VaR Prediction for Emerging Stock Markets: GARCH Filtered Skewed t Distribution and GARCH Filtered EVT Method

Ibrahim Ergen*
Supervision Regulation and Credit, Policy Analysis Unit
Federal Reserve Bank of Richmond
Baltimore, MD

January, 2010

Abstract

Using a long time series of stock index data from twelve emerging markets, this study compares the performances of ten different market risk models by predicting one day ahead Value-at-Risk and backtesting these predictions. In addition to simple benchmark models and well known models from the existing literature, such as the GARCH-EVT model of McNeil and Frey (2000), a new two step methodology is proposed. This method involves estimation of a GARCH model for returns with quasi maximum likelihood estimation in the first step and subsequent modeling of filtered returns using Azzalini and Capitanio (2003) type skewed t distribution in the second step (GARCH-St model). Previous studies employing GARCH models with skewed t distributed errors preferred to estimate all model parameters jointly in a single step using maximum likelihood estimation. However, this

*The author thanks Jordan Nott for excellent research assistance. The author also thanks Edward Prescot, Robert Carpenter and other participants of applied research seminars in Federal Reserve Bank of Richmond for their valuable comments. The views expressed in this paper are those of the author, and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

approach produces biased parameter estimates if the model is misspecified. Also, it is not directly comparable to the GARCH-EVT model. Therefore, two-step estimation is preferred in this study, which ensures unbiased parameter estimates for the GARCH process. Also, the model becomes directly comparable to GARCH-EVT model since their first steps are exactly the same. The classical single step joint estimation is also implemented and it is shown that the two-step methodology provides better backtesting performance. Overall, depending on the choice of estimation window size, either the GARCH-EVT model or the GARCH-St model outperforms all other models.

1 Introduction

Financial markets have frequently experienced very high volatility and extreme price movements in the last two decades. Both the introduction of VaR as a market risk measure and the use of models that incorporate fat tails in its estimation were motivated by this high volatility and extreme market movements. Economically, VaR provides a single number that the loss of a portfolio cannot exceed at a certain confidence level within a specified time. Statistically, it is just a high quantile of the loss distribution.

VaR became the prominent market risk measure in the financial industry mostly due to the amendment made to Basel I guidelines for banking supervision and regulation in 1996. Until this amendment, banks were required to hold regulatory capital only against their credit risk. Additional regulatory capital against market risk was added with this amendment. Moreover, this additional capital has to cover the losses of a bank's portfolio 99 percent of the time. More specifically, the Basel Committee suggested the following formula

$$MRC = k \times max(VaR_{t,0.99}, \frac{1}{60}\sum_{i=1}^{60} VaR_{t-i,0.99}) \tag{1}$$

for the calculation of market risk capital (MRC) where k is a constant to be chosen by the national supervision agencies based on the backtesting performance of bank models. The Basel Committee also allowed the financial institutions to choose their market risk models, subject to review of regulators. Subsequently, the choice of a market risk model with high accuracy in out-of-sample VaR prediction became an important subject of study both for banks and regulators.

A model that underestimates the VaR results in less capital holdings than required, which in turn increases the insolvency risk in case of an extreme loss. On the other hand, a model that overestimates the VaR causes more capital holdings than required which creates inefficiency in the banking system.

This study compares the backtesting performance of ten different market risk models using data from twelve emerging market stock indices. Accurate measurement of market risk for stock positions requires modeling of the stylized facts observed in financial time series including fat tails, skewness, and dynamic volatility. Fat tail modeling is particularly important since VaR is a measure of tail risk. It was a known fact that asset return distributions had fat tails during the 1960s as shown by Mandelbrot (1963) and Fama (1965). However, with the introduction of mathematical finance literature in the 1970s, this fact was ignored for a long time, and most risk management methodologies were developed based on the assumption of geometric Brownian motion (GBM) for asset prices. Upon the observation of catastrophic market events during the 1990s, the weaknesses of these models were uncovered. Awareness of rare but devastating market events led more researchers to point out the need to go back and study the implications of fat tails[1].

Basically, GBM implies normally distributed logarithmic returns, but empirical data show that the probability of extreme losses is much larger than implied by a normal distribution. Also, the normal distribution cannot capture the skewness in asset returns. As a solution, historical simulation that uses the empirical distribution function was suggested. Although it can account for the skewness, it still cannot capture the fat tails fully because it assigns zero probability to losses larger than those observed in the sample.

The most successful methods in modeling fat tails and skewness together have been the Extreme Value Theory (EVT) methods and the skewed t distributions. Embrechts (2000), McNeil and Frey (2000), and Gencay and Selcuk (2004), among others, showed that EVT methods fit the tails of heavy-tailed financial time series better than more conventional distributional approaches. Gencay et. al, (2003) provides an excellent overview of the EVT and its practical use in producing quantile estimates. Technically, EVT is a limit law for extremes just like central limit theory is a limit law for the mean. Using these laws,

[1]See for example Longin (2000), Gencay and Selcuk (2004), McNeil (1999), Bali (2003).

it is possible to find limiting distributions to model only the tails of the sample instead of imposing a distribution to the entire sample. Also, since EVT methods are implemented separately to the two tails, the skewness is implicitly accounted for.

In order to model dynamic volatility and fat tails together, McNeil and Frey (2000) developed a two-step procedure for VaR estimation known as the GARCH-EVT model. In the first step, a GARCH model is fit to the return data. In the second step, the EVT method is applied to the implied residuals extracted from this fit. GARCH models are very strong in incorporating the dynamic nature of volatility and its persistence in a parsimonious way, contributing to the accuracy of VaR measurement. If dynamic volatility is not accounted for, VaR predictions would be very loose during tranquil periods and they would be unnecessarily conservative during periods of turmoil. In a comprehensive review of VaR prediction methods, Kuester, Paolella and Mittnick (2006) found that the GARCH-EVT model performs best based on backtesting analysis.

The skewed t distribution is another good candidate for modeling skewness and fat tails simultaneously. There are two basic methods to generate skewed t distributions from a baseline symmetric student's t distribution. The first method is to split a baseline symmetric t distribution into semi-halves around its mean and then scale them differently with scaling factors based on indicator functions. Skewed t distributions of Hansen (1994), Fernandez and Steel (1998), and Paolella (1997) are among this first class. They appeared in several comparative VaR prediction studies including Kuester, Paolella and Mittnick (2006) and Giot and Laurent (2004). The second method involves perturbation of a symmetric student's t density as in Azzalini and Capitanio (2003). In this case, instead of indicator functions, the cumulative distribution function of the baseline symmetric t distribution is used to generate scaling factors. This approach has the potential to make the generated skewed t distribution smoother and more flexible for financial applications. Therefore, I will use the skewed t distribution of Azzalini and Capitanio (2003) in this study. Also, this distribution is not used in large scale comparative VaR prediction studies before.

The new model suggested in this study (GARCH-St) combines the GARCH model with the skewed t distribution of Azzalini and Capitanio (2003) in a two step methodology following the same spirit of the GARCH-EVT model employed by McNeil and Frey (2000). In

the first step, the GARCH model is estimated with quasi maximum likelihood method and in the second step a skewed t distribution is fit to the implied residuals from the GARCH fit. In fact, it is possible to estimate the GARCH model parameters and the skewed t distribution parameters jointly in a single step. Previous studies using GARCH models with different forms of skewed t distributions preferred this classical approach[2]. However, the joint estimation approach puts a lot of structure on the data generating process; the model becomes very sensitive to misspecification and loses its flexibility. In particular, if the residuals of GARCH process are not really skewed t distributed, then the parameter estimates would be biased. On the other hand, it is theoretically shown that the quasi maximum likelihood estimation of a GARCH model which relies on normal distribution assumption for the residuals is shown to produce unbiased and consistent parameter estimates[3]. Also, the GARCH-St model with joint estimation of all parameters is not directly comparable to the GARCH-EVT model since the latter adopts a two step procedure with quasi maximum likelihood estimation. On the other hand, GARCH-St model with a two-step estimation method is directly comparable to the GARCH-EVT model, because with this approach the first steps of both models are the same. Any difference in performance is entirely attributable to the choice of the EVT method or skewed t distribution in the second step. Therefore, two step estimation of the GARCH-St model is preferred in this study. The classical single step joint estimation method is also implemented and shown to underperform compared to the two-step methodology.

There are several contributions of this study to the existing literature on VaR prediction. Azzalini and Capitanio's (2003) skewed t distribution is used in a large scale VaR prediction study for the first time. In addition to simple benchmark models and well known models from the existing literature, such as the GARCH-EVT model of McNeil and Frey (2000), a new model is proposed that involves volatility modeling via a GARCH process and subsequent modeling of the filtered returns with Azzalini and Capitanio's (2003) skewed t distribution. The stock index data covers a comprehensive set of emerging markets over long time including the global financial meltdown of 2008-2010. For each of the 12 emerging countries the data extends from June 30, 1995 to March 18, 2011. Finally, the models'

[2]See Kuester et. al. (2006), Giot and Laurent (2004) and Mittnick and Paolella (2000).
[3]See Gourieroux (1997).

backtesting performances of models are evaluated by scoring them across all countries and several quantiles without going into country level details. Models are chosen in a systematic way so that these scores enable attribution analysis regarding the relative importance of the skewness, fat tail and dynamic volatility modeling.

The remaining sections of the paper are organized as follows. In section 2, the statistical analysis of the data is presented, and credibility of the geometric Brownian motion assumption is determined by Jarque Bera normality tests and Ljung Box serial independence tests. In section 3, out of sample prediction methodology is presented. Some background knowledge on extreme value theory and skewed t distribution is provided and all market risk models used in VaR prediction are reviewed. In section 4, the performance of models in VaR prediction are compared by backtesting methods suggested by Christoffersen (1998). In section 5, conclusions are discussed.

2 Data and Statistical Analysis

Standard & Poors and International Financial Corporation Investable (S&P/IFCI) equity price index data with daily frequency are obtained from Bloomberg for 12 emerging markets: Turkey, Malaysia, Indonesia, Philippines, Korea, China, Taiwan, Thailand, Mexico, Brazil, Chile, and Peru. Stock index data for all countries are dollar denominated prices. The S&P/IFCI indices are market capitalization weighted averages of only those equities that can be traded by international investors. The dataset runs from June 30, 1995[4] to March 18, 2011 for all countries. This time span covers the period of the Asian financial crisis in 1997, the Russian default in 1998, the Turkish banking crisis in 2001, the Brazilian crisis of 2002 and the global financial meltdown of 2008-2010.

The daily percentage logarithmic returns are calculated as

$$r_{t,i} = 100 \ \log(\frac{S_{t,i}}{S_{t-1,i}}), \tag{2}$$

where $S_{t,i}$ is the level of the equity price index at the end of the day t for country i.

[4]June 30, 1995 is the beginning of SP/IFCI series when IFC started to collect and report stock price index data from emerging markets.

Working with logarithmic returns is a standard approach in the financial literature, because they are normally distributed and serially independent under the geometric Brownian motion (GBM) assumption for the asset price. Most risk management methodologies that are widely used by the financial industry relied on the GBM assumption for a long time. In this section, the logarithmic returns are analyzed to check the plausibility of normality and serial independence assumptions. For ease of exposition, the logarithmic returns and logarithmic losses are referred to simply as "returns" and "losses" throughout the paper.

The results of Jarque Bera normality tests and Ljung Box serial independence tests are presented in table 1 together with other basic statistics. Some countries exhibit negative skewness while others display positive skewness. The fat tail phenomenon is evident from the excess kurtosis statistics. The loss distributions for all countries exhibit higher kurtosis than implied by a normal distribution[5]. Also, the Jarque-Bera test statistic (JBTS) and associated probability values inside parentheses indicate that the null hypothesis of normality is rejected for all countries at all reasonable significance levels. Finally, the Ljung-Box test is administered at the first lag, for both raw (LBQ(1) column) and squared returns(sqr.LBQ(1) column). The results indicate that there is strong autocorrelation, especially for the squared returns, and the serial independence hypothesis is rejected for all countries.

Numerical data analysis shows that alternative tools are needed other than the conventional risk models assuming normality and serial independence. To address these issues, EVT method, skewed t distribution, and GARCH model are used in this study. Section 3 summarizes out-of-sample forecasting methodology and all the models used for VaR prediction.

3 VaR Estimation

VaR is a high quantile of the loss distribution. More formally,

$$VaR_\alpha = inf\{x \in \Re : F_X(x) \geq \alpha\} = F^{\leftarrow}(\alpha), \tag{3}$$

[5]The kurtosis statistics reported in table 1 are the excess kurtosis over three, the kurtosis for a normal distribution.

	mean	median	Stdev	Skewness	Kurtosis	JBTS	LBQ(1)	sqr.LBQ(1)
Turkey	0.036	0	3.149	−0.06	6.12	6409 (0)	6.359 (0.012)	326.248 (0)
Indonesia	0.008	0.025	2.875	−0.95	24.838	106102 (0)	88.153 (0)	186.809 (0)
Malaysia	0	0	1.687	0.682	34.52	204055 (0)	51.563 (0)	734.108 (0)
Philippines	−0.011	0	1.723	0.47	12.712	27784 (0)	132.853 (0)	123.036 (0)
Korea	0.017	0.01	2.527	0.135	12.784	27960 (0)	35.096 (0)	67.046 (0)
China	0.02	0.017	2.088	−0.113	6.023	6214 (0)	38.198 (0)	278.553 (0)
Taiwan	0.006	0	1.735	−0.056	2.515	1084 (0)	6.956 (0.008)	37.437 (0)
Thailand	−0.015	0	2.096	0.169	8.085	11199 (0)	53.513 (0)	338.064 (0)
Brazil	0.048	0.101	2.35	−0.308	6.878	8155 (0)	35.158 (0)	161.881 (0)
Chile	0.021	0.008	1.257	−0.239	11.001	20736 (0)	105.18 (0)	268.828 (0)
Mexico	0.045	0.064	1.828	0.108	9.994	17088 (0)	46.63 (0)	205.001 (0)
Peru	0.055	0.04	1.586	−0.262	6.504	7281 (0)	30.269 (0)	98.324 (0)

Table 1: Summary Statistics for S&P/IFCI Returns

where X denotes the loss, F_X is the distribution function of losses, and α is the quantile at which VaR is calculated. $F^{\leftarrow}$ is known as the generalized inverse of F_X, or the quantile function associated with F_X. If F_X is a monotonically increasing function, then $F^{\leftarrow}$ reduces to the regular inverse function.

Loss data are calculated by negating the returns given in (2). For out-of-sample VaR prediction, sliding samples of 500 and 1000 observations are used[6]. For example, when the window size is 500 observations, only the loss observations $\{\ell_{t-1}, \ell_{t-2}, ...\ell_{t-500}\}$ are used in order

[6]In GARCH literature, at least 500 observations are suggested in order to have stable parameter estimates. I choose to follow this minimum requirement in order to have a longer backtesting period; I check the results' robustness by using the 1000 observations window size as well.

to make a prediction for day t where $t \in \{501, 502, ..., T\}$. In the sample period used in this study, there are $T = 4,099$ daily observations for each country. This results in an out-of-sample test period of 3,599 (3,099) days when a window with 500 (1000) observations is used. VaR predictions are made at the 0.95^{th}, 0.975^{th}, 0.99^{th} and 0.995^{th} quantiles. In what follows, the methods used for VaR estimation are presented briefly.

3.1 Fully Parametric Static Models

With static models, the loss data is assumed to be coming from a fully parametric distribution, i.e, $\ell_t \sim iid\ F(\theta)$. The parameters θ specifying the loss distribution are constant over time, hence these methods are referred as static models. The only source of change in distribution parameters is the rolling window forecasting approach.

Normal Distribution (The Benchmark): In this model, losses are assumed to be normally distributed; that is, $\ell_t \sim iid\ N(\mu, \sigma^2)$. Using the loss data in the sliding window $\{\ell_{t-1}, \ell_{t-2}, ..., \ell_{t-500}\}$, MLE estimators of μ and σ are calculated and VaR_α is estimated by $VaR_\alpha = \hat{\mu} + \hat{\sigma}\ \Phi^{-1}(\alpha)$ where Φ is the standard normal cdf function. Normal distribution is included as the benchmark model.

Student's *t* Distribution Model: In this model, the losses are assumed to be student's t distributed; that is, $\ell_t \sim iid\ t(\mu, \sigma^2, \nu)$. The student's t distribution can account for fat tails. Therefore, any increase in the precision with respect to the benchmark model can be attributed to fat-tail modeling. The parameters $\{\mu, \sigma, \nu\}$ are estimated by MLE methods, and VaR_α is calculated by $VaR_\alpha = \hat{\mu} + \hat{\sigma}\ T_{\hat{\nu}}^{-1}(\alpha)$, where $T_{\hat{\nu}}$ is the cdf function of standard t-distribution with $\hat{\nu}$ degrees of freedom.

Skewed Normal Distribution Model: In this model, the losses are assumed to be coming from the skewed normal distribution of Azzalini (1985); that is, $\ell_t \sim iid\ Sn(\xi, \delta, \lambda)$ where ξ is the location, δ is the scale and λ is the skewness parameter. The density function is given by

$$f_x(x) = 2\frac{1}{\delta}\phi\left(\frac{x-\xi}{\delta}\right)\Phi\left(\lambda(\frac{x-\xi}{\delta})\right) \tag{4}$$

where ϕ and Φ are the standard normal density and cdf functions respectively. If $\lambda > 0$, the distribution is right skewed. Otherwise it is left skewed. As it increases, the skewness also increases and when $\lambda = 0$ the distribution reduces to symmetric normal distribution.

The skewed normal model can account only for the skewness of loss distribution. Therefore, any increase in precision compared to the benchmark model can be attributed to skewness modeling. The empirical procedure is to estimate $\{\xi, \delta, \lambda\}$ with MLE methods and calculate the VaR with numerical methods as $VaR_\alpha = F^{-1}_{\hat{\xi},\hat{\delta},\hat{\lambda}}(\alpha)$, where $F_{\hat{\xi},\hat{\delta},\hat{\lambda}}$ is the cdf function of the skewed normal distribution associated with the estimated MLE parameters.

Skewed *t* Distribution Model: In this model, the losses are assumed to be coming from the skewed t-distribution of Azzalini and Capitanio (2003); that is, $\ell_t \sim iid\ St(\xi, \delta, \lambda, \nu)$ where ξ is the location, δ is the scale, λ is the skewness and ν is the shape parameter, i.e. degree of freedom. The density function is given by

$$f_x(x) = 2\ \frac{1}{\delta}\ t_\nu\left(\frac{x-\xi}{\delta}\right)\ T_{\nu+1}\left(\lambda\left(\frac{x-\xi}{\delta}\right)\sqrt{\frac{\nu+1}{(\frac{x-\xi}{\delta})^2+\nu}}\ \right) \tag{5}$$

where t_ν and $T_{\nu+1}$ are the density and cdf functions of standard symmetric student's t distributions with ν and $\nu+1$ degrees of freedom respectively. When $\lambda = 0$, the above density reduces to the density of a symmetric student's t distribution.

This model can account for both skewness and the fat tails of loss distributions. Therefore, any gain in precision with respect to student's t distribution can be attributed to skewness modeling on top of the fat-tail modeling. On the other hand, any gain in precision with respect to skewed normal can be attributed to fat-tail modeling on top of the skewness modeling. The empirical procedure is to estimate the parameters $\{\xi, \delta, \lambda, \nu\}$ by MLE methods and calculate the VaR by numerical methods as $VaR_\alpha = F^{-1}_{\hat{\xi},\hat{\delta},\hat{\lambda},\hat{\nu}}(\alpha)$, where $F_{\hat{\xi},\hat{\delta},\hat{\lambda},\hat{\nu}}$ is the cdf function of the skewed t distribution associated with the estimated MLE parameters.

3.2 EVT Model

EVT is a strong method to study the tail behavior of loss distributions. Instead of imposing a single distribution on the entire sample, threshold exceedances methodology models only the tail region of the distribution. More specifically, it is a method of estimating the distribution of exceedances above a high threshold. Let X be the loss variable, F_X its distribution function, u a high threshold, and x_o the

right endpoint of the support of X[7]. Then, the distribution of exceedances is defined as

$$F_u(y) = \Pr(X - u \leq y | X > u) \qquad \text{for} \qquad 0 < y < x_o - u \ . \tag{6}$$

The following result on the limiting distribution of exceedances $F_u(y)$ is the key point in EVT.

$$\lim_{u \to x_o} \sup_{0 \leq y \leq x_o} |F_u(y) - G_{\xi,\beta(u)}(y)| = 0 \ , \tag{7}$$

where $G_{\xi,\beta(u)}$ is the distribution function for the generalized pareto distribution (GPD) and is given by:

$$G_{\xi,\beta}(y) = \begin{cases} 1 - (1 + \frac{\xi y}{\beta})^{\frac{-1}{\xi}} & \text{if } \xi \neq 0, \\ 1 - \exp\left(\frac{-y}{\beta}\right) & \text{if } \xi = 0 \end{cases} \tag{8}$$

$$\text{where } \beta > 0 \text{ and } (1 + \frac{\xi y}{\beta}) > 0.$$

where β is the scale and ξ is the shape parameter of GPD respectively. The limit expression (7) basically reveals that the distribution of exceedances uniformly converges to GPD as the threshold converges to the right end point of the support of X regardless of what the original distribution of X was. This limit result can be exploited and it can be assumed that the exceedances over a high threshold are distributed as GPD.

In order to use the theory in practice, a high threshold is chosen, and the observations that are larger than the threshold are filtered from the data. Let N_u observations exist over the threshold u, and they are labeled as $X_1, X_2, ...X_{N_u}$. Then, the exceedances are calculated as $Y_j = X_j - u$, and GPD is fit to this exceedance data by maximizing the following likelihood function

$$LogL(\xi, \beta, Y) = -N_u \log(\beta) - (1 + \frac{1}{\beta}) \sum_{j=1}^{j=N_u} \log(1 + \xi \frac{Y_j}{\beta}) \ ,$$

subject to the parameter constraints $\beta > 0$ and $(1 + \xi \frac{Y_j}{\beta}) > 0$.

The choice of the threshold is an important step in this procedure. The assumption of GPD holds in the limit as the threshold approaches

[7] If there is no finite right endpoint for the support of X, then $x_o = \infty$, which is the case for most distributions of financial data.

infinity and therefore, a low threshold results in bias. On the other hand, a very high threshold results in very few observations exceeding the threshold which results in high standard errors for parameter estimates. Since it is not practical to decide on an appropriate threshold for each of the 3,599 day-ahead out-of-sample forecasts, the $95th$ quantile of the sample data is used as a high threshold. Sensitivity to threshold is checked by doing the analysis again using the $90th$ quantile as the threshold and results are found robust[8].

MLE estimates of the GPD fit can be used to calculate VaR at higher quantiles than the threshold. Note that for a loss observation greater than the threshold, $x > u$, we have

$$\begin{aligned}
\Pr(\ell > x) &= \Pr(\ell > u)\ \Pr(\ell > x | \ell > u) \\
&= \Pr(\ell > u)\ \Pr(\ell - u > x - u | \ell > u) \\
&= (1 - Pr(\ell \leq u))\ (1 - \Pr(\ell - u \leq x - u | \ell \geq u)) \\
&= (1 - F_\ell(u))\ (1 - F_u(x - u)).
\end{aligned}$$

By exploiting the limit expression (7), F_u can be assumed as a GPD distribution function. Then,

$$\Pr(\ell > x) = (1 - F_\ell(u))\ \Big(1 + \hat{\xi}\frac{x - u}{\hat{\beta}}\Big)^{-1/\hat{\xi}}\ .$$

Also, if $x = VaR_\alpha$, $Pr(\ell > x) = 1 - \alpha$ by definition. Then,

$$1 - \alpha = (1 - F_\ell(u))\ \Big(1 + \hat{\xi}\frac{VaR_\alpha - u}{\hat{\beta}}\Big)^{-1/\hat{\xi}}\ .$$

VaR_α can be obtained by arranging the terms as:

$$VaR_\alpha = u + \frac{\hat{\beta}}{\hat{\xi}}\ \Big((\frac{1 - \alpha}{1 - F_\ell(u)})^{-\xi} - 1\Big)\ .$$

Lastly, with the threshold chosen at 95th quantile of the sample, $F_\ell(u) = \Pr(x \leq u)$ can be estimated non-parametrically as 0.95.

3.3 Dynamic Models

All of the static models explained above assume that the returns are independent and identically distributed. The model parameters

[8]These results are available upon request from the author.

therefore stay constant within the sample period. However, financial returns exhibit strong serial dependence, especially in their second moment. GARCH models are very strong in incorporating volatility persistence of returns. Without such models, spikes in volatility cannot be captured quickly and VaR predictions may not reflect sudden changes in volatility.

The following model, which allows a first order autoregressive mean component in addition to Bollerslev's (1986) GARCH(1,1) structure for conditional variance, is estimated for all of the methods described below.

$$\begin{aligned} r_t &= \mu_t + \sigma_t z_t \\ \mu_t &= c_1 + c_2 r_{t-1} \\ \sigma_t^2 &= w + \alpha(r_{t-1} - \mu_{t-1})^2 + \beta\sigma_{t-1}^2, \end{aligned}$$

where r_t is the return and z_t is the shock to the data generating process for day t .

Gaussian GARCH Model: The main assumption in this model is that of conditional normality. The return on day t is normally distributed conditional on the information on day $t-1$. Therefore, the shocks $z_t \sim iid\ N(0,1)$.

The model parameters, $\hat{\theta} = \{\hat{c}_1, \hat{c}_2, \hat{w}, \hat{\alpha}, \hat{\beta}\}$, are estimated by MLE methods first. Once, $\hat{\mu}_t$ and $\hat{\sigma}_t$ are obtained by substituting the MLE parameter estimates into the conditional mean and variance equations, the VaR can be calculated as:

$$VaR_\alpha(\ell_t) = -\hat{\mu}_t + \hat{\sigma}_t\ \Phi^{-1}(\alpha),$$

where Φ is the standard normal cdf. The mean has a negative sign because the GARCH model is fit to the returns not to the losses.

GARCH-EVT Model: In this model, no distributional assumption is specified for the shocks z_t. The GARCH-EVT model first proposed by McNeil and Frey (2000) follows a two-step methodology to estimate the VaR. First, the model parameters $\hat{\theta} = \{\hat{c}_1, \hat{c}_2, \hat{w}, \hat{\alpha}, \hat{\beta}\}$ are estimated by quasi maximum likelihood estimation (QMLE). QMLE refers to fitting a Gaussian GARCH model, although the shocks z_t are not assumed to be Gaussian. This may seem unreasonable but, Gourieroux (1997) showed that the QMLE still delivers consistent and asymptotically normal parameter estimates. The only loss is the efficiency of the parameter estimates. The output of the first step is the conditional mean and conditional variance predictions, $\hat{\mu}_t, \hat{\sigma}_t$.

In the second step, shocks are extracted from the fit and the EVT methodology is applied to the shocks as described in section 3.2. The output of the second step is a quantile estimate for the distribution of shocks, z_t given by

$$VaR_\alpha(z_t) = u_z + \frac{\hat{\beta}_z}{\hat{\xi}_z}\left((\frac{1-\alpha}{1-F_\ell(u_z)})^{\hat{\xi}_z} - 1\right),$$

Finally, the VaR can be calculated using the results of both steps as

$$VaR_\alpha(\ell_t) = -\hat{\mu}_t + \hat{\sigma}_t VaR_\alpha(z_t).$$

GARCH-t Model: The standard method to estimate a GARCH-t model would be the maximization of likelihood function obtained by assuming the shocks z_t are student's t distributed with zero mean and unit variance. In that case, the distributional parameters for the student's t distribution would be estimated jointly with the parameters governing the mean and volatility dynamics of the GARCH process. However, I advocate a two-step estimation methodology in this study in order to make the model directly comparable to the GARCH-EVT model and get rid of too much restrictive structure which may result in misspecification problems. For example, if the error distribution is misspecified as student's t distribution, then the parameter estimates for the GARCH process would possibly be corrupted. However, QMLE estimation of the GARCH model parameters in a two-step method methodology provides unbiased and consisntent estimates even if the error distribution is misspecified as Gaussian (Gourieroux, 1997). Therefore, GARCH model is estimated with QMLE in the first step and then, in the second step, student's t distribution is fit to the implied residuals extracted from the GARCH fit. Therefore, the first step and the extracted residuals which are the inputs for the second step are exactly the same with the GARCH-EVT model. Any difference in forecasting performance is only attributable to the choice of EVT method or t distribution in the second step. With the two-step approach, VaR can be calculated as

$$VaR_\alpha(\ell_t) = -\hat{\mu}_t + \hat{\sigma}_t VaR_\alpha(z_t).$$

where $VaR_\alpha(z_t)$ is the quantile of the student's t distribution fit to the implied residuals.

In order to check the sensitivity of the model to estimation methodology, the standard single step GARCH-t estimation is also implemented. In this approach, the distributional parameters μ, σ, ν for the

student's t distribution are also part of the likelihood function and are estimated jointly with the parameters governing the GARCH process. In order to make sure the distribution of the shocks have zero mean and unit variance, it is assumed that

$$\mu = 0,$$
$$\sigma = \frac{\nu - 2}{\nu}.$$

Then VaR is calculated as

$$VaR_\alpha(\ell_t) = -\hat{\mu}_t + \left(\sqrt{\frac{\nu - 2}{\nu}}\right) \hat{\sigma}_t \; T_\nu^{-1}(\alpha),$$

where T_ν is the cdf function for the standard student's t distribution with ν degrees of freedom.

GARCH-Sn Model: The standard method for joint estimation of the parameters of a GARCH-Sn model would be the maximization of likelihood function obtained by assuming the shocks z_t are skewed normal distributed with zero mean and unit variance. However, in this study, the model parameters are estimated following the same two step methodology as in GARCH-EVT and GARCH-t models. The only difference is that a skewed normal distribution is fit to the extracted residuals in the second step. Therefore, any difference in performance is attributable to the choice of skewed normal distribution in the second step.

VaR is calculated using the property that linear transformations of skewed normal variables are also distributed as skewed normal. For example, if $z_t \sim \; iid \; Sn(\xi, \delta, \lambda)$ and the loss variable is given by $\ell_t = -\mu_t - \sigma_t z_t$, then we have $\ell_t \sim Sn(-\mu_t - \sigma_t \xi, \sigma_t \delta, -\lambda)$.

In order to check the sensitivity of the model to estimation methodology, the standard single step GARCH-Sn estimation is also implemented. In this approach, the distributional parameters ξ, δ, λ for the skewed normal distribution are also part of the likelihood function and are estimated jointly with the parameters governing the GARCH process. In order to make sure the distribution of the shocks have zero mean and unit variance, it is assumed that

$$\xi = -\delta \sqrt{\frac{2}{\Pi}} \frac{\lambda}{\sqrt{1 + \lambda^2}} \tag{9}$$

$$\delta = \sqrt{\frac{1}{1 - \frac{2}{\Pi} \frac{\lambda^2}{1+\lambda^2}}} \tag{10}$$

Once, the MLE estimates are calculated, VaR can be obtained as $VaR_\alpha(\ell_t) = -\hat{\mu}_t + \hat{\sigma}_t \; F^{-1}_{\hat{\xi},\hat{\delta},\hat{\lambda}}(\alpha)$ where $F_{\hat{\xi},\hat{\delta},\hat{\lambda}}$ is the cdf function for the skewed normal distribution associated with the estimated MLE parameters.

GARCH-St Model: The standard method to estimate the parameters of a GARCH-St model jointly would be the maximization of likelihood function obtained by assuming the shocks z_t are skewed t distributed with zero mean and unit variance. However, to make all models uniformly comparable, model parameters are estimated in two steps as in the GARCH-EVT, GARCH-t and GARCH-Sn models. In the first step, the parameters of the GARCH process are estimated with QMLE. In the second step, the standardized residuals z_t are extracted from the fit and Azzalini and Capitanio's (2003) skewed t distribution is fit to these residuals.

VaR is calculated using the property that linear transformations of skewed t distributed variables are also skewed t distributed. For example, $z_t \sim iid\ St(\xi, \delta, \lambda, \nu)$ and loss is given by $\ell_t = -\mu_t - \sigma_t z_t$, then we have $\ell_t \sim St(-\mu_t - \sigma_t\xi, \sigma_t\delta, -\lambda, \nu)$.

In order to check the sensitivity of the model to estimation methodology, the standard single step GARCH-St estimation is also implemented. In this approach, the distributional parameters $\xi, \delta, \lambda, \nu$ for the skewed t distribution are also part of the likelihood function and are estimated jointly with the parameters governing the GARCH process. In order to make sure the distribution of the shocks have zero mean and unit variance, it is assumed that

$$\xi = \sqrt{\frac{\nu}{\Pi}} \; \frac{\delta \; \lambda}{\sqrt{1+\lambda^2}} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \tag{11}$$

$$\delta = \sqrt{\left(\frac{\nu - 2}{\nu\left(1 - \frac{\nu-2}{\Pi} \; \frac{\lambda^2}{1+\lambda^2}\left(\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\right)\right)}\right)} \tag{12}$$

Once, the MLE estimates are calculated, VaR can be obtained as $VaR_\alpha(\ell_t) = -\hat{\mu}_t + \hat{\sigma}_t \; F^{-1}_{\hat{\xi},\hat{\delta},\hat{\lambda},\hat{\nu}}(\alpha)$ where $F_{\hat{\xi},\hat{\delta},\hat{\lambda},\hat{\nu}}$ is the cdf function for the skewed t distribution associated with the estimated MLE parameters.

4 Backtesting Results

Measuring the performance of VaR predictions continuously over time is known as backtesting. This can be done by monitoring the VaR violations, i.e losses that are larger than the predicted VaR. The VaR violation process can be thought of as a Bernoulli experiment because each day there is a $(1-\alpha)$ chance that the daily loss can be larger than the predicted VaR_α just by definition. Therefore, if a violation variable $V_t \sim Bernoulli(p)$ is defined as

$$V_t = \begin{cases} 1 & \text{if } \ell_t \geq VaR_{t,\alpha}, \\ 0 & \text{if } \ell_t < VaR_{t,\alpha} \end{cases} \tag{13}$$

we should have $H_o : p = 1-\alpha$. This can be tested against the alternative $H_1 : p \neq 1-\alpha$ using a two-sided likelihood ratio test.

4.1 Unconditional Coverage Ratio Test

Defining n_1 as the total number of violations and n_0 as the total number of non-violation days

$$n_1 = \sum_{t=501}^{t=4099} V_t; \text{ and } n_0 = \sum_{t=501}^{t=4099} (1-V_t)$$

the MLE estimator of the violation probability can be written as

$$\hat{p}_{mle} = n_1/(n_0+n_1)$$

Then, the likelihood ratio statistic is given by

$$LR_{uc} = -2\left(lnL(1-\alpha, V) - lnL(\hat{p}_{mle}, V)\right) \tag{14}$$

where $L(p,V) = (p)^{n_1}(1-p)^{n_0}$.

Under the null hypothesis, the LR statistic follows a chi-square distribution with one degree of freedom. Detailed backtesting results for all countries, models, and quantiles are provided in Table 2. The expected number of violations for all quantiles is given in the first rows of the tables. The number of VaR violations for all methods is also presented in their respective rows. Probability values from two-sided likelihood ratio tests of the null hypothesis are also provided inside parentheses.

Table 3 summarizes of hypothesis testing results without going into country specific details. In columns associated with different quantiles, the models are scored with the number of countries where the null hypothesis can't be rejected. A probability value of 0.05 is used as the rejection criteria. The maximum score is 12 since there are 12 countries in the dataset. In the "Score" column, the total across all 4 quantiles are given, so the maximum score can be 48. In the "Rank" column, models are ranked based on their scores. When there is a tie between two models, the average probability values are used as tie breakers. Table 4 presents exactly the same information as in Table 3 this time with 1000 observation window size.

Several comments are in order based on the information in Table 3. For static models, if the evaluation is performed entirely based on the total score, it seems that skewness and fat tail modeling are equally important and modeling only one of them doesn't improve the forecasting precision significantly. For example, both the skewed normal model and the student's t model improve the benchmark (normal) model marginally, increasing the score from 12 to 15. On the other hand, joint modeling of skewness and fat tails improves forecasting precision significantly. For example, the skewed t model and the EVT model improve the benchmark model significantly, increasing the score to 29 and 30 respectively. However, this observation is not valid at every quantile level. If we were to exclude the 95th quantile from this attribution analysis, then the benchmark model score would be only 1. The skewed normal model increases it to 3, whereas student's t model increases it to 13. Therefore, for the purposes of catastrophic risk management at higher quantiles, fat tail modeling is much more important than skewness modeling. Excluding 95th quantile, the skewed t distribution would increase the score of the student's t distribution from 13 to 23. Therefore, skewness modeling also improves the performance at higher quantiles, but only if fat tails are accounted for first.

The same conclusions can be made for dynamic models as well. Incorporating dynamic volatility into the model significantly increases the backtesting performance but only if the fat tails are taken into account first. For example, GARCH-EVT increases the score of EVT from 30 to 45, GARCH-St increases the score of skewed t from 29 to 42 and GARCH-t increases the score of the student's t model from 15 to 28. However, Gaussian GARCH model didn't improve the backtesting performance of the normal model and Garch-Sn model performed only

(a) Turkey

	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$	$\alpha = 0.995$
Target	180	90	36	18
Normal	173 (0.593)	108 (0.062)	69 (0)	51 (0)
Student's t	201 (0.114)	103 (0.174)	47 (0.078)	23 (0.257)
Evt	190 (0.446)	84 (0.519)	41 (0.412)	22 (0.36)
Sn	162 (0.163)	106 (0.096)	65 (0)	53 (0)
St	197 (0.199)	101 (0.248)	47 (0.078)	23 (0.257)
Garch	180 (0.997)	112 (0.023)	66 (0)	43 (0)
Garch−t	190 (0.446)	98 (0.398)	42 (0.327)	26 (0.076)
Garch−Evt	190 (0.446)	91 (0.913)	40 (0.509)	28 (0.029)
Garch−Sn	163 (0.188)	97 (0.459)	58 (0.001)	45 (0)
Garch−St	193 (0.324)	98 (0.398)	44 (0.195)	28 (0.029)

(b) Indonesia

	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$	$\alpha = 0.995$
Target	180	90	36	18
Normal	175 (0.704)	130 (0)	88 (0)	70 (0)
Student's t	221 (0.002)	127 (0)	55 (0.003)	26 (0.076)
Evt	193 (0.324)	103 (0.174)	54 (0.005)	34 (0.001)
Sn	162 (0.163)	115 (0.01)	80 (0)	67 (0)
St	204 (0.071)	110 (0.039)	49 (0.039)	24 (0.177)
Garch	181 (0.936)	120 (0.002)	75 (0)	56 (0)
Garch−t	201 (0.114)	109 (0.049)	49 (0.039)	21 (0.489)
Garch−Evt	179 (0.942)	85 (0.592)	39 (0.619)	23 (0.257)
Garch−Sn	160 (0.12)	103 (0.174)	66 (0)	47 (0)
Garch−St	184 (0.758)	101 (0.248)	40 (0.509)	19 (0.814)

(c) Malaysia

	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$	$\alpha = 0.995$
Target	180	90	36	18
Normal	169 (0.398)	120 (0.002)	83 (0)	65 (0)
Student's t	211 (0.021)	119 (0.003)	49 (0.039)	26 (0.076)
Evt	202 (0.098)	103 (0.174)	53 (0.008)	30 (0.01)
Sn	167 (0.316)	121 (0.002)	81 (0)	67 (0)
St	203 (0.084)	108 (0.062)	48 (0.056)	25 (0.118)
Garch	154 (0.042)	98 (0.398)	57 (0.001)	47 (0)
Garch−t	164 (0.216)	90 (0.998)	40 (0.509)	18 (0.999)
Garch−Evt	172 (0.54)	79 (0.232)	38 (0.739)	20 (0.642)
Garch−Sn	158 (0.087)	102 (0.209)	55 (0.003)	42 (0)
Garch−St	170 (0.443)	94 (0.67)	41 (0.412)	21 (0.489)

(d) Philippines

	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$	$\alpha = 0.995$
Target	180	90	36	18
Normal	187 (0.592)	128 (0)	82 (0)	59 (0)
Student's t	218 (0.005)	121 (0.002)	51 (0.018)	21 (0.489)
Evt	207 (0.043)	102 (0.209)	46 (0.108)	26 (0.076)
Sn	194 (0.288)	123 (0.001)	84 (0)	63 (0)
St	210 (0.025)	108 (0.062)	48 (0.056)	19 (0.814)
Garch	175 (0.704)	110 (0.039)	65 (0)	41 (0)
Garch−t	197 (0.199)	104 (0.144)	37 (0.866)	22 (0.36)
Garch−Evt	189 (0.492)	89 (0.917)	34 (0.736)	22 (0.36)
Garch−Sn	165 (0.246)	111 (0.03)	58 (0.001)	42 (0)
Garch−St	191 (0.403)	102 (0.209)	35 (0.868)	19 (0.814)

(e) South Korea

	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$	$\alpha = 0.995$
Target	180	90	36	18
Normal	194 (0.288)	130 (0)	86 (0)	66 (0)
Student's t	230 (0)	127 (0)	56 (0.002)	34 (0.001)
Evt	189 (0.492)	97 (0.459)	54 (0.005)	38 (0)
Sn	174 (0.647)	117 (0.006)	79 (0)	60 (0)
St	203 (0.084)	105 (0.118)	48 (0.056)	32 (0.003)
Garch	217 (0.006)	139 (0)	68 (0)	46 (0)
Garch−t	226 (0.001)	121 (0.002)	47 (0.078)	21 (0.489)
Garch−Evt	183 (0.816)	84 (0.519)	39 (0.619)	26 (0.076)
Garch−Sn	186 (0.645)	109 (0.049)	58 (0.001)	39 (0)
Garch−St	206 (0.051)	98 (0.398)	35 (0.868)	14 (0.326)

(f) China

	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$	$\alpha = 0.995$
Target	180	90	36	18
Normal	190 (0.446)	130 (0)	86 (0)	66 (0)
Student's t	222 (0.002)	113 (0.018)	46 (0.108)	21 (0.489)
Evt	197 (0.199)	104 (0.144)	46 (0.108)	30 (0.01)
Sn	168 (0.356)	115 (0.01)	83 (0)	57 (0)
St	213 (0.014)	111 (0.03)	39 (0.619)	26 (0.076)
Garch	215 (0.009)	127 (0)	73 (0)	50 (0)
Garch−t	227 (0.001)	115 (0.01)	44 (0.195)	21 (0.489)
Garch−Evt	197 (0.199)	108 (0.062)	47 (0.078)	24 (0.177)
Garch−Sn	187 (0.592)	116 (0.008)	69 (0)	44 (0)
Garch−St	220 (0.003)	114 (0.014)	44 (0.195)	19 (0.814)

(g) Taiwan

	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$	$\alpha = 0.995$
Target	180	90	36	18
Normal	194 (0.288)	122 (0.001)	90 (0)	58 (0)
Student's t	228 (0)	112 (0.023)	47 (0.078)	17 (0.812)
Evt	202 (0.098)	102 (0.209)	46 (0.108)	24 (0.177)
Sn	188 (0.541)	121 (0.002)	80 (0)	58 (0)
St	215 (0.009)	113 (0.018)	43 (0.255)	18 (0.999)
Garch	214 (0.011)	138 (0)	73 (0)	51 (0)
Garch−t	221 (0.002)	119 (0.003)	44 (0.195)	25 (0.118)
Garch−Evt	210 (0.025)	107 (0.077)	45 (0.146)	28 (0.029)
Garch−Sn	206 (0.051)	131 (0)	66 (0)	47 (0)
Garch−St	226 (0.001)	121 (0.002)	45 (0.146)	21 (0.489)

(h) Thailand

	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$	$\alpha = 0.995$
Target	180	90	36	18
Normal	168 (0.356)	110 (0.039)	62 (0)	45 (0)
Student's t	196 (0.226)	105 (0.118)	34 (0.736)	16 (0.631)
Evt	187 (0.592)	103 (0.174)	37 (0.866)	21 (0.489)
Sn	169 (0.398)	105 (0.118)	64 (0)	49 (0)
St	199 (0.152)	109 (0.049)	33 (0.611)	15 (0.466)
Garch	160 (0.12)	102 (0.209)	60 (0)	39 (0)
Garch−t	177 (0.821)	94 (0.67)	35 (0.868)	17 (0.812)
Garch−Evt	185 (0.701)	93 (0.748)	38 (0.739)	21 (0.489)
Garch−Sn	161 (0.14)	100 (0.293)	59 (0)	40 (0)
Garch−St	183 (0.816)	98 (0.398)	40 (0.509)	17 (0.812)

(i) Brazil

	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$	$\alpha = 0.995$
Target	180	90	36	18
Normal	213 (0.014)	153 (0)	89 (0)	66 (0)
Student's t	255 (0)	150 (0)	60 (0)	34 (0.001)
Evt	211 (0.021)	105 (0.118)	49 (0.039)	31 (0.005)
Sn	190 (0.446)	125 (0)	78 (0)	63 (0)
St	215 (0.009)	111 (0.03)	45 (0.146)	28 (0.029)
Garch	212 (0.017)	136 (0)	80 (0)	54 (0)
Garch−t	221 (0.002)	126 (0)	60 (0)	36 (0)
Garch−Evt	192 (0.362)	94 (0.67)	43 (0.255)	23 (0.257)
Garch−Sn	165 (0.246)	103 (0.174)	67 (0)	46 (0)
Garch−St	193 (0.324)	100 (0.293)	47 (0.078)	23 (0.257)

(j) Chile

	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$	$\alpha = 0.995$
Target	180	90	36	18
Normal	197 (0.199)	133 (0)	87 (0)	64 (0)
Student's t	230 (0)	131 (0)	69 (0)	35 (0)
Evt	198 (0.174)	114 (0.014)	49 (0.039)	26 (0.076)
Sn	168 (0.356)	118 (0.004)	82 (0)	58 (0)
St	209 (0.03)	119 (0.003)	58 (0.001)	28 (0.029)
Garch	201 (0.114)	124 (0.001)	63 (0)	46 (0)
Garch−t	207 (0.043)	112 (0.023)	52 (0.012)	26 (0.076)
Garch−Evt	196 (0.226)	101 (0.248)	48 (0.056)	23 (0.257)
Garch−Sn	169 (0.398)	102 (0.209)	55 (0.003)	39 (0)
Garch−St	194 (0.288)	103 (0.174)	45 (0.146)	26 (0.076)

(k) Mexico

	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$	$\alpha = 0.995$
Target	180	90	36	18
Normal	179 (0.942)	123 (0.001)	76 (0)	60 (0)
Student's t	218 (0.005)	119 (0.003)	54 (0.005)	35 (0)
Evt	196 (0.226)	106 (0.096)	47 (0.078)	35 (0)
Sn	167 (0.316)	108 (0.062)	74 (0)	54 (0)
St	192 (0.362)	102 (0.209)	43 (0.255)	26 (0.076)
Garch	204 (0.071)	128 (0)	74 (0)	50 (0)
Garch−t	207 (0.043)	120 (0.002)	54 (0.005)	27 (0.048)
Garch−Evt	188 (0.541)	98 (0.398)	45 (0.146)	23 (0.257)
Garch−Sn	176 (0.762)	110 (0.039)	62 (0)	44 (0)
Garch−St	196 (0.226)	105 (0.118)	44 (0.195)	20 (0.642)

(l) Peru

	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$	$\alpha = 0.995$
Target	180	90	36	18
Normal	184 (0.758)	124 (0.001)	85 (0)	64 (0)
Student's t	228 (0)	123 (0.001)	53 (0.008)	28 (0.029)
Evt	207 (0.043)	110 (0.039)	49 (0.039)	30 (0.01)
Sn	166 (0.28)	118 (0.004)	79 (0)	59 (0)
St	216 (0.007)	109 (0.049)	47 (0.078)	28 (0.029)
Garch	191 (0.403)	117 (0.006)	70 (0)	48 (0)
Garch−t	213 (0.014)	104 (0.144)	44 (0.195)	23 (0.257)
Garch−Evt	193 (0.324)	97 (0.459)	40 (0.509)	25 (0.118)
Garch−Sn	183 (0.816)	110 (0.039)	63 (0)	46 (0)
Garch−St	210 (0.025)	105 (0.118)	40 (0.509)	21 (0.489)

Table 2: Detailed Country Level Backtesting Results

Model	0.95	0.975	0.99	0.995	Score	Rank	Avg.p.val
Normal	11	1	0	0	12	(9)	11.8
Student's t	2	2	4	7	15	(7)	9.7
Evt	9	10	6	5	30	(3)	16.9
Sn	12	3	0	0	15	(8)	9.6
St	6	5	10	8	29	(4)	15.2
Garch	7	2	0	0	9	(10)	8.6
Garch-t	5	5	8	10	28	(5)	24.7
Garch-Evt	11	12	12	10	45	(1)	40.7
Garch-Sn	12	6	0	0	18	(6)	12.5
Garch-St	9	10	12	11	42	(2)	36.2

Table 3: Scoring based on unconditional coverage test with window size=500

Model	0.95	0.975	0.99	0.995	Score	Rank	Avg.p.val
Normal	7	6	1	0	14	(10)	12.8
Student's t	6	8	8	11	33	(6)	32.9
Evt	9	11	9	10	39	(4)	34.9
Sn	7	9	2	1	19	(8)	10.6
St	8	10	11	11	40	(3)	41.8
Garch	9	7	0	0	16	(9)	12.5
Garch-t	7	8	10	12	37	(5)	41.1
Garch-Evt	11	12	12	12	47	(2)	56.2
Garch-Sn	7	11	2	0	20	(7)	18.1
Garch-St	11	12	12	12	47	(1)	59.5

Table 4: Scoring based on unconditional coverage test with window size=1000

slightly better than its static counterpart.

Also note that the conventional risk measurement methods that rely on some sort of normality assumption – the normal, skewed normal, Gaussian GARCH and GARCH-Sn– result in significant underestimation of the risk. This is because none of these models address fat tail modeling.

When using 1000 observations as the window size, the backtesting performance of all models increases. However, the other comments made above are still valid. Skewness modeling and dynamic volatility modeling can improve the results significantly but only when fat tails are taken into account first. Fat tail modeling is still the most important issue for robust risk management and its importance increases with the quantile of interest.

Overall, when using a window size of 500 observations, the GARCH-EVT model of McNeil and Frey (2000) outperforms all other models with a score of 45 and the new GARCH-St model suggested in this study follows it closely with a score of 42. When the window size is increased to 1000 observations, both models have a score of 47 and the new GARCH-St model ranks first based on the tie breaker average probability value. The excellent performance of GARCH-EVT model is consistent with earlier studies such as Kuester et al. (2006). However, in this study GARCH-St model provides a much better backtesting performance compared to earlier studies. First, previous studies mostly relied on Fernandez and Steel (1998) and Hansen (1994) skewed t distributions. Azzalini and Capitanio's (2003) skewed t distribution was not used in large scale comparative studies of VaR prediction methods before. Secondly, earlier studies estimated GARCH model parameters and the distributional parameters jointly in a single step. In this study, a two step methodology is proposed with QMLE estimation of the GARCH parameters in the first step and subsequent modeling of the error terms with a skewed t distribution. In fact, the sensitivity of the proposed model is checked against a single step joint estimation alternative and results are reported in Table 5(6) for a window size of 500 (1,000) observations. Joint estimation in a single step is achieved by maximization of the exact likelihood function with MLE instead of the QMLE method used in the two-step methodology. Results confirm that two-step estimation of the model outperforms the joint estimation method. The only exception is the GARCH-t model with 500 observations window.

The intuition behind this result is related to model misspecifica-

Model	0.95	0.975	0.99	0.995	Score	Avg.p.val
Garch-t (joint estimation)	4	6	8	11	29	24.9
Garch-t (2-step estimation)	5	5	8	10	28	24.7
Garch-Sn (joint estimation)	7	2	0	0	9	8.3
Garch-Sn (2-step estimation)	12	6	0	0	18	12.5
Garch-St (joint estimation)	7	7	11	11	36	26.5
Garch-St (2-step estimation)	9	10	12	11	42	36.2

Table 5: Sensitivity to estimation methodology; window size=500

Model	0.95	0.975	0.99	0.995	Score	Avg.p.val
Garch-t (joint estimation)	4	8	10	11	33	38.1
Garch-t (2-step estimation)	7	8	10	12	37	41.1
Garch-Sn (joint estimation)	9	7	0	0	16	11.9
Garch-Sn (2-step estimation)	7	11	2	0	20	18.1
Garch-St (joint estimation)	9	11	12	12	44	53.4
Garch-St (2-step estimation)	11	12	12	12	47	59.5

Table 6: Sensitivity to estimation methodology; window size=1,000

tion.Therefore, it makes sense to estimate a Gaussian GARCH model with QMLE as the first step and model fat tails by fitting the skewed t distribution to the implied residuals in the second step.

Note that joint estimation of too many parameters governing a complicated data generating process is challenging and may require a long history of data to arrive at reasonable parameter estimates. Also, such estimation is not robust to model misspecification since the likelihood function relies on too many restricting assumptions. In particular, if the distribution of errors are misspecified as skewed t distribution, then MLE estimation is corrupted and can't provide consistent and unbiased parameter estimates for the GARCH process. However, if the error distribution is misspecified as Gaussian, then unbiased and consistent parameter estimates can still be achieved from the QMLE method.

In fact, the outstanding performance of the GARCH-EVT model lies with the two step approach as well as the success of the EVT in modeling fat tails. As shown in this study, the skewed t distribution of Azzalini and Capitanio (2003) is another good alternative if implemented following the same two-step methodology.

4.2 Independence Test

Another desirable property of VaR predictions is that a violation of VaR today shouldn't have an impact on a violation of the VaR tomorrow, i.e. the binary violation variable V_t should be serially independent over time. Otherwise, the violations create clusters which imply that during turmoil the VaR is underestimated and during tranquil times it is needlessly conservative. The unconditional coverage ratio test ignores this issue and evaluates only the average performance of models. To check for independence of the VaR violations, Christoffersen's (1998) independence test is employed. Let the violation process V_t be a first order binary Markow chain process with the Markow switching matrix given by

$$\pi = \begin{bmatrix} \pi_{00} & 1-\pi_{00} \\ \pi_{10} & 1-\pi_{10} \end{bmatrix}, \quad \pi_{ij} = \Pr(V_{t+1} = j | V_t = i)$$

In order to test for independence, it is required to test for the null hypothesis of $H_0 : \pi_{00} = \pi_{10}$ against the alternative $H_1 : \pi_{00} \neq \pi_{10}$. Denoting n_{ij} as the empirical number of transitions from state i to state j,

$$n_{ij} = \sum_{t=501}^{T=4099} I(V_{t+1} = j | V_t = i),$$

where $I()$ is the indicator function, the maximum likelihood estimator of the transition matrix can be given as

$$\hat{\pi}_{00,mle} = \frac{n_{00}}{n_{00} + n_{01}} \qquad \text{and} \qquad \hat{\pi}_{10,mle} = \frac{n_{10}}{n_{10} + n_{11}}.$$

Under the null hypothesis $\pi_{00} = \pi_{10}$ and therefore they can be estimated as in the unconditional coverage test by

$$\pi_{00,null} = \pi_{10,null} = (n_0)/(n_0 + n_1)$$

Then the likelihood ratio statistic can be calculated as

$$LR_{ind} = -2\left(lnL(\pi_{null}, V|V_1) - lnL(\pi_{mle}, V|V_1)\right).$$

where the likelihood function is given by

$$L(\pi, V|V_1) = \hat{\pi}_{00}^{n_{00}}\ (1 - \hat{\pi}_{00})^{n_{01}}\ \hat{\pi}_{10}^{n_{10}}\ (1 - \hat{\pi}_{10})^{n_{11}}$$

Under the null hypothesis, the LR statistic follows a chi-square distribution with one degree of freedom. A summary of the independence test results is provided in Table 7 without going into country specific details[9]. In columns associated with different quantiles, models are scored using the number of countries for which the null hypothesis can't be rejected[10]. In the score column, the total score across all quantiles is given.

The main strength of the GARCH models in market risk measurement is revealed by independence test results in Table 7. Since they capture the volatility spikes quickly and model the strong persistence in volatility successfully, GARCH models can quickly adjust the level of VaR up and down depending on market conditions. Therefore, they are able to produce VaR predictions in such a way that violations do not cluster over time. However, this doesn't provide any information regarding the precision of the VaR forecasts. The model might be overestimating or underestimating the risk. Passing the independence

[9]Detailed country level tables similar to Table 2 are available upon request from the author.

[10]A probability value of 0.05 is used as the rejection criteria.

Model	0.95	0.975	0.99	0.995	Score	Rank	Avg.p.val
Normal	0	0	0	1	1	(9)	0.2
Student's t	0	0	1	7	8	(8)	4.2
Evt	0	0	2	6	8	(6)	6.1
Sn	0	0	0	1	1	(10)	0.2
St	0	0	2	6	8	(7)	6.1
Garch	7	10	12	11	40	(3)	33.3
Garch-t	9	8	12	12	41	(1)	34.3
Garch-Evt	7	9	12	12	40	(2)	36.3
Garch-Sn	6	8	12	12	38	(5)	34.5
Garch-St	7	8	12	11	38	(4)	36.3

Table 7: Scoring based on independence tests with window size=500

test means that the performance of the model is uniform over time. Finally, minor differences between different GARCH models can be considered noise. Independence test results with a window size of 1,000 observation are presented in Table 8. The same conclusion can be made.

Model	0.95	0.975	0.99	0.995	Score	Rank	Avg.p.val
Normal	0	1	0	1	2	(10)	1.1
Student's t	0	1	2	7	10	(7)	5.4
Evt	0	0	2	6	8	(8)	5.7
Sn	0	1	0	2	3	(9)	1.2
St	0	1	3	7	11	(6)	8.9
Garch	10	10	12	10	42	(5)	35.5
Garch-t	11	11	10	11	43	(3)	37.2
Garch-Evt	11	9	11	12	43	(1)	38.8
Garch-Sn	12	9	11	11	43	(4)	33.4
Garch-St	12	9	11	11	43	(2)	38.1

Table 8: Scoring based on independence tests with window size=1000

4.3 Conditional Coverage Ratio Test

While the unconditional coverage ratio test doesn't check for independence, the independence test doesn't account for the correct coverage ratio. Although the two parameters are set to be equal, i.e $\pi_{00} = \pi_{10}$, they are jointly unconstrained. To check for the correct coverage ratio as in Christoffersen (1998), the same Markow switching matrix introduced in the last section is assumed for the violation variable. Then, the null hypothesis is that $H_o : \pi_{00} = \pi_{10} = \alpha$. Under the alternative, both parameters are unconstrained.

Note that the null is the null hypothesis of the unconditional coverage ratio test and the alternative is the alternative hypothesis of the independence test. Therefore, the LR statistic can be written as

$$LR_{cc} = -2\left(lnL(1-\alpha, V) - lnL(\pi_{mle}, V|V_1)\right)$$

Further, note that the null hypothesis of the independence test is the same as the alternative hypothesis of the unconditional coverage ratio test. Therefore, the likelihood ratio statistic can be written as the sum of the LR statistics of the unconditional coverage ratio test and the independence test.

$$\begin{aligned} LR_{cc} &= -2\left(lnL(1-\alpha, V) - lnL(\pi_{mle}, V|V_1)\right) \\ LR_{cc} &= -2\left(lnL(1-\alpha, V) - lnL(\hat{p}_{mle}, V) + lnL(\pi_{null}, V|V_1) - lnL(\pi_{mle}, V|V_1)\right) \\ &= LR_{uc} + LR_{ind} \end{aligned}$$

and it is distributed as chi-squared with two degrees of freedom since both parameters are constrained under the null hypothesis.

Summary results are presented in Table 9 for a window size of 500 observations and in Table 10 for a window size of 1000 observations. Static models have the lowest scores due to their unacceptable performance with respect to the independence component. Within the dynamic models, again the GARCH-EVT model ranks first with a window size of 500 observations and the GARCH-St model ranks first with a window size of 1,000 observation window size.

The sensitivity of the two-step GARCH-St methodology is checked against the alternative single step joint estimation with the conditional coverage ratio test as well. The results are reported in Table 11 (12) for window size of 500 (1000) observations. The previous finding with the unconditional coverage ratio test is still valid. Two-step implementation of models outperforms their single step joint estimation counterparts.

Model	0.95	0.975	0.99	0.995	Score	Rank	Avg.p.val
Normal	0	0	0	0	0	(10)	0.0
Student's t	0	0	1	5	6	(7)	3.6
Evt	0	0	1	4	5	(8)	2.9
Sn	0	0	0	0	0	(9)	0.0
St	0	0	2	6	8	(5)	5.1
Garch	5	2	0	0	7	(6)	4.0
Garch-t	5	4	9	11	29	(3)	22.2
Garch-Evt	9	10	12	10	41	(1)	36.9
Garch-Sn	6	6	0	0	12	(4)	9.8
Garch-St	6	8	12	11	37	(2)	34.4

Table 9: Scoring based on conditional coverage ratio test with window size=500

Model	0.95	0.975	0.99	0.995	Score	Rank	Avg.p.val
Normal	0	0	0	0	0	(10)	0.1
Student's t	0	0	2	7	9	(8)	5.5
Evt	0	1	3	7	11	(7)	6.4
Sn	0	0	0	0	0	(9)	0.2
St	0	1	3	7	11	(6)	10.4
Garch	8	7	2	0	17	(5)	9.7
Garch-t	7	8	10	12	37	(3)	34.0
Garch-Evt	11	10	11	12	44	(2)	48.6
Garch-Sn	7	10	4	1	22	(4)	15.4
Garch-St	11	10	12	12	45	(1)	48.8

Table 10: Scoring based on conditional coverage ratio test with window size=1000

5 Conclusion

Despite the criticism after the recent financial crisis, VaR remains the main regulatory tool for estimation of market risk capital. Estimation of VaR for banks' trading books requires the choice of a market risk model among numerous alternatives. This brings an additional risk known as model risk. In this study, ten different market risk models

Model	0.95	0.975	0.99	0.995	Score	Avg.p.val
Garch-t (joint estimation)	3	5	8	11	27	22.7
Garch-t (2-step estimation)	5	4	9	11	29	22.2
Garch-Sn (joint estimation)	5	2	0	0	7	4.0
Garch-Sn (2-step estimation)	6	6	0	0	12	9.8
Garch-St (joint estimation)	4	8	11	11	34	25.4
Garch-St (2-step estimation)	6	8	12	11	37	34.4

Table 11: Sensitivity to estimation methodology using conditional covergae ratio test; window size=500

Model	0.95	0.975	0.99	0.995	Score	Avg.p.val
Garch-t (joint estimation)	3	8	10	12	33	31.3
Garch-t (2-step estimation)	7	8	10	12	37	34.0
Garch-Sn (joint estimation)	8	7	2	0	17	9.3
Garch-Sn (2-step estimation)	7	10	4	1	22	15.4
Garch-St (joint estimation)	10	11	12	12	45	42.7
Garch-St (2-step estimation)	11	10	12	12	45	48.8

Table 12: Sensitivity to estimation methodology using conditional covergae ratio test; window size=1000

in VaR prediction are compared using 12 emerging market stock indices. Models are scored based on their backtesting performance using Christoffersen (1998) tests.

Conventional risk measurement methods that rely on normality assumptions, whether it be the normal model, the skewed normal model, the Gaussian GARCH model or the GARCH-Sn model, result in significant underestimation of VaR. This is because these models fail to address fat tail distributions. These results are consistent with the previous literature.

Backtesting of one-day-ahead VaR predictions in several quantiles revealed that the most important issue in VaR prediction for stock indices is the fat tail modeling. Skewness and dynamic volatility modeling can also significantly improve the model performance but only if the fat tails are taken into account first. Also, the importance of fat tail modeling increases as the quantile at which VaR is estimated increases.

Overall, when using a window size of 500 observations, the GARCH-EVT model of McNeil and Frey (2000) outperforms all other models and the GARCH-St model follows it closely. When the window size is increased to 1000 observations, the GARCH-St model ranks first outperforming all other models.

In this study, the GARCH-St model performs better than previous studies that used GARCH models with skewed t distributed errors. This is due to the choice of Azzalini and Capitanio's (2003) skewed t distribution and, more importantly, the proposed two-step implementation. It is shown that the two step GARCH-St methodology outperforms the single step alternative which jointly estimates the GARCH model parameters and the distributional parameters for the error term. Such an approach is too restrictive and very sensitive to misspecification of the model. On the other hand, QMLE estimation of the model in the first step provides consistent and unbiased estimates for the parameters of the GARCH process even if the distribution for the error term is misspecified. Then, fat tails and skewness are accounted in the second step by fitting a skewed t distribution to the extracted residuals of the GARCH fit.

In fact, the outstanding performance of the GARCH-EVT model lies with the two step approach as well as the success of the EVT in modeling fat tails. As shown in this study, the skewed t distribution of Azzalini and Capitanio (2003) is another good alternative if implemented following the same two-step methodology.

References

[1] Azzalini, A., 1985. *"A Class of Distributions Which Includes the Normal Ones"* Scandinavian Journal of Statistics 12, pgs.171-178.

[2] Azzalini, A. and Capitanio A., 2003. *"Distributions Generated by Perturbation of Symmetry With Emphasis on a Multivariate Skew t Distribution"* Journal of Royal Statististical Society B 65, pgs.367-389.

[3] Bali, T.G., 2003. *"An Extreme Value Approach to Estimating Volatility and Value at Risk"* Journal of Business 76, pgs.83-108.

[4] Basle Committee on Banking Supervision, 1996. *"Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Models Approach to Market Risk Capital Requirement"* Manuscript, Bank for International Settlements.

[5] Bollerslev, T., 1986. *"Generalized Autoregressive Conditional Heteroscedasticity"* Journal of Econometrics 1, pgs.307-327.

[6] Christoffersen, P. F., 1998. *" Evaluating Interval Forecasts"* International Economic Review 39, pgs. 841-862

[7] Embrechts, P., Kluppelberg C., and Mikosch T., 1997. *"Modeling Extremal Events for Insurance and Finance"* Springer Verlag, Berlin.

[8] Engle, R., 1982. *"Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of U.K. Inflation"* Econometrica 50, pgs.987-1008.

[9] Fama, E., 1965. *"The Behavior of Stock Market Prices"* Journal of Business 38, pgs.34-105.

[10] Fernandez, C. and Steel, M.F.J., 1998. *"On Bayesian Modeling of Fat Tails and Skewness"* Journal of the American Statistical Association, 93, No. 441 pgs.359-371.

[11] Hansen, B. E., 1994. *"Autoregressive Conditional Density Estimation"* International Economic Review 35, No.3 pgs.705-730.

[12] Longin, F. M., 2000 *"From value at risk to stress testing: The extreme value approach"* Journal of Banking and Finance 24, pgs.1097-1130.

[13] Gencay, R. and Selcuk, F., 2004. *"Extreme Value Theory and Value-at-Risk: Relative Performance in Emerging Markets"* International Journal of Forecasting 20, pgs.287-303.

[14] Genay, R., Seluk, F. and Uluglyaci, A. 2003. *"High volatility, thick tails and extreme value theory in value-at-risk estimation"* Insurance: Mathematics and Economics, 33, pgs. 337-356

[15] Gourieroux, C., 1997. *"ARCH Models and Financial Applications"* Springer Verlag, Berlin.

[16] Hill, B.M., 1975. *"A Simple General Approach to Inferenece about the Tail of a Distribution"* Annals of Statistics 3 pgs.1163-1174.

[17] Kuester, K., Mittnik, S. and Paolella M.S., 2006. *"Value-at-Risk Prediction: A Comparison of Alternative Strategies"* Jounal of Financial Econometrics 4, pgs.53-89

[18] Mandelbrot, B., 1963. *"The Variation of Certain Speculative Prices"* The Journal of Business 36, No.4 pgs.394-419.

[19] McNeil, A.J. and Frey, R., 2000. *"Estimation of Tail Related Risk Measures for Heteroscedastic Financial Time Series"* Journal of Empirical Finance 7, pgs.271-300.

[20] McNeil, A.J., 1999. *"Extreme Value Theory for Risk Managers"* Internal Modeling CAD II, Risk Books, pgs.99-113.

[21] Smith, R.L., 1987. *"Estimating Tails of Probability Distributions"* Annals of Statistics 15, pgs.1174-1207.