Expert Systems with Applications 37 (2010) 4928–4934

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Forecasting S&P-100 stock index volatility: The role of volatility asymmetry and distributional assumption in GARCH models

Hung-Chun Liu [a], Jui-Cheng Hung [b,*]

[a] Department of Finance, Minghsin University of Science and Technology, No. 1, Xinxing Rd., Xinfeng Hsinchu 30401, Taiwan, ROC
[b] Department of Finance, Lunghwa University of Science and Technology, No. 300, Sec. 1, Wanshou Rd., Guishan Shiang, Taoyuan County 33306, Taiwan, ROC

ARTICLE INFO

Keywords:
Volatility
GARCH
Asymmetry
Distribution
SPA test

ABSTRACT

This study investigates the daily volatility forecasting for the Standard & Poor's 100 stock index series from 1997 to 2003 and identifies the essential source of performance improvements between distributional assumption and volatility specification using distribution-type (GARCH-*N*, GARCH-*t*, GARCH-*HT* and GARCH-*SGT*) and asymmetry-type (GJR-GARCH and EGARCH) volatility models through the superior predictive ability (SPA) test. Empirical results indicate that the GJR-GARCH model achieves the most accurate volatility forecasts, closely followed by the EGARCH model. Such evidence strongly demonstrates that modeling asymmetric components is more important than specifying error distribution for improving volatility forecasts of financial returns in the presence of fat-tails, leptokurtosis, skewness and leverage effects. Furthermore, if asymmetries are neglected, the GARCH model with normal distribution is preferable to those models with more sophisticated error distributions.

1. Introduction

Since the 1987 stock market crash, modeling and forecasting financial market volatility has received a great deal of attention from academics, practitioners and regulators due to its central role in several financial applications, including option pricing, asset allocation and hedging. In addition, the financial world has witnessed the bankruptcy or near bankruptcy of various institutions that incurred huge losses due to their exposure to unexpected market moves for more than a decade. These financial disasters have further highlighted the significance of volatility forecasting in risk management (calculating Value-at-Risk). Given these facts, the quest for accurate forecasts appears to still be ongoing in recent years.

Whilst it has long been recognized that returns volatility is often characterized by a number of stylized facts, such as persistence, volatility clusters, time-varying volatility, and leptokurtic data behavior, the introduction of the generalized conditional heteroskedasticity (GARCH) model (Bollerslev, 1986; Engle, 1982) has created a new approach for accommodating these temporal dependencies for financial econometricians, becoming a popular tool for volatility modeling and forecasting. However, despite the success of the GARCH model, it has been criticized for failing to capture asymmetric volatility.[1] This limitation has been overcome by introducing more flexible volatility treatments by accommodating the asymmetric responses of volatility to positive and negative shocks. This more recent class of asymmetric GARCH models includes the Exponential GARCH (EGARCH) of Nelson (1991) and the threshold GARCH by Glosten, Jagannathan, and Runkle (1993) (GJR-GARCH). In addition, substantial empirical evidence exists of a non-normal distribution of financial returns.[2] For this reason, previous studies have proposed various error distributions to alleviate the deficiency of normal assumption. For example, some studies have found evidence in favor of fat-tailed distributions, such as the student-*t* (Bollerslev, 1987) and *HT* (heavy-tailed) distributions (Hung, Lee, & Liu, 2008; Politis, 2004). Other studies have found evidence that the *SGT* (skewed generalized *t*, *SGT*) distribution (Theodossiou, 1998) can flexibly treat fat-tails, leptokurtosis and skewness in returns distribution.

To the best of our knowledge, the existing literature regarding the GARCH genre of volatility forecasting techniques can be divided into two categories. The first category exploits symmetric GARCH models with alternative distributional assumptions to forecast volatility in various financial markets. Examples include Wilhelmsson (2006) and Chuang et al. (2007). Wilhelmsson (2006) investigated

* Corresponding author. Tel.: +886 2 82093211x6425; fax: +886 3 6102367.
E-mail addresses: hongrc@mail.lhu.edu.tw, hongrc@mail.ypu.edu.tw (J.-C. Hung).

[1] For stock prices, negative shocks (bad news) generally have large impacts on their volatility than positive shocks (good news). Such a phenomenon is also meant as a synonym of leverage effect.

[2] The leptokurtosis is reduced but not eliminated even if returns are standardized using time-varying estimates for the means and variances.

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.12.022

the predictive ability of the GARCH(1,1) model when estimating S&P-500 index future returns with various error distributions. Allowing for a leptokurtic error distribution is demonstrated to significantly improve variance forecasts for alternative forecast horizons. Notably, allowing for time-varying skewness or kurtosis of the distribution does not further improve forecasts. Recently, Chuang, Lu, and Lee (2007) compared the volatility forecasting performance of linear GARCH models based on a group of distributional assumptions in the context of equity and foreign exchange markets. Using moving window estimation, they showed that a complex distribution does not always outperform a simpler one.

The second category examines the predictive ability of different GARCH models with various volatility specifications. More specifically, some studies have found evidence supporting the Quadratic-GARCH (Engle & Ng, 1993) model for the case of stock returns predictions (Franses & van Dijk, 1996; Wei, 2002). Furthermore, other studies have found evidence in favor of the GJR-GARCH model (Brailsford & Faff, 1996; Taylor, 2004). Nevertheless, other researchers, including Heynen and Kat (1994), Chong, Ahmad, and Abdullah (1999) Loudon, Watt, and Yadav (2000), have came to the conclusion that the EGARCH model achieves superior performance in predicting stock market volatility. Furthermore, Awartani and Corradi (2005) and Evans and McMillan (2007) provided supportive evidence that GARCH models that allow for asymmetries in volatility produce more accurate volatility predictions. Overall, these studies explain that asymmetries play a crucial role in volatility forecasting.

For emerging stock market data, Gokcan (2000) found that the GARCH(1,1) model outperforms the EGARCH model, even if the stock market return series exhibit skewed distributions. McMillan, Speight, and Apgwilym (2000) suggested that EGARCH does not outperform the simple GARCH model in forecasting stock index volatility. Using the ARCH, GARCH, GJR-GARCH and EGARCH models, Balaban (2004) indicated that the standard GARCH model produces the best overall performance for forecasting monthly US dollar-Deutsche mark exchange rate volatility, whereas the GJR-GARCH model seems a poor alternative. Along the same lines, Ng and McAleer (2004) indicated that the GARCH(1,1) and GJR(1,1) models are superior to the RiskMetrics model in forecasting stock market volatility; however, neither GARCH(1,1) nor GJR(1,1) dominates the other. Despite extensive literature on volatility model evaluation, no consensus exists regarding the optimum model for providing which model has the optimal performance in forecasting volatility. In addition, none of them investigate the effect of either distributional assumption or volatility specification on the accuracy of forecasting financial market volatility. It is interesting to examine the predictive content of the above different possible sources of performance improvements.

To substitute a proxy for the true volatility is also crucial when evaluating the out-of-sample forecasting performances of competing GARCH models. The vast majority of previous articles have used squared daily returns as the proxy for *ex post* latent volatility (see, e.g. Awartani & Corradi, 2005; Brailsford & Faff, 1996; Brooks & Persand, 2002; Sadorsky, 2006). However, the squared returns represent a very noisy estimator of the daily variance and may result in poor GARCH model performance in forecasting volatility. For this reason, recent developments[3] have argued that the failure of GARCH models to produce accurate forecasts is not due to failures of the models *per se*, but rather to the inappropriate use of *ex post* squared returns with the same frequencies as the forecast evaluation measure of "true volatility". To prevent the drawing of incorrect conclusions, Andersen and Bollerslev (1997, 1998) and Andersen (2000) have advocated the use of high-frequency data (intraday) for extracting daily volatility, in cases where the so-called 'realized volatility' was obtainable by summing squared intraday returns. Thus, this study uses realized volatility rather than squared returns as a proxy for the latent underlying volatility process in evaluating the predictive ability of the considered volatility models.

Among the previous studies on volatility forecasting using GARCH-type models, none have explored the relative importance of the distributional assumption and asymmetric specification for improving volatility forecasting performance. This study complements previous works by examining the relative forecasting performance of distribution-type (GARCH-*N*, GARCH-*t*, GARCH-*HT* and GARCH-*SGT*) and asymmetry-type (GJR-GARCH and EGARCH) GARCH models. The empirical investigation is for the Standard and Poor's 100 (S&P-100) stock index series over the period 6 October 1997–15 September 2003, comprising a total of 1494 trading days. The forecasting performance of asymmetry-type and distribution-type GARCH models for the last 493 days of the data set is the focus of the empirical study. The forecasts are generated by a rolling-window of 1000 observations through the last 493 daily observations. The forecast is compared using two different loss functions, including the mean squared error (MSE) and the mean absolute error (MAE) statistics. The fact that a specific loss criterion is smallest for a particular model does not indicate the superiority of that model among its competitors for other samples within the data set and for future samples. For this reason, this study adopts the powerful superior predictive ability (SPA) test advocated by Hansen (2005) rather than the reality check test of White (2000), to examine the relative performance between distribution-type and asymmetry-type models.

Empirical results indicate that the GJR-GARCH model obtains the most accurate volatility forecasts. There is strong evidence that modeling asymmetric component is much more important than specifying the error distribution for improving the volatility forecasts of financial returns in the presence of fat-tails, leptokurtosis, skewness and the leverage effect. Furthermore, if asymmetric properties are neglected, the GARCH model with normal distribution is more preferable than those models with more sophisticated error distributions, suggesting that allowing for a flexible error distribution does not lead to significant improvements in volatility forecasts.

The remainder of this study is organized as follows: Section 2 outlines the GARCH-based volatility forecasting models. Section 3 describes the forecasting and evaluation methods. Data description and main findings are then reported in Section 4. Finally, summary and conclusion are presented in the last section.

2. The GARCH-based volatility forecasting models

Let $r_t = 100(\ln P_t - \ln P_{t-1})$ denote the continuously compounded rate of returns from time $t-1$ to t, where P_t is the price level of underlying assets at time t, and denotes the information set of all observed returns up to time $t-1$ by Ω_{t-1}.

2.1. GARCH genre of volatility models

The symmetric GARCH(1,1) model with a basic mean can be formulated as follows:

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t, \quad z_t|_{\Omega_{t-1}} \overset{iid}{\sim} F(0,1) \tag{1}$$

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

where μ and σ_t^2 denote the conditional mean and variance of returns, respectively. ε_t is the innovation process, while $F(0,1)$ is a density function with a mean of zero and a unit variance. Furthermore, ω, α and β are nonnegative parameters with the restriction

[3] See Andersen and Bollerslev (1997, 1998) and McMillan and Speight (2004) for dedicated discussions.

of $\alpha + \beta < 1$ to ensure the positive of conditional variance and stationarity as well.

Two simple classes of models that can cope with asymmetric volatility in response to asymmetric shocks are the GJR-GARCH model proposed by Glosten et al. (1993) and the so-called Exponential GARCH (EGARCH), advocated by Nelson (1991). The GJR-GARCH model differs from model (2) by:

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

where the indicator function I_{t-1} takes the value of unity if $\varepsilon_{t-1} < 0$, and 0 otherwise. The indicator variable differentiates between positive and negative shocks, so that asymmetric effects in the data are captured by δ. Thus, in the GJR-GARCH model, positive news has an impact of α, and negative news has an impact of $\alpha + \delta$, with negative (positive) news having a greater effect on volatility if $\delta > 0 (\delta < 0)$. Besides, ω, α and β are nonnegative parameters with the restriction of $\alpha + \beta + 0.5\delta < 1$, whereas the estimate of the sum $\alpha + 0.5\delta$ should still be positive (Ling & McAleer, 2002).

The EGARCH model of Nelson (1991) provides an alternative asymmetric model as follows:

$$\log(\sigma_t^2) = \omega + \alpha\left[\gamma\frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} - \sqrt{2/\pi}\right] + \beta\log(\sigma_{t-1}^2) \tag{4}$$

where the coefficient γ captures the asymmetric impact of news with negative shocks having a greater impact than positive shocks of equal magnitude if $\gamma < 0$, the volatility clustering effect is captured by a significant α. Finally, the use of the log form allows the parameters to be negative without conditional variance becoming negative.

2.2. Distributional assumptions

Another common finding in the GARCH literature is the non-normal characteristic of the empirical distribution of financial returns. To model such phenomenon when estimating the GARCH model, Theodossiou (1998) developed the skewed generalized t (*SGT*) distribution which provides a flexible tool for modeling the empirical distribution of financial data exhibiting fat-tails, leptokurtosis and skewness. The *SGT* probability density function (pdf) for the standardized residuals (z_t) of returns can be defined as follows:

$$F(z_t; N, \kappa, \lambda) = C\left(1 + \frac{|z_t + \delta|^\kappa}{((N+1)/\kappa)(1 + \mathrm{sign}(z_t + \delta)\lambda)^\kappa\theta^\kappa}\right)^{-(N+1)/\kappa} \tag{5}$$

where

$$C = 0.5\kappa \cdot \left(\frac{N+1}{\kappa}\right)^{-1/\kappa} \cdot B\left(\frac{N}{\kappa}, \frac{1}{\kappa}\right)^{-1} \cdot \theta^{-1} \tag{6}$$

$$\theta = (g - \rho^2)^{-1/2} \tag{7}$$

$$\rho = 2\lambda \cdot B\left(\frac{N}{\kappa}, \frac{1}{\kappa}\right)^{-1} \cdot \left(\frac{N+1}{\kappa}\right)^{1/\kappa} \cdot B\left(\frac{N-1}{\kappa}, \frac{2}{\kappa}\right) \tag{8}$$

$$g = (1 + 3\lambda^2) \cdot B\left(\frac{N}{\kappa}, \frac{1}{\kappa}\right)^{-1} \cdot \left(\frac{N+1}{\kappa}\right)^{2/\kappa} \cdot B\left(\frac{N-2}{\kappa}, \frac{3}{\kappa}\right) \tag{9}$$

$$\delta = \rho \cdot \theta \tag{10}$$

where z_t is the standardized residual with zero mean and unit variance; N is a tail-thickness parameter with constraint $N > 2$; κ is a leptokurtosis parameter with $\kappa > 0$; λ governs the skewness of the density obeying the constraint $|\lambda| < 1$; sign is the sign function and $B(\bullet)$ denotes the beta function. Notably, the *SGT* distribution nests several well-known distributions. Specifically, it gives for $\kappa = 2$ and $\lambda = 0$ the student-t distribution and for $N = \infty$, $\kappa = 2$ and $\lambda = 0$ the normal distribution.

However, several recent articles have reported more new results in favor of the heavy-tailed (*HT*) distribution, e.g. Politis (2004) and Hung et al. (2008). With a *HT* distribution, the pdf of the innovations becomes:

$$F(z_t, a_0, 1) = \frac{(1 + a_0 z_t^2)^{-1.5}\exp\left(-\frac{z_t^2}{2(1+a_0 z_t^2)}\right)}{\sqrt{2\pi}(\Phi(a_0^{-0.5}) - \Phi(-a_0^{-0.5}))} \tag{11}$$

where 1 denotes the standard deviation of z_t, and Φ denotes the cumulative probability density function of the standard normal distribution. The shape parameter, a_0, reflects the degree of the heavy tails with constraint $0 < a_0 < 1$. When $a_0 \to 0$, the *HT* will reduce to a standard normal distribution, while the distribution has a thicker tails than the normal when $a_0 \to 1$ (see Politis, 2004; for more details).

In the empirical investigation, four conditional distributions for the standardized residuals of returns innovations were considered: (i) a standard normal, (ii) a standardized student-t, (iii) a *SGT*, and (iv) a *HT* distribution. Accordingly, six competing model specifications in modeling volatility of the S&P-100 index returns can be constructed in our comparative analysis: GARCH-*N*, GARCH-*t*, GARCH-*HT*, GARCH-*SGT*, GJR-GARCH (or simply GJR-*N*) and EGARCH (or EGARCH-*N*). The parameter vector $\Theta = [\mu, \omega, \alpha, \beta, \ldots]$ is obtained from the maximization of the sample log-likelihood function, using QMLE (Quasi maximum likelihood estimation, QMLE) as follows:

$$LL(\Theta) = \sum_{t=1}^{T} \log F(\Theta) \tag{12}$$

where $F(\bullet)$ is the likelihood function of the GARCH models with various distributional assumptions.

3. Forecasting and evaluation methods

In this section, the methods are presented for an out-of-sample forecasting study in which we compare the forecasting performances of the asymmetry-type (GJR-GARCH and EGARCH) and distribution-type (GARCH-*N*, GARCH-*t*, GARCH-*HT*, GARCH-*SGT*) volatility models described in previous section. The forecasting procedure of these GARCH models is implemented as follows. The volatility models are estimated 493 times based on 493 samples of 1000 observations. The first sample starts 6 October 1997 and ends 28 September 2001. A forecast of volatility is generated for 1 October 2001 based on the estimated model for the first sample. The second sample, starting 7 October 1997 and ending 1 October 2001, is used to forecast the volatility of 2 October 2001. These estimation and forecasting steps can repeated 493 times for the available sample from 1 October 2001 to 15 September 2003. More specifically, we produce the 493 one-step-ahead volatility forecasts.

Since the true volatility can not be observed, it is important to choose the proxy of true volatility when evaluating out-of-sample volatility forecasting performance. As many empirical studies (Andersen & Bollerslev, 1997, 1998; Andersen, Bollerslev, & Lange, 1999; Hansen & Lunde, 2006) suggested, using realized volatility (RV) rather than the squared return as the proxy for the true volatility provides a consistent ranking of competing volatility models. The RV is constructed by taking the sum of squared intraday returns that yields an unbiased measure of the conditional variance. However, the accuracy of this volatility estimator is affected by market microstructure noise, particular when data are sampled at very high frequency. Therefore, to obtain a robust result, the realized volatilities of 5-, 15-, 30- and 60-min intervals are used to examine the forecasting performance among the competing models.[4]

3.1. Superior predictive ability test

Recent work has focused on a testing framework for determining whether a particular model is outperformed by another model. A further development of the White's reality check test (White, 2000) is known as the superior predictive ability (SPA) test and is proposed by Hansen (2005) where it is also shown that SPA has good power properties and is robust. In this study, we adopt the SPA test to examine the relative performance improvements between asymmetry-type and distribution-type GARCH models.

We consider $l+1$ different models M_k for $k=0,1,\ldots,l$ and which are discussed in previous section. M_0 is the benchmark model and the null hypothesis is that none of the models $k=1,2,\ldots,l$ outperforms the benchmark in terms of the specific loss function chosen. For each model M_k, we generate n volatility forecast $\widehat{h}_{k,t}$ for $t=1,2,\ldots,n$. For every forecast, we generate the loss function $L_{t,k}$ describing as follows. Let $L_{t,k}\equiv L(\widehat{\sigma}_t^2,\hat{h}_{k,t})$ denote the loss[5] if one makes the prediction $\hat{h}_{k,t}$ with k-th model when the realized volatility turns out to be $\widehat{\sigma}_t^2$. The performance of model k relative to the benchmark model (at time t), can be defined as:

$$f_{k,t}=L_{t,0}-L_{t,k} \text{ for } k=1,2,\ldots,l; \quad t=1,2,\ldots,n \tag{13}$$

Assuming stationarity for $f_{k,t}$, we can define the expected relative performance of model k relative to the benchmark as $\mu_k=E[f_{k,t}]$ for $k=1,2,\ldots,l$. If model w outperforms the benchmark, then the value of μ_w will be positive. Therefore, we can analyze whether any of the competing models significantly outperform the benchmark, testing the null hypothesis that $\mu_k\leqslant 0$, for $k=1,2,\ldots,l$. Consequently, the null hypothesis that none of models is better than the benchmark (i.e. no predictive superiority over the benchmark itself) can be formulated as:

$$H_0:\mu_{\max}\equiv\max_{k=1,\ldots,l}\mu_k\leqslant 0 \tag{14}$$

against the alternative that the best model is superior to the benchmark. By the law of large number we can consistently estimate μ_k with the sample average $\bar{f}_{k,n}=n^{-1}\sum_{t=1}^{n}f_{k,t}$ and then obtain the test statistic:

$$T_n\equiv\max_{k=1,\ldots,l}n^{1/2}\cdot\bar{f}_{k,n} \tag{15}$$

If the null hypothesis has been rejected, we have evidence that among the competing models, at least one is significantly better than the benchmark.

The most difficult problem is to derive the distribution of the statistic T_n under H_0, because the distribution is not unique. Hansen (2005) emphasizes that the RC test applies a supremum over the non-standardized performances T_n and, more dangerously, a conservative asymptotic distribution that makes it very sensitive to the inclusion of poor models. The author argues that since the distribution of the statistic is not unique under the null hypothesis, it is necessary to obtain a consistent estimate of the p-value, as well as a lower and an upper bound. Therefore, he applies a supremum over the standardized performances and tests the null hypothesis:

$$H_0:\mu_{\max}^s\equiv\max_{k=1,\ldots,l}\frac{\mu_k}{\sqrt{var(n^{1/2}\bar{f}_{k,n})}}\leqslant 0 \tag{16}$$

using the statistic:

$$T_n^s=\max_k\frac{n^{1/2}\bar{f}_{k,n}}{\sqrt{\widehat{var}(n^{1/2}\bar{f}_{k,n})}} \tag{17}$$

where $\hat{var}(n^{1/2}\bar{f}_{k,n})$ is an estimate of the variance of $n^{1/2}\bar{f}_{k,n}$ obtained by the bootstrap. Therefore, Hansen (2005) suggests additional refinements to the RC test and some modifications of its asymptotic distribution that result in tests less sensitive to the inclusion of poor models and with a better power. He also argues that the p-values of the RC are generally inconsistent (i.e. too large) and the test can be asymptotically biased. To overcome these drawbacks, Hansen (2005) shows that it is possible to derive a consistent estimate of the p-value along with an upper and a lower bound. Such a test is called SPA test and it includes the RC as a special case. The upper bound (SPA_u) is the p-value of a conservative test (i.e. it has the same asymptotic distribution as the RC test) where it is implicitly assumed that all the competing models ($k=1,2,\ldots,l$) are as good as the benchmark in terms of expected loss. Hence, the upper-bound p-value of the liberal test where the null hypothesis assumes that those models with worse performance than the benchmark are poor models in the limit. With the SPA test it is possible to assess which models are worse than the benchmark and asymptotically we can prevent them from affecting the distribution of the test statistic. The conservative test (and thus the RC test) is quite sensitive to the inclusion of poor and irrelevant models in the comparison, while the consistent (SPA_c) and the liberal test (SPA_l) are not.

4. Data description and empirical results

4.1. Data

The data for this study consists of Standard & Poor's 100 (S&P-100) stock market daily closing price index during the period from October 6, 1997 to September 15, 2003, which constitutes a total of 1494 observations. The data employed was retrieved from the database of Yahoo Finance website (http://finance.yahoo.com/). The forecasting performance of various volatility models for the last 493 days (from October 1, 2001 to September 15, 2003) of the data set is the focus of our out-of-sample evaluation and comparison.

Preliminary analysis of daily returns of S&P-100 cash index for the whole sample period is presented in Table 1. From panel A, the average daily returns are positive and very small compared with the variable's standard deviation. The daily returns display

Table 1
Preliminary analysis of S&P-100 daily returns.

Mean %	Std. Dev.	Skewness	Kurtosis	J-B	$Q^2(12)$	LM(12)
Panel A. Summary statistics						
0.006	1.4328	−0.0384	1.920*	229.706*	221.940*	113.028*
PP		Bandwidth		KPSS		Bandwidth
Panel B. Unit root tests						
−39.828*		16		0.295		17
Panel C. Engle and Ng (1993) test for asymmetric volatility						
Test statistic ($\sim\chi^2(3)$)						11.813*

* *Note:* 1. Denotes significantly at the 1% level. 2. J-B represents the statistics of Jarque and Bera (1987)'s normal distribution test. 3. $Q^2(12)$ denotes the Ljung-Box Q test for 12th order serial correlation of the squared returns. 4. LM test also examines for autocorrelation of the squared returns. 5. PP and KPSS are the test statistics for stationarity of return series. The PP-test rejects the null hypothesis of non-stationarity if the test statistic is negative and the absolute value of the test statistic exceeds the critical value of the respective significance level: 1%: −3.435; 5%: −2.864; 10%: −2.568. The KPSS-test rejects the null hypothesis of stationarity if the test statistic exceeds the critical value of the respective significance level: 1%: 0.739; 5%: 0.463; 10%: 0.347.

[4] The realized volatility of S&P-100 for 5, 15, 30 and 60 min used in this study are the same as Marcucci (2005) and are available from the journal website: "Studies in Nonlinear Dynamic & Econometrics ".

[5] The loss functions used in this paper are MSE and MAE, and the function forms are respectively $L_{k,t}=(\hat{\sigma}_t^2-\hat{h}_{k,t})^2$ and $L_{k,t}=|\hat{\sigma}_t^2-\hat{h}_{k,t}|$.

Table 2
Alternative model estimates.

Parameter	GARCH-*N*	GARCH-*t*	GARCH-*HT*	GARCH-*SGT*	GJR-GARCH	EGARCH
μ	0.0703^{b} [0.0334]	0.0628^{a} [0.0377]	0.0609 [0.0394]	0.0477 [0.0362]	0.0081 [0.0296]	0.0014 [0.0386]
ω	0.2070^{c} [0.0219]	0.1266^{c} [0.0435]	0.0916^{c} [0.0101]	0.1127^{c} [0.0311]	0.1600^{c} [0.0083]	0.0289^{c} [0.0098]
α	0.1224^{c} [0.0198]	0.0939^{c} [0.0223]	0.0740^{c} [0.0069]	0.0941^{c} [0.0213]	-0.0384^{c} [0.0055]	0.0905^{c} [0.0072]
β	0.7739^{c} [0.0093]	0.8441^{c} [0.0347]	0.8547^{c} [0.0075]	0.8521^{c} [0.0300]	0.8261^{c} [0.0061]	0.9452^{c} [0.0141]
δ	–	–	–	–	0.2649^{c} [0.0158]	–
γ	–	–	–	–	–	-1.8161^{c} [0.1642]
N	∞	8.2370^{c} [1.8994]	–	5.8305^{c} [1.5980]	–	–
κ	2	2	–	2.4088^{c} [0.3363]	–	–
λ	0	0	–	-0.0718^{a} [0.0429]	–	–
a_0	–	–	0.0535^{c} [0.0096]	–	–	–
$Q^2(12)$	15.486	12.804	12.120	11.847	13.122	10.332
LL	−1644.68	−1630.65	−1630.29	−1628.73	−1613.88	−1601.72

Notes: 1. Standard errors for the estimators are included in parentheses. 2. a, b, c indicate significantly at the 10%, 5% and 1% level, respectively. 3. $Q^2(12)$ is the Ljung-Box Q test for serial correlation in the squared standardized residuals with 12 lags. 4. LL refers to the log-likelihood value.

evidence of skewness and kurtosis. Namely, the returns series is skewed towards the left, while the returns series is characterized by a distribution with tails that are significantly thicker than for a normal distribution. J-B test statistic further confirms that the daily return is non-normal distributed. Moreover, the Q^2 and LM-test statistics display linear dependence of squared returns and strong ARCH effects. Accordingly, these preliminary analyses of the data encourage the adoption of a sophisticated distribution, which embody fat-tailed features, and of conditional models to allow for time-varying volatility. Panel B of Table 1 reports Phillips and Perron (1988) (PP) unit root tests and KPSS (Kwiatkowski, Phillips, Schmidt, & Shin, 1992) unit root tests. The test results indicate no evidence of non-stationarity in the S&P-100 cash index returns. Finally, the test statistic of Engle and Ng (1993) indicates returns volatility exhibiting asymmetric behavior, suggesting the adoption of more flexible volatility treatments by accommodating the asymmetric responses of volatility to positive and negative shocks.

4.2. Model estimates

In this study, the parameters are estimated by quasi maximum likelihood estimation (QMLE) in terms of the BFGS optimization algorithm using the econometric package of WinRATS 7.0. Model estimates and diagnostic tests for S&P-100 index returns during the in-sample period are provided in Table 2.

As shown in Table 2, the parameters, ω, α, β, and, δ in the conditional variance equations, are all positive and found to be highly significant. Meanwhile, the symmetric GARCH component exhibits the existence of strong volatility persistence in the stock returns, as the $\alpha + \beta \approx 1$. A notable point in Table 2 is that the parameter δ of the conditional volatility equation in GJR-GARCH model is positive and highly significant, implying that negative shocks (bad news) exert larger impact on S&P-100 volatility than positive shocks (good news) of the same magnitude. Similarly, the estimated value, γ, of the EGARCH model is significantly negative, indicating asymmetric impact of news on returns volatility. Furthermore, the estimated values for the specific shape parameters N, κ, λ, and, a_0, are at least statistically significant at the 10% level, confirming the presence of fat-tails, leptokurtosis and skewness features in the returns series. Turing the discussion to diagnostic tests, the Ljung-Box Q statistic indicates that the linear or non-linear GARCH specifications in these models are sufficient to correct the serial correlation of the returns series in the conditional variance equation. Subsequently, a comparative evaluation of the predictive performance of the competing models will be carried out in the next section.

4.3. SPA test results of asymmetry-type and distribution-type models

The first column in Panel A (Panel B) of Table 3 presents the means of the MSE (MAE) function in which the true volatility is proxied by 30-min intervals.[6] In terms of MSE and MAE, the asymmetry-type GARCH models consistently produce smaller values than the distribution-type ones for this sample. In particular, GJR-GARCH model generates smallest loss function values among the various competitors, which is consistent with Marcucci (2005). In addition, Awartani and Corradi (2005) also found that the GARCH-*N* model is outperformed by the EGARCH and GJR-GARCH models across different forecast horizons when using MSE in evaluating the volatility forecasting performance of S&P-500 index. Notably, the MSE give relative more weight to forecast errors associated with high values of realized volatility as compared to the MAE. Moreover, the MSE loss function is similar to the quadratic and regulatory loss functions commonly used by the value-at-risk practitioners as the efficiency measurement. Therefore, using MSE to evaluate the forecast errors here has some implications for risk management.

As indicated by Inoue and Kilian (2004), out-of-sample loss function comparison does not give unusual meaningful information about the relative predictive ability of the various competitors, although it provides an initial overview of model performance. Besides, a forecasting model with the smallest loss function value does not imply the superiority of that model among its competitors (Hansen & Lunde, 2005; Koopman, Jungbacker, & Hol, 2005). Such a conclusion cannot be made on the basis of just one criterion and just one sample. To avoid the problem of data-snooping, this study employs the SPA test[7] of Hansen (2005) to assess the relative forecasting performance between asymmetry-type and distribution-type GARCH models. The second and third columns in Panel A and Panel B of Table 3 present respectively the *p*-values of the Hansen's consistent test (SPA_c) and the liberal test (SPA_l) for the two classes of volatility models. The individual $\bar{f}_{k,n}$ is reported for each model that is considered as a benchmark model M_k. The *p*-values are computed using the stationary bootstrap of Politis and Romano (1994) generating 1000 bootstrap re-samples with dependency parameter $q = 0.5$.[8]

[6] The forecasting results for 5, 15 and 60 min cases are quite similar to that obtained using realized volatility calculated with 30-min returns and aggregated, and are not reported here.

[7] The program code of SPA test is downloaded from the website of Peter Reinhard Hansen. The *p*-values in SPA test are calculated by the stationary bootstrap of Politis and Romano (1994) as in White (2000) and Hansen and Lunde (2005).

[8] The number of bootstrap re-samples B and the dependency parameter *q* is selected to be B = 1000, 2000, 3000 and *q* = 0.25, 0.5, 0.75 while implementing SPA test and only the result of B = 1000 and *q* = 0.5 is reported in Tables 3 and 4. The other results are available upon request.

Table 3
SPA test results of asymmetry-type and distribution-type volatility models.

Benchmark (M_0)	MSE	Rank	SPA$_c$	Rank	SPA$_l$	Rank
Panel A. Performance based on mean squared error (MSE)						
GARCH-*N*	2.874	3	0.029	3	0.023	4
GARCH-*t*	2.898	4	0.025	4	0.025	3
GARCH-*HT*	2.930	5	0.007	6	0.007	6
GARCH-*SGT*	2.967	6	0.010	5	0.010	5
GJR-GARCH	2.308	1	0.802	1	0.525	1
EGARCH	2.367	2	0.200	2	0.200	2
Benchmark	MAE	Rank	SPA$_c$	Rank	SPA$_l$	Rank
Panel B. Performance based on mean absolute error (MAE)						
GARCH-*N*	1.111	3	0.004	4	0.003	4
GARCH-*t*	1.116	4	0.003	5	0.003	5
GARCH-*HT*	1.122	5	0.002	6	0.002	6
GARCH-*SGT*	1.133	6	0.006	3	0.006	3
GJR-GARCH	0.996	1	0.876	1	0.532	1
EGARCH	1.016	2	0.125	2	0.125	2

Notes: 1. The latent volatility is proxied by the realized volatility calculated with 30-min returns and aggregated. 2. SPA$_c$ and SPA$_l$ denote the reality check *p*-values of the Hansen's consistent test and Hansen's liberal test, respectively. The null hypothesis is that none of the models is better than the benchmark. 3. The number of bootstrap replications to calculate the *p*-values is 1000 and the dependency parameter *q* is 0.5.

All models considered in this study are consecutively taken as benchmark models in order to evaluate whether a particular model (benchmark) is significantly outperformed by other competing models. A high *p*-value indicates that the null hypothesis ($H_0 : \mu_{\max} \equiv \max_{k=1,\dots,l}\mu_k \leqslant 0$) cannot be rejected, which means that the benchmark model is not outperformed by the competing models. For example, in Panel A of Table 3, when the GJR-GARCH is the benchmark Hansen's consistent *p*-value is 0.802, indicating the null hypothesis is not rejected. Namely, no model performs better than the GJR-GARCH model. On the other hand, when the GARCH-*HT* is selected as the benchmark, the null hypothesis is significantly rejected (SPA$_c$ = 0.007). Accordingly, there exists a better model that dominates the GARCH-*HT* model.

In Table 3, for both loss functions employed, the *p*-values of SPA$_c$ and SPA$_l$ clearly show that the distribution-type GARCH models are outperformed by asymmetry-type ones, suggesting a need for volatility models that take account of asymmetric dynamics. In particular, among the asymmetry-type GARCH models, the GJR-GARCH model achieves its superiority to the EGARCH model since the former model always produces higher *p*-value than the latter model for all cases. In fact, Hansen and Lunde (2005) found that GARCH model (GARCH-*N* and GARCH-*t*) is apparently outperformed by models that account for a leverage effect in the case of IBM returns. Moreover, Awartani and Corradi (2005) and Evans and McMillan (2007) also concluded that GARCH-class of models that allow for asymmetries in the volatility process produce more accurate volatility predictions. Therefore, the results here are consistent with Hansen and Lunde (2005) and partially in line with Awartani and Corradi (2005) and Evans and McMillan (2007). In sum, our empirical results clearly indicate that modeling the asymmetric component is much more important than specifying the error distribution for improving volatility forecasts of financial returns in the presence of fat-tails, leptokurtosis, skewness and leverage effect.

4.4. SPA test results of distribution-type models

Wilhelmsson (2006) employed nine error distributions on S&P-500 to explore whether volatility forecasting performance of GARCH-based model can be improved by specifying more flexible error distributions. To further compare the results with Wilhelmsson (2006), the asymmetry-type models in Table 3 are thus removed to re-examine the relative performance among GARCH models with different error distributions and the results are reported in Table 4. The results show that the GARCH-*N* model is not outperformed by the competing models with returns innovations that allow for fat-tailed, leptokurtic and skewed characteristics.[9] These findings are consistent with the results of Chuang et al. (2007), which showed that a complex distribution does not always outperform a simpler one. Nevertheless, our results contradict earlier findings of Wilhelmsson (2006), which argued that the GARCH model with a *t*-distribution improves volatility forecasts.[10] The possible interpretations of the mixed conclusions are that previous researches adopted a noisy true volatility proxy, such as squared returns, with distinct criterions to evaluate the forecasting performances, and, in particular, lacking for a powerful and robust test to reveal statistical significance among competing models.

Table 4
SPA test results of distribution-type volatility models.

Benchmark (M_0)	MSE	Rank	SPA$_c$	Rank	SPA$_l$	Rank
Panel A. Performance based on mean squared error (MSE)						
GARCH-*N*	2.874	1	0.947	1	0.531	1
GARCH-*t*	2.898	2	0.053	2	0.053	2
GARCH-*HT*	2.930	3	0.004	4	0.004	4
GARCH-*SGT*	2.967	4	0.007	3	0.007	3
Benchmark	MAE	Rank	SPA$_c$	Rank	SPA$_l$	Rank
Panel B. Performance based on mean absolute error (MAE)						
GARCH-*N*	1.111	1	0.939	1	0.519	1
GARCH-*t*	1.116	2	0.061	2	0.061	2
GARCH-*HT*	1.122	3	0.007	4	0.007	4
GARCH-*SGT*	1.133	4	0.013	3	0.013	3

Notes: 1. The latent volatility is proxied by the realized volatility calculated with 30-min returns and aggregated. 2. SPA$_c$ and SPA$_l$ denote the reality check *p*-values of the Hansen's consistent test and Hansen's liberal test, respectively. The null hypothesis is that none of the models is better than the benchmark. 3. The number of bootstrap replications to calculate the *p*-values is 1000 and the dependency parameter *q* is 0.5.

5. Summary and conclusion

Volatility is a fundamental variable in both theoretical and practical applications owing to its central role in option pricing and risk management. This study explores the relative importance of the distributional assumption and the asymmetric specification in improving volatility forecasting performance. This study empirically investigates the one-step-ahead forecasting performance of asymmetry-type and distribution-type GARCH models for the S&P-100 stock index over the period 1 October 2001–15 September 2003. Using the formal SPA test advocated by Hansen (2005), several salient points have emerged from the current study. First, the GJR-GARCH generates the most accurate volatility forecasts, closely followed by the EGARCH model, indicating that asymmetric specification of volatility dynamics needs to be taken into account. Second, the analytical results clearly indicate that modeling the asymmetric component is much more important than specifying the error distribution for improving volatility forecasts of financial returns in the presence of fat-tails, leptokurtosis, skewness and the leverage effect. Finally, in cases involving distribution-type GARCH models, out-of-sample volatility forecasting performance is found not to be significantly improved using student-*t*, *HT* or *SGT* error

[9] Franses and Ghijsels (1999) obtained similar results that GARCH model with normal distribution outperforms GARCH model with a *t*-distribution when assessing with mean and median squared error loss functions.

[10] Hamilton and Susmel (1994) found that the regime-switching GARCH model with a *t*-distribution outperforms other models when evaluating with a logarithm loss criterion.

distributions. Restated, if asymmetric properties are neglected, the GARCH model with normal distribution is preferable to those models with more sophisticated error distributions.

Acknowledgements

The authors are grateful to the editor, Jay Liebowitz, and anonymous referees for their valuable comments and suggestions on an earlier draft that have improved the article. All remaining errors are our responsibility. Hung-Chun Liu also partially acknowledges the National Science Council of the Republic of China, Taiwan for financially supporting this research under Contract No. NSC98-2410-H-159-001.

References

Andersen, T. G. (2000). Some reflections on analysis of high-frequency data. *Journal of Business and Economic Statistics, 18*, 146–153.
Andersen, T. G., & Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance, 4*, 115–158.
Andersen, T. G., & Bollerslev, T. (1998). Answering the Skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review, 39*(4), 885–905.
Andersen, T. G., Bollerslev, T., & Lange, S. (1999). Forecasting financial market volatility: Sample frequency vis-à-vis forecast horizon. *Journal of Empirical Finance, 6*, 457–477.
Awartani, B. M. A., & Corradi, V. (2005). Predicting the volatility of the S&P-500 stock index via GARCH models: The role of asymmetries. *International Journal of Forecasting, 21*, 167–183.
Balaban, E. (2004). Comparative forecasting performance of symmetric and asymmetric conditional volatility models of an exchange rate. *Economics Letters, 83*, 99–105.
Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*, 307–327.
Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics, 69*, 542–547.
Brailsford, T. J., & Faff, R. W. (1996). An evaluation of volatility forecasting techniques. *Journal of Banking and Finance, 20*, 419–438.
Brooks, C., & Persand, G. (2002). Model choice and Value-at-Risk performance. *Financial Analysts Journal, 58*, 87–97.
Chong, C. W., Ahmad, M. I., & Abdullah, M. Y. (1999). Performance of GARCH models in forecasting stock market volatility. *Journal of Forecasting, 18*, 333–343.
Chuang, I. Y., Lu, J. R., & Lee, P. H. (2007). Forecasting volatility in the financial markets: A comparison of alternative distributional assumptions. *Applied Financial Economics, 17*, 1051–1060.
Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of variance of UK inflation. *Econometrica, 50*, 987–1008.
Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance, 48*, 1749–1778.
Evans, T., & McMillan, D. G. (2007). Volatility forecasts: The role of asymmetric and long-memory dynamics and regional evidence. *Applied Financial Economics, 17*, 1421–1430.
Franses, P. H., & Ghijsels, H. (1999). Additive outliers, GARCH and forecasting variance. *International Journal of Forecasting, 15*, 1–9.
Franses, P. H., & van Dijk, R. (1996). Forecasting stock market volatility using (non-linear) GARCH models. *Journal of Forecasting, 15*, 229–235.
Glosten, L., Jagannathan, R., & Runkle, D. (1993). On the relation between the expected value and the volatility nominal excess return on stocks. *Journal of Finance, 46*, 1779–1801.
Gokcan, S. (2000). Forecasting volatility of emerging stock market: linear versus non-linear GARCH models. *Journal of Forecasting, 19*, 499–504.
Hamilton, J., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics, 64*, 307–333.
Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business and Economic Statistics, 23*(4), 365–380.
Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics, 20*, 873–889.
Hansen, P. R., & Lunde, A. (2006). Consistent ranking of volatility models. *Journal of Econometrics, 131*, 97–121.
Heynen, R. C., & Kat, H. M. (1994). Volatility prediction: A comparison of stochastic volatility, GARCH(1,1) and EGARCH models. *Journal of Derivatives, 2*, 50–65.
Hung, J. C., Lee, M. C., & Liu, H. C. (2008). Estimation of Value-at-Risk for energy commodities via fat-tailed GARCH models. *Energy Economics, 30*, 1173–1191.
Inoue, A., & Kilian, L. (2004). In sample or out of sample tests for predictability: Which one should we use? *Econometric Reviews, 23*, 371–402.
Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistics Review, 55*, 163–172.
Koopman, S. J., Jungbacker, B., & Hol, E. (2005). Forecasting daily volatility of the S&P-100 stock index using historical, realized and implied volatility measurements. *Journal of Empirical Finance, 12*, 445–475.
Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics, 54*, 159–178.
Ling, S., & McAleer, M. (2002). Stationarity and the existence of moments of a family of GARCH processes. *Journal of Econometrics, 106*, 109–117.
Loudon, G. F., Watt, W. H., & Yadav, P. K. (2000). An empirical analysis of alternative parametric ARCH models. *Journal of Applied Econometrics, 2*, 117–136.
McMillan, D. G., & Speight, A. E. H. (2004). Daily volatility forecasts: Reassessing the performance of GARCH models. *Journal of Forecasting, 23*, 449–460.
McMillan, D. G., Speight, A. E. H., & Apgwilym, O. (2000). Forecasting UK stock market volatility. *Applied Financial Economics, 10*, 435–448.
Marcucci, J. (2005). Forecasting stock market volatility with regime-switching GARCH models. *Studies in Nonlinear Dynamics and Econometrics, 9*, 1–53.
Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica, 59*, 347–370.
Ng, H. G., & McAleer, M. (2004). Recursive modelling of symmetric and asymmetric volatility in the presence of extreme observations. *International Journal of Forecasting, 20*, 115–129.
Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika, 75*, 335–346.
Politis, D. N. (2004). A heavy-tailed distribution for arch residuals with application to volatility prediction. *Annals of Economics and Finance, 5*, 283–298.
Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association, 89*, 1303–1313.
Sadorsky, P. (2006). Modeling and forecasting petroleum futures volatility. *Energy Economics, 28*, 467–488.
Taylor, J. W. (2004). Volatility forecasting with smooth transition exponential smoothing. *International Journal of Forecasting, 20*, 273–286.
Theodossiou, P. (1998). Financial data and the skewed generalized *t* distribution. *Management Science, 44*, 1650–1661.
Wei, W. (2002). Forecasting stock market volatility with non-linear GARCH models: A case for China. *Applied Economics Letters, 9*, 163–166.
White, H. (2000). A reality check for data snooping. *Econometrica, 68*, 1097–1126.
Wilhelmsson, A. (2006). GARCH forecasting performance under different distribution assumptions. *Journal of Forecasting, 25*, 561–578.