

UNIVERSITY OF VAASA
FACULTY OF BUSINESS STUDIES
DEPARTMENT OF ACCOUNTING AND FINANCE

Sun Yimo

**EVALUATION OF VOLATILITY MODELS: EVIDENCE FROM CHINESE
EQUITY MARKETS**

Master's Thesis in

Finance

VAASA 2017

TABLE OF CONTENTS

1.INTRODUCTION	8
1.1.Purpose of This Thesis	9
1.2.Structure of This Thesis	9
2.PREVIOUS STUDIES	11
3.VOLATILITY	15
3.1.The Definition of Volatility	15
3.2.Conditional and Unconditional Variance	16
3.3.Volatility Clustering	16
3.4.Mean Reverting	17
3.5.Asymmetric Impact	17
3.6.Long Memory	17
3.7.Exogenous Variables	18
3.8.Leptokurtosis	19
4.VOLATILITY FORECASTING MODELS	20
4.1.Reasons for Choosing	20
4.2.Basic Notion	21
4.3.Random Walk (RW)	23
4.4. RiskMetrics Exponentially Weighted Moving Average (RiskMetrics EWMA)	23
4.5.Generalized Autoregressive Conditional Heteroskedasticity (GARCH)	24
4.6.GARCH-In-Mean (GARCH-M)	25
4.7.Exponential GARCH (EGARCH)	26
4.8.Threshold GARCH (TGARCH)	26
4.9.Asymmetric Power ARCH (APRCH)	27
4.10. Fractionally integrated GARCH (FIGARCH)	28
5.EVALUATION OF VOLATILITY FORECAST	29

5.1.The Root Mean Square Error (RMSE)	29
5.2.The Mean Absolute Error (MAE)	30
5.3.The Mean Absolute Percentage Error (MAPE)	30
5.4.The Theil Inequality Coefficient (Theil)	31
5.5.The Mean Mixed Error (MME)	31
5.6.The Mincer-Zarnowitz (MZ) Regression	32
5.7.Forecast Encompassing	33
6.DATA	34
7.METHODOLOGY	36
8.EMPIRICAL RESULTS	39
8.1.Preliminary Data	39
8.2.Symmetric Forecast Errors	40
8.3.Asymmetric Forecast Errors	43
8.4.The Mincer-Zarnowitz Regression	46
8.5.Forecast Encompassing	46
9.SUMMARY AND CONCLUSION	49
REFERENCES	52
APPENDICES	60

LIST OF TABLES

Table 1. Descriptive statistics of return series.	38
Table 2. Unit root tests.	39
Table 3. ARCH Lagrange Multiplier tests.	39
Table 4. Relative forecast error statistics for SSE-index.	40
Table 5. Relative forecast error statistics for SZSE-index.	40
Table 6. Ranks of relative forecast error statistics for SSE index.	41
Table 7. Ranks of relative forecast error statistics for SZSE index.	41
Table 8. MME statistics and numbers of over- and under- predictions for SSE index.	44
Table 9. MME statistics and numbers of over- and under- predictions for SZSE index.	44
Table 10. Coefficients of determination for Mincer-Zarnowitz.	45
Table 11. Encompassing test with the RW model as the benchmark.	46
Table 12. Encompassing test with the RiskMetrics model as the benchmark.	47

LIST OF FIGURES

Figure 1. The trend graph of daily closing prices for SSE and SZSE.	59
Figure 2. The trend graph of returns for SSE and SZSE.	59
Figure 3. Normal Quantile-Quantile Plots for the SSE and SZSE.	59
Figure 4. Graphs of the true and forecasted volatility for the SSE index.	60
Figure 5. Graphs of the true and forecasted volatility for the SZSE index.	61

UNIVERSITY OF VAASA**Faculty of Business Studies**

Author:	Sun Yimo
Topic of the Thesis:	Evaluation of Volatility Models: Evidence from Chinese Equity Markets
Name of the Supervisor:	Anupam Dutta
Degree:	Master of Science in Economics and Business Administration
Bachelor's/Master's Programme:	Finance
Year of Entering the University:	2015
Year of Completing the Thesis:	2017

pages:62

ABSTRACT

This thesis aims to find the most appropriate model to estimate and forecast volatility in Chinese stock markets, and to investigate the differences between simple historical models and GARCH-type models.

The studied model collection includes seven models: Random Walk, RiskMetrics EWMA, GARCH, GARCH-in-mean, EGARCH, TGARCH and APARCH. The forecast performances of those models are then evaluated in seven different criteria including symmetric loss functions and asymmetric loss functions. Other measurements such as the forecast encompassing test is conducted to check whether GARCH-type models carry additional information than simple historical models. The whole evaluation process is conducted with two Chinese stock markets' indices, namely the SSE composite index and the SZSE component index. The selected sample period with updated data spans from 04 March 2006 through 30 December 2016.

The empirical evidence shows that the Random Walk model has the worst performance among all studied models. Model Performance is highly sensitive to the choice of forecast error statistics. The asymmetric loss function suggests systematically over-prediction exists in the forecasts which might be caused by the choice of forecast period. GARCH models carry more information than the Random Walk model. But no significant evidence is found in this study to support that GARCH models carry additional information than the RiskMetrics EWMA model.

KEYWORDS: Volatility forecasting, Chinese stock markets, GARCH

1. INTRODUCTION

China is one of the biggest and important emerging markets in the world. Because of its rapidly developing economic during recent years and its different political system compares to major western countries, China has been continuously drawing researchers' attention on every aspect of its economic.

Financial market is probably the most comprehensive market within a country's economic. Along with the development of national economic, China's financial market is rapidly developing as well. New financial instruments have been allowed to open to market players to satisfy their various financing demands more than before.

Volatility has its important role in financial market. The applications of volatility such as risk management, assets allocation, derivatives pricing, hedging and policy making all need a proper measure of past volatility or even estimation of future volatility. However, volatility itself cannot be observed directly from the market. Therefore, how to find a proper model to forecast volatility has attracted concerns of both market players and academic researchers. Numerous previous studies were conducted to evaluate the forecasting power of different models whereas different data and various evaluation methods led to inconsistent results.

Seven volatility forecast models are evaluated in this thesis. All of them come from two major volatility model categories: Historical volatility models and GARCH models.

1.1. Purpose of This Thesis

The main purpose of this thesis is to evaluate volatility forecasting performance of different types of volatility forecast models in Chinese stock market, specifically with two major indices: SSE Composite Index (SHA: 000001) and SZSE Component Index (SHE: 399001). The second target is to augment previous studies and contribute more empirical results to the relevant literatures. If consistent results can be drawn by using the latest returns data, we can confirm the validity of some previous studies.

Even though GARCH-type models are expected to outperform those simple historical models because theoretically they are able to capture more styled facts of stock return volatility than historical models. However, the empirical results from previous studies do not always support this idea.

There are two main hypotheses we are going to test in this thesis:

H₁: The Random Walk model provides the worst forecast among all studied volatility models in this thesis.

H₂: GARCH-type models provide better volatility forecasts than the RiskMetrics EWMA model in Chinese stock market.

1.2. Structure of This Thesis

The remained content of this thesis is organized as following. Chapter two reviews several previous studies which relating to our topic. Chapter three to chapter five cover the

essential theoretical framework relating to the research issue. Chapter three gives basic conceptions and definitions about volatility and chapter four introduces the forecasting models which we used in this thesis. Chapter five explains the methods we used to evaluate the performance of the estimated results as well the procedure of evaluation. Chapter six briefly describes the data collection. Chapter seven shows the methodology which applied in this study while chapter eight reports the major empirical results. Finally, chapter nine states the conclusion of this thesis and some relevant suggestions for further research.

2. PREVIOUS STUDIES

Since volatility is very useful to both practitioners and researchers, people never stop investigating the most efficient and accurate method to describe and predict it. By a comprehensive review of Poon and Granger (2003), it concludes four major types of volatility forecast methods. The first type is historical volatility models (HISVOL), which includes random walk model (RW), historical averages of squared returns, or absolute returns, time series models based on historical volatility using moving averages (MA), exponential weights moving average (EWMA), and autoregressive models. The second type is GARCH models. Any model based on ARCH, GARCH, EGARCH families are included in this category. The third type is option implied standard deviation based on the Black-Scholes model and other various generations. The fourth type is stochastic model forecasts. Due to the topic and methodology of this paper, only the first two types of previous studies will be discussed.

On one hand, some people prefer a simple method like EMWA or HISVOL. Taylor (1987) compares EMWA with ARCH and simple historical average. It is one of the earliest studies in ARCH class forecasts. He shows that EWMA type of non-stationary series have the advantage of having fewer parameter estimates and respond to variance change quicker than ARCH. Tse (1991) confirms Taylor's result by comparing EMWA with ARCH/GARCH models on the Tokyo Stock Exchange in the period 1986 through 1989. By using dummy variables in mean equation to control for 1987 crash, Tse finds ARCH/GARCH models are slow to react to sudden volatility change but EWMA reacts to changes very quickly. Tse and Tung (1992) again use simple historical average, EWMA and GARCH to forecast volatility in the Singapore stock market and the results show the EWMA method is superior.

Boudoukh, Richardson and Whitelaw (1997) finds that EWMA and MDE have comparable performance and they are better than simple historical average and GARCH model. Figlewski (1997) points out that the length of estimation horizon has a positive relationship with the accuracy of forecast results. When forecasting with the longest estimation period, in his paper which is 60 months, simple historical average performs better than GARCH.

McMillan, Speight and Apgwilym (2000) provides a comparative evaluation of the ability of ten econometric models to forecast the volatility of the UK indices. Then they use ME, MAE, RMSE for symmetry loss function to evaluate the performance of those models. They conclude that actual volatility is proxied by mean adjusted squared returns, which is likely to be extremely unclear and noisy. Evaluation conducted on variance so forecast error statistics are quite close. Under the symmetric loss evaluation, the random walk model is vastly superior monthly volatility forecasts. While GARCH, moving average, and exponential smoothing models provide marginally superior daily volatility forecasts, RW, MA, and recursive smoothing models provide moderately superior weekly volatility forecasts.

On the other hand, there are quite many researchers who support ARCH/GARCH models as well. These models are considered more sophisticated than simple historical average and EWMA because they could capture more features of time-series data.

Akigray (1989) conducts ARCH(2), GARCH(1,1), EWMA and simple historical average models on EW indices and concludes that GARCH is most accurate and produced best

forecast especially in high volatility periods. He finds GARCH consistently outperforms other models in all subperiods and under all evaluation procedures.

Cumby, Figlewski and Hasbrouck (1993) compare between EGRACH and simple historical mean model. And their conclusion is that in both regressions and directional tests of out-of-sample forecasting ability, EGARCH seems to be superior than simple historical mean model. Though overall explanatory power is weak. Yu (2002) evaluates the performance of several alternative models using daily New Zealand data. The result shows the GARCH(3,2) model has the best performance within ARCH family, and it is sensitive to the choice of assessment measures. The EWMA model does not perform well according to any assessment method.

So and Yu (2006) use VaR estimation to evaluate volatility models. They conduct the assessment on the EWMA model and three GARCH-type models based on standardized normal and t distribution assumptions on residuals. It turns out that both stationary and fractionally integrated GARCH outperform the EWMA model in estimating 1% VaR.

Besides above two sides of researchers, there are some other voices. Some who hold the view that no single method is superior than others. The performance depends on the specific situation. After comparing simple historical models and ARCH models, Brailsford and Faff (1996) conclude there is no single model is clearly superior, the performance is sensitive to the choice of error statistics. Brooks (1998) also compares simple historical models with ARCH models by using DJ Composite index. In his paper, performances are similar across models especially when 1987's crash is excluded. Sophisticated models like GARCH do not dominate. Moreover, McMillan and Kambouroudis (2009) conduct their research on 31 different stock markets. The result

shows that no model totally outperforms all others in all markets. APARCH model performs better in G7 and Europe countries while the EWMA model does better in Asian market.

Some researchers have done similar researches by using Chinese market data. Song, Liu and Peter (1998) suggest that both Shanghai and Shenzhen indices returns may be best explained by the GARCH-M(1,1) specification with the mean equations of ARMA(6,6) for Shanghai and ARMA(10,10) for Shenzhen. Frank, Tunaru and Wu (2004) report that the GARCH(1,1) fits well on Shenzhen market while the TAGRACH(1,1) does better on Shanghai market index. Zhang and Pan (2006) explores both EWMA and GARCH-type models for predicting the daily volatility of Shanghai and Shenzhen market by using 3 different distribution assumptions. They find that the results are sensitive to the evaluation method. However, the RW model performs the worst in any situation.

Among all these various studies, we can conclude that there is no simple answer to our question. It seems no model is absolutely superior than others. Sophisticated models do not always perform better than simple ones. Results may vary from different input settings. Other factors such as asset type, assessment measure, forecast horizon, market location are all matters.

3. VOLATILITY

3.1. The Definition of Volatility

Hull (2012: 303 - 304) states that volatility, σ , of a stock is a measure of uncertainty about the returns provided by the stock. And the volatility of a stock price can be defined as the standard deviation of the return provided by the stock in one year when the return is expressed using continuous compounding.

Volatility is unobservable. We can only ever estimate and forecast volatility, and this only within the context of an assumed statistical model. So, there is no absolute 'true' volatility: what is 'true' depends only on the assumed model (Alexander 2008: 92 - 93).

It should be noted the reason that create volatility may not be the new information reaching the market. There is no research supporting that. By the research of Roll (1986) on orange juice futures, volatility is to a large extent caused by trading itself.

In this thesis, we follow the definition of volatility given by Alexander (2008: 90): "The volatility of an asset is an annualized measure of dispersion in the stochastic process that is used to model the log returns."

3.2. Conditional and Unconditional Variance

To understand a GARCH model, it is critical to understand the differences between conditional and unconditional variance of a time series of returns. For the unconditional variance, we assume the unconditional returns distributed constant though the whole data period considered. On the other hand, the conditional variance changes at every point of time because the history of returns changes all the time. (Alexander 2008: 131 - 132)

3.3. Volatility Clustering

First mentioned by Mandelbrot (1963), it is the observation that "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes." This observation is also confirmed by other researchers. Fama (1965) reports that history repeats itself in patterns of past price behavior will tend to recur in the future. Chou (1988) and Schwert (1989) also confirm this observation in their papers.

According to Engle and Patton (2001), a good volatility model must be able to incorporate some stylized facts about volatility. volatility clustering is usually approached by modeling the price process with an ARCH-type model. The ARCH model was first introduced by Engle (1982) to test if the ARCH effect exists. Then the GARCH model which is more general and allows much more flexible lag structure than the ARCH model was introduced by Bollerslev (1986).

3.4. Mean Reverting

This characteristic of volatility means that in the long run, there is a normal level to which volatility will eventually return. A period of high volatility will follow a period of low volatility. It also means, current information has no effect on the very long run forecast. (Engle & Patton 2001)

3.5. Asymmetric Impact

Some volatility models assume that the conditional volatility of the asset is symmetrically affected by negative and positive innovations. For equity markets this assumption is quite unlikely to be true (Engle & Patton 2001; Andersen et al. 2006). Volatility increases more following a market drop than following a rise in prices (Alexander 2008: 198 - 199). This phenomenon also ascribed to a leverage effect. Company's debt-to-equity ratio rises when the stock price falls. Then increases the volatility of returns to shareholders. (Engle & Patton 2001)

To capture this asymmetric response, several GARCH-type models have been introduced such as EGARCH (Nelson 1991), AGARCH (Engle & Ng 1993), GIR-GARCH (Glosten, Jagannathan & Runkle 1993) and TGARCH (Zakoian 1994).

3.6. Long Memory

How to properly capture volatility persistence is a key issue in volatility models. Normal GARCH model which is a short memory model features an exponential decay in the autocorrelation of conditional variances. However, in practice it has been noted that

exponential decay might be too fast for financial series. A long memory series has autocorrelation coefficients that decline slowly at a hyperbolic rate. When the effect of volatility shocks decay slowly, long memory occurs. (Poon 2005)

Many researchers such as Dacorogna, Müller, Nagler, Olsen and Pictet (1993), Ding, Granger and Engle (1993) and Anderson and Bollerslev (1997) have confirmed this financial volatility feature in their papers. To capture long memory, a new class of models have been introduced to allow this phenomenon. Baillie, Bollerslev and Mikkelsen (1996) have introduced the fractionally integrated GARCH model which is known as FIGARCH.

3.7. Exogenous Variables

What we have talked about above, those volatility characteristics, only limited to the time-series itself and the market around it. However, it is natural to think that there are some exogenous variables also influence volatility such as other assets, other markets and deterministic events.

Evidence has been found by many researchers. Engle, Ito and Lin (1990) examine the impact of news in one market on the time path of volatility in other markets. Engle, Ng and Rothschild (1990) use the Factor-ARCH model to study the dynamic relationship between asset risk premium and volatilities in a multivariate system. Moreover, Bollerslev and Melvin (1994) present empirical evidence that the bid-ask spread in the foreign exchange market has a positive relationship with the underlying exchange rate volatility.

3.8. Leptokurtosis

It is a well-known fact that density distributions of financial assets are not normally distributed. They often have a negative skewness and excess kurtosis than normal distribution. This feature can be also described as longer left tail and higher peak. It implies that, for most of the time, financial asset returns fluctuate in a range smaller than a normal distribution. However, there are few occasions where financial asset returns swing in a much wider scale than that permitted by a normal distribution (Poon 2005). Mandelbrot (1963) as well as Fama (1963) are the earliest researchers who provide evidence for this characteristic. A properly structured volatility model should take this feature into consideration.

4. VOLATILITY FORECASTING MODELS

4.1. Reasons for Choosing

Among numerous volatility models, we choose seven potential candidates according to previous studies. Some of them were studied before to be the best or the worst, and some were controversial. It is valuable to check the results of this thesis with previous papers' results. For the first type of models which only based on historical standard deviation, we choose the RW model and the RiskMetrics EWMA model. For GARCH-type models, we choose GARCH, GARCH-M, EGARCH, APARCH and TGARCH.

We choose the RW model because according to Zhang, Pan (2006), the RW model performs the worst under any situation. We want to confirm this conclusion by checking if it also performs the worst in our evaluation. We also choose the EWMA model from this group of models because according to McMillan and Kambouroudis (2009), the EWMA model is the best volatility model for forecasting in Asian markets. Additionally, the EWMA model with its representative the RiskMetrics EWMA model is very popular in practical financial applications.

Among GRACH-type models, we choose the GARCH model because it is the most general ARCH-type model and according to Frank, Tunaru and Wu (2004), that the GARCH(1,1) fits well on Shenzhen Component Index. We also choose the GARCH-M model because according to Song, Liu and Peter (1998), it is the best model for both Shanghai and Shenzhen markets. Besides, we choose the EGARCH model, the APARCH model and the TGARCH model to capture the asymmetric feature.

4.2. Basic Notion

First, we define our return equation as follows:

$$(1) \quad r_t = \text{Ln}\left(\frac{p_t}{p_{t-1}}\right)$$

Where r_t is the continuously compounded return of our concerned asset at time t , p_t is the asset closing price at time t and p_{t-1} is the closing price of the asset at time $t-1$.

Then, because the actual volatility cannot be observed directly, we need to find a proper proxy for it. There are two major methods for estimating the true volatility. The first measure was introduced by Andersen and Bollerslev (1998), it is called realized volatility. According to Andersen, Bollerslev, Diebold and Labys (1999), realized volatility is an unbiased and consistent estimator of daily volatility if returns have zero mean and uncorrelated to each other. To acquire realized volatility, we need high frequency trading data such as intra-day trading data. However, in this thesis we only have access to daily data. High frequency data is not available. So, we turn to the alternative measure. Another common approach for estimating true volatility, using squared daily return:

$$(2) \quad \sigma_t^2 = r_t^2$$

Though, the use of squared returns may favour the RiskMetrics model. When testing GARCH-type models, it is more common to express the actual volatility as squared of error term from a conditional mean regression when intra-day data is not available.

We model return series as follows:

$$(3) \quad r_t = x_t' \beta + u_t$$

$$u_t | \mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$$

$$(4) \quad \sigma_t^2 = u_t^2$$

Where x_t are independent variables which affect conditional mean of r_t , u_t denotes error term, \mathcal{F}_{t-1} is an information set follows normal distribution with zero mean and σ_t^2 conditional variance.

Here in this paper, we use this u_t^2 as the proxy of true volatility and the conditional mean model is an autoregressive moving average (ARMA) model.

Unconditional mean μ and unconditional variance σ^2 are defined as:

$$(5) \quad \mu = E[r_t]$$

$$(6) \quad \sigma^2 = E[r_t - \mu]^2$$

Moreover, time-varying conditional mean μ_t and conditional variance are shown as:

$$(7) \quad \mu_t = \mu_{t|t-1} = E[r_t | \mathcal{F}_{t-1}]$$

$$(8) \quad \sigma_t^2 = \sigma_{t|t-1}^2 = Var[r_t | \mathcal{F}_{t-1}] = E[(r_t - \mu_{t|t-1})^2 | \mathcal{F}_{t-1}] = E[\mu_t^2 | \mathcal{F}_{t-1}]$$

4.3. Random Walk (RW)

Random Walk model is the simplest structured model in this thesis. It bases on only historical returns. We forecast volatility as follows:

$$(9) \quad \hat{\sigma}_t = \sigma_{t-1}$$

Where $\hat{\sigma}_t$ is the estimated value and σ_{t-1} is the actual volatility from time $t-1$ which is defined in equation (2).

4.4. RiskMetrics Exponentially Weighted Moving Average (RiskMetrics EWMA)

The EWMA model is an extension of the simple historical model, which allows us to give different weight on different time by using a ‘decay factor’ λ . Recent observations could have a stronger impact on the forecast than older observations and the weights decline exponentially over time. The model is structured as follows:

$$(10) \quad \hat{\sigma}_t^2 = (1 - \lambda) \sum_{i=1}^{\infty} \lambda^i (r_{t-i} - \mu_t)^2$$

The RiskMetrics EWMA approach is a special case of normal EWMA model where the decay factor λ is fixed to be 0.94 for daily forecasts. This name comes from the popular risk measurement software producer RiskMetrics who provides fix decay factors. It should be noted that one-day ahead forecast will only arrive at the second expression. The model is structured as follows:

$$(11) \quad \hat{\sigma}_t^2 = \lambda \hat{\sigma}_{t-1}^2 + (1 - \lambda) \sigma_{t-1}^2 = (1 - \lambda) \sum_{i=1}^{\infty} \sigma_{t-i}^2$$

Where $\hat{\sigma}_t^2$ is the forecast of variance at time t , σ_{t-1}^2 is the ‘true’ volatility which is defined in Equation (4) at time $t-1$.

One major advantage of this model is that only one variable needs to be estimate, others are constant or observable. So, by using this model, it is relatively easy to track day-to-day volatility changes. This is also why it has its popularity in finance industry. However, this model also be criticized for several reasons. It is unable to capture the asymmetric impact which is created by the negative relationship between return and volatility in equity markets. Besides, this it is not able to provide long-horizon forecasts.

4.5. Generalized Autoregressive Conditional Heteroskedasticity (GARCH)

GARCH is more general and allows much more flexible lag structure than the ARCH model. It was first introduced by Bollerslev (1986) to overcome some difficulties brought by ARCH. The model which allows p lags of past forecast conditional variance and q lags of the squared error is defined as follows:

$$(12) \quad h_t = \omega + \sum_{i=1}^q \alpha_i u_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}$$

Where h_t denotes for the conditional variance, and $\omega > 0$, $\alpha_i, \beta_j \geq 0$ to ensure our conditional variance is strictly positive, $\sum \alpha_i + \sum \beta_j < 1$.

In this thesis we use a GARCH(1,1) model, which is found to be sufficient to capture the volatility clustering feature in the data. The model is structured as follows:

$$(13) \quad h_t = \omega + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1}$$

As well as in equation (11), h_t denotes for conditional variance, and $\omega > 0$, $\alpha_1, \beta_1 \geq 0$, $\alpha_1 + \beta_1 < 1$.

4.6. GARCH-In-Mean (GARCH-M)

It is natural to assume that investors need extra profit to compensate their additional risk. And this assumption is supported by risk aversion theory. One way to express this idea in a volatility model is to let the risk to become part of determination of stock returns. An ARCH-M model was introduced by Engle, Lilien and Robins (1987) to allow the conditional variance to be a determinant of the mean. Because GARCH is a more common approach than ARCH, so researchers often use a GARCH-M model instead of an ARCH-M model. A GARCH-M(1,1) is structured as follows:

$$(14) \quad r_t = \mu + \theta h_t + u_t$$

$$(15) \quad h_t = \omega + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1}$$

Where μ and θ are constants. The parameter θ is called the risk premium parameter. Positive θ means that the return of stock has a positive relationship with its volatility. In other words, when conditional variance rises, the expected return of stock will rise as well.

4.7. Exponential GARCH (EGARCH)

As we mentioned before, GARCH model has some constraints on its coefficients to ensure positive value of variance (Nelson & Cao 1992). In that way, it is not able to capture the asymmetric effect of both positive and negative shocks on volatility. Nelson (1991) suggests a new approach called EGARCH to solve these issues. The model is given as:

$$(16) \quad \ln(h_t^2) = \omega + \sum_{j=1}^p \beta_j \ln(h_{t-j}^2) + \sum_{i=1}^q \alpha_i \left\{ \left| \frac{\varepsilon_{t-i}}{h_{t-i}} \right| - \sqrt{\frac{2}{\pi}} \right\} - \gamma_i \frac{\varepsilon_{t-i}}{h_{t-i}}$$

Where γ is the leverage parameter or asymmetric response parameter. The sign of γ should be positive which implies a negative shock increases future volatility. In this thesis, we use an EGARCH(1,1) model which is:

$$(17) \quad \ln(h_t^2) = \omega + \beta_1 \ln(h_{t-1}^2) + \alpha_1 \left\{ \left| \frac{\varepsilon_{t-1}}{h_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right\} - \gamma \frac{\varepsilon_{t-1}}{h_{t-1}}$$

4.8. Threshold GARCH (TGARCH)

Threshold GARCH model is also known as the GJR model. It is also used to capture the asymmetric effect by adding an additional term to account for possible asymmetries. This model was first introduced by Glosten, Jagannathan and Runkle (1993) and Zakoian (1994). In the general TGARCH model, the specification of the conditional variance is constructed as follows:

$$(18) \quad h_t^2 = \omega + \sum_{i=1}^q (\alpha_i + \gamma_i d_{t-i}) \varepsilon_{t-1}^2 + \sum_{j=1}^p \beta_j h_{t-j}^2$$

Where, α_i , γ_i and $\beta_j \geq 0$. The coefficient γ is the asymmetry term, or leverage term. If $\gamma = 0$, then the model becomes the standard GARCH model. When there is a positive shock, the effect on volatility is α_i . When there is a negative shock, the effect on volatility is $\alpha_i + \gamma_i$. So, when γ_i is significantly positive, meaning that negative shocks have a larger effect on volatility than positive shocks. In this paper, a TGARCH(1,1) model is used:

$$(19) \quad h_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \gamma d_{t-1} \varepsilon_{t-1}^2 + \beta_1 h_{t-1}^2$$

where d_{t-1} is a dummy variable:

$$(20) \quad d_{t-1} = \begin{cases} 1 & \text{if } \varepsilon_{t-1} < 0 \\ 0 & \text{if } \varepsilon_{t-1} \geq 0 \end{cases}$$

4.9. Asymmetric Power ARCH (APRCH)

This is one of the most flexible ARCH-type models. According to Laurent and Peters (2002), the APRCH model derivatives at least seven other GARCH specifications. The model was first suggested by Ding et al. (1993) and an APARCH(1,1) model is specified as follows:

$$(21) \quad h_t^{\delta/2} = \omega + \alpha_1 (|\varepsilon_{t-1}| - \gamma \varepsilon_{t-1})^\delta + \beta_1 h_{t-1}^{\delta/2}$$

Where γ accounts for the leverage effect. $\gamma > 0$ means that past negative shocks have stronger impact on current volatility than past positive shocks. δ is the parameter

playing the role of a Box-Cox(1964) power transformation of the conditional standard deviation process.

4.10. Fractionally integrated GARCH (FIGARCH)

As we know, the GARCH model is a short memory model based on r_t^2 where the impact of past shocks decay exponentially. However, numerous empirical studies present that the impact of past shocks may decay at a slower hyperbolic rate over long lags.

According to Granger and Joyeux (1980) and Hosking (1981), fractionally integrated series exhibit long memory property. After this work, Ding, Granger and Engle (1993) propose a fractionally integrated model based on $|r_t|^d$ where d is a fraction. A FIGARCH(1, d,1) model was first used in Baillie, Bollerslev and Mikkelsen (1996):

$$(22) \quad h_t = \omega + [1 - \beta_1 L - (1 - \phi_1 L)(1 - L)^d] \varepsilon_t^2 + \beta_1 h_{t-1}$$

Where L denotes the lag operator. And for the one-step-ahead forecast, we can specify the model as:

$$(23) \quad \hat{h}_{t+1} = \omega(1 - \beta_1)^{-1} + [1 - (1 - \beta_1 L)^{-1}(1 - \phi_1 L)(1 - L)^d] \varepsilon_t^2$$

5. EVALUATION OF VOLATILITY FORECAST

Among previous studies, we can see various of evaluation approaches were used to assess volatility models. And there is no single standard for choosing access approaches because no study has shown that one approach is obviously superior than others.

As Granger (1999) states, the comparison is straightforward when the error distribution of one model dominates another model, but this rarely happens. However, in most cases, comparisons are based on some average figures, such as the root mean square error and the mean absolute error. Regression-type of measures can also be used to evaluate the accuracy such as the Mincer-Zarnowitz regression and the forecast encompassing.

Because there is no clear evidence which loss function is more appropriate for the evaluation of volatility models, we will choose several different measurements to conduct our evaluation.

5.1. The Root Mean Square Error (RMSE)

The root mean square error statistic is also known as the root-mean-square deviation (RMSD). It represents the sample standard deviation of the differences between forecast values and observed values. It is a good measure of accuracy when comparing forecasting errors of different models but the same variable. (Hyndman & Anne 2006). Since it is a scale-dependent statistic, it cannot be used to compare variables with different scales. The RMSE is defined as follows:

$$(24) \quad RMSE = \sqrt{n^{-1} \sum_{t=1}^n (\sigma_t - h_t)^2}$$

Where h_t denotes the estimated volatility generated by our volatility models at time t , σ_t is the actual volatility which we regard as the squared error from the conditional mean model for returns.

5.2. The Mean Absolute Error (MAE)

Like the RMSE, the mean absolute error is also a symmetric loss function that used to measure how close forecasts are to the observe values. The model is given as:

$$(25) \quad MAE = n^{-1} \sum_{t=1}^n |\sigma_t - h_t|$$

It is also a scale-dependent accuracy measure, therefore cannot be used to make comparisons between series using different scales.

5.3. The Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD). It expresses accuracy as a percentage and is defined as follows:

$$(26) \quad MAPE = 100 * n^{-1} \sum_{t=1}^n \left| \frac{\sigma_t - h_t}{\sigma_t} \right|$$

The MAPE statistic has some major drawbacks in practice. Such as it will be invalid if there is any zero value and it will systematically prefer a forecasting method whose forecasts are too low. (Tofallis 2015)

5.4. The Theil Inequality Coefficient (Theil)

The Theil inequality coefficient is also known as the Theil's U-statistic. The metric is defined as follows:

$$(27) \quad U = \frac{\sqrt{\sum_{t=T_i}^T \left(\frac{\gamma_{t+s} - f_{t,s}}{\gamma_{t+s}} \right)^2}}{\sqrt{\sum_{t=T_i}^T \left(\frac{\gamma_{t+s} - f_{b_{t,s}}}{\gamma_{t+s}} \right)^2}}$$

Where $f_{b_{t,s}}$ is the forecast obtained from a benchmark model. When $U < 1$, it means the considered model is more accurate than the benchmark, and vice versa for $U > 1$. Although the model is commonly used and useful, it is not without problems. If $f_{b_{t,s}}$ equals to γ_{t+s} , the denominator will be zero. Moreover, the U-statistic will also be influenced by outliers in a similar vein to MSE and has little intuitive meaning. (Brooks 2014)

5.5. The Mean Mixed Error (MME)

Based on the positive relationship between volatility and expected returns, it is natural to consider that investors will not attribute same importance to both under-predictions and over-predictions of volatility of similar magnitude. Since those symmetric loss functions are not able to support this idea, Brailsford and Faff (1996) construct an error statistic which penalizes under-predictions more heavily called the mean mixed error (MME). And this error statistic can also be redefined to weight over-predictions more heavily. They are constructed as follows:

$$(28) \quad MME(U) = n^{-1} \left[\sum_{t=1}^O |\sigma_t^2 - h_t^2| + \sum_{t=1}^U \sqrt{|\sigma_t^2 - h_t^2|} \right]$$

$$(29) \quad MME(O) = n^{-1} \left[\sum_{t=1}^U |\sigma_t^2 - h_t^2| + \sum_{t=1}^O \sqrt{|\sigma_t^2 - h_t^2|} \right]$$

where O is the number of over-predictions and U is the number of under-predictions.

If a forecast model is considered systematically over-prediction or under-prediction according to MME error statistic, then we could conclude this model is ‘biased’. Otherwise, when a forecast model provides 50 percent over-predictions and 50 percent under-predictions, it is considered a ‘unbiased’ model. (Brailsford & Faff 1996).

5.6. The Mincer-Zarnowitz (MZ) Regression

One popular method called Mincer-Zarnowitz (MZ) regression which regresses squared returns on the estimated volatility and a constant. And its logarithmic version which regresses log squared returns on the log estimated volatility and a constant.

$$(30) \quad \sigma_t^2 = a + b h_t^2 + u_t$$

$$(31) \quad \log(r_t^2) = a + b \log(h_t^2) + u_t$$

According to Pagan and Schwert (1990), Engle and Patton (2001), the logarithmic version is less sensitive to extreme samples. However, according to Hansen and Lunde (2005), MZ regression is not an ideal evaluation for volatility models because it does not penalize a biased forecast.

5.7. Forecast Encompassing

The MZ regression provides us some evidence whether the chosen models result reasonable forecasts for stock return volatility. Following Kambouroudis and McMillian (2016), we extend this regression for our purpose in this thesis, to examine whether GARCH-type models perform a significantly better forecast than the RiskMetrics model by carrying additional information.

The method of forecast encompassing was first developed by Chong and Hendry (1986). It allows us to check if additional information was carried by the considered model comparing to the benchmark model. Specifically, if the competing model carries no additional information, then the benchmark model is said to ‘encompass’ the competing model. (Kambouroudis & McMillian 2016). The extension of the MZ model used for our forecast encompassing is structured as follows:

$$(32) \quad \sigma_t^2 = \omega + \rho_1 h_{1,t}^{2f} + \rho_2 h_{2,t}^{2f} + \varepsilon_t$$

Where the subscripts 1 and 2 denote the forecast models. 1 refers to the benchmark model and 2 refers to the competing model). The null hypothesis is that the benchmark model encompasses the competing model. When ρ_2 is significantly differ from zero, the null hypothesis should be rejected, and we should admit that the competing model contains information that the benchmark model does not. If ρ_2 is equal to zero, we say that the benchmark model encompasses the competing model.

6. DATA

There are two major stock exchanges in mainland China's equity market. The Shanghai Stock Exchange (SSE) and the Shenzhen Stock Exchange (SZSE). Both have their official index, respectively, the Shanghai Stock Exchange Composite Index and the Shenzhen Stock Exchange Component Index.

The Shanghai Stock Exchange, established on November 1990, was permitted by People's Bank of China. It is a non-profit organization which directly governed by the China Securities Regulatory Commission (CSRC). Its main functions include providing marketplace and facilities for the securities trading, formulating business rules, accepting and arranging listings, organizing and monitoring securities trading, regulating members and listed companies, and managing and disseminating market information. By the end of 2016, there were in total 1156 listed companies, 9279 listed securities (including 1200 stock securities), total market capitalization 29 trillion RMB.

The Shenzhen Stock Exchange was founded in December 1990. It is a self-regulated legal entity under the supervision of CSRC. It also organizes, supervises securities trading and performs duties prescribed by laws, regulations, rules and policies. Its main functions include providing the venue and facilities for securities trading, formulating operational rules, receiving listing applications and arranging securities listing, organizing and supervising securities trading, supervising members; regulating listed companies, managing and disseminating market information and other functions as approved by the CSRC. The main difference between SZSE and SSE is, SZSE is more specialized in developing China's multi-tiered capital market system. It provides finance platform for many start-up companies and small cap companies. By the end of 2016, there were 1870

listed companies, including 478 on Main-Board, 822 on Small and Medium Enterprise Board and 570 on Growth Enterprise Board. The total capitalization was 22 trillion RMB.

The SSE Composite contains all stocks (A-shares and B-shares) that are traded at the Shanghai Stock Exchange. It is calculated using Paasche weighted composite price index formula. We consider daily observations and our sample period goes from 04 January 2006 to 30 December 2016, in total 11 years, 2674 samples. The data source is SSE official website.

The SZSE Component Index is an index of 500 A-share stocks that are traded at the Shenzhen Stock Exchange. It is designed to represent the performance of the multi-tier Shenzhen stock market and calculated using a capitalization weighted free-float measure. We consider daily observations and our sample period goes from 04 January 2006 to 30 December 2016, in total 11 years, 2674 samples, same as SSE index. The data source is SZSE official website.

7. METHODOLOGY

First, we use our raw data which are daily closing prices to calculate the logarithms returns. After that, we generate the true volatility series. The proxy for it is the squared error from an ARMA(1,1) model for return series. This process is operated corresponding with the procedure of Kambouroudis and McMillan (2015).

When checking our return series, several tests are conducted to check the normality, the stationarity, autocorrelation and the heteroscedasticity. the Jarque–Bera test is employed to check whether our sample data have the skewness and kurtosis that could match a normal distribution. It is defined as:

$$(33) \quad JB = \frac{n-k+1}{6} \left(S^2 + \frac{1}{4} (C - 3)^2 \right)$$

Where S is the sample skewness, C is the sample kurtosis, and k is the number of regressors. The null hypothesis is a joint hypothesis of the skewness being zero and the excess kurtosis being zero which means the sample follows a normal distribution. Moreover, the Quantile-Quantile (QQ) plot technique is also conducted to cross-check the normality.

After checking the normality, the Augmented Dickey-Fuller (ADF) test and the Phillips-Perron (PP) unit root are conducted to check the stationarity. The null hypothesis is that there is a unit root presents in the sample. Davidson and MacKinnon (2004) report that the PP test performs worse than the ADF test in a finite sample set, so we just use it to confirm the result from the ADF test.

The Ljung-Box test is performed to check whether any serial correlation exists. The test is defined as:

$$(34) \quad Q = n(n + 2) \sum_{k=1}^h \frac{\widehat{\rho}_k^2}{n-k}$$

Where n is the sample size, $\widehat{\rho}_k$ is the sample autocorrelation at lag k , and h is the number of lags being tested. The null hypothesis is the correlations in the population from which the sample is taken are zero which means the data are independently distributed.

Before applying GARCH models. It is important to examine the heteroscedasticity in the residuals. Following Engle (1982), the ARCH Lagrange Multiplier (LM) test is conducted on the residuals of conditional mean. First, apply the ARMA (1,1) model to the return series as an initial regression. Secondly, test the null hypothesis that there is no ARCH effect up to order q in the residual series.

The out-of-sample forecast method is used in this thesis. The whole sample period is divided into two parts, an estimation part and an evaluation part. Data in estimation period are used to estimate model parameters. Estimated parameters are then used to perform one-step ahead forecast. After the first forecast, the estimation sample move one step ahead, and the coefficients are estimated again. This process iterates until the end of the evaluation period.

In this thesis, the estimation period contains samples between 04 January 2006 and 30 December 2014, roughly nine years. The forecast period contains samples begins at 05 January 2015 until the end which is 30 December 2016, roughly two years.

The models analyzed in this thesis include the Random Walk model, the RiskMetrics EWMA model, the GARCH(1,1) model, the GARCH-in-mean(1,1) model, the EGARCH(1,1) model, the APARCH(1,1) model and the TGARCH(1,1) model. In the RiskMetrics EWMA model, $\lambda = 0.94$ is applied as in the model of J.P Morgan daily data. For all GARCH models, the Maximum Likelihood Method is employed to estimate the equation parameters. The specifications we use for all these models can be found in Chapter four.

After the model estimation and the out-of-sample volatility forecast stage, seven evaluation methods will be conducted to all forecast results. Symmetric and asymmetric loss functions are used to try to find out the best performing model. Forecast encompassing method is used to examine whether GARCH-type models carry additional information than the Random Walk model or the RiskMetrics EWMA model.

8. EMPIRICAL RESULTS

8.1. Preliminary Data

After calculating the logarithms returns, two sets of daily return series are generated. The descriptive statistics are as follows:

Table 1. Descriptive statistics of return series.

	SSE	SZSE
Mean	0.000362	0.000464
Median	0.001028	0.000958
Maximum	0.090343	0.091614
Minimum	-0.092562	-0.097500
Std. Dev.	0.017687	0.020046
Skewness	-0.618773	-0.536871
Kurtosis	6.797022	5.551395
Jarque-Bera	1775.644 (0.0000)	853.0957 (0.0000)
Ljung-Box $Q(5)$	21.614 (0.0000)	22.034 (0.0000)

Notes: Numbers in brackets indicate p-value.

According to Jarque-Bera test, both returns reject the null hypothesis that their density distributions follow normal distribution. And the Quantile-Quantile plot also has confirmed this result. We can easily observe the fat tail feature of both returns from Figure 3. The Ljung-Box $Q(5)$ statistics result shows both return series exhibit autocorrelation up to five lags.

Table 2. Unit root tests.

	SSE	SZSE
ADF statistics	-50.49136 (0.0000)	-49.0324 (0.0000)
PP statistics	-50.55191 (0.0000)	-49.0324 (0.0000)

Notes: Numbers in brackets indicate p-value. Tests are conducted without a time trend. The lag length for the tests is based on the Schwarz information criterion.

Table 2 reports the result of unit root tests. It indicates that for both return series, the null hypothesis of a unit root can be rejected at 1% significant level. So, both returns are stationary.

The ARCH LM tests, reported in Table 3, show that the null hypothesis is rejected at 5 lags and the ARCH effect exists in both residuals. Therefore, the variances of both return series are non-constant.

Table 3. ARCH Lagrange Multiplier tests.

	SSE	SZSE
Obs*R-squared	223.9462	205.4995
Prob. Chi-Square(5)	(0.0000)	(0.0000)

Notes: Numbers in brackets indicate p-value.

8.2. Symmetric Forecast Errors

The relative forecast error statistics of all our studied models for both SSE and SZSE markets are reported in Table 4 and Table 5. The ranks based on these relative forecast error statistics are reported in Table 6 and Table 7. It should be noted that these statistics are standardized by the worst model in the same column.

For SSE index, the GARCH-in-mean(1,1) model performs best under the RMSE measure. Under the same measure, the RW model performs the worst. When we turn to the MAE measure, the best performing model changes to the EGARCH(1,1) model, and the worst model is still the RW model. The RiskMetrics model performs the best under two measures, the MAPE measure and the Theil measure, while the RW model performs the worst under the MAPE and the EGARCH(1,1) model performs the worst under the Theil. If we take an average, then the RW model is the worst model and the RiskMetrics model performs beyond the average. However, there is no clear winner.

Table 4. Relative forecast error statistics for SSE-index.

	RMSE	MAE	MAPE	Theil
RW	1.0000	1.0000	1.0653	0.9991
RiskMetrics	0.7415	0.8113	0.8272	0.9676
GARCH(1,1)	0.7406	0.8080	0.9963	0.9851
GARCH-M(1,1)	0.7405	0.8080	0.9973	0.9842
EGARCH(1,1)	0.7406	0.7878	0.8986	1.0000
APARCH(1,1)	0.7415	0.7988	0.9404	0.9915
TGARCH(1,1)	0.7413	0.8077	1.0000	0.9920

Table 5. Relative forecast error statistics for SZSE-index.

	RMSE	MAE	MAPE	Theil
RW	1.0000	1.0000	0.4744	0.9838
RiskMetrics	0.7266	0.8008	1.0000	0.9109
GARCH(1,1)	0.7246	0.7876	0.9435	0.9379
GARCH-M(1,1)	0.7241	0.7867	0.9470	0.9332
EGARCH(1,1)	0.7266	0.7718	0.9364	0.9410
APARCH(1,1)	0.7313	0.7789	0.9762	0.9400
TGARCH(1,1)	0.7262	0.7213	0.7549	1.0000

Notes: Table 4 and Table5 report the result of four forecast error statistics. The in-sample estimation period is from 04 January 2006 to 30 December 2014, and the out-of-sample forecast period is from 05 January 2015 to 30 December 2016.

For SZSE index, the GARCH-in-mean(1,1) model also performs best under the RMSE measure and the RW model performs the worst under the same measure. When we turn to the MAE measure, the best performing model changes to the TGARCH(1,1) model, and the worst model is still the RW model. However, under the MAPE measure, the RW model becomes the best model and the EWMA model becomes the worst. The EWMA model then performs the best under the Theil measure and the TGARCH(1,1) performs the worst. If we take an average, then the RW model performs the worst while the TGARCH(1,1) model is the best model for SZSE index.

Table 6. Ranks of relative forecast error statistics for SSE index.

	RMSE	MAE	MAPE	Theil
RW	7	7	7	6
RiskMetrics	5	6	1	1
GARCH(1,1)	2	4	4	3
GARCH-M(1,1)	1	5	5	2
EGARCH(1,1)	3	1	2	7
APARCH(1,1)	6	2	3	4
TGARCH(1,1)	4	3	6	5

Table 7. Ranks of relative forecast error statistics for SZSE index.

	RMSE	MAE	MAPE	Theil
RW	7	7	1	6
RiskMetrics	5	6	7	1
GARCH(1,1)	2	5	4	3
GARCH-M(1,1)	1	4	5	2
EGARCH(1,1)	4	2	3	5
APARCH(1,1)	6	3	6	4
TGARCH(1,1)	3	1	2	7

Notes: Table 6 and Table 7 report the ranking result of four forecast error statistics. The in-sample estimation period is from 04 January 2006 to 30 December 2014, and the out-of-sample forecast period is from 05 January 2015 to 30 December 2016.

The findings are consistent with the previous study conducted by Zhang and Pan (2006) that the RW model performs the worst under both markets and almost every measure. And the result also confirms with Brailsford and Faff (1996) that there is no single model is clearly superior, the performance is sensitive to the choice of error statistics.

8.3. Asymmetric Forecast Errors

Following Brailsford and Faff (1996), the mean mixed error statistics are calculated. Table 8 and Table 9 report the results for both SSE and SZSE indices. The ‘Actual’ column shows the actual statistic value of the MME measure and remains four decimal places. The ‘Relative’ column shows the MME value standardized by dividing the value of the worst performing model in the same column. The ‘Rank’ column runs a rank based on the actual MME statistics, the smaller the error is, the better the model performs. The last two columns present how many times the studies model over- and under-predict the true volatility. The MME(U) model penalizes under-prediction more heavily while the MME(O) model penalizes over-prediction more heavily as we mentioned in section 5.

As we can see from the tables, except for the RW model, all other models, no matter if it is a linear or non-linear model, all systematically over-predict the future volatility. This is probably because of the period which we choose. During 2015, there was a stock market crisis happened in China. This over-prediction problem implies for all our models, there is something not consistent with the real world. It might be the error assumption which here in this thesis we assume a normal distribution on regression errors, but in real world it is not like this because the fact of fat tail distribution. From this aspect, these volatility models are not able to properly capture this feature.

For SSE index, the MME(U) statistics suggest a consistent result with the RMSE error statistics that the RW model is the worst model and the GARCH-in-mean(1,1) model is the best. However, when the MME(O) is used, the result goes to the opposite. The MME(O) suggests that the RW model performs the best while the GARCH-in-mean(1,1) model performs the worst. It is obvious that the result is highly dependent on the choice of error statistics.

For SZSE index, the MME(U) model also evaluates the RW model the worst, but the best model is not consistent with the result from above symmetric loss functions. It suggests the RiskMetrics is the best performing model not the GARCH-in-mean. Though the GARCH-in-mean(1,1) model still performs not bad. It ranks the second place. Like what happened to SSE index, when we change the error statistics to the MME(O), the result goes to an opposite side. According to the MME(O) model, the RW model performs the best and the RiskMetrics model performs the worst.

Overall, the results from the asymmetric loss functions are partially consistent with the results from the previous symmetric loss functions. It should be confirmed that the RW model is the worst performing model under most situations. The exception under the MME(O) measure is because that all other models provides systematic over-predictions. This issue has a significant impact on the result when we penalize over-predictions more heavily by using the MME(O). It is not clear to say which model it the best model until we conduct further analysis.

Table 8. MME statistics and numbers of over- and under- predictions for SSE index.

	MME(U)			MME(O)			O	U
	Actual	Relative	Rank	Actual	Relative	Rank		
RW	0.0090	1.0000	7	0.0088	0.7240	1	232	256
RiskMetrics	0.0060	0.6609	4	0.0119	0.9783	4	360	128
GARCH	0.0059	0.6554	2	0.0122	0.9998	6	372	116
GARCH_M	0.0059	0.6550	1	0.0122	1.0000	7	373	115
EGARCH	0.0060	0.6688	5	0.0118	0.9712	3	369	119
APARCH	0.0061	0.6778	6	0.0115	0.9472	2	367	121
TGARCH	0.0060	0.6595	3	0.0122	0.9976	5	372	116

Table 9. MME statistics and numbers of over- and under- predictions for SZSE index.

	MME(U)			MME(O)			O	U
	Actual	Relative	Rank	Actual	Relative	Rank		
RW	0.0102	1.0000	7	0.0103	0.7583	1	245	243
RiskMetrics	0.0069	0.6751	1	0.0136	1.0000	7	366	122
GARCH	0.0070	0.6887	3	0.0135	0.9920	5	370	118
GARCH_M	0.0070	0.6864	2	0.0135	0.9929	6	370	118
EGARCH	0.0073	0.7123	5	0.0128	0.9436	4	360	128
APARCH	0.0072	0.7081	4	0.0128	0.9419	3	359	129
TGARCH	0.0078	0.7677	6	0.0110	0.8089	2	347	141

Notes: Table 8 and Table 9 report the result of the Mean Mixed Error statistics. The in-sample estimation period is from 04 January 2006 to 30 December 2014, and the out-of-sample forecast period is from 05 January 2015 to 30 December 2016.

8.4. The Mincer-Zarnowitz Regression

Table 9 presents the results for the simple MZ regression which we defined in Equation 30. The coefficient of determination R^2 is reported for regressing the true volatility on a constant and volatility forecast for each of our studied models. The greater the R^2 is, the more accurate the forecast is. We can see from the table, for both SSE and SZSE indices, the RW model gives the worst forecast. This result consists with our previous findings. However, it is not clear which model is the best. For SSE index, the RiskMetrics model gives the best forecast whereas for SZSE index, the TGARCH model performs the best. Furthermore, the GARCH-in-mean model has an average good performance in both SSE and SZSE markets.

Table 10. Coefficients of determination for Mincer-Zarnowitz.

	SSE index		SZSE index	
	R^2	Rank	R^2	Rank
RW	0.039035	7	0.033921	7
RiskMetrics	0.109256	1	0.120661	4
GARCH	0.107658	3	0.120766	3
GARCH_M	0.107889	2	0.122040	2
EGARCH	0.107312	4	0.117408	5
APARCH	0.105648	5	0.108823	6
TGARCH	0.105461	6	0.128332	1

Notes: this table reports the result of the Mincer-Zarnowitz regression. The in-sample estimation period is from 04 January 2006 to 30 December 2014, and the out-of-sample forecast period is from 05 January 2015 to 30 December 2016.

8.5. Forecast Encompassing

Table 10 reports the result of forecast encompassing test using the RW model as the benchmark. For both markets, all GARCH-type models exhibit very strong significance

to reject the null hypothesis which is the RW model encompasses them. This result provides a clear point that GARCH-type models contains additional information than the RW model. Furthermore, GARCH models are superior than the RW model in forecasting volatility in Chinese stock markets.

Table 11 reports the result of forecast encompassing test by using the RiskMetrics model as the base model. For both markets, only the TGARCH model has a p-Value that is significant at 5% level that can reject the null hypothesis. All other GARCH-type models do not have a p-Value big enough to reject the null hypothesis which is the RiskMetrics model encompasses them. According to this result, in our case, GARCH-type models do not carry additional information than the RiskMetrics EWMA model except for the TGARCH model and the reason needs further investigation. The TGARCH model performs well for SZSE index, this result consists with our pervious finding in the MZ regression and the MAE statistics.

Table 11. Encompassing test with the RW model as the benchmark.

	SSE		SZSE	
	Coefficient	p-Value	Coefficient	p-Value
GARCH	0.8375	0.0000*	0.9296	0.0000*
GARCH_M	0.8354	0.0000*	0.9150	0.0000*
EGARCH	0.8572	0.0000*	0.8607	0.0000*
APARCH	0.8218	0.0000*	0.7753	0.0000*
TGARCH	0.8495	0.0000*	1.1022	0.0000*

Notes: this table reports the result of forecast encompassing test by using the Random Walk model as benchmark. The in-sample estimation period is from 04 January 2006 to 30 December 2014, and the out-of-sample forecast period is from 05 January 2015 to 30 December 2016. The asterisk indicates significant at 1% level.

Table 12. Encompassing test with the RiskMetrics model as the benchmark.

	SSE		SZSE	
	Coefficient	p-Value	Coefficient	p-Value
GARCH	-6.0325	0.1029	0.5308	0.6735
GARCH_M	-5.2863	0.1592	1.0457	0.3810
EGARCH	-0.0426	0.9648	-0.1264	0.8698
APARCH	-0.9613	0.4032	-0.5949	0.2535
TGARCH	-4.8356	0.0277**	1.2752	0.0382**

Notes: this table reports the result of forecast encompassing test by using the RiskMetrics model as benchmark. The in-sample estimation period is from 04 January 2006 to 30 December 2014, and the out-of-sample forecast period is from 05 January 2015 to 30 December 2016. The double asterisks indicate significant at 5% level.

9. SUMMARY AND CONCLUSION

This thesis aims to find the most appropriate model to estimate and forecast volatility in Chinese stock markets, and to investigate the differences between simple historical models and GARCH-type models. The studied model collection includes seven models: Random Walk, RiskMetrics EWMA, GARCH, GARCH-in-mean, EGARCH, TGARCH and APARCH. The forecast performances of those models are then evaluated in seven different criteria including symmetric loss functions and asymmetric loss functions, and other measurements such as the forecast encompassing test is conducted to check whether GARCH-type models carry additional information than simple historical models. The whole evaluation process is conducted with two Chinese stock markets' indices, namely the SSE composite index and the SZSE component index. The selected sample period with updated data spans from 04 March 2006 through 30 December 2016.

Our empirical evidence based on seven different forecast error measures shows that the RW model has the worst performance among all models we choose. For the SSE index, the RMSE statistics suggest the GARCH-in-mean model is the best model; the MAE statistics suggest the EGARCH model performs the best; the MAPE statistics and the Theil's U-statistics suggest the RiskMetrics model is the best; the MME(U) measure suggests the GARCH-in-mean model to be the best model; the MZ regression prefers the RiskMetrics model. For the SZSE index, the RMSE statistics suggest the GARCH-in-mean model is the best model; the MAE statistics suggest the TGARCH model performs the best; The Theil's U-statistics prefer the RiskMetrics EMWA model. The asymmetric loss function MME(U) suggests the RiskMetrics EMWA model is the best. The MZ regression prefers the TGARCH model.

The asymmetric loss function MME suggests that the phenomenon of systematically over-prediction exists in our forecasts for both markets. This might be caused by the forecast period we choose, when there was a major stock market crisis hit during 2015, intensively volatile price trend could lead to forecast failure since the assumption of regression errors following normal distribution.

Furthermore, the forecast encompassing tests document that all GARCH models carry additional information than the RW model. It is obvious that GARCH models can explain more of stock market volatility, hence they are superior to the RW model. However, in our analysis, we couldn't find the evidence that GARCH models are carrying more information than the RiskMetrics EWMA model, so we cannot conclude that GARCH models are superior to the RiskMetrics model.

The findings of our study are consistent with some previous empirical studies such as Zhang and Pan (2006), Brailsford and Faff (1996) and Brooks (1998). Consistent with Zhang and Pan (2006), we find that the forecast performance of the RW model is the worst under almost every criteria and market. Consistent with Brailsford and Faff (1996), we find that the ranking of our studied models is highly sensitive to the choice of forecast error statistics. We also find the ranking would change when different market index applies. Therefore, it should depend on the particular purpose when choosing forecast error statistics. Consistent with Brooks (1998), performances are similar across models, most of them only contains marginally differences.

As a suggestion for further study. This study can be extended in at least three following ways. Firstly, go deeper on the evaluation methods. Since our models only contain marginally differences in their forecast errors. It is natural to consider whether they do

have significant differences in forecasting abilities or the various results only come by chance. Hansen (2005) uses a test called the Superior Predictive Ability test (SPA) to test whether any alternative forecast is better than the base forecast. By using this test, it might give a robust conclusion based on the results from the first evaluation stage. Hansen et al. (2005) also introduce a method called the Model Confidence Set to examine whether there is a possibility that more than one model in a set of models having ability to provide accurate forecast.

Secondly, change a proxy of true volatility. We use squared error from a conditional mean model for returns as the proxy of actual conditional volatility in this thesis. However, many studies have been conducted to show that this might not be the best way to express the true volatility. The realized volatility calculated from intra-day data is recommended as a better proxy for true volatility. (McMillan & Speight 2004; Hansen & Lunde 2005). Hence, if high frequency data is available, it is worthy to check the accuracy of above models by using realized volatility.

Finally, try more volatility models. In this thesis, we use seven different models, but there are hundreds of other available GARCH-type models existing among various studies. Nowadays, stochastic volatility (SV) models and option-based volatility models become more and more popular. In the SV modelling framework, volatility is subject to a source of innovations that may or may not be associated with those that drive returns. Option-based models depend on the implied volatility. These models should also be evaluated to see whether they would have better performance.

REFERENCES

- Akgiray, Vedat. (1989). Conditional heteroscedasticity in time series of stock returns: Evidence and forecasts. *Journal of business*, 55-80.
- Alexander Carol (2008). *Market Risk Analysis Volume II - Practical Financial Econometrics*. West Sussex, England: John Wiley & Sons Ltd.
- Andersen, T. G., Bollerslev, T., Christoffersen, P. F., & Diebold, F. X. (2006). Volatility and correlation forecasting. *Handbook of economic forecasting*, 1, 777-878.
- Andersen, Torben G., and Tim Bollerslev. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review*, 885-905.
- Anderson, T. G., and T. Bollerslev. (1997). Answering the Critics: Yes, ARCH Models Do Provide Good Volatility Forecasts. *NBER Working paper*, 6023.
- Baillie, Richard T. (1996). Long memory processes and fractional integration in econometrics. *Journal of econometrics*, 73(1), 5-59.
- Baillie, Richard T., Tim Bollerslev, and Hans Ole Mikkelsen. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 74(1), 3-30.

- Bollerslev, Tim, and Michael Melvin. (1994). Bid—ask spreads and volatility in the foreign exchange market: An empirical analysis. *Journal of International Economics*, 36(3), 355-372.
- Bollerslev, Tim. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3), 307-327.
- Boudoukh, J., Richardson, M., & Whitelaw, R. F. (1997). Investigation of a class of volatility estimators. *The Journal of Derivatives*, 4(3), 63-71.
- Brailsford, Timothy J., and Robert W. Faff. (1996). An evaluation of volatility forecasting techniques. *Journal of Banking & Finance*, 20(3), 419-438.
- Brooks, Chris. (1998). Predicting stock index volatility: can market volume help?. *Journal of Forecasting*, 17(1), 59-80.
- Brooks, Chris. (2014). *Introductory econometrics for finance*. Cambridge university press.
- Chong, Yock Y., and David F. Hendry. (1986). Econometric evaluation of linear macro-economic models. *The Review of Economic Studies*, 53(4), 671-690.
- Chou, Ray Yeutien (1988). Volatility Persistence and Stock Valuations: Some Empirical Evidence using GARCH. *Journal of Applied Econometrics*. 3, 279- 294.
- Cumby, Robert, Stephen Figlewski, and Joel Hasbrouck. (1993). Forecasting volatilities and correlations with EGARCH models. *The Journal of Derivatives*, 1(2), 51-63.

- Dacorogna, Michael M., Ulrich A. Müller, Robert J. Nagler, Richard B. Olsen & Olivier V. Pictet (1993). A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance*, 12(4), 413-438.
- Ding, Zhuanxin, Clive WJ Granger, and Robert F. Engle. (1993). A long memory property of stock market returns and a new model. *Journal of empirical finance*, 1(1), 83-106.
- Ding, Zhuanxin, Clive WJ Granger, and Robert F. Engle. (1993). A long memory property of stock market returns and a new model. *Journal of empirical finance*, 1(1), 83-106.
- Engle III, Robert F., Takatoshi Ito, and Wen-Ling Lin. (1988). Meteor Showers or Heat Waves? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market. *Econometrica*, 58(3), 525-542.
- Engle, Robert F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.
- Engle, Robert F., and Andrew J. Patton. (2001). What good is a volatility model. *Quantitative finance*, 1(2), 237-245.
- Engle, Robert F., and Victor K. Ng. (1993). Measuring and testing the impact of news on volatility. *The journal of finance*, 48(5), 1749-1778.

- Engle, Robert F., David M. Lilen, and Russell P. Robins. (1987). Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica: Journal of the Econometric Society*, 391-407.
- Engle, Robert F., Victor K. Ng, and Michael Rothschild. (1990). Asset pricing with a factor-ARCH covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics*, 45(1-2), 213-237.
- Fabozzi, Frank J., Radu Tunaru, and Tony Wu. (2004). Modeling volatility for the Chinese equity markets. *Annals of Economics and Finance.*, 5, 79-92.
- Fama, Eugene F. (1963). Mandelbrot and the stable Paretian hypothesis. *The journal of business*, 36(4), 420-429.
- Fama, Eugene F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business*. 38:1, 34-105.
- Figlewski, Stephen. (1997). Forecasting volatility. *Financial markets, institutions & instruments*, 6(1), 1-88.
- Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5), 1779-1801.

- Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5), 1779-1801.
- Granger, Clive WJ, and Roselyne Joyeux. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis*, 1(1), 15-29.
- Granger, Clive WJ. (1999). *Empirical modeling in economics: Specification and evaluation*. Cambridge University Press.
- Hansen, Peter R., and Asger Lunde. (2005). A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of applied econometrics*, 20(7), 873-889.
- Hosking, Jonathan RM. (1981). Fractional differencing. *Biometrika*, 165-176.
- Hull, John C. (2012). *Options, futures, and other derivatives*. 8th edition. Pearson Education India.
- Hyndman, Rob J., and Anne B. Koehler. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4), 679-688.
- Kambouroudis, Dimos S., and David G. McMillan. (2016). Does VIX or volume improve GARCH volatility forecasts?. *Applied Economics*, 48(13), 1210-1228.

- Mandelbrot, Benoit. (1963). New methods in statistical economics. *Journal of political economy*, 71(5), 421-440.
- Mandelbrot, Benoit. (1963). The Variation of Certain Speculative Prices. *The Journal of Business*, 36(4), 394-419.
- McMillan, D., Speight, A., & Apgwilym, O. (2000). Forecasting UK stock market volatility. *Applied Financial Economics*, 10(4), 435-448.
- McMillan, David G., and Alan EH Speight. (2004). Daily volatility forecasts: Reassessing the performance of GARCH models. *Journal of Forecasting*, 23(6), 449-460.
- McMillan, David G., and Dimos Kambouroudis. (2009). Are RiskMetrics forecasts good enough? Evidence from 31 stock markets. *International Review of Financial Analysis*, 18(3), 117-124.
- Nelson, Daniel B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347-370.
- Nelson, Daniel B., and Charles Q. Cao. (1992). Inequality constraints in the univariate GARCH model. *Journal of Business & Economic Statistics*, 10(2), 229-235.
- Pagan, Adrian R., and G. William Schwert. (1990). Alternative models for conditional stock volatility. *Journal of econometrics*, 45(1), 267-290.

- Poon, Ser-Huang. (2005). *A practical guide to forecasting financial market volatility*. John Wiley & Sons.
- Roll, Richard. (1984). Orange juice and weather. *The American Economic Review*, 74(5), 861-880.
- Schwert, G. William (1989). Why does stock market volatility change over time? *Journal of Finance*. 44, 1115-1153.
- So, Mike KP, and L. H. Philip. (2006). Empirical analysis of GARCH models in value at risk estimation. *Journal of International Financial Markets, Institutions and Money*, 16(2), 180-197.
- Song, Haiyan, Xiaming Liu, and Peter Romilly. (1998). Stock returns and volatility: an empirical study of Chinese stock markets. *International Review of Applied Economics*, 12(1), 129-139.
- Taylor, Stephen. J. (1987). Forecasting the volatility of currency exchange rates. *International Journal of Forecasting*, 3(1), 159-170.
- Tofallis, Chris. (2015). A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, 66(8), 1352-1362.
- Tse, Yiu Kuen and Siew H. Tung. (1992). Forecasting volatility in the Singapore stock market. *Asia Pacific Journal of Management*, 9(1), 1-13.

Tse, Yiu Kuen. (1991). Stock returns volatility in the Tokyo Stock Exchange. *Japan and the World Economy*, 3(3), 285-298.

Yu, Jun. (2002). Forecasting volatility in the New Zealand stock market. *Applied Financial Economics*, 12(3), 193-202.

Zakoian, Jean-Michel. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and control*, 18(5), 931-955.

Zakoian, Jean-Michel. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and control*, 18(5), 931-955.

Zhang, Zhichao, and Hongyu Pan. (2006). Forecasting financial volatility: evidence from Chinese Stock Market. *Durham Business School Working Paper*. 06/02.

APPENDICES

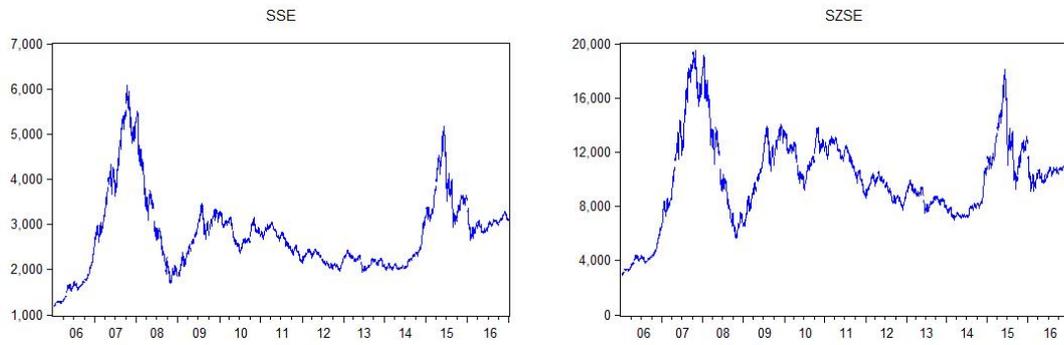
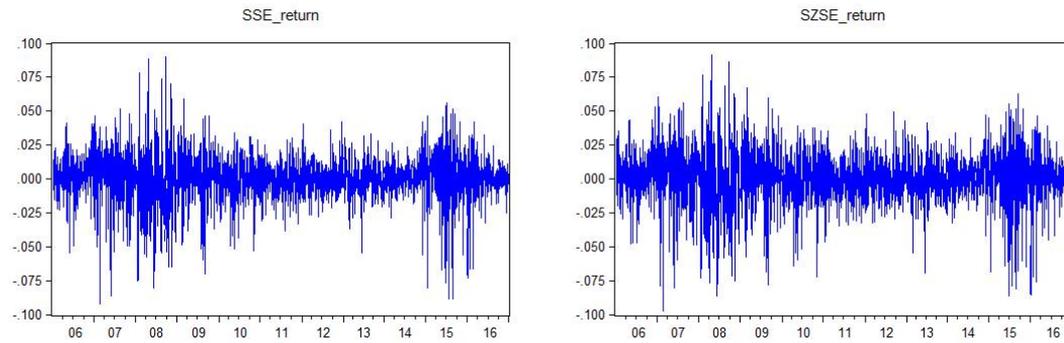
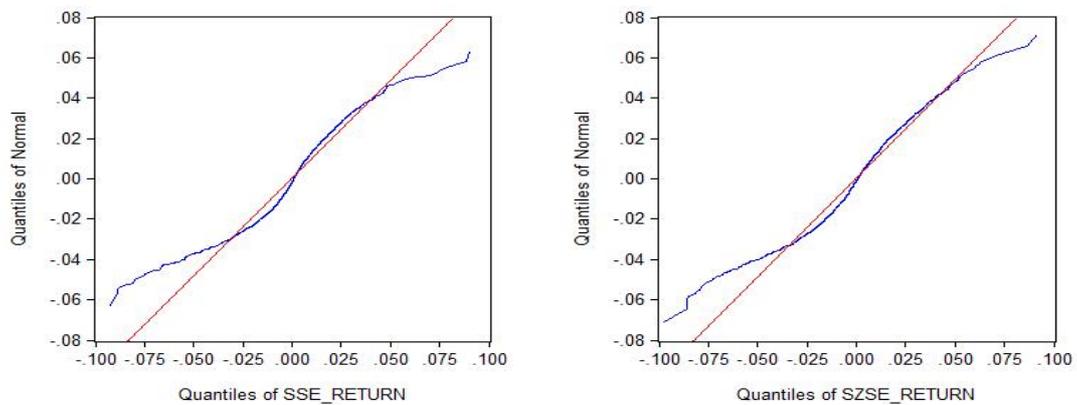
Figure 1. The trend graph of daily closing prices for SSE and SZSE.**Figure 2.** The trend graph of returns for SSE and SZSE.**Figure 3.** Normal Quantile-Quantile Plots for the SSE and SZSE.

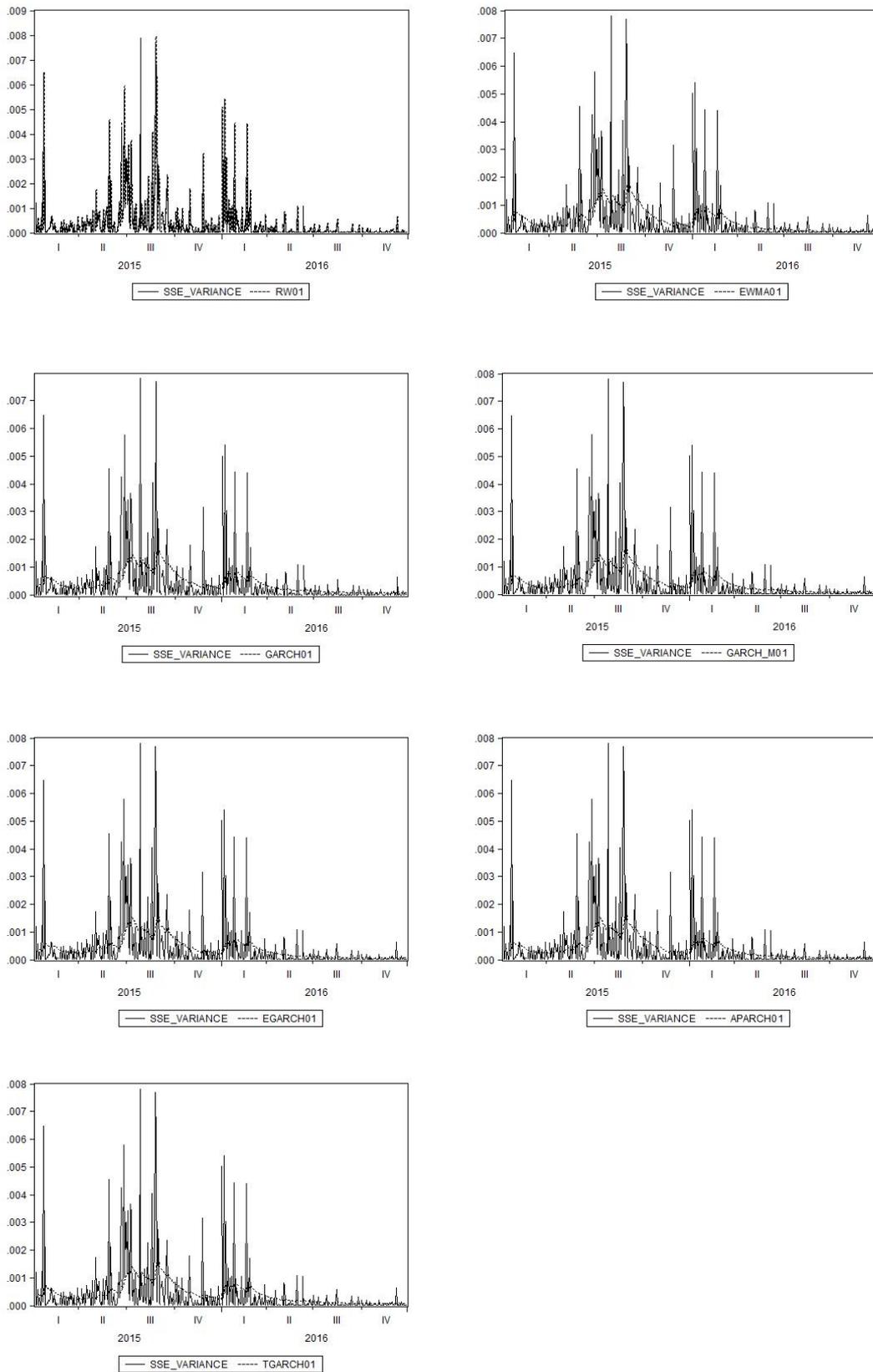
Figure 4. Graphs of the true and forecasted volatility for the SSE index.

Figure 5. Graphs of the true and forecasted volatility for the SZSE index.

