Degree: ***BSc Investment and Financial Risk Management***

***Bayes Business School***

Title: ***Forecasting Energy prices volatilities***

Name: ***Liyang Zhang***

Supervisor's name: ***Marchese Malvina***

Submission date: ***6th April 2021***

"I certify that I have complied with the guidelines on plagiarism outlined in the Course Handbook in the production of this dissertation and
that it is my own, unaided work".

Signature……………………Liyang Zhang……………………………….

# Forecasting Energy prices volatilities

Liyang Zhang

# Abstract

This paper employs several univariate GARCH family models to forecast the return volatility in two conventional gasoline markets. It compares the out-of-sample fitting abilities of different models and evaluates their out-of-sample predictive performance via the SPA test. The existence of symmetric leverage and long memory effects in conventional gasoline markets are also examined. Finally, a basic Markov-Switching GARCH model is established to alleviate the influence of extreme points caused by financial events in the energy market to improve the accuracy of return volatility forecasting.

# contents

# 1. Introduction

The price of crude oil and its refined products are closely related with global economic development and activities. And researchers, policymakers, and market actors have all voiced concern about the deviation of their price. And till now much research with the purpose of capturing the volatility of returns in oil related products market has been done, and among them the family of Autoregressive Conditional Heteroscedasticity (ARCH) models of Bollerslev (1986) and Engle (1982) has been widely applied to energy market. Later much research has found that different volatility forecasting models outperformed in different oil refined product markets. For example, Sadorsky (2006) has found that the standard GARCH (1,1) model well fits for crude oil volatility, while GJR fits conventional gasoline better. Instead of presuming the volatility effect is decayed in short time, Kang et al. (2009) used CGARCH and IGARCH model to capture long-memory univariance volatility in crude oil price and evaluate the forecast performance through Diebold and Mariano (1995) test, and concluded their model perform better than just using GARCH and IGARCH.

However, Hansen (2005) introduced a novel model comparison approach, the superior predictive ability (SPA) test, which has been shown to be more resilient than DM test and reality check (RC) test of White (2000). Another significant benefit of the SPA test is that it allows simultaneous evaluating of the performance of several models under a given loss function, whereas the DM test can only be used for performance comparing of two models. Meanwhile, based on CBOE's data, we find the price of WTI index in the start of 2014 to mid of 2016 was approximately 61.9% lower than the average level after financial crisis in 2008, which indicates the second most volatile period in oil market in 21 centuries. Among the historic of crude oil and its refined product markets, there are several time points which witnessed irregular price change and made the volatility forecast more difficult. Therefore, to forecast the volatility of return of conventional gasoline with less intervention of these break points, this paper will add a crude oil return as an explanatory variable, which has a highly correlated relationship with crude oil and better performed in both in sample fitting and out sample forecast when compared with just add ARMA terms.

The remainder of the paper is organized as follows. In Section 2, we introduced the sample data series we used, and described the statistical characteristics through explaining the pre-diagnostic test results. In Section 3, we explained the univariance models we used with their special features. In Section 4, we fitted our data with these models and compared their in-sample estimation performances. Then we use the most appropriate models to make forecast. In Section 5, we introduced the loss functions we used and evaluate the performance of our forecast trough SPA test. In Section 6, a summary of the out-sample volatility forecasting was provided.

## 2. Data and descriptive analysis

### 2.1 Data introduction

We use the daily price data (in U.S dollars per Gallon) of New York Harbor Conventional Gasoline Regular Spot Price FOB and U.S. Gulf Coast Conventional Gasoline Regular Spot Price FOB and West Texas Intermediate (WTI) from January 02, 1986, to June 12, 2017. The first two data are our forecast targets while the crude oil data is used to explain the forecast through offsetting the influence of market events. And we set our out-sample as the last four years, i.e., June 2, 2014, to June 12, 2017, which are used to evaluate the out-of-sample volatility forecasts accuracy. During period 2014-2017, the Shale Oil Revolution and lower demand towards oil products greatly affected the price of crude oil which decreased from about USD 116 to USD 40 per barrel. The crude oil market has a strong correlation with conventional gasoline market thus, could captured the extreme change in gasoline market, which help us to exam the performances of different volatility models of conventional Gasoline in such a great volatility testing time horizon. To analysis our data series which are obtained from The Thomson Reuters, we first assumed $P_t$ denote the price of conventional gasoline on day t. And in this paper, we use log returns to convert daily return into nominal percentage return. Therefore, for conventional gasoline, i.e., $r_t$ = 100 ln ($P_t$/ $P_{t-1}$) for t= 1,2,…, T, where $r_t$ denotes the returns for conventional gasoline at time t. Proved by Sadorsky(2006) and Kang et al. (2009), the daily actual volatility (variance) can be approximate by daily squared returns, i.e., $r^2$. **Figure 1** depicts a graphical depiction of the prices, returns, and volatility for conventional gasoline, whereas he left panel of the graphic shows data for the Harbor market, while the right panel shows data for the Gulf Coast market. As mentioned in section 1, the return volatility of conventional volatility has some extremely high values in particular years, which might provide a barrier for our forecast if we only based on autocorrelation, therefore, we tested volatility change pattern in independent variable crude oil and found they are comparable to each other.

**2.2 Statistic analysis**

Table 1 summarizes descriptive statistics for the conventional gasoline returns which have similar outcomes, both have a small mean return but a relative much higher standard deviation. In comparison to a normal distribution, their negative Skewness and high Skewness indicate that there is a greater likelihood that return decrease during a specific time and extreme price swings occur more frequently. At the 1% level of significance, the Jarque–Bera statistic reveals that the null hypothesis of normality is rejected. Both the Dickey-Fuller and Dickey-Fuller GLS unit-root tests show that the two returns are stationary, which is a prerequisite for forecasting. The F statistics in Lagrange Multiplier test had a p value of 0, which rejects the null that no arch effect in the data and led us to models that can capture volatility. The Q statistics in Ljung-Box rejects the null hypothesis that there are no autocorrelations of up to 30 orders for returns and their squared values. As a result, return and volatility are serially correlated and forecastable.

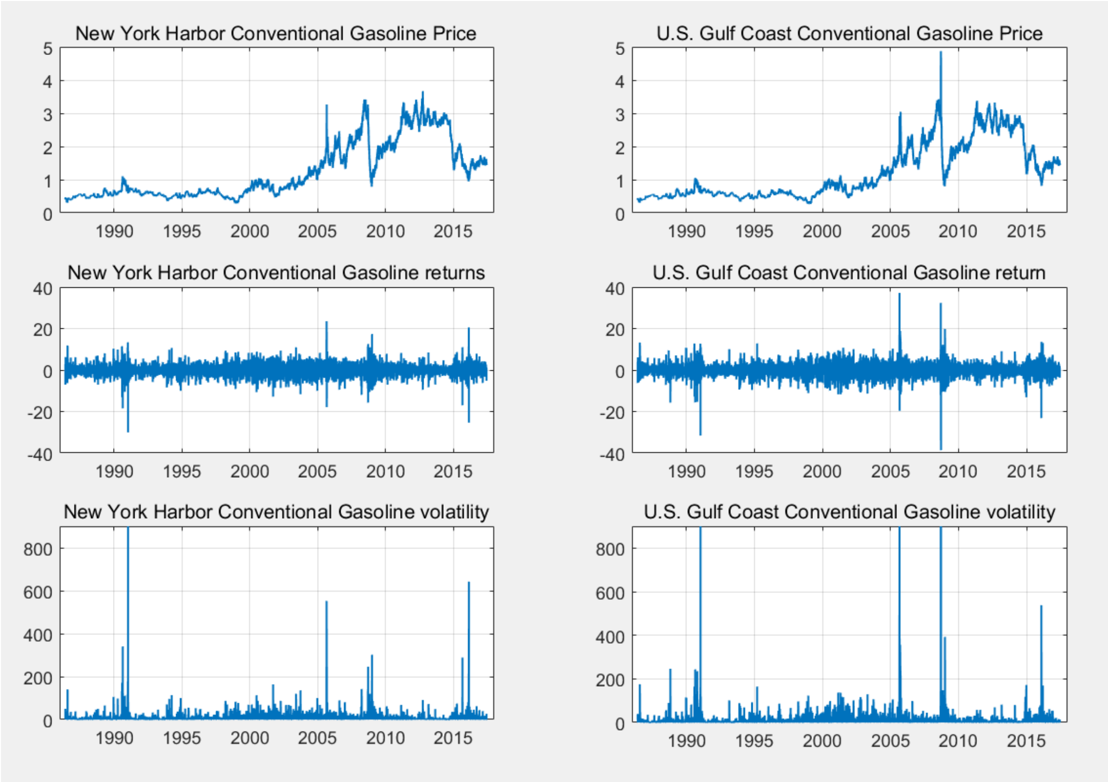


Figure 1

Table 1

| Describe statistic table | NYCGG | USCGG |
|---|---|---|
| Mean (%) | 0.0144 | 0.014956 |
| Std. Dev. (%) | 2.611921 | 2.879402 |
| Maximum | 23.50513 | 37.17351 |
| Minimum | -30.05855 | -38.6755 |
| Skewness | -0.230685 | -0.109483 |
| Kurtosis | 10.7083 | 17.54438 |
| Jarque-Bera | 19404.85* | 68853.92* |
| Q (30) | 98.965* | 69.678* |
| Q^2(30) | 1441.8* | 1435.1* |
| ADF | -20.99671* | -22.7093* |
| D-F | -2.171472* | -2.497455* |
| ARCH (22) | 3.506288* | 2.873801* |
| Observations | 7810 | 7810 |

Note: Note: The Jarque–Bera statistic tests for the null hypothesis of normality in the sample returns distribution. Q (30) is the Ljung–Box statistic of the return series for up to the 30th order serial correlation. ADF and D-F are the statistics of the augmented Dickey–Fuller and Dickey-Fuller GLS unit root tests, respectively, based on the lowest AIC value. ARCH (22) denotes the F-statistic for Lagrange Multiplier test.

* Indicates rejection at the 1% significance level.

# 3. Analysis

## 3.1 GARCH (1,1)

Based on the work of Sadorsky (2006), the GARCH(1,1) model proposed by Bollerslev(1986) fit good with the crude oil volatility. And Agnolucci (2009) also proved that GARCH(1,1) also worked well in crude oil option market. Therefore, we will first attempt GARCH(1.1) for our crude oil refined product forecast. The model is given by

$$r_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t = \sigma_t z_t \tag{3.1}$$

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{3.2}$$

$\mu_t$ represent the conditional mean, which we assume equal to 0 as the mean return is small compared with their volatility. And $\sigma_t^2$ stand for the conditional variance. The three parameters $\alpha, \beta\ and\ \omega$ are restricted greater than 0 to guarantee the positive value of $\sigma_t^2$. And parameter ω is a constant, while α+β captures the persistence of volatility shocks, which need to be smaller than 1. We also assume the $z_t$ follows Student-t distribution as in **table 1** we found their returns are not followed normal distribution.

## 3.2 TGARCH (1,1)

Although according to Agnolucci (2009) research, the leverage effect can hardly be observed in the oil series. The refined product may have some exceptions, therefore, instead of just using GARCH (1,1) model, which assume volatility responds symmetrically to the previous volatility change. We also use TGARCH (1,1), which was introduced by Zakoian(1994) and can deal with both negative and positive shock separately.

$$\sigma_t^2 = \omega + a\varepsilon_{t-1}^2 + \gamma\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2\Gamma_{t-1} \tag{3.3}$$

Where $\sigma_t^2$, $\alpha, \beta\ and\ \omega$ stand for the same things as they are in GARCH (1,1) model. $\Gamma_t$=1 if $\varepsilon_t$<0 and 0 otherwise. Thus, the sign of $\varepsilon_t$ have differential effects on the conditional variance. If leverage parameter $\gamma$ >0, negative shock will add volatility and if $\gamma \neq 0$, the impacts of news are asymmetric.

### 3.3 EGARCH (1,1)

To better fit with our situations that we have explanatory variable crude oil, an EGARCH model could perform well. And EGARCH model can also solve the non-normal distribution problem in some extend, which can be helpful for our data forecast.

$$\log(\sigma_t^2) = \omega + \alpha \left|\frac{u_t - 1}{h_t - 1}\right| + \frac{\gamma(u_t - 1)}{h_t - 1} + \beta \log(\sigma_{t-1}^2) \quad (3.4)$$

The $\gamma$ is the parameter that used to present the leverage effect, where a $\gamma>0$ means the future volatility change is bigger for positive shock than same degree negative shock.

### 3.4 FIGARCH (1, d, 1)

However, the above two models we mentioned above are all based on the assumption that volatility autocorrelation follows an exponential decays rate. And Aloui and Mabrouk (2010) research on energy market strongly support that long-memory of volatility exist in energy markets. Therefore, we applied fractionally integrated ARCH(FIGARCH) model, which allows volatilities' autocorrelation function to degrade at a hyperbolic rate. This means FIGARCH can better forecast the influence of previous volatility shock. The FIGARCH(1,d,1) model can be expressed in the following way:

$$\sigma_t^2 = \omega(1-\beta L)^{-1} + [1-(1-\beta L)]^{-1} \varphi(L)(1-L)^d \varepsilon_t^2 \quad (3.5)$$

L is the lag operator and d is the fractional difference parameter, which measure the long memory and range from 0 to 1. When d=0 the FIGARCH (1,0,1) is exactly the GARCH(1,1) model, which has a volatility geometric decay associated with. Whereas the volatility shocks last for long time associated with d = 1.

## 4. Estimation results

### 4.1 in-sample performance

To make the models' long-term predictions more accurate we added AR MA terms and additional variable crude oil in our models, which is highly correlated with both New York Harbor and U.S. Gulf Coast. Moreover, to select the most appropriate AR MA terms up to 4, we compared the Akaike information criterion among different combinations. And we find, for New York Harbor, EGARCH and FIGARCH models best performed with ar(1) ma(3) terms added, while TGARCH and GARCH models fitted better with ar(1) ar(3) ma(1) ma(3) terms. Similarly, for US Gulf Coast, all models worked optimal with ar(2) ar(3) ma(3) ma(4) terms.

After fitted data with the univariate volatility models discussed in section 3, we get the parameters estimators shown in **Table 2**. First, we can see that all the models tend to reject the estimate of $\mu$, which is reasonable as mentioned in section 2 that our data have a tiny mean and can be approximated regarded as 0. Then started with GARCH (1,1) model, all other parameters are highly significant with p value equal to 0. And $\alpha+\beta$ are quite close to 1, which denote high level of volatility persistent in conventional gasoline return. And then the asymmetric parameters $\gamma$ in TGARCH and EGARCH models are all at 95% significant level but quiet small when compared with other parameters, which reflect a weak asymmetric effect in this market.

Lastly, the estimate value of parameter d in both FIGARCH models are 0.36 and 0.43 respectively and they are all highly significant, this denotes an significant evident of persistence of long memory in return volatility change. Meanwhile, the diagonal tests for squared residuals includes logarithm maximum likelihood function value, Ljung and Box Q-statistics and ARCH-LM test are shown in the lower part of **Table 2** as well. Among all these univariance models the value of log(L) are quite similar and cannot reject null hypothesis of no serial correlation at 1% significant level (better than before it fitted). And the ARCH effects are well captured in U.S. Gulf Coast market through all the models, while in New York Harbor market only FIGARCH outperforms and captures some of ARCH effect. Overall, there is no specific model that outperforms other in the in-sample estimation.

Table 2

| | NY | US | NY | US | NY | US | NY | US |
|---|---|---|---|---|---|---|---|---|
| | GARCH(1,1) | | TGARCH(1,1) | | EGARCH(1,1) | | FIGARCH(1,1) | |
| μ | 0.003599 | 0.002617 | 0.004698 | 0.004206 | 0.011201 | 0.004557 | 0.002945 | 0.003088 |
| z-Statistic | 0.4207 | 0.690616 | 1.105674 | 1.182559 | 0.650383 | 1.307443 | 0.169355 | 0.811986 |
| P-value | 0.8589 | 0.4898 | 0.2689 | 0.237 | 0.5154 | 0.1911 | 0.8655 | 0.4168 |
| β_co | 0.759558 | 0.787806 | 0.760951 | 0.7891 | 0.760693 | 0.793047 | 0.760573 | 0.788509 |
| z-Statistic | 96.03175 | 94.6267 | 96.35618 | 94.97839 | 96.57389 | 96.30127 | 94.94361 | 94.27811 |
| P-value | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ω | 0.112751 | 0.099347 | 0.098621 | 0.094039 | -0.110579 | -0.129613 | 0.25424 | 0.260051 |
| z-Statistic | 6.781087 | 6.693263 | 6.654826 | 6.792734 | -11.39412 | -12.32164 | 3.797918 | 4.286992 |
| P-value | 0 | 0 | 0 | 0 | 0 | 0 | 0.0001 | 0 |
| α | 0.105323 | 0.116392 | 0.111272 | 0.134157 | 0.188483 | 0.2239 | | |
| z-Statistic | 11.62328 | 12.39956 | 10.32563 | 11.33285 | 13.51737 | 14.97629 | | |
| P-value | 0 | 0 | 0 | 0 | 0 | 0 | | |
| β | 0.869656 | 0.868812 | 0.882773 | 0.87584 | 0.975366 | 0.972201 | 0.34818 | 0.374936 |
| z-Statistic | 85.42375 | 94.73892 | 93.59577 | 100.397 | 239.9673 | 242.211 | 3.282893 | 3.940875 |
| P-value | 0 | 0 | 0 | 0 | 0 | 0 | 0.001 | 0.0001 |
| γ | | | -0.033971 | -0.051487 | 0.034987 | 0.034218 | | |
| z-Statistic | | | -2.519158 | -3.59621 | 3.80558 | 3.723645 | | |
| P-value | | | 0.0118 | 0.0003 | 0.0001 | 0.0002 | | |
| φ | | | | | | | 0.149745 | 0.122678 |
| z-Statistic | | | | | | | 1.660588 | 1.576176 |
| P-value | | | | | | | 0.0968 | 0.115 |
| d | | | | | | | 0.355555 | 0.434862 |
| z-Statistic | | | | | | | 8.651708 | 9.445322 |
| P-value | | | | | | | 0 | 0 |
| Log(L) | -13763.81 | -14226.94 | -13741.26 | -14225.33 | -13738.32 | -14208.07 | -13751.25 | -14211.83 |
| Q(10) | 10.243 | 12.927 | 16.523 | 14.728 | 26.326 | 15.374 | 24.675 | 12.927 |
| P-value | 0.175 | 0.044 | 0.011 | 0.022 | 0.001 | 0.018 | 0.002 | 0.044 |
| Q(20) | 15.152 | 23.958 | 24.032 | 28.16 | 31.528 | 30.831 | 29.347 | 24.254 |
| P-value | 0.585 | 0.09 | 0.089 | 0.03 | 0.025 | 0.014 | 0.044 | 0.084 |
| ARCH(10) | 4.59059 | 1.546474 | 4.98982 | 1.579392 | 11.17739 | 2.341401 | 1.52785 | 0.769535 |
| P-value | 0 | 0.1163 | 0 | 0.106 | 0 | 0.0095 | 0.1226 | 0.6582 |
| ARCH(20) | 2.940744 | 1.213917 | 3.206891 | 1.202712 | 6.2767 | 1.818374 | 1.371453 | 1.248581 |
| P-value | 0 | 0.2309 | 0 | 0.2405 | 0 | 0.0142 | 0.1243 | 0.2031 |

## 4.2 Out sample Forecasting graph comparation

Although, in sample estimation fitness is a kid of measurement standard, the most significant judgment criteria is the out-sample forecast performance. After forecasted from GARCH family models, we will get the predicted returns, and to get the volatility, we use square return as mentioned in Section 2. And the comparisons of out-sample volatility forecast forecasted from June 2, 2014, to June 12, 2017, for New York Harbor and U.S. Gulf Coast conventional gasoline through the 4 models are shown in Figure 1 and Figure 2 respectively.

From these two graphs, they show all the models' forecast can capture most of the real conventional gasoline volatility. But we cannot do a precise judgement only based on observing the graph. Besides, the forecasted GARCH terms (conditional variance), shown in Figure 3 and Figure 4, of each model converged to a fixed number very fast. This may cause bias in precise forecast.

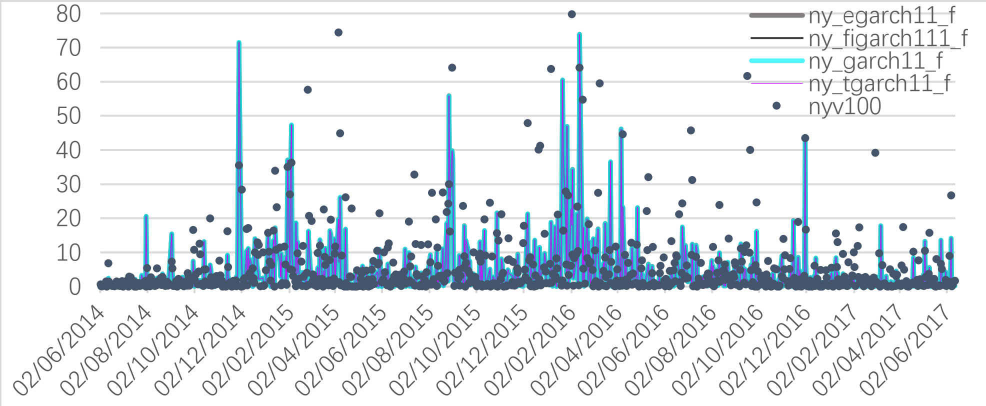


Figure1

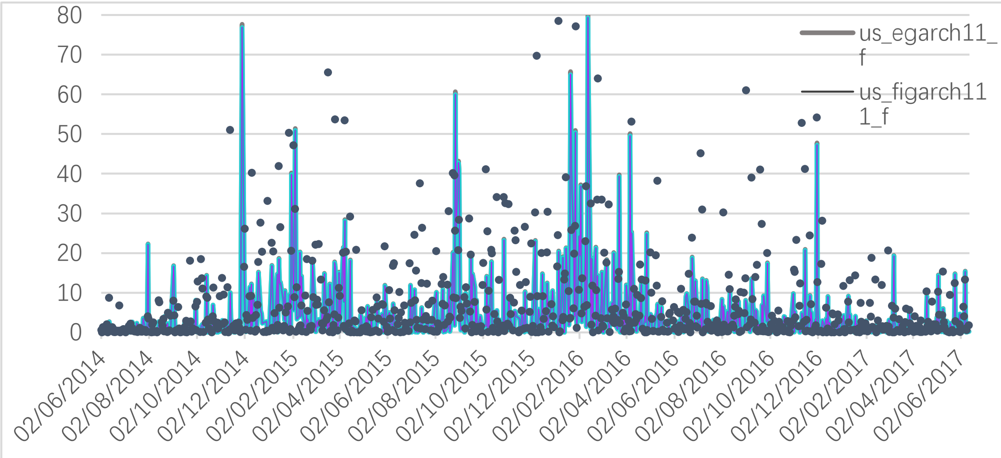


Figure 2

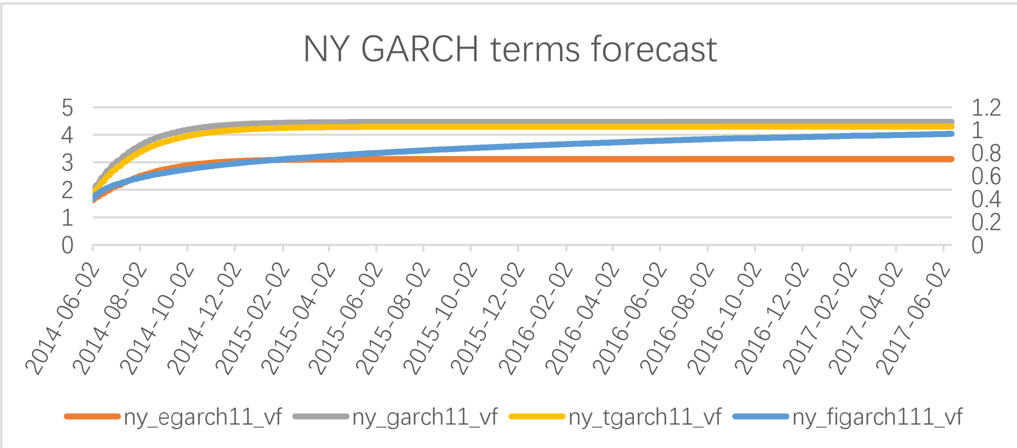


Figure 3

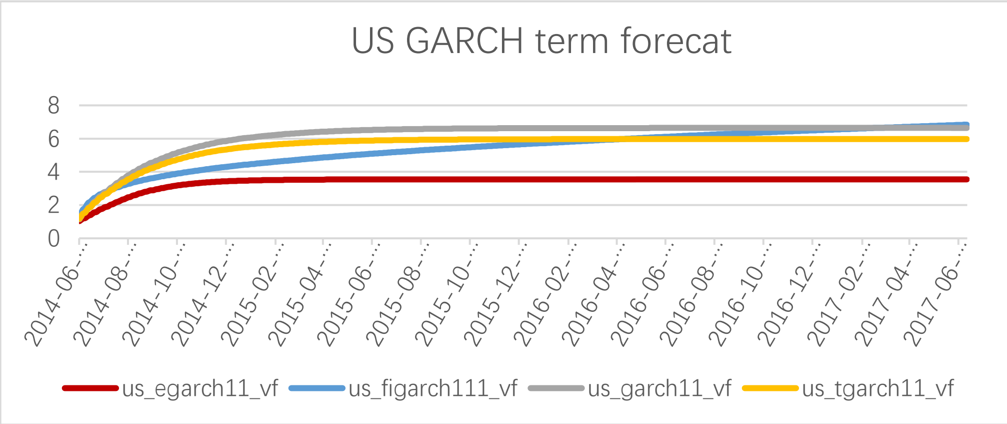


Figure 4

## 5. Out sample Forecasting performance

### 5.1 Loss Functions introduction

There are a lot of forecasting test for judging the accuracy of forecast. And the loss function is the foundation of conducting these robust tests. As investigated by Bollerslev et al. (1994) and Lopez (2001), it is unclear whether the loss function is better suited for quantifying the forecasting accuracy of volatility models. We employ the following five loss functions or accuracy statistics as the forecasting error criterion, rather than making a single decision:

$$MSE = n^{-1} \sum_{t=1}^{n} (\sigma_t^2 - \hat{\sigma}_t^2)^2 \tag{5.1}$$

$$MAE = n^{-1} \sum_{t=1}^{n} |\sigma_t^2 - \hat{\sigma}_t^2| \tag{5.2}$$

$$HMSE = n^{-1} \sum_{t=1}^{n} (1 - \sigma_t^2 / \hat{\sigma}_t^2)^2 \tag{5.3}$$

$$R^2LOG = n^{-1} \sum_{t=1}^{n} [ln\left(\frac{\sigma_t^2}{\hat{\sigma}_t^2}\right)]^2 \tag{5.4}$$

$$QLIKE = n^{-1} \sum_{t=1}^{n} (ln(\sigma_t^2) + \frac{\sigma_t^2}{\hat{\sigma}_t^2}) \tag{5.5}$$

where n is the number of predicted samples; MSE and MAE denote the mean square error and mean absolute error, respectively; The MSE modified for heteroskedasticity are HMSE; QLIKE is the loss resulted by conducting a Gaussian likelihood; a $\mathrm{R^2LOG}$ is close to the $\mathrm{R^2}$ of the Mincer–Zarnowitz regressions, which is also an evaluation criterion for volatility forecast. And each of these loss function can well perform towards special purposes. For example, in value at risk management applications, the MSE criteria is one of the most important loss functions as it tends to show the accurate of forecast for a higher volatility level.

**5.2 SPA test**

And to help us make a more reliable decision of better performance model, we did the superior predictive ability (SPA) test for determining whether a specified model is outperforming under a given loss function. As an extension of Diebold and Mariano (1995) White test, spa test can be used to compare more than two models' performance at one time and is made more robust by including bootstrap. In SPA test, it always compares the forecast performance between the benchmark model and other models, which give a conclusion of whether benchmark can be rejected as a good forecast model. In our case we will firstly use GARCH (1,1) model as the bench mark and the rest FIGARCH (1,d,1), EGARCH (1,1), TGARCH(1,1) will allow us to test whether including long memory or asymmetries in the univariance models improved the forecast accuracy. And then we will change the benchmark to other models to see if they are the winner of volatility forecast model.

The intrinsic logic of SPA test is it compare the expected value of loss function among benchmark and other models, and its' null suggest the benchmark will outperform than others. In more specific, under a fixed loss function, SPA test account for the difference $d_{j,k,t}$ of the losses between benchmark $L_{j,0,t}$ and other models $L_{j,k,t}$(k = 1, 2 ,... ; t=1,2,…T, (where k indicate the other models and t indicate the time of the loss).

Ho: $\lambda_{j,k} = E\left(d_{j,k,t}\right) \leq 0 \; for \; all \; K = 1, \ldots.,m.$

And the statistic test can be written as:

$T_{SPA} = max(\sqrt{T} \max_{i=1,\ldots,K} \frac{\bar{d}_{j,k}}{\widehat{\omega_k}}$ , 0), where $\bar{d}_{j,k} = \frac{1}{T}\sum_{T=1}^{T} d_{j,k,t}$ is the difference in sample loss between the benchmark and the competing model k. $\widehat{\omega_k}$ is the estimator of $\omega_k = \lim_{T\to\infty} \sqrt{\sqrt{T} variance(\bar{d}_{j,k})}$. Further in-depth discussion can be found in the study of Marchese et al. (2020) and Hansen (2005), which covered a lot of technique details of SPA test.

The following Table 3 is the results of the SPA p-value under different loss functions, where the left-hand side list is the choice of benchmarks. $P_c, P_l \; and \; P_u$ indicate consistent p-values as well as their lower and upper bounds. We found no models are rejected at 1% of significance level under any loss function, which means we cannot do a directly find a bad model who is always outperformed by others. However, a higher p value of SPA test means the benchmark model is not outperformed by all of the other models is not rejected. When benchmark model is FIGARCH (1,d,1) the SPA give a p value equal to 1 under 3 out of 5 loss functions, which means FIGARCH (1,1) forecast performance at least equal good with other models in New York Harbor market. Similarly, with p value equal to 1 under 4 out of 5 loss functions, EGARCH (1, d, 1) performed at least equally good in U.S. Gulf Coast market through out all the models. Therefore, we have confident to believe asymmetry and long memory should be accept in forecasting conventional gasoline volatility. Additionally, compared the p value of SPA test between GARCH (1,1) and TGARCH (1,1) models, we find the value for TGARCH is bigger under 4 out of 5 loss functions, which also support the persistence of leverage effect.

In summary, none of the four univariance GARCH models examined in this research are obviously better than other models. However, models captured long memory and asymmetric leverage effect will have a more likely possibility to outperform. And the p value is largely affected by the loss functions used and the market choose. Therefore, we need to evaluate the model forecast performance cautiously.

Table 3

SPA test for conventional gasoline

| Benchmark | L-MSE | | | L-MAE | | | L-HMSE | | | L-R^2Log | | | L-QLIKE | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NY | PL | PC | PU | PL | PC | PU | PL | PC | PU | PL | PC | PU | PL | PC | PU |
| GARCH(1,1) | 0.5192 | 0.5619 | 0.5671 | 0.0015 | 0.0015 | 0.0015 | 0.1120 | 0.1120 | 0.1217 | 0.2356 | 0.2356 | 0.2411 | 0.1150 | 0.1150 | 0.1155 |
| TGARCH(1,1,1) | 0.6719 | 0.8552 | 0.8615 | 0.0008 | 0.0009 | 0.0009 | 0.1769 | 0.1807 | 0.1867 | 0.2186 | 0.3131 | 0.3137 | 0.1655 | 0.1745 | 0.1726 |
| EGARCH(1,1) | 0.2257 | 0.2257 | 0.2265 | 0.0303 | 0.0305 | 0.0305 | 1 | 1 | 1 | 0.4364 | 0.5485 | 0.5535 | 1 | 1 | 1 |
| FIGARCH(1,1,1) | 1 | 1 | 1 | 1 | 1 | 1 | 0.1923 | 0.3595 | 0.3559 | 1 | 1 | 1 | 0.1957 | 0.3591 | 0.3618 |
| US | | | | | | | | | | | | | | | |
| GARCH(1,1) | 0.1495 | 0.2655 | 0.2655 | 0.0123 | 0.0188 | 0.0235 | 0.0364 | 0.0364 | 0.0397 | 0.0383 | 0.0383 | 0.0388 | 0.0369 | 0.0369 | 0.0407 |
| TGARCH(1,1,1) | 0.1410 | 0.2525 | 0.2572 | 0.0486 | 0.0486 | 0.0772 | 0.1227 | 0.2160 | 0.2951 | 0.1232 | 0.2141 | 0.2936 | 0.1278 | 0.2214 | 0.2950 |
| EGARCH(1,1) | 1 | 1 | 1 | 0.0160 | 0.0160 | 0.0160 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| FIGARCH(1,1,1) | 0.1652 | 0.1652 | 0.1667 | 1 | 1 | 1 | 0.0360 | 0.0360 | 0.0360 | 0.0338 | 0.0338 | 0.0358 | 0.0340 | 0.0340 | 0.0356 |

note: $L_{MSE}, L_{MAE}, L_{HMSE}, L_{R^2}Log \; and \; L_{QLIKE}$ denote the loss function we introduced above, respectively. The number of sieve bootstrap samples we used is 10,000 with average block length equal to 100. ↵

## 6. Conclusion

In this paper, we use univariance GARCH family models to forecast the volatility in conventional gasoline markets. And we choose 5 loss functions as the framework to conduct the superior predictive ability (SPA) test to anticipate volatility in the conventional gasoline market. We improved the forecast accuracy of the models by adding explanatory variable and optimal AR and MA terms. And the forecast evaluation results show long memory and asymmetric leverage effect exist in this market. However, all these models will forecast a quickly converge to constant GARCH term, which could be a bias for long-term forecast. Therefore, a method that can generate dynamic GARCH term could be more optimal, e.g., Markov-Switching. In the appendix we attempted the most naïve Markov-Switching GARCH (1,1) model with crude oil added as an explanatory variable. Although it won't consider the long memory effect, which has been proved exist in this market, it improved the conditional volatility forecast and gave us a direction to improve our models, like using Markov-Switching FIAPARCH model to both capture asymmetric and long-memory effect. This paper only studied the conventional gasoline market volatility based on univariance models, looking forwards, it could be interesting to capture the correlation among different refined products in energy market to forecast their co-movement through multivariance GARCH family models, such as the constant conditional correlation GARCH models and the dynamic conditional correlation GARCH models with long memory and asymmetric of Marchese et al. (2020) and the regime switching asymmetric DCC approach of Pan et al. (2014).

## References


[1] Agnolucci, P., 2009. Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models. Energy Economics 31, 316-321.

[2] Aloui, C., Mabrouk, S., 2010. Value-at-risk estimations of energy commodities via longmemory, asymmetry and fat-tailed GARCH models. Energy Policy 38 (5), 2326-2339

[3] Ardia, D., Bluteau, K., Boudt, K. and Trottier, D., 2016. Markov-Switching GARCH Models in R: The MSGARCH Package. SSRN Electronic Journal. https://CRAN.R-project.org/package=MSGARCH.

[4] Bollerslev, T., Engle, R.F., Nelson, D., 1994. ARCH models. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. IV. Elsevier Science BV, Amsterdam, pp. 2961-3038.

[5] Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307-327.

[6] Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 13, 253-263.

[7] Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987-1007.

[8] Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business & Economic Statistics 23, 365-380

[9] Kang, S.H., Kang, S.M., Yoon, S.M., 2009. Forecasting volatility of crude oil markets. Energy Economics 31, 119-125.

[10] Lopez, J., 2001. Evaluating the predictive accuracy of volatility models. Journal of Forecasting, 20(2), pp.87-109.

[11] Marchese, M., Kyriakou, I., Tamvakis, M. and Di Iorio, F., 2020. Forecasting crude oil and refined products volatilities and correlations: New evidence from fractionally integrated multivariate GARCH models. *Energy Economics*, 88.

[12] Pan, Z., Wang, Y. and Yang, L., 2014. Hedging crude oil using refined product: A regime switching asymmetric DCC approach. *Energy Economics*, 46, pp.472-484.

[13] Sadorsky, P., 2006. Modeling and forecasting petroleum futures volatility. Energy Economics

28, 467–488.

[14] White, H., 2000. A reality check for data snooping. Econometrica 68, 1097-1126.

[15] Zakoian, J., 1994. Threshold heteroskedastic models. Journal of Economic Dynamics and Control, 18(5), pp.931-955.

# Appendix

## MSGARCH Tables and Figures

Table 4 NY model specification

```
Specification type: Markov-switching
Specification name: sGARCH_std sGARCH_std
Number of parameters in each variance model: 3 3
Number of parameters in each distribution: 1 1
------------------------------------------
Fixed parameters:
None
------------------------------------------
Across regime constrained parameters:
None
------------------------------------------
Fitted parameters:
          Estimate Std. Error  t value  Pr(>|t|)
alpha0_1    0.1080     0.1403   0.7699 2.207e-01
alpha1_1    0.0717     0.1315   0.5451 2.928e-01
beta_1      0.9086     0.0267  34.0095    <1e-16
nu_1       17.9304    20.7156   0.8655 1.934e-01
alpha0_2    0.0462     0.0252   1.8357 3.320e-02
alpha1_2    0.0664     0.0804   0.8250 2.047e-01
beta_2      0.9299     0.0054 171.9958    <1e-16
nu_2        4.3627     0.4184  10.4263    <1e-16
P_1_1       0.9963     0.0027 371.9979    <1e-16
P_2_1       0.0025     0.0034   0.7138 2.377e-01
------------------------------------------
Transition matrix:
      t+1|k=1 t+1|k=2
t|k=1  0.9963  0.0037
t|k=2  0.0025  0.9975
------------------------------------------
Stable probabilities:
State 1 State 2
 0.3981  0.6019
------------------------------------------
LL: -16732.6904
AIC: 33485.3808
BIC: 33553.9843
------------------------------------------
```

Table 5 US model specification

```
Specification type: Markov-switching
Specification name: sGARCH_std sGARCH_std
Number of parameters in each variance model: 3 3
Number of parameters in each distribution: 1 1
------------------------------------------
Fixed parameters:
None
------------------------------------------
Across regime constrained parameters:
None
------------------------------------------
Fitted parameters:
          Estimate Std. Error  t value  Pr(>|t|)
alpha0_1    0.0450     0.0193   2.3349 9.775e-03
alpha1_1    0.0380     0.0294   1.2954 9.759e-02
beta_1      0.9590     0.0017 563.6287    <1e-16
nu_1        4.0992     0.7909   5.1829 1.092e-07
alpha0_2    0.0364     0.0210   1.7337 4.148e-02
alpha1_2    0.1245     0.1711   0.7279 2.333e-01
beta_2      0.8745     0.0015 564.7777    <1e-16
nu_2       39.5904    60.8896   0.6502 2.578e-01
P_1_1       0.9820     0.0280  35.0184    <1e-16
P_2_1       0.0248     0.0083   2.9814 1.434e-03
------------------------------------------
Transition matrix:
      t+1|k=1 t+1|k=2
t|k=1  0.9820  0.0180
t|k=2  0.0248  0.9752
------------------------------------------
Stable probabilities:
State 1 State 2
 0.5801  0.4199
------------------------------------------
LL: -16729.3697
AIC: 33478.7393
BIC: 33547.3429
```

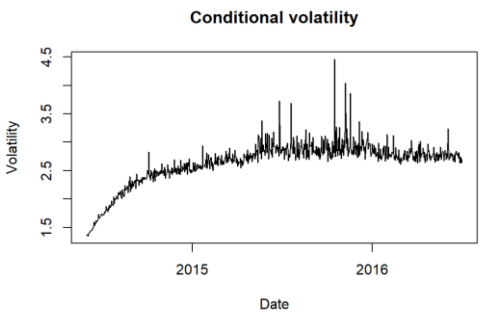


Figure 5 NY GARCH term forecast

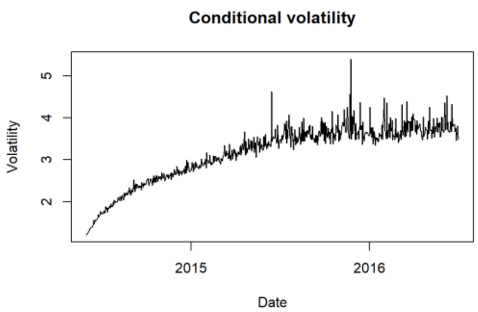


Figure 6 US GARCH term forecast

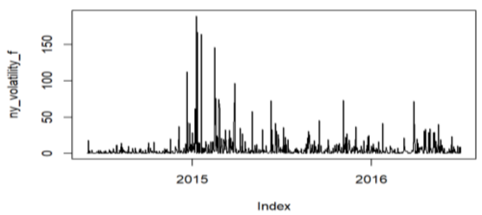


Figure 7 NY daily volatility forecast

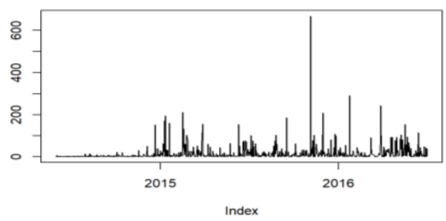


Figure 8 US daily volatility forecast

## MSGARCH code in R

```
### msgarch(1,1)__with crude oil as explanatory variable

##import variable(import excel msgarch in R)

library(zoo)

library(MSGARCH)

library(readxl)

msgarch <- read_excel("C:/Users/Olivier Frex/Desktop/msgarch.xlsx")

### code for import data from excel file "msgarch"

co <- read_excel("C:/Users/Olivier Frex/Desktop/co.xlsx")

### code for import data from excel file "co"

X.Date<-as.Date(msgarch$y)

arpny<-zoo(msgarch$x,X.Date)

arpus<-zoo(msgarch$z,X.Date)

Y.Date<-as.Date(co$y)

arpcor<-zoo(co$o,Y.Date)

## msgarch(1,1) estimate and forecast for NY

spec <- CreateSpec(variance.spec = list(model = c("sGARCH")), distribution.spec =

list(distribution = c("std")), switch.spec = list(do.mix = FALSE, K = 2))

f_custom_optim <- function(vPw, f_nll, spec, data, do.plm){

out <- stats::optim(vPw, f_nll, spec = spec, data = arpcor,

do.plm = FALSE, method = "Nelder-Mead")

return(out)

}

fit <- FitML(spec, data = arpny, ctr = list(OptimFUN = f_custom_optim))

summary(fit)

### code for showing estimator

pred <- predict(object = fit, nahead =764L, do.return.draw = TRUE)

plot(pred[["vol"]])

### code for plotting garch term forecast

predreturn=pred[["draw"]]
```

```
ny_volatility_f=predreturn[1:764,1]^2
plot(ny_volatility_f)
### code for plotting volatility forecast

## msgarch(1,1) estimate and forecast for US
spec <- CreateSpec(variance.spec = list(model = c("sGARCH")), distribution.spec =
list(distribution = c("std")), switch.spec = list(do.mix = FALSE, K = 2))
f_custom_optim <- function(vPw, f_nll, spec, data, do.plm){
out <- stats::optim(vPw, f_nll, spec = spec, data = arpcor,
do.plm = FALSE, method = "Nelder-Mead")
return(out)
}
fit <- FitML(spec, data = arpus, ctr = list(OptimFUN = f_custom_optim))
summary(fit)
### code for showing estimator
pred <- predict(object = fit, nahead =764L, do.return.draw = TRUE)
plot(pred[["vol"]])
### code for plotting garch term forecast
predreturn=pred[["draw"]]
us_volatility_f=predreturn[1:764,1]^2
plot(us_volatility_f)
### code for plotting volatility forecast
```

SPA code in MATLAB

```
#### MSE, MAE, HMSE...... are 764x4 tables with egarch11 figarch111 garch11 tgarch11 list in each column accordingly.

###MSE
#FIGARCH
models=[table2array(MSE(:,1)) table2array(MSE(:,3)) table2array(MSE(:,4))]
bench=[table2array(MSE(:,2))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')

#GARCH
models=[table2array(MSE(:,1)) table2array(MSE(:,2)) table2array(MSE(:,4))]
bench=[table2array(MSE(:,3))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')

#TGARCH
models=[table2array(MSE(:,1)) table2array(MSE(:,2)) table2array(MSE(:,3))]
bench=[table2array(MSE(:,4))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')

#EGARCH
models=[table2array(MSE(:,2)) table2array(MSE(:,3)) table2array(MSE(:,4))]
bench=[table2array(MSE(:,1))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
```

```
###MAE
#FIGARCH
models=[table2array(MAE(:,1)) table2array(MAE(:,3)) table2array(MAE(:,4))]
bench=[table2array(MAE(:,2))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')

#GARCH
models=[table2array(MAE(:,1)) table2array(MAE(:,2)) table2array(MAE(:,4))]
bench=[table2array(MAE(:,3))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')

#TGARCH
models=[table2array(MAE(:,1)) table2array(MAE(:,2)) table2array(MAE(:,3))]
bench=[table2array(MAE(:,4))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')

#EGARCH
models=[table2array(MAE(:,2)) table2array(MAE(:,3)) table2array(MAE(:,4))]
bench=[table2array(MAE(:,1))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')

###HMSE
#FIGARCH
models=[table2array(HMSE(:,1)) table2array(HMSE(:,3)) table2array(HMSE(:,4))]
bench=[table2array(HMSE(:,2))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
```

```
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')


#GARCH
models=[table2array(HMSE(:,1)) table2array(HMSE(:,2)) table2array(HMSE(:,4))]
bench=[table2array(HMSE(:,3))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')


#TGARCH
models=[table2array(HMSE(:,1)) table2array(HMSE(:,2)) table2array(HMSE(:,3))]
bench=[table2array(HMSE(:,4))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')


#EGARCH
models=[table2array(HMSE(:,2)) table2array(HMSE(:,3)) table2array(HMSE(:,4))]
bench=[table2array(HMSE(:,1))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')



###R^2LOG
#FIGARCH
models=[table2array(R2LOG(:,1)) table2array(R2LOG(:,3)) table2array(R2LOG(:,4))]
bench=[table2array(R2LOG(:,2))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')


#GARCH
models=[table2array(R2LOG(:,1)) table2array(R2LOG(:,2)) table2array(R2LOG(:,4))]
```

```
bench=[table2array(R2LOG(:,3))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')

#TGARCH
models=[table2array(R2LOG(:,1)) table2array(R2LOG(:,2)) table2array(R2LOG(:,3))]
bench=[table2array(R2LOG(:,4))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')

#EGARCH
models=[table2array(R2LOG(:,2)) table2array(R2LOG(:,3)) table2array(R2LOG(:,4))]
bench=[table2array(R2LOG(:,1))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')



###QLIKE
#FIGARCH
models=[table2array(QLIKE(:,1)) table2array(QLIKE(:,3)) table2array(QLIKE(:,4))]
bench=[table2array(QLIKE(:,2))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')

#GARCH
models=[table2array(QLIKE(:,1)) table2array(QLIKE(:,2)) table2array(QLIKE(:,4))]
bench=[table2array(QLIKE(:,3))]
[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
```

```
#TGARCH

models=[table2array(QLIKE(:,1)) table2array(QLIKE(:,2)) table2array(QLIKE(:,3))]

bench=[table2array(QLIKE(:,4))]

[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')

[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')


#EGARCH

models=[table2array(QLIKE(:,2)) table2array(QLIKE(:,3)) table2array(QLIKE(:,4))]

bench=[table2array(QLIKE(:,1))]

[C, RCP]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')

[C,U,L]=bsds(bench,models,10000,100,'STUDENTIZED','STATIONARY')
```