The International Journal of Business and Finance Research
Vol. 11, No. 2, 2017, pp. 39-50
ISSN: 1931-0269 (print)
ISSN: 2157-0698 (online)

www.theIBFR.com

CAN SKEWED GARCH-TYPE DISTRIBUTIONS IMPROVE VOLATILITY FORECASTS DURING GLOBAL FINANCIAL CRISIS?

Jack J.W. Yang, National Yunlin University of Science and Technology
Chien-Tsung Li, National Yunlin University of Science and Technology

ABSTRACT

This paper is related to the work of Patton (2011), who proposed the required robust loss functions MSE and QLIKE for imperfect fluctuations in the proxy variables, as well as the use of GW and MCS test for statistical analysis. In the same volatility model, the use the GW test pairing for comparing volatility forecasts of skewed and non-skewed error distributions. With the exception of EGARCH, the results produce no clear evidence of better prediction by a non-skewed distribution. In the same volatility model, the comparison of six different error distribution functions for volatility forecast showed no consistent result. In addition to the APARCH model with skewed Student-t distribution, the remaining results favored in non-skewed error distribution function for better prediction. In the comparison of all 30 models for forecasting volatility, better prediction models were all based on APARCH with six different error distribution functions. However, with a 90% confidence level, according to MCS tests, they all were included in the set of better volatility prediction models. A return with skewness, leptokurtic, and thick tail does not necessarily have the best performance in volatility prediction in the skewed error distribution.

JEL: C58, G01

KEYWORDS: Volatility Forecast, Mode Confidence Set (MCS), Global Financial Crisis, Giacomini and White Test (GW Test)

INTRODUCTION

Fluctuation phenomena, such as volatility clustering, kurtosis and leverage effects are important for financial markets. During financial downturns, such as around the 2007-2008 tsunami, accurate volatility forecasts can be risk averse to asset prices, portfolios and risk management. Fluctuations are unobservable and must be estimated from the model. Good volatility models, in addition to being used for estimating, should also be used for prediction.

The financial crises of 2008 caused great disturbance in every country's stock market and such a volatile situation would be worth a closer look. Brownlees et al. (2011) argued the global financial crisis of 2008 raised question about the appropriateness of using financial models, especially the standard tools for estimation and volatility forecasting. Therefore, that they wanted to further investigate the model to verify its accuracy in volatility as the financial crises transmitted across various economic entities through time. The goals is to make valuable prediction of fluctuation. The research result was affirmative. Therefore, our study used the five GARCH categories and six error distribution functions to study volatility forecasts of Taiwan's weighted stock price index during the financial tsunami period and classified them for volatility prediction comparison. The remainder of the paper is organized as follows. The first section introduced the issue. The second section introduces the data and methodology used in the study. The third section illustrates the results and discussions. The final section provides some concluding comments.

LITERATURE REVIEW

In the GARCH model, Engle (1982) first proposed the so-called ARCH model to estimate the return volatility of financial markets, followed by the generalized ARCH model, which was the so-called GARCH model proposed by Bollerslev (1986). Derivatives of GARH models were subsequently proposes. For example: Glosten et al. (1993) developed the GJR-GARCH model, Nelson (1991) developed EGARCH model, Ding et al. (1993) derived the APARCH model, and Engle and Lee (1999) simplified it to propose the CGARCH model. These volatility models were used to complete the study. It is necessary to study the forecasting capabilities of GARCH, and consider the influence of different error distribution functions on the fluctuation forecast. Wilhelmsson (2006) used a single GARCH with nine different error distribution functions to study the volatility prediction of the S & P500 futures index, in which the result showed that Student's t distribution forecast performance was better and a leptokurtic distribution function was better than a normal distribution function. Skewness of the distribution function, with higher dynamic difference, was not good for the fluctuation prediction.

Chuang et al. (2007) investigated a single GARCH model with 13 different error distribution functions. They used seven national stock price indices and six exchange rates as data to study the effect of different error distributions on volatility predictions under the same model. The estimation method was based on rolling estimation. The result showed the volatility prediction of complex residual distribution functions was not necessarily better than the simple one in volatility prediction of the same period. Lin et al. (2010) used the S&P 500 index to study GARCH with four different error distribution functions: Normal, Student-t, heavy-tailed (HT) and skewed generalized-t distribution (SGT), and the GJR-GARCH and EGARCH with normal distribution functions to yield a total of six types, to study the volatility forecasts. It is more important to consider whether the asymmetric distribution function displays leptokurtic, skewness, thick tails or leverage effects when considering the precision of a forecasting volatility model. If asymmetric distributions were not considered, the GARCH model with normal distribution showed better performance in volatility forecast than the other three. Brownlees' study found that, in the case of such severe fluctuations during the financial crisis, although the student-t distribution did take into account the thick tail, the use of Normal distribution and Student-t function was not different or better for volatility forecasts.

Lee and Pai (2010) study the GARCH model, using three different error distribution functions (Normal, Student-t and SGT) on US REIT volatility forecast. By using DM-tests on the loss functions, MSE and MAE, the study found that GARCH-SGED produced better forecast results on volatility than the other two. Therefore, skewness and tail thickness of an error distribution could impact the volatility forecast. Haque et al.(2014) analyzed the forecasting performance of the GARCH model with four error distribution functions, including Normal, Student-t, SGT and HT. They also examined asymmetric models such as GJR-GARCH and EGARCH, to study volatility forecasts for the Karachi Stock Exchange 100 Index (KSE-100) and Value at Risk (Var). Research found that GARCH-HT and GARCH-SGT had better prediction on volatility. However Var, GARCH-T and GARCH-SGT had better prediction results. Acuña et.al.(2015) used IBM stock to study the APARCH model, with five types of error distribution function, including: Normal, Student- t, Generalized Error, skew Student-t and Pearson type-IV distributions. The research used three minimal loss functions of MSE, MAE, and LOG LOSS showing that the skewed error distribution in the APARCH model had better volatility forecasts than the non-skewed distribution. Wilhelmsson (2006) and Haque et.al.(2014) argued that the GARCH-type volatility forecast model, which took into account the influence of different error distributions, were not well studied in the extant literature. Therefore, this article fills this gap in the literature.

DATA AND METHODOLOGY

Data

The research used the Taiwan weighted stock price index. The study was conducted from Jan. 2, 2007 to Dec. 31, 2008, including 496 daily observations. The data employed was retrieved from the database of Taiwan Stock Exchange website. The segmentation method was the same as used by Brownlees et al.(2011) and Ewing et al.(2007) to study volatility during financial tsunami and the Great Depression. The period for in-sample volatility prediction was from Jan. 2, 2007 to Sept. 13, 2008, while the period for out-sample volatility was from Sept. 16, 2008 to Dec. 31, 2008. The index return is calculated as r_t=100(lnP_t- lnP_{t-1}), in which P_t represented the index in the t-period. While volatility proxies are commonly shown as squared returns or realized volatility, and according to Poon and Granger (2003) and Patton (2011) realized volatility, when compared with the squared returns, produces a less noisy estimate. Therefore, when forecasting volatility, comparison of loss functions, using the realized volatility as proxy variables would be much better than the squared returns. In this paper, five minutes of squared returns of intraday data were used as volatility proxy variables. The estimation method adopted the rolling window estimation as proposed by Poon and Granger (2003) and Brownlees (2011), which assumed there were T samples in all samples, a number of R = T−P in the sample, and an extra prediction number of P. The method requires disregarding the oldest entry of data to add a new one, to keep the estimated number at R in the samples. It implies that during prediction of the second forecast we use the estimate of 2, 3……R+1 to predict R + 2 until all the data were completed. However, during the financial tsunami, when policies were constantly introduced as measures, the estimation coefficient in the samples might vary, and thus, the method should produce better prediction results.

Volatility Models

The study involved a univariate GARCH-class model and produced one-step-ahead volatility forecasts, in which GARCH-type models included GARCH, EGARCH, GJR-GARCH, APARCH, and CGARCH. The models are described below. The GARCH model was derived by Bollerslev (1986) after Engle (1982) presented the ARCH model. The GARCH (1,1) model was basically described as follows:

$$r_t = u_t + \varepsilon_t \tag{1}$$

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{2}$$

in which, u_t was the conditional mean, and $\varepsilon_t = \sigma_t z_t$ and $z_t \sim$N(0,1), σ_t^2 were conditional variables. Parameter limits included ω>0, α>0, β>0 and α+β<1. α+β represented the persistence of volatility shock. GARCH estimation could estimate the phenomenon of the volatility cluster. The GJR-GARCH model was developed by Glosten et al. (1993). The original GARCH model was supplemented by an asymmetric condition to capture the effect of negative shocks on fluctuation.

$$r_t = u_t + \varepsilon_t$$
$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma I_{t-1}\varepsilon_{t-1}^2 \tag{3}$$

Because the good news $\varepsilon_{t-1} > 0$ and bad news $\varepsilon_{t-1} < 0$ have different effects on the conditional variance of returns, to estimate the effect, it is generally assumed with $\varepsilon_{t-1} < 0$, where I_{t-1}=1, or else I_{t-1}=0. Therefore, the estimation effect on volatility due to good news would be α , while the bad news would have impact of α+γ. When γ > 0, it represents a leverage effect in volatility, and when γ≠ 0, there is sign of asymmetry.

GRJ-GARCH is a nested model of GARCH. The statistical test method is different from the general non-nested model when making accurate comparisons of two models. EGARCH (exponential GARCH) was proposed by Nelson (1991). The linear GARCH model required estimated parameters to be non-negative, but in contrast, there is no such limit for the EGARCH. The EGARCH (1,1) model specification is shown below:

$$r_t = u_t + \varepsilon_t$$
$$\log\sigma_t^2 = \omega + \alpha\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| + \beta\log\sigma_{t-1}^2 + \gamma\frac{\varepsilon_{t-1}}{\sigma_{t-1}} \quad (4)$$

The left-form is the logarithm of the conditional variation. It is estimated that if the leverage effect exists, under the condition of $\gamma<0$, for the magnitude of leverage effect, its γ value should be converted through the index. When the EGARCH is used for volatility prediction, it is necessary to be careful that prediction values cannot be non-negative. When $\gamma\neq$ is 0, there is an asymmetric volatility. APARCH was proposed by Ding et al.(1993) and the configuration of APARCH(1,1) model is shown below:

$$r_t = u_t + \varepsilon_t$$
$$\sigma_t^\delta = \omega + \alpha(|\varepsilon_{t-1}| - \gamma\varepsilon_{t-1})^\delta + \beta\,\sigma_{t-1}^\delta \quad (5)$$

in which $\delta > 0$ and $|\gamma| \leq 1$. γ could reflect the phenomenon of asymmetry. APARCH includes several special examples, derivative from the ARCH model. When $\delta=2$ and $\gamma=0$, the GARCH(1,1) was used, while $\delta=2$, GJR-GARCH would be used. Engle and Lee(1999) considered the short-term and the long-term factors on volatility and proposed the following CGARCH model, as shown below:

$$r_t = u_t + \varepsilon_t$$
$$\sigma_t^2 - m_t = \alpha(\varepsilon_{t-1}^2 - m_{t-1}) + \beta(\sigma_{t-1}^2 - m_{t-1}) \quad (6)$$

$$m_t = \omega + \gamma(m_{t-1} - \omega) + \delta(\varepsilon_{t-1}^2 - \sigma_{t-1}^2) \quad (7)$$

In this model m_t, denotes time-varying long-term volatility, derived from (ε_{t-1}^2-σ_{t-1}^2). It depends on γ to converge to a value of ω, where γ was generally between 0.99 and 1. $(\sigma_t^2 - m_t)$ denotes the transitory component, in which β value was commonly converged to 0 and is dependent of the magnitude of $\alpha+\beta$ value. If the estimation produced $0<\alpha+\beta<\gamma<1$, it showed more persistence of influence from long-term factors than from short-term factors.

<u>Error Distribution Functions</u>

The error distribution functions were classified into the skewed and non-skewed error distribution functions. The skewed error distribution functions included: skewed Normal, skewed Student-t, and skewed Generalized error distribution(GED). The non-skewed error distribution functions included: Normal, Student-t, and GED. The functions were described below. The error distribution of time series was generally assumed to be a normal probability density distribution function of the independently and identically distribution (IID) described as follows

$$f(z_t) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(z_t-u)^2}{2\sigma^2}\right] \quad (8)$$

Information on return rates showed the phenomenon of thick tails. The Student–t distribution would be more appropriate than the normal distribution in describing the feature. If x_v, with degree of freedom as

ν, were the Student t-distribution and if $\nu>2$, then it was possible to obtain $\mathrm{Var}(x_v) = \frac{\nu}{\nu-2}$. By the use of $z_t = \frac{x_v}{\sqrt{v/_{v-2}}}$, the probability density function of z_t would be

$$f(z_t|\,v) = \frac{\Gamma[(v+1)/2]}{\Gamma(\frac{2}{v})\sqrt{(v-2)\pi}}(1+\frac{z_t}{v-2}^2)^{-(v+1)/2} \quad , \; v>2 \tag{9}$$

in which Γ(•) was the Gamma function.

The probability density function of the generalized error distribution function is described as follows:

$$f(z_t) = \frac{v\exp(-\frac{1}{2}\left|\frac{z_t}{\theta}\right|^v)}{\theta 2^{(1+\frac{1}{v})}\Gamma(\frac{1}{v})} \tag{10}$$

in which Γ(•) was the Gamma function, $\theta = (2^{\frac{-2}{v}}\frac{\Gamma(\frac{1}{v})}{\Gamma(\frac{3}{v})})^{\frac{1}{2}}$. If ν=2, then it would converges to a normal distribution function. When ν<2, there exists evidence of thick tails.

In consideration of the random variable z_t, the probability density function of skewed normal distribution is described as:

$$f(z_t)=2\phi(z_t)\Phi(\alpha\, z_t) \tag{11}$$

in which, $\phi(z_t)$ is the probability density function of normal distribution $\Phi(\alpha z_t)$ is the cumulative distribution function of standard normal distribution and α was related with the form factor.

The probability density function of skewed Student- t distribution is illustrated as:

$$f(z_t \mid \nu, \lambda)=\begin{cases} bc(1+\frac{1}{v-2}\left(\frac{bZ_t+a}{1-\lambda}\right)^2)^{-\frac{v+1}{2}} & if\; z_t < -\frac{a}{b} \\ bc(1+\frac{1}{v-2}\left(\frac{bZ_t+a}{1+\lambda}\right)^2)^{-\frac{v+1}{2}} & if\; z_t \geq -\frac{a}{b} \end{cases} \tag{12}$$

where 2<ν, -1<λ<1 , $a = 4\lambda c\frac{v-2}{v-1}$ 、 $b=1+3\lambda^2-a^2$ and $c = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi(v-2)\Gamma(\frac{v}{2})}}$.

The probability density function of the standardized skewed generalized error distribution function is specified as:

$$f\left(z_t \mid v,\lambda\right) = v(2\theta\cdot\Gamma\left(\frac{1}{v}\right))^{-1}\cdot\exp(-\frac{|z_t-\delta|^v}{[[1-sign(z_t-\delta)\cdot\lambda]^v\theta^v}) \tag{13}$$

where $\theta = \Gamma(\frac{1}{v})^{\frac{1}{2}}\cdot\Gamma\left(\frac{3}{v}\right)^{-\frac{1}{2}}\cdot S(\lambda)^{-1}$, $\delta=2\cdot\lambda\cdot A\cdot S(\lambda)^{-1}$、 $S(\lambda)=\sqrt{1+3\lambda^2-4A^2\cdot\lambda^2}$ and $A=\Gamma\left(\frac{2}{v}\right)\cdot\Gamma(\frac{1}{v})^{\frac{1}{2}}\cdot\Gamma\left(\frac{3}{v}\right)^{-\frac{1}{2}}$.

v was related with height and thick tails. λ was the parameter of skewness and -1<λ<1. λ>0, indicates negative skewness (right-modal), and vice versa for positive skewness if λ<0. Sign() is the symbol for function.

Comparison of Prediction Performance

The true volatility σ^2 is unobservable. Choosing a good volatility to represent true volatility is important to assess the volatility prediction model. The volatility proxy variables in this research are the realized volatility σ_t^2 proposed by Andersen and Bollerrslev (1997) and were mainly based on 5-minute intraday data. Realized volatility was an unbiased estimate of the conditional variance, which was the sum of all squared returns in the five-minute intraday data. It would be important to select the appropriate loss function to use the volatility model to obtain the volatility $\widehat{\sigma}^2$ and proxy variables to achieve a realized volatility σ_t^2. Patton (2011) demonstrated issues associated with volatility proxies. His research was realted to the use of imperfect volatility proxies, such as squared returns, realized volatility, and intra-day range volatility to assess model prediction performance. He foujnd if the loss function is not robust, it could cause variation in the sequencing of volatility prediction models due to selection of a non-robust loss function. But, if a robust loss function were selected, then it would not be necessary to consider the effect on loss function due to different unit. He concluded that many previous studies with inconsistent results and conclusions may be driven by selection of a non-robust loss function. Further, the necessary and sufficient condition for the selection of the robust loss function should be homogeneous when using imperfectly volatility proxy variables. Brownlees believed that almost all loss functions used in the literatures needed to be excluded, except

$$\text{MSE}=L(\sigma_t^2,\widehat{\sigma}^2)=(\sigma_t^2-\widehat{\sigma}^2)^2 \quad (14)$$

$$\text{QLIKE}=L(\sigma_t^2,\widehat{\sigma}^2)=\log\widehat{\sigma}^2+\frac{\sigma_t^2}{\widehat{\sigma}^2} \quad (15)$$

QLIKE was not affected by the extreme value of a tail. Therefore, the main loss function selected in the research was QLIKE with MSE function as the supplement. Diebold (2015) argued that incorrect conclusions occur when comparing the minimum of loss functions, without regard to the significance of statistical precision, to conclude a model as the best or improved. The use of statistical precision comparison models must fulfill the criteria of usage for the model. Common mistakes occur when using the Diebold-Mariano test, including the use rolling estimation in a nested model for comparison and inadequate fulfillment of stationary configuration of loss function. Therefore, in the comparison of loss functions in the research, paired model comparison and multi-model comparison were used. The former employed the Giacomini and White (2006) test (GW test) and the latter employed the model confidence set (MCS test) by Hansen et al.(2011).

Giacomini and White Test (GW Test)

Under the same loss function $L_{t+\tau}(\sigma_{t+\tau}^2,\widehat{\sigma}_{t+\tau}^2)$, the comparison of forecasting volatility in the loss function of two models, which was referring to as the prediction value of $\Delta L_{t+\tau}=\left|\sigma_{t+\tau}^2-\widehat{\sigma}_{1,t+\tau}^2\right|-\left|\sigma_{t+\tau}^2-\widehat{\sigma}_{2,t+\tau}^2\right|$, $\widehat{\sigma}_{i,t+\tau}^2$ of the model i (i=1,2) in the t+τ period for statistical significance, was a common test for prediction performance of models. The GW test assesses whether the conditional prediction ability of the two models is statistically different, and the prediction estimation method adopts the rolling estimation structure. It can be used in nested model comparison without the prerequisite of considering the problem of model mismatch and stationary state of loss function in the two models. However, it is not applicable when recursive estimation was used as equation. In the study, the GW test was used to compare volatility prediction performance improvement in the same model due to skewness. The null hypothesis was

$$H_0: E[\Delta L_{t+\tau} \mid \phi_t]=0 \quad (16)$$

The alternative hypothesis was

$$H_1 : E[\Delta L_{t+\tau} \mid \phi_t] \neq 0 \quad (17)$$

in which ϕ_t was an information set. Rejection of H_0 indicates a variation in prediction performance precision in the two models. When $E[\Delta L_{t+\tau} \mid \phi_t]>0$, model 2 had more precise prediction performance than model 1, and vice versa.

Model Confidence Set Test (MCS Test)

The MCS test was used for prediction equivalence comparison of multiple models. The advantage was that the benchmark model was not required thus, it allowed comparison of more than one prediction performance. The MCS test was used for comparison of the robust volatility prediction multiple models. The principle is as follows: M_0 included a finite numbered model with the model number of 1,2,…,m_0. $L_{i,t}$ indicated the loss function value of the *i* model at t-point. For all $i,j \in M_0$, the difference in prediction loss function of any two models is:

$$d_{ij,t} = L_{i,t} - L_{j,t} \quad (18)$$

$$M^* \equiv \{i \in M_0 \mid E(d_{ij,t}) \leq 0, \forall j \in M_0\} \quad (19)$$

M^* of the selection procedure was based on elimination of the model with statistically poor performance in each and every comparison. The null hypothesis for such elimination was:

$$H_{0,M} : E(d_{ij,t}) = 0, \forall i,j \in M; \ M \subset M_0 \quad (20)$$

The alternative hypothesis was:

$$H_{1,M} : E(d_{ij,t}) \neq 0, \forall i,j \in M; \ M \subset M_0 \quad (21)$$

The MCS test procedure was based on the equivalence test δ_M and the elimination rule e_M. However, δ_M was used for test $H_{0,M}$. And, e_M implied that when $H_{0,M}$ was rejected, the model with poor prediction performance would be eliminated from M. This step is repeated until no other model was eliminated. Until then, $\widehat{M}^*_{1-\alpha}$ was obtained and combined into a set, indicating a collection of models with good prediction performance at 1-α confidence level. For $M \subset M_0$, δ_M and e_M at α significance level, they must fulfill the following three criteria:

$$\lim_{n\to\infty} \sup P(\delta_M = 1 \mid H_{0,M}) \leq \alpha \quad (22)$$

$$\lim_{n\to\infty} P(\delta_M = 1 \mid H_{1,M}) = 1 \quad (23)$$

$$\lim_{n\to\infty} P(e_M \in M^* \mid H_{1,M}) = 0 \quad (24)$$

Lastly, the surviving objects in the model of $\widehat{M}^*_{1-\alpha}$, which were not eliminated, were models with good prediction performance. P-values with higher statistical significance α, indicated better performance of the model. For the statistics of p-value, Hansen et al. (2011) recommended a bootstrap method for constructing t-statistics, including two types. The range statistic T_R and semi-quadratic Statistic T_{SQ} are defined as:

$$T_R = \max_{i,j \in M} |t_{ij}| \tag{25}$$

$$T_{SQ} = \max_{i \in M} t_i \tag{26}$$

in which $t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\widehat{\text{var}}(\bar{d}_{ij})}}$; $t_i = \frac{\bar{d}_i}{\sqrt{\widehat{\text{var}}(\bar{d}_i)}}$; $\bar{d}_{ij} = n^{-1} \sum_{t=1}^{n} d_{ij,t}$; $\bar{d}_i = m^{-1} \sum_{j \in M} \bar{d}_{ij}$.

RESULTS AND DISCUSSIONS

The descriptive statistics, as seen in Table 1, showed that at a 1% statistical significant level, we reject the normal distribution and revealed a unit root. The kurtosis, was leptokurtic. Skewness showed positive shift.

Table 1：Summary Statistics

Obs	Mean	Std. Dev	Kurtosis	Skewness	J-B	Q(20)	ADF
496	-0.11	1.78	4.44	-0.24	47.56***	36.58***	-21.48***

**** Denoted significantly at the 1% level.*

In the paired comparison of the same GARCH performance between the skewed and non-skewed error distribution, the GW test on the loss functions of MSE and QLIKE were shown in Table 2 and 3. From Table 2 and Table 3, we see that at the 10% significance level, the loss functions MSE and QLIKE showed the EGARCH with denied null hypothesis. This finding suggests that the remaining candidates, such as GARCH, GJR-GARCH, APARCH and CGARCH, could not prove that under a fixed GARCH-type, the prediction performance of a skewed residual distribution was definitely better than a non-skewed distribution during the financial tsunami. However, the EGARCH model with non-skewed residual distribution showed a better significance.

Table 2: GW Test of Out-of-Sample Forecasts (MSE)

	SNORM		SSTD		SGED	
	T-Stat	P-Val	T-Stat	P-Val	T-Stat	P-Val
GARCH	4.191	0.123	4.699	0.095	3.811	0.149
EGARCH	5.086	0.079	5.716	0.057	6.069	0.048**
GJR-GARCH	3.271	0.195	3.375	0.185	3.593	0.166
APARCH	0.941	0.625	1.190	0.552	2.522	.0.283
CGARCH	2.192	0.334	0.074	0.964	0.757	0.685

This table presents the statistic and p-values of the GW test performed for pairs of models, with the no-skew distribution as a benchmark. The t-stat indicates the significance of a model's performance relative to the benchmark model.

Table 3: GW Test of Out-of-Sample Forecasts (QLIKE)

	SNORM		SSTD		SGED	
	T-stat	P-val	T-stat	P-val	T-stat	P-val
GARCH	4.738	0.094	4.941	0.085	4.427	0.109
EGARCH	0.744	0.689	6.031	0.049**	6.700	0.035**
GJR-GARCH	0.753	0.686	3.016	0.221	3.462	0.177
APARCH	0.483	0.785	3.125	0.210	1.146	0.564
CGARCH	0.276	0.871	0.478	0.787	0.266	0.876

This table presents the statistic and p-values of the GW test performed for pairs of models, with the no-skewed distribution as a benchmark. The t-stat indicates the significance of a model's performance relative to the benchmark model.

In the financial tsunami, Table 4 and Table 5 report statistics using the MCS statistical test for loss functions MSE and QLIKE and six types of error distribution functions. Results show that under the same GARCH-type model, at a 90% confidence level, the QLIKE loss function, with exception of exclusion of skewed

EGARCH as a better model, the remainders were included. In GARCH, EGARCH, and GJR-GARCH, the non-skewed error distribution for volatility prediction, under the MCS test, was more likely to be included as the better model with a more relaxed confidence level. GARCH, EGARCH and GJR-GARCH showed better volatility prediction in the Student-t distribution. While APARCH had better volatility forecast in the skewed Student-t distribution, the CGARCH had better volatility forecast in normal distribution. Therefore, in the fixed model, because of the model selection, the best prediction of volatility appears in the different error distribution function without a consistent result.

Table 4: MCS Test of Out-of-Sample Forecasts with Different Distributions

	MSE									
	GARCH		**EGARCH**		**GJR-GARCH**		**APARCH**		**CGARCH**	
	Rank	**Pmcs**	**Rank**	**Pmcs**	**Rank**	**Pmcs**	**Rank**	**Pmcs**	**Rank**	**Pmcs**
Nor	3	0.214	3	0.346	1	1.000	4	0.389	1	1.000
Std	1	1.000	1	1.000	2	0.952	2	0.455	6	0.618
Ged	2	0.762	2	0.346	3	0.952	3	0.389	2	0.831
Snor	6	0.158	5	0.105	4	0.305	6	0.234	3	0.831
Sstd	4	0.214	4	0.105	5	0.305	1	1.000	5	0.618
Sged	5	0.214	6	01.05	6	0.305	5	0.256	4	0.831

Performance based on the different distributions of the GARCH -type and MCS tests (Pmcs) obtained by the same GARCH- type in different distributions. The superscripts ∗, ∗∗, and ∗ ∗ ∗ represent the significance level of 10%, 5%, and 1%.

Table 5: MCS Test of Out-of-Sample Forecasts with Different Distributions

	QLIKE									
	GARCH		**EGARCH**		**GJR-GARCH**		**APARCH**		**CGARCH**	
	Rank	**Pmcs**	**Rank**	**Pmcs**	**Rank**	**Pmcs**	**Rank**	**Pmcs**	**Rank**	**Pmcs**
Nor	3	0.177	3	0.111	3	0.259	5	0.234	1	1.000
Std	1	1.000	1	1.000	1	1.000	2	0.234	6	0.364
Ged	2	0.586	2	0.111	2	0.259	3	0.234	3	0.876
Snor	6	0.165		0.046**	4	0.240	6	0.149	2	0.876
Sstd	4	0.177		0.046**	3	0.240	1	1.000	5	0.364
Sged	5	0.177		0.046**	6	0.240	4	0.234	4	0.876

Performance based on the different distributions of the GARCH -type and MCS tests(Pmcs) obtained by the same GARCH- type in different distributions. The superscripts ∗, ∗∗, and ∗ ∗ ∗ represent the significance level of 10%, 5%, and 1%.

Table 6 shows that under the MCS statistical tests for all models with loss functions QLIKE and MSE, at 90% confidence level, all models were included in the set with better volatility prediction performance. The best volatility prediction model for the financial tsunami was the skewed Student t-distribution of the APARCH model.

Table 6: MCS Test of All 30 Out-of-Sample Forecasting Models

		QLIKE			MSE2		
		Loss(*10^3)	**Rank**	**Pmcs**	**Loss(*10^7)**	**Rank**	**Pmcs**
GARCH	Nor	-6.042	8	0.381	7.725	19	0.560
	Std	-6.048	4	0.381	7.645	16	0.568
	Ged	-6.048	5	0.381	7.652	17	0.568
	Snor	-6.029	16	0.381	7.869	23	0.445
	Sstd	-6.034	11	0.381	7.821	22	0.560
	Sged	-6.030	15	0.381	7.861	24	0.560
EGARCH	Nor	-5.996	28	0.381	7.468	14	0.569
	Std	-6.021	19	0.381	7.385	10	0.569
	Ged	-6.010	24	0.381	7.427	11	0.569
	Snor	-5.987	29	0.167	7.733	20	0.326
	Sstd	-6.000	27	0.191	7.677	18	0.386
	Sged	-5.987	30	0.161	7.737	21	0.329
GJR-GARCH	Nor	-6.033	13	0.381	7.243	7	0.569
	Std	-6.047	6	0.381	7.250	8	0.569
	Ged	-6.040	9	0.381	7.253	9	0.569
	Snor	-6.026	17	0.381	7.463	12	0.560
	Sstd	-6.033	12	0.381	7.467	13	0.568
	Sged	-6.026	18	0.381	7.486	15	0.560
APARCH	Nor	-6.035	10	0.381	7.067	4	0.569
	Std	-6.070	2	0.381	6.826	2	0.569
	Ged	-6.056	3	0.381	6.921	3	0.569
	Snor	-6.031	14	0.355	7.167	6	0.425
	Sstd	-6.122	1	1.000	6.535	1	1.000
	Sged	-6.047	7	0.381	7.142	5	0.514
CGARCH	Nor	-6.017	20	0.381	7.964	25	0.560
	Std	-6.002	26	0.381	8.096	30	0.560
	Ged	-6.014	22	0.381	7.994	26	0.560
	Snor	-6.015	21	0.381	7.997	27	0.560
	Sstd	-6.004	25	0.381	8.083	29	0.560
	Sged	-6.013	23	0.361	8.027	28	0.560

This table presents the Loss functions statistic and p-values of MCS test that is performed for 30 models.

CONCLUSION

Volatility variables play an important role in options and risk management. In this paper, we used the Taiwan's weighted stock price index during the 2007 to 2008 financial tsunami period to study five volatility models and six error distributions. When the extreme condition occurs, could the skewed error distribution function be better than non-skewed function in predicting volatility? The five volatility models included: the standard GARCH, the EGARCH with leverage effect, the GJR-GARCH for explaining the negative impact on volatility, the APARCH for reflecting the asymmetric phenomenon, and the CGARCH for demonstrating variation between short-term and long-term impact on volatility. In a paired study of a volatility model with skewed and non-skewed error distribution, by using GW test to examine the loss function, we found that, an exception to the EGARCH with STD and GED as error distribution and a better volatility forecast by a non-skewed function. No other data produced a better prediction volatility model with skewed error distribution during the time of financial crisis.

In research on a single volatility model with 6 error distribution functions, the MCS-test was used to examine the loss function at a 90% confidence level. We found that skewness was no better than a non-skewed function, and even showed the opposite in the EGARCH model. In the study of all 30 models, using MCS-test at 90% confidence level, to determine if there is any chance that a non-skewed error distribution could result in worse prediction, none of the results support this notion. Therefore, although Taiwan's weighted stock price index showed a high narrow peak and a left skewness in its distribution during a financial crisis period, it could not significantly improve the volatility forecast. The EGARCH predicted a worst case scenario for the financial market. During an extreme event, such as a financial crisis, a volatility model with skewed error distribution function could not significantly improve the prediction result.

Therefore, it was not necessarily true that a return with skewness, leptokurtic, and thick tails would have the best performance in volatility prediction in the skewed error distribution.

We note the following limitations of the work here. In consideration of the impact on MSE loss function during an extreme event, the MSE loss function was listed on the statistical test, but only for reference without further detail description. Future research, might include more volatility models or consider whether different volatility proxy variables could affect the selection choice of the error distribution function. Every market trading mechanism is different and in the future, the impact of a financial crisis on volatility forecast results can be further studied.

REFERENCES

Acuña,D. Allende-Cid,H.and Allende,H.(2015) "The Effect of Innovation Assumptions on Asymmetric GARCH Models for Volatility Forecasting, " *in Pardo,A.and Kittle,J.eds Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* ,527-534, Springer International Publishing Press.

Andersen, T. G.and Bollerslev, T. (1997). "Intraday Periodicity and Volatility Persistence in Financial Market, " *Journal of Empirical Finance* , 4(2), 115-158.

Bollerslev,T.(1986)."Generalized Autoregressive Conditional Heteroskedasticity,"*Journal of Econometrics*, 31(3), 307-327.

Brownlees,C.,Engle,R.F.and Kelly,B.(2011). "A Practical Guide to Volatility Forecasting through Clam and Storm, "*The Journal of Risk*,14(2), 3-22.

Chuang, I. Y., Lu, J. R. and Lee, P. H. (2007). "Forecasting Volatility in the Financial Markets: A Comparison of Alternative Distributional Assumptions, " *Applied Financial Economics*,17(13), 1051–1060.

Diebold, F.X. (2015). "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on theUse and Abuse of Diebold-Mariano Tests. " *Journal of Business & Economic* Statistics, 33(1), 1-8.

Ding, Z., C. Granger, W. J. and Engle, R. F. (1993). "A Long Memory Property of Stock Market Returns and a New Model, " *Journal of Empirical* Finance, 1(1), 83-106.

Engle, R. F.(1982). "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation, "*Econometrica*, 50(4), 987-1008.

Engle, R. F. and Lee, G. G. J. (1999). "A Long–run and Short–run Component Model of Stock Return Volatility, " *in Engle, R. F. and White H. eds. Cointegration, Causality and Forecasting*, 475-497. Oxford: Oxford University Press.

Ewing, B.T.,Thompson,M.A. and Yanochik,M.A (2007), "Using Volume to Forecast Stock Market Volatility around the Time of the 1929 Crash," *Applied Financial Economics*, 17(14), 1123-1128.

Giacomini, R. and White,H. (2006). "Tests of Conditional Predictive Ability, " *Econometrica*,74(6), 1545–1578.

Glosten, L. R., Jagannathan, R. and Runkle, D. E.(1993). "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks , " *Journal of Finance*, 48(5), 1779-1801.

Hansen, P.R., Lunde, A. and Nason, J.M. (2011). "The Model Confidence Set," *Econometrica* ,79(2), 453–497.

Haque, A. and Naeem, K. (2014). "Forecasting volatility and Value-at-Risk of Karachi Stock Exchange 100 Index: Comparing distribution-type and symmetry-type models. " *European Online Journal of Natural and Social Sciences*, 3(2), 208–219.

Lee,H.Y. and Pai,T.Y. (2010) "REIT volatility prediction for skew-GED distribution of the GARCH model. " *Expert Systems with Applications,* 37(7), 4737-4741.

Liu,H.C. and Hung, J.C. (2010). "Forecasting S_P-100 Stock Index Volatility：The Role of Volatility Asymmetry and Distributional Assumption in GARCH Models, "*Expert Systems with Applications* , 37(7), 4928-4934.

Nelson, D. (1991). "Conditional Heteroskedasticity in Asset Returns : A New Approach, " *Econometrica*, 59(2),347-370.

Patton, A (2011). "Volatility Forecast Comparison Using Imperfect Volatility Proxies, " *Journal of Econometrics* ,160(1), 246–256.

Poon, S. and Granger, C. (2003). "Forecasting Volatility in Financial Markets, " *Journal of Economics Literature*, 41(2), 478−539.

Wilhelmsson, A.(2006). "GARCH Forecasting Performance under Different Distribution Assumptions, " *Journal of Forecasting*, 25(8), 561–578.

BIOGRAPHY

Jack J.W. Yang is a professor at the Department of Finance, National Yunlin University of Science and Technology, Taiwan, Email: yangjw@yuntech.edu.tw

Chien-Tsung Li is a Doctoral Candidate of Finance and Finance Department lecturer at National Yunlin University of Science and Technology Graduate Institute of Finance. Email: g9724808@yuntech.edu.tw (Corresponding author)