	95 **	-45	63
4Y	-71 *	-30	53	-60	-30	46	**14**	71 **	-45	51
5Y	-68 *	-34	33	-54	-35	32	**29**	91 **	-37	44
7Y	-66 *	-22	32	-51	-25	30	**8**	103 **	-43	34
10Y	-72 *	-35	50	-60	-37	50	**16**	95 **	-53	46
20Y	-70 *	-34	35	-68	**-34**	35	92 **	67	-61	38
25Y	-66	-71	58	-73 *	-71	41	82	122 **	-63	**41**

Notes: Figures in percent.

* Indicates lowest MRB in a tenor; ** indicates highest MRB in a tenor.

The lowest absolute average MRB in a tenor is set in bold.

Table 4. Accuracy ranking of VaR models based on backtesting

Rank	*Model*	*No. of zero exceptions out of 12 tenors*	*No. of green zones out of 12 tenors*	*No. of yellow zones out of 12 tenors*	*No. of red zones out of 12 tenors*
1	AR-IGARCH, normal	12	12	0	0
1	AR-IGARCH, Student-t	12	12	0	0
3	AR-GARCH, Student-t	10	12	0	0
3	AR-GARCH-M, Student-t	10	12	0	0
5	Dynamic EVT	9	12	0	0
6	AR-GARCH, normal	0	11	1	0
6	AR-GARCH-M, normal	0	11	1	0
8	Static EVT	0	9	3	0
9	AR-EGARCH, normal	0	7	5	0
10	RiskMetrics	0	1	10	1

Table 5. Summary results of likelihood ratio tests

Tenor	*RiskMetrics*			*AR-GARCH, normal*			*AR-GARCH, Student-t*			*AR-EGARCH, normal*			*AR-GARCH-M, normal*		
	LRuc	LRind	LRcc	LRuc	LRind	LRcc	Lruc	LRind	LRcc	LRuc	LRind	LRcc	LRuc	LRind	LRcc
1M	A	R	na	A	-	-	A	-	-	R	na	na	A	-	-
3M	R	na	na	A	-	-	A	-	-	R	na	na	A	-	-
6M	A	-	-	A	-	-	Na	na	na	A	-	-	A	-	-
1Y	R	na	na	A	-	-	na	na	na	A	-	-	A	-	-
2Y	R	na	na	A	-	-	na	na	na	A	-	-	A	-	-
3Y	A	A	A	A	-	-	na	na	na	A	-	-	A	-	-
4Y	A	-	-	A	-	-	na	na	na	A	-	-	A	-	-
5Y	R	na	na	A	-	-	na	na	na	A	-	-	A	-	-
7Y	A	A	A	A	-	-	na	na	na	A	-	-	A	-	-
10Y	R	na	na	A	-	-	na	na	na	A	-	-	A	-	-
20Y	A	-	-	A	-	-	na	na	na	A	-	-	A	-	-
25Y	A	A	A	A	-	-	na	na	na	R	na	na	A	-	-

Table 5. Summary results of likelihood ratio tests (continued)

Tenor	*AR-GARCH-M, Student-t*			*AR-IGARCH, normal*			*AR-IGARCH, Student-t*			*Static EVT*			*Dynamic EVT*		
	Lruc	LRind	LRcc	LRuc	LRind	LRcc	LRuc	LRind	LRcc	LRuc	LRind	LRcc	LRuc	LRind	LRcc
1M	A	-	-	na	na	na	na	na	na	R	na	na	A	-	-
3M	A	-	-	na	na	na	na	na	na	A	-	-	A	-	-
6M	Na	na	na	na	na	na	na	na	na	A	-	-	A	-	-
1Y	Na	na	na	na	na	na	na	na	na	A	-	-	na	na	na
2Y	Na	na	na	na	na	na	na	na	na	A	-	-	na	na	na
3Y	Na	na	na	na	na	na	na	na	na	A	-	-	na	na	na
4Y	Na	na	na	na	na	na	na	na	na	A	-	-	na	na	na
5Y	Na	na	na	na	na	na	na	na	na	A	-	-	na	na	na
7Y	Na	na	na	na	na	na	na	na	na	A	-	-	na	na	na
10Y	Na	na	na	na	na	na	na	na	na	A	-	-	na	na	na
20Y	Na	na	na	na	na	na	na	na	na	A	-	-	na	na	na
25Y	Na	na	na	na	na	na	na	na	na	A	-	-	na	na	na

Note: "A" means acceptance of null hypothesis or passing the LR test, "R" means rejection of the null hypothesis or the model failed the LR test, "na" means that the LR test is not applicable due to zero exception or the previous LR test is rejected, "-" means that the LR test cannot be conducted due to undefined values in the test statistic.

Table 6. Summary of average loss functions of VaR models per tenor

Tenor	*Risk Metrics*	*AR-GARCH, normal*	*AR-GARCH, Student-t*	*AR-EGARCH, normal*	*AR-GARCH-M, normal*	*AR-GARCH-M, Student-t*	*AR-IGARCH, normal*	*AR-IGARCH, Student-t*	*Static EVT*	*Dynamic EVT*
1M	2.69	1.15	0.38	3.46 **	1.15	0.38	0.00 *	0.00 *	3.46	0.38
3M	3.08 **	1.15	0.38	3.08	1.54	0.38	0.00 *	0.00 *	2.69	0.38
6M	2.69 **	1.54	0.00 *	1.54	1.54	0.00 *	0.00 *	0.00 *	2.31	0.77
1Y	3.85 **	0.38	0.00 *	0.77	0.38	0.00 *	0.00 *	0.00 *	0.77	0.00 *
2Y	3.08 **	0.38	0.00 *	1.92	0.38	0.00 *	0.00 *	0.00 *	0.77	0.00 *
3Y	2.31 **	1.15	0.00 *	0.77	1.15	0.00 *	0.00 *	0.00 *	0.77	0.00 *
4Y	1.54 **	0.38	0.00 *	1.15	0.38	0.00 *	0.00 *	0.00 *	0.77	0.00 *
5Y	3.08 **	0.38	0.00 *	0.77	0.38	0.00 *	0.00 *	0.00 *	0.38	0.00 *
7Y	2.69 **	0.38	0.00 *	1.15	0.38	0.00 *	0.00 *	0.00 *	0.38	0.00 *
10Y	3.08 **	0.38	0.00 *	0.77	0.38	0.00 *	0.00 *	0.00 *	0.77	0.00 *
20Y	2.31 **	0.77	0.00 *	2.69	0.77	0.00 *	0.00 *	0.00 *	1.54	0.00 *
25Y	2.31 **	2.31	0.00 *	3.08	2.31	0.00 *	0.00 *	0.00 *	1.54	0.00 *

Notes: Figures in percent.

** Indicates highest average loss function in a tenor; * indicates lowest average loss function in a tenor.

For the MRC figures, Table 7 summarizes the results. Although RiskMetrics has the lowest MRC in 11 tenors, it does not make it an efficient model because of its poor accuracy. The low MRC is due to its low *VaR* forecasts, which understate risk. On the other hand, AR-IGARCH (Student-*t*) model has the highest MRC in ten tenors. Although it is the top-one model in terms of accuracy, it is not considered efficient as it mainly overstates risk, resulting in high capital charges, which is not practical for banks to implement. Among the top-five models based on backtesting, AR-IGARCH (normal) has the lowest MRC in seven tenors (medium- to long-term) while dynamic EVT has the lowest MRC in three tenors (short-term).

Comparing further the MRC of the top-five models, the minimum and maximum MRC in each day in the out-of-sample period are counted (see Table 8). Results show that the models having the most number of days with maximum MRC are the AR-IGARCH (Student-*t*) in ten tenors and the AR-IGARCH (normal) in two tenors. This result confirms that the AR-IGARCH (Student-*t*) model generally overstates risk, resulting in zero exception. On the other hand, the models having the most number of days with minimum MRC are the AR-IGARCH (normal) in seven tenors, dynamic EVT in three tenors, and AR-GARCH-M (Student-*t*) in two tenors. Again, the AR-IGARCH (normal) model is more practical to use in medium- to long-term tenors whereas dynamic EVT model is more practical to use in short-term tenors due to their low MRC.

Finally, MRSB statistics are presented in Table 9. Among the ten *VaR* models, AR-IGARCH (normal) has the most number of lowest *absolute* MRSB (three tenors), followed by AR-GARCH (normal), AR-GARCH-M (Student-*t*), static EVT, and dynamic EVT (two tenors each).

Comparing the two EVT *VaR* models, dynamic EVT beats static EVT based on the following results:

- Dynamic EVT has fewer *VaR* exceptions than static EVT.
- The MRB statistics show that static EVT generally understates market risk while dynamic EVT is more conservative. Static EVT produces relatively lower *VaR* forecasts.
- Backtesting results reveal that dynamic EVT is more accurate than static EVT as the former is in the green zone in 12 tenors while the latter is in the green zone in nine tenors only and in the yellow zone in three tenors. Dynamic EVT passed all LR tests indicating model accuracy while static EVT failed in one LR test indicating that the model is not correctly specified. Also, the average loss function of static EVT is higher than that of dynamic EVT, implying that static EVT is less accurate than dynamic EVT.

Table 7. Summary of average market risk charge (MRC) of VaR models in each tenor

Tenor	Risk Metrics	AR-GARCH, normal	AR-GARCH, Student-t	AR-EGARCH, normal	AR-GARCH-M, normal	AR-GARCH-M, Student-t	AR-IGARCH, normal	AR-IGARCH, Student-t	Static EVT	Dynamic EVT
1M	1.38 *	1.62	3.04	1.61	1.59	2.97	4.70	5.06 **	1.54	2.58 ‡
3M	1.34 *	1.73	2.95	1.57	1.70	2.92	5.90 **	5.20	1.60	2.76 ‡
6M	1.51 *	1.77	2.41	1.87	1.77	2.41	2.52	4.06 **	1.51	2.09 ‡
1Y	0.99 *	2.04	5.75	1.20	2.01	5.74	2.67 ‡	7.51 **	1.28	4.79
2Y	0.85 *	1.47	4.78	1.09	1.47	4.80	2.85 ‡	5.60 **	1.23	4.57
3Y	0.79 *	0.85	3.45	0.91	0.85	3.24	2.46 ‡	4.07 **	1.20	3.42
4Y	0.68 *	1.54	3.36	0.89	1.54	3.19	2.49 ‡	3.73 **	1.24	3.32
5Y	0.82 *	1.30	2.58	0.92	1.28	2.56	2.49 ‡	3.65 **	1.26	2.79
7Y	0.89 *	1.63	2.72	1.04	1.56	2.69	2.21 ‡	4.11 **	1.22	2.76
10Y	0.89 *	1.61	3.65	1.02	1.55	3.65	2.81 ‡	4.70 **	1.21	3.56
20Y	0.84 *	1.66	3.41	1.02	1.66	3.40 ‡	4.80 **	4.19	1.03	3.47
25Y	0.89	0.79	3.64	0.81	0.79 *	3.24	4.15	5.06 **	0.86	3.24 ‡

Notes: Figures in percent.

** Indicates highest average MRC in a tenor; * indicates lowest average MRC in a tenor; ‡ indicates lowest average MRC in a tenor considering only the top-five models.

Table 8. Comparison of the top-five models by counting the number of days out of 260 with minimum/maximum MRC

	AR-IGARCH, normal		*AR-IGARCH, Student-t*		*AR-GARCH, Student-t*		*AR-GARCH-M, Student-t*		*Dynamic EVT*	
Tenor	*Min*	*Max*	*Min*	*Max*	*Min*	*Max*	*Min*	*Max*	*Min*	*Max*
1M	26	0	0	260	0	0	0	0	234	0
3M	30	205	0	55	0	0	0	0	230	0
6M	81	0	0	260	0	0	0	0	179	0
1Y	260	0	0	252	0	6	0	2	0	0
2Y	260	0	0	260	0	0	0	0	0	0
3Y	260	0	0	246	0	0	0	0	0	14
4Y	260	0	0	230	0	0	0	0	0	30
5Y	165	0	0	208	0	0	95	0	0	52
7Y	241	0	0	241	0	0	19	0	0	19
10Y	260	0	0	260	0	0	0	0	0	0
20Y	30	225	0	0	0	0	188	0	42	35
25Y	55	0	0	260	0	0	137	0	68	0

Table 9. Summary of average mean relative scaled bias (MRSB) of VaR models in each tenor

Tenor	*Risk Metrics*	*AR-GARCH, normal*	*AR-GARCH, Student-t*	*AR-EGARCH, normal*	*AR-GARCH-M, normal*	*AR-GARCH-M, Student-t*	*AR-IGARCH, normal*	*AR-IGARCH, Student-t*	*Static EVT*	*Dynamic EVT*
1M	37 **	-9	-6	-5	-10	-8	-13	28	-4 *	-9
3M	36	-15	-7	-12	-13	-8	-3 *	45 **	-16	-7
6M	-18 **	0 *	6	10	1	6	-10	-6	7	3
1Y	6	-7	6	3	-9	5	-2 *	4	7	-13 **
2Y	11	5	-6	17 **	5	-5 *	-16	-16	14	-10
3Y	30 **	-5	-2	-2	-5	-8	-2 *	-6	3	-3
4Y	25 **	13	-11	-4	13	-15	-7	0	0 *	-13
5Y	23 **	-1	-8	4	-3	-8	-9	-11	14	-1 *
7Y	16	-3 *	-4	16	-7	-5	-10	-15	17 **	-3
10Y	-14	29	-16	-4	32 **	-16	-9	9	-7	-3 *
20Y	4	1	-7	1 *	1	-8	14 **	-1	1	-6
25Y	16 **	-12	8	-12	-12	-4 *	7	9	-8	9

Notes: Figures in percent.

** Indicates highest absolute average MRSB in a tenor; * indicates lowest absolute average MRSB in a tenor.

- Although static EVT has lower average MRC than dynamic EVT, static EVT is not necessarily more efficient as its low MRC is mainly due to its low *VaR* forecasts. Comparing the average absolute MRSB, dynamic EVT has relatively lower average MRSB than static EVT. Hence, dynamic EVT is more efficient than static EVT.

6. Conclusion

Less than a year from now, the BSP hopes to achieve full implementation of Basel II in the Philippines. With this initiative, it can be expected that local universal and commercial banks (UBs/KBs) will start developing internal risk measurement models, not just to comply, but also to reduce risk-based capital charges. This study considers two value-at-risk models using the POT approach of EVT, which can serve as potential internal market risk models: static and dynamic. These models are applied to the benchmark rates (secondary market rates) of the 12 tenors of government securities. The findings are summarized below:

(1) The static EVT model the is a straightforward application of EVT to the delta yield series while the dynamic EVT is the application of EVT to the residuals of the fitted AR-GARCH model of the delta yield series. Results show that the dynamic EVT model is more efficient than static EVT, i.e., dynamic EVT predicts losses more effectively and has less *VaR* exceptions than static EVT. It is also a better EVT model because it can cover some risk measurement issues such as non-normality of distribution, heavy-tails, autocorrelation, nonconstant variance, and volatility clustering.

(2) AR-IGARCH (normal) and dynamic EVT are the most conservative *VaR* models based on low mean relative bias statistics. Specifically, AR-IGARCH (normal) is most conservative in medium- to long-term tenors while dynamic EVT is most conservative in short-term tenors. Both are able to capture risks by generating *VaR* figures that are high enough to estimate the losses.

(3) The accuracy of the dynamic EVT model is comparable to the AR-GARCH-type *VaR* models, i.e., dynamic EVT's accuracy is not significantly different from the econometric *VaR* models based on the number of exceptions and backtesting results. Dynamic EVT falls into the green zone of backtesting in all 12 tenors, indicating high probability of model accuracy. Results of likelihood ratio tests and loss functions also indicate that dynamic EVT is one of the most accurate *VaR* models.

(4) Based on the mean relative scaled bias statistics, dynamic EVT is efficient as it is both conservative and accurate. Furthermore, among the top-performing *VaR* models, dynamic EVT is one of the most efficient for having low market-risk charges specifically in short-term tenors. Hence it is practical to use and has a potential to lower the required market risk capital.

With these results, it can be concluded that the study has successfully assessed extreme value theory's potential as a risk measure. EVT-based *VaR* model like dynamic EVT is able to address some known issues in risk measurement. It can also measure risk conservatively, accurately, and efficiently, therefore making it a candidate internal market risk-measurement model for banks.

References

Bangko Sentral ng Pilipinas (BSP) [2006a] "Circular No. 538: Revised Risk-Based Capital Adequacy Framework", BSP Regulations. BSP.

Bangko Sentral ng Pilipinas (BSP) [2006b] "Circular No. 544: Guidelines on Market Risk Management", BSP Regulations. BSP.

Bangko Sentral ng Pilipinas. Website, www.bsp.gov.ph.

Bank for International Settlements. Website, www.bis.org.

Basel Committee on Banking Supervision [1996] *Supervisory framework for the use of "backtesting" in conjunction with the internal model approach to market risk capital requirements*. Bank for International Settlements.

Basel Committee on Banking Supervision [2004] *International convergence of capital measurement and capital standards*. Bank for International Settlements.

Bollerslev, T. [1986] "Generalized autoregressive conditional heteroscedasticity", *Journal of Econometrics* **31**: 307-327.

Christoffersen, P. [1998] "Evaluating interval forecast", *International Economic Review* **3**(4): 841-862.

Christoffersen, P. and D. Pelletier [2003] "Backtesting value-at-risk: a duration-based approach", CIRANO Scientific Series.

Danielsson, J. and C. G. de Vries [2000] "Value-at-risk and extreme returns", *Annales D'Economie et de Statistique* **60**: 239-270.

Embrechts, P., S. Resnick, and G. Samorodnitsky [1999] "Extreme value theory as a risk management tool", *North American Actuarial Journal* **3**: 30-41.

Engel, J. and M. Gizycki [1999] "Conservatism, accuracy and efficiency: comparing value-at-risk models", Australian Prudential Regulation Authority, Reserve Bank of Australia, Working Paper 2.

Engle, R. F. [1982] "Autoregressive conditional heteroscedastic models with estimates of the variance of United Kingdom inflation", *Econometrica* **50**: 987-1007.

Gencay, R., F. Selcuk, and A. Ulugulyagci [2003] "High volatility, thick tails and extreme value theory in value-at-risk estimation", *Elsevier, Insurance: Mathematics & Economics* **33**: 337-356.

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

Jorion, P. [2000] *Value at risk: the new benchmark for managing financial risk*. USA: McGraw-Hill.

Lopez, J. [1998] "Methods for evaluating value-at-risk estimates", *Federal Reserve Bank of New York, Economic Policy Review*, October, Research Paper No. 9802.

McNeil, A. J. [1998] "Calculating quantile risk measures for financial time series using extreme value theory", Department of Mathematics, ETH, Swiss Federal Technical University, ETH E-Collection, http://e-collection.ethbib.ethz.ch/.

McNeil, A. J. [1999] "Extreme value theory for risk managers" in: *Internal modelling and CAD II*. Risk Books: 93-113.

McNeil, A. J. and R. Frey [2000] "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach", *Journal of Empirical Finance* 7: 271-300.

Money Market Association of the Philippines. Website, www.mart.com.ph.

Morgan J. P. [1996] *RiskMetrics technical document*. Fourth edition. New York: J.P. Morgan.

Philippine Dealing & Exchange Corp. Website, www.pdex.com.ph.

Professional Risk Manager's International Association. Website, www.prmia.org.

Tsay, R. [2005] *Analysis of financial time series*. Second edition. John Wiley & Sons.

APPENDIX

The Pickands-Balkema-de Haan Theorem: *For a large class of underlying distributions we can find a function $\beta(u)$ such that*

$$\lim_{u \to x_0} \sup_{0 \le y < x_{0-u}} \left| F_u(y) - G_{\xi,\beta(u)}(y) \right| = 0$$

The important result of the theorem is that, for a large class of underlying distributions F, the excess distribution function F_u converges to the generalized Pareto as the threshold u is progressively increased. The class of underlying distributions for which the theorem holds comprises essentially all common distributions of statistics [McNeil and Frey 2000]. These distributions can be grouped according to the value of the parameter ξ in the limiting GPD approximation to the excess distribution. When $\xi > 0$, then the underlying F corresponds to heavy-tailed distributions whose tails decay like power functions such as the Pareto, Student-*t*, Cauchy, Burr, loggamma, and Fréchet distributions. If $\xi = 0$, then F corresponds to distributions whose tails decay exponentially such as the normal, exponential, gamma, and lognormal. The case $\xi < 0$ is the group of short-tailed distributions with finite right endpoint, like the uniform and beta distributions.