

Stony Brook University



OFFICIAL COPY

The official electronic file of this thesis or dissertation is maintained by the University Libraries on behalf of The Graduate School at Stony Brook University.

© All Rights Reserved by Author.

Regime switching and long memory in High Frequency financial data

A Dissertation presented

by

Xiao Zhang

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

(Quantitative Finance)

Stony Brook University

August 2015

Stony Brook University

The Graduate School

Xiao Zhang

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation

Zari Rachev

Professor, Department of Applied Mathematics and Statistics

Xinyun Chen

Assistant Professor, Department of Applied Mathematics and Statistics

Aaron Kim

Assistant Professor, Department of Applied Mathematics and Statistics

Keli Xiao

Assistant Professor, College of Business

Petar M. Djuric

Professor, Department of Electrical and Computer Engineering

This dissertation is accepted by the Graduate School

Charles Taber

Dean of the Graduate School

Abstract of the Dissertation

Regime switching and long memory in High Frequency financial data

by

Xiao Zhang

Doctor of Philosophy

in

Applied Mathematics and Statistics

(Concentration - Quantitative Finance)

Stony Brook University

2015

This thesis is to incorporate Markov Regime Switching model with Fractionally Integrated process in order to capture abrupt change and regime persistence simultaneously in long memory dynamic volatility process. We adapt truncated ARCH ∞ to estimation scheme of our model. We carry out standard back testing procedure to validate Regime Switching FIGARCH VaR based forecasts, on S&P 500 and SHSZ 300 data in 1 minute and 5 minute frequencies.

In the Chapter 1, Regime Switching model, and parameter estimation steps based on truncated ARCH infinite and Hamilton filter will be given. Topics like stationary conditions of RS-FIGARCH and standard Normal Tempered Stable distributions as fat-tailed innovations of time series are also covered.

In Chapter 2, Fractionally Integrated GARCH is reviewed, and incorporated with Regime switching model. Modified likelihood ratio based test proposed by (Kasahara, 2013) is introduced as test against multiple regimes.

Chapter 3 is to discuss VaR-based back testing procedure, using China's ShanghaiShenzhen 300 Index log return series and S&P 500 log return series, both in 1 minute and 5 minute frequencies. Backtesting results are given, 99%, 99.5% and 99.9% VaRs are compared with log returns illustratively and violations of Kupiec

test at 0.01 and 0.05 significance levels are present as well. We claim that Regime switching FIGARCH not only can be used in risk management but has potential to be used in portfolio optimization.

Table of Contents

Contents

1	Introduction	1
2	Markov Regime Switching review	2
2.1	Regime switching GARCH model	4
2.2	Smoothed probability	9
2.3	Regime switching stationary conditions	12
2.4	Test with null as single regime against multiple regimes	15
2.5	Autocorrelation function of regime switching model .	17
3	Normal Tempered Stable distribution introduction	19
4	FIGARCH introduction	24
4.1	Estimation of FIGARCH	29
4.2	regime switching FIGARCH	31
5	Backtesting	35
6	Empirical results in high frequency	40
6.1	Goodness of fit test	41

6.2	Some VaR forecasts versus log return series	42
7	Conclusion	45
8	Appendix	47

List of Figures/Tables/Illustrations

List of Figures

1	smoothed probabilities, ShanghaiShenzhen 300 index daily, from 1/7/2002 to 4/8/2010	10
2	smoothed probabilities, ShanghaiShenzhen 300 index with 1 minute frequency, from 2015-05-28 to 2015-06-09	11
3	Illustration of stdNTS under different parameters . . .	23
4	box plot of estimation results for RS-FIGARCH . . .	32
5	box plot of estimation results for RS-FIGARCH . . .	33
6	box plot of estimation results for RS-FIGARCH . . .	33
7	box plot of estimation results for RS-FIGARCH . . .	34
8	Data is between 2008-09-02 and 2008-09-15, in the middle of Lehmann Brother bankruptcy event (at time index 350)	43

9	ACF of long memory process and short memory process	53
10	QQ plot of residuals versus Gaussian distribution	54
11	QQ plot of residuals versus Gaussian distribution	55
12	Standardized residuals extracted from RS-GARCH versus fitted stdNTS	56
13	Standardized residuals versus fitted stdNTS	57
14	QQ plots of residuals versus stdNTS, estimate SHSZ 1minute data, from 2015-5-28 to 2015-6-9	58
15	VaR forecasting results of RS-FIGARCH and FIGARCH	59
16	volatility forecasting results of GARCH, FIGARCH and RS-FIGARCH	60
17	99% VaR forecasting results of GARCH, FIGARCH, RS-FIGARCH, GARCH-NTS, FIGARCH-NTS, stdNTS-RS-GARCH	61
18	99% VaRs of different models versus realized returns	61
19	99.5% VaRs of different models versus realized returns	62
20	99.9% VaRs of different models versus realized returns	62
21	99% VaRs of different models versus realized returns	63

22	99.5% VaRs of different models versus realized returns	64
23	99.9% VaRs of different models versus realized returns	65
24	99% VaRs of different models versus realized returns	66
25	99.5% VaRs of different models versus realized returns	67
26	99.9% VaRs of different models versus realized returns	68

List of Tables

1	Non-rejection region for Kupiec test of different confidence levels, significance level of the test is 0.05 . . .	37
2	Non-rejection region for Kupiec test of different confidence levels, significance level of the test is 0.01 . . .	37
3	normality test of residuals extracted from RS-GARCH	47
4	normality test of residuals extracted from RS-FIGARCH	47
5	Goodness of fit test statistics	48
6	violation numbers for 99% VaR in different horizons, significance level is 0.01	48
7	Backtesting results (p-values) of SP 1min with 200 forecasting horizon, significance level is 0.05	49
8	violation numbers for 99% VaR in 100 forecasting horizon, significance level is 0.05	50

9	violation numbers for 99% VaR in 200 forecasting horizon, significance level is 0.05	50
10	violation numbers for 99% VaR in 400 forecasting horizon, significance level is 0.05	51
11	violation numbers for 99.5% VaR in 100 forecasting horizon, significance level is 0.01	51
12	violation numbers for 99.5% VaR in 200 forecasting horizon, significance level is 0.01	52
13	violation numbers for 99.5% VaR in 400 forecasting horizon, significance level is 0.01	52

List of Abbreviations

Quasi-likelihood ratio	QLR
Normal Tempered Stable	NTS
Data generating process	DGP
Autoregressive Moving Average	ARMA
Generalized AutoRegressive Conditional Heteroskedasticity	GARCH
Fractionally Integrated Autoregressive Moving Average	FARIMA
Fractionally Integrated Generalized AutoRegressive Conditional Heteroskedasticity	FIGARCH
Regime Switching GARCH	RS-GARCH
Autocorrelation function	ACF
Quasi maximum likelihood estimator	QMLE
Quasi modified likelihood ratio	QMLR
Broyden Fletcher Goldfarb Shanno algorithm	BFGS algorithm
Kolmogorov Smirnov test	KS test
Anderson-Darling test	AD test
ShanghaiShenzhen 300 Index	SHSZ 300 or SHSZ

Acknowledgements

First and foremost I would like to thank my doctoral advisor Prof. Dr. Zari Rachev. He was the one who took me into the area of quantitative finance in the first place and gave me the main impetus for my doctoral thesis. Within my four and half years within Quantitative Finance program he introduced to me many interesting topics and entrusted me challenging projects. Not only is he my academic guide but also a person hold in high regard.

Many thanks go also to Prof. Dr. Aaron Kim, who kindly agreed to be the co-supervisor of my doctoral thesis. He guided me in the process of implementing and showed me a lot of valuable insights of my results. I also want to express a lot of appreciations to my dissertation committee members, Petar Djuric, Xinyun Chen, Stoyan Stoyanov, Keli Xiao for their precious comments during my presentation. I would also like to thank all the colleagues at Stony Brook University, include but not limited to Yu Mu, Xiang Shi, Frank Cheng, Kinsley Yang, Jae-hyung Choi, Barret Shao, Xiao Yu Ke Zhang and Riyu Yu. This doctoral thesis is mainly dedicated to my parents and my wife.

Chapter 1

1 Introduction

Regime switching model has been widely used in modeling financial time series, especially volatility process, see (Hamilton, 1994), (Gray, 1996), (Klaassen, 2002), (Marcucci, 2005). In this paper, we introduce Markov regime switching Fractionally Integrated Generalized AutoRegressive Conditional Heteroskedasticity (RS-FIGARCH henthforce) model for volatility process, to capture both long memory effect and multiple regimes.

The remaining parts of this thesis are arranged as follows: Section 2 reviews literatures in regime switching models and fractional integrated GARCH (FIGARCH). stationary condition on regime switching model, especially regime switching GARCH and the moments. Section 5 shows empirical results. Section 6 presents VaR based backtesting.

2 Markov Regime Switching review

Markov regimes switching class of models are first introduced by (Hamilton, 1989). Regime switching models assume that a process switches among different regimes, and the switching behavior is driven by an unobserved first order Markov chain with finite state space $S = 1, \dots, k$ and transition matrix $P_{k \times k}$,

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{bmatrix} \quad (1)$$

where the entries of P represent the transition probabilities $p_{ij} = p(S_t = j | S_{t-1} = i)$, $i, j = 1, \dots, k$.

Matrix P should primarily satisfy both $0 < p_{ij} < 1$ and $\sum_{j=1}^k p_{ij} = 1$. And if any diagonal entry p_{ii} is close to 1, then we can say state i of the regime switching model is quite persistent, and if all the diagonal entries are close to 1, then every state is persistent. To mathematically describe the overall persistence level of a regime switching model, we can approach either from steady state distribution of the Markov Chain or from the second largest eigenvalue of the transition probability matrix. For the second approach, see (Francq et al., 2001), (Haas et al., 2004). The steady state distribution is defined as follows. Let $\pi_t = (\pi_{1t}, \dots, \pi_{kt})'$ denote the distribution of the Markov chain at time t . For $j = 1, \dots, k$,

$$\pi_{j,t+1} = p(S_{t+1} = j) = \sum_{i=1}^k p(S_t = i) p(S_{t+1} = j | S_t = i) = \sum_{i=1}^k \pi_{ij} p_{ij},$$

or

$$\pi_{t+1} = P \pi_t, \quad \pi_{t+\tau} = P^\tau \pi_t, \quad \tau \leq 1,$$

in matrix form where

P^τ is the transition matrix after τ -step. Moreover, under general conditions (\mathbf{P} is irreducible and aperiodic (Hamilton, 1994), there exists a stationary distribution $\boldsymbol{\pi} = (\pi_1, \dots, \pi_k)'$ independent of $\boldsymbol{\pi}_t$, such that

$$\boldsymbol{\pi} = P\boldsymbol{\pi} = \lim_{\tau \rightarrow \infty} \mathbf{P}^\tau \boldsymbol{\pi}_t.$$

If regimes are persistent, this will be reflected in rather large diagonal elements of \mathbf{P} , and the steady state probabilities are corresponding to the non-diagonal elements. For example, given a model with two regimes, $\delta := p_{11} + p_{22} - 1$ measures the persistence of the regimes, and the unconditional probabilities that process is in regime 1 and 2 are: $\pi_1 = (1 - p_{22}) / (2 - p_{11} - p_{22})$ and $\pi_2 = (1 - p_{11}) / (2 - p_{11} - p_{22})$, respectively. When $p_{11} + p_{22} = 1$, the regime switching process turns into mixture models with independent components, where at any given time, the transition probabilities no longer depend on previous state but only determined by deterministic weights identical to diagonal transition probabilities. In other word, $\mathbf{P} = \mathbf{P}_\infty = (\boldsymbol{\pi}, \dots, \boldsymbol{\pi})$, there is no persistence in $\{S_t\}$ and we have an i.i.d. mixture model with constant weights ($\pi_{jt} = \pi_j$ for all $t, j = 1, \dots, k$). Thus, the quantity of $p_{11} + p_{22}$ should be bounded away from 1 in practical use to prevent degenerate cases where regime transition is irrelevant to regime itself.

Another quantity of interest is the expected regime duration, which has the properties of geometric distribution. Once we are in regime j , we expect it to last for $\sum_{d=1}^{\infty} dp_{ij}^{d-1} (1 - p_{ij}) = (1 - p_{ij})^{-1}$. Detailed proof can be found in (Bauwens et al., 2012).

2.1 Regime switching GARCH model

In quantitative finance, modeling volatility process is always a primary task for option pricing, dynamic hedging, etc. When a quantitative analyst or time series researchers attempt to explain the changes in conditional variance, conditional heteroscedastic models are generally considered. The basic genre of models that has been extensively used is Autoregressive Conditional heteroscedastic (ARCH) model of Engle (1982), and it is extended to Generalized Autoregressive Conditional Heteroscedastic (GARCH) model by (Bollerslev, 1986a), Integrated GARCH (IGARCH) model by Engle and Bollerslev (1986) and Fractionally Integrated GARCH (FIGARCH) model by (Baillie et al., 1996), etc. Under the framework of ARCH, large shocks tend to follow large shocks and the small shocks tend to follow small shocks, which is a stylized fact known as volatility clustering. Although the ARCH model is simple, it often requires many parameters to describe the volatility process of an asset return. As a result, several alternative models were introduced to accommodate the shortages of ARCH models. (Bollerslev, 1986a) proposes a very useful extension of ARCH model, known as GARCH model. GARCH(p,q) process is described as follows,

$$\epsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \quad (2)$$

To incorporate regime switching with GARCH, we are going to briefly review different specifications of RS-GARCH models in literature. First of all, to model the error term of time series, ϵ_t , we have

$$\epsilon_t = \sigma_{S_t, t} \eta_t, \quad (3)$$

where η_t follows standard Gaussian distribution, and the regime-specific conditional variances for RS-GARCH are

$$\sigma_{jt}^2 = \omega_j + \alpha_j \epsilon_{t-1}^2 + \beta_j \sigma_{j,t-1}^2, \quad \omega_j > 0, \quad \alpha_j, \beta_j \geq 0, \quad j = 1, \dots, k. \quad (4)$$

We can derive the full sample log-likelihood function as follows. We start from the quantity

$$f(\epsilon_t | I_{t-1}) = \sum_{i=1}^k f(\epsilon_t, S_t = i | I_{t-1}) = \sum_{i=1}^k f(\epsilon_t | S_t = i, I_{t-1}) p(S_t = i | I_{t-1}) \quad (5)$$

where I_t is the information up to time t , and k stands for the number of regimes of the process. Conditional densities $f(\epsilon_t | S_t = i, I_{t-1})$ are summed up over all possible regimes, weighted by the ex-ante probabilities $p(S_t = i | I_{t-1}), i = 1, \dots, k$.

$$p(S_t = i | I_{t-1}) = \sum_{j=1}^k p(S_t = i | S_{t-1} = j, I_{t-1}) p(S_{t-1} = j | I_{t-1}),$$

for $i, j = 1, \dots, k$. Since S_t follows first-order Markov property,

$$p(S_t = i | S_{t-1} = j, I_{t-1}) = p(S_t = i | S_{t-1} = j) = p_{ji}.$$

Because the error terms follow Gaussian distribution in Equation 3, then the conditional density function is given by

$$f(\epsilon_t, S_t = i | I_{t-1}) = \frac{1}{\sqrt{2\pi h_{i,t}}} \exp \left\{ \frac{-\epsilon_{i,t}^2}{2h_{i,t}} \right\} \quad (6)$$

The quantity $p(S_{t-1} = j | I_{t-1})$ of Equation 5 is of much use not only in estimation of RS-GARCH but also forecasting steps, it can be derived following Hamilton filter (Hamilton, 1994):

$$p(S_{t-1} = j | I_{t-1}) = p(S_{t-1} | \epsilon_{t-1}, I_{t-2}) = \frac{p(\epsilon_{t-1} | S_{t-1}, I_{t-2}) p(S_{t-1} | I_{t-2})}{p(\epsilon_{t-1} | I_{t-2})}$$

$$= \frac{p(\varepsilon_{t-1}|S_{t-1}, I_{t-2}) \sum_{S_{t-2}=i}^k p(S_{t-2}|I_{t-2}) p(S_{t-1}|S_{t-2})}{p(\varepsilon_{t-1}|I_{t-2})} \quad (7)$$

Note that in the Equation 7, the term matters is $p(S_{t-2}|I_{t-2})$. The other terms are either conditional densities or deterministic part based on the constant transition probability assumption. Starting from unconditional probabilities estimated based on pre-sample, the filtered probability $p(S_{t-2}|I_{t-2})$ can be calculated with first-order recursive steps.

Thus, we can build up the likelihood function as

$$L = \sum_{t=1}^T f(\varepsilon_t|S_t) p(S_t|I_{t-1}) \quad (8)$$

In order to calculate Quasi maximum likelihood estimator (QMLE henceforth) of Equation 8, we implement with constrained optimization of Equation 8 primarily through 'fmincon' function in optimization toolbox of MATLAB. The interior point algorithm is used as hill-climbing algorithm, and Broyden Fletcher Goldfarb Shanno (BFGS) algorithm is chosen to calculate Hessian updates in each step. To ensure the convergence of the optimization steps, we set maximum function evaluations as 4000, and if it still does not converge, we switch our search algorithm as 'steepest' and optimize the likelihood function again. Sequential estimation is used in order to obtain robust estimation results, especially for regime switching model that has many parameters. Although, global QMLE of Likelihood function 8 is also favored in this thesis, since we can perform a grid random selection for starting points and choose the maximum value among each optimization evaluation. Optimization toolbox and Statistics toolbox are required to implement the

model and backtesting results.

Cai (1994) and Hamilton and Susmel (1994) first incorporate Regime switching with ARCH, and they also claim that the maximum likelihood (ML) estimation of this specification is intractable because of the path dependence of conditional variances in equation (4), and as a result they employed RS-ARCH instead of RS-GARCH. The path dependence the problem comes from the fact that S_{t-1} and subsequent $\sigma_{S_{t-1},t-1}^2$ in Equation 4 is not observable, and depend on the previous regime or state S_{t-2} , and thus depend on the entire regime history. Thus, the evaluation of the likelihood for a sample of T observations requires the integration over all k^T possible regime paths, implying its calculation infeasibility. Nonetheless, there are some approaches since (Hamilton, 1994). (Bauwens et al., 2012) brings up a Gibbs sampling algorithm with Bayesian estimation to calculate sample state variables, and (Henneke et al., 2011) show how an RS-GARCH model built on model 3 can be estimated via Bayesian Markov chain Monte Carlo (MCMC) methods. However, both implementation options rely on ultra fast machine to deal with large-scale resampling method, and for now they are only for scientific purpose but not practical approachable.

There are another two approaches used more frequently in the literature. The first approach used to circumvent path dependence problem is given by (Gray, 1996). He replaces $\sigma_{S_{t-1},t-1}^2$ in equation 4 with the conditional variance of ϵ_{t-1} given the observable information up to time $t - 2$. With this information, the conditional distribution of ϵ_{t-1} is thus a k-component mixture with conditional variance:

$$h_{t-1} := Var_{t-2}(\epsilon_{t-1}) = \sum_{j=1}^k p_{t-2}(S_{t-1} = j) \sigma_{j,t-1}^2, \quad (9)$$

where $p_{t-2}(S_{t-1} = j)$ represents the conditional regime probabilities $prob(S_{t-1} =$

$j|t - 2)$, $j = 1, \dots, k$, calculated under specified residual distribution for regime-specific process. Quantity 9 is then used instead of $\sigma_{S_{t-1}, t-1}^2$ in the regime-specific GARCH equations. Similar solutions have also been proposed by Dueker (1997) and (Klaassen, 2002). Ane and Ureche-Rangau (2006) extend the model to allow for asymmetric volatility dynamics, and Chen and Hung (2010) consider its use in option pricing. Even though this approach reduced a lot of complex computation issues, it is a tweaked version of GARCH, which introduces bias into the model and also has poor forecast results. In (Marcucci, 2005), he compares volatility forecasting of (Klaassen, 2002) model with various models such as GARCH with student-T innovation, GARCH-Gaussian, etc. He applies both statistical loss functions and VaR-based backtesting, and it turns out (Klaassen, 2002) does not have favorable results.

A second approach to RS-GARCH models, proposed by (Haas et al., 2004), can be viewed as a direct generalization of the single-regime GARCH model. In this specification, each regime-specific conditional variance depends only on its own lag. To see the analogy between Equations 4 and 2, recall that a GARCH(1,1) process $\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$ with $\beta < 1$ can be written as ARCH(∞),

$$\sigma_t^2 = \frac{\omega}{1 - \beta} + \alpha \sum_{i=1}^{\infty} \beta^{i-1} \epsilon_{t-i}^2, \quad (10)$$

where α can be described as the reaction parameter reflecting the magnitude of a unit shock immediate impact on the next period variance and β measures the memory in the volatility process. The total impact of a unit shock on future variances is $\frac{\alpha}{1-\beta}$. Adopting equation 10 implies an ARCH(∞) representation for each regime, so that this specification allows a clear-cut interpretation of the regime-specific volatility processes. That is, the parameters α_j and β_j measure the imme-

mediate responsiveness to a shock and the memory of the volatility process in regime j , respectively. To rule out explosive volatility processes, an intuitive requirement is therefore $\alpha + \beta < 1$, but in Regime Switching-GARCH (RS-GARCH henceforth), both weakly stationary and covariance stationary conditions required in Equations 4 are broader necessary according to (Bauwens et al., 2007). Bayesian estimation of this specification has been developed by (Bauwens et al., 2007) and (Ardia, 2009), with the latter extending the model to allow for asymmetric volatility dynamics as well as Student- t innovations. When incorporated with GARCH structure for the component-specific volatility dynamics and Gaussian innovations, this process reduces to the mixture normal GARCH process ((Haas et al., 2004)). (Gray, 1996) and (Haas et al., 2004) approaches differ with the lagged variance terms in which the regime-specific GARCH recursions is specified.

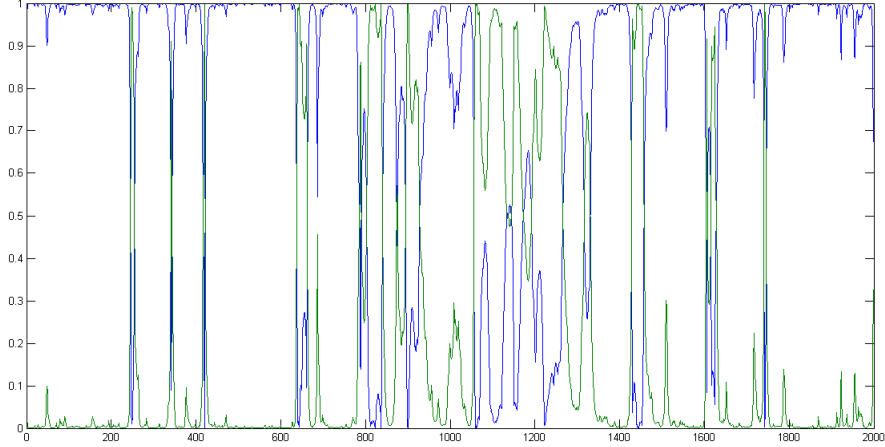
2.2 Smoothed probability

According to (Hamilton, 1989), and being applied in (Gray, 1996), (Klaassen, 2002), (Marcucci, 2005), (Piplack et al., 2009), to deliver the inference about regimes at each time of the full sample, the ex post probability $p(S_t|I_T)$ is generated by:

$$p(S_t|\Phi_T) = p(S_t|\varepsilon_T, \Phi_{T-1}) = \frac{p(\varepsilon_T|S_t, \Phi_{T-1})p(S_t|\Phi_{T-1})}{\sum_{S_t=1}^K p(\varepsilon_T|S_t, \Phi_{T-1})p(S_t|\Phi_{T-1})}. \quad (11)$$

With smoothed regime probabilities we can do ex post analysis and (Hamilton, 1989) proposes a selection criteria of what is the state at given time within the sample. He claims that whenever $p(S_t = 1|\Phi_T) < 0.5$ it means at time t , the state is considered as 2, and vice versa. However, (Xie, 2009) argues sometimes

Figure 1: smoothed probabilities, ShanghaiShenzhen 300 index daily, from 1/7/2002 to 4/8/2010



there is significant 'intermediate' regime in which we can hardly discriminate among states from smoothed probabilities, and he also suggests to choose state 1 when $p(S_t = 1|\Phi_T) > 0.8$ and state 2 when $p(S_t = 1|\Phi_T) < 0.2$ and otherwise an intermediate regime. To illustrate this idea, we can compare the smoothed regime graph estimated from a sample in daily frequency with that estimated from a sample in high frequency data. We show smoothed regimes in Figure 1 and Figure 2. We can see in Figure 1, none of two regimes are not quite dominant compared to the smoothed regime depicted in blue line from Figure 2.

To solve Equation eq:smoothprob mathematically, we need to first consider when $t = T$. Then term $p(S_t|\Phi_{T-1})$ of the above equation turns to $p(S_t|\Phi_{t-1})$, in the meantime, $p(\epsilon_t|S_t, \Phi_{t-1})$ is simply the result from estimation step of Equation 7, and the calculation of Equation 11 can thus be calculated in a first order forward

smoothed probabilities

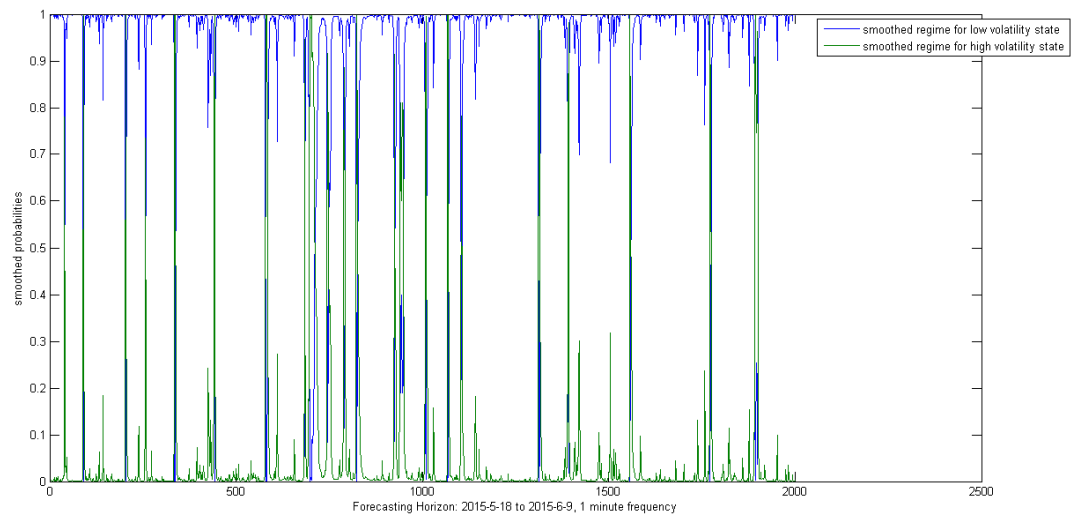


Figure 2: smoothed probabilities, ShanghaiShenzhen 300 index with 1 minute frequency, from 2015-05-28 to 2015-06-09

manner for times $t + 1, t + 2, \dots, T$. When $t < T$, we have more diligent steps to fulfill. To solve the ingredient $p(\varepsilon_T | S_t, \Phi_{T-1})$

$$p(S_T | \varepsilon_t, \Phi_{T-1}) = \sum_{S_{T-1}=1}^K p(S_T | S_{T-1}, \Phi_{T-1}) p(S_{T-1} | S_t, \Phi_{T-1}),$$

the second ingredient of the right hand side of the above equation can be simply derived by the transition probabilities matrix directly, while the first ingredient can be calculated as:

$$\begin{aligned} p(S_T | S_t, \Phi_{T-1}) &= \sum_{S_{T-1}=1}^K p(S_T | S_{T-1}, \Phi_{T-1}) p(S_{T-1} | S_t, \Phi_{T-1}) \\ &= \sum_{S_{T-1}=1}^K p_{ij} p(S_{T-1} | S_t, \Phi_{T-1}), \end{aligned}$$

the term $p(S_{T-1} | S_t, \Phi_{T-1})$ is given by:

$$\begin{aligned} p(S_{T-1} | S_t, \Phi_{T-1}) &= p(S_{T-1} | S_t, \varepsilon_{t-1}, \Phi_{T-2}) \\ &= \frac{p(\varepsilon_{T-1} | S_{T-1}, \Phi_{T-2}) p(S_{T-1} | S_t, \Phi_{T-2})}{\sum_{S_{T-1}=1}^K p(\varepsilon_{T-1} | S_{T-1}, \Phi_{T-2}) p(S_{T-1} | S_t, \Phi_{T-2})}, \end{aligned}$$

and thus follows the iteration manner.

Using the notations from (Hamilton, 1989), with filtered regime probability $P[s_t = i | \varepsilon_t, \dots, \varepsilon_0], i = 1, 2, \dots, k$. Then it can be calculated by following the algorithm provided by Hamilton(1989) iteratively as well.

2.3 Regime switching stationary conditions

To see the mild regularity conditions of geometrically ergodicity for Markov Chain, see (Meyn and Tweedie, 1993). Following assumptions of RS-GARCH come from (Bauwens et al., 2010):

- 1 The error term ϵ_t is i.i.d. with a continuous density on the whole real line which is centered on zero. Furthermore, $E(|\epsilon_t^2|^\delta) < \infty$ for some $\sigma > 0$.
- 2 $\alpha_i > 0, \beta_j > 0$, and the Markov chain is homogeneous.
- 3

$$\sum_{i=1}^n \pi_i E[\log(\alpha_i \epsilon_t^2 + \beta_i)] < 0$$

The first assumption is satisfied for a wide range of distributions for the error term, e.g. the normal and the Student t distributions. For $\sigma = 1$, we set the variance to unity and if $\sigma < 1$, the parameters of the conditional scaling factor of the data are estimated. The second assumption is slightly stronger than the non-negativity conditions of (Bollerslev, 1986b) for the GARCH(1,1) model. Under this assumption all the regimes are accessible and the discrete Markov chain is ergodic. The third assumption implies that at least one of the regimes is stable. Without loss of generality, we assume that in the first regime ($S_t = 1$) the process is strictly stationary, so that $E[\log(\alpha_1 \epsilon_t^2 + \beta_1)] < 0$. In order to discover the higher order moments of RS-GARCH model, we can define the n by n matrix as below, following the work of (Bauwens et al., 2010),

$$\Omega = \begin{pmatrix} E(\alpha_1 u_t^2 + \beta_1)^m p_{11} & \cdots & E(\alpha_n u_t^2 + \beta_n)^m p_{k1} \\ \vdots & & \vdots \\ E(\alpha_1 u_t^2 + \beta_1)^m p_{1k} & \cdots & E(\alpha_n u_t^2 + \beta_n)^m p_{nn} \end{pmatrix}$$

Let $\rho(\bullet)$ denote the spectral radius of a matrix, in another word, its largest eigenvalue in modulus. With assumptions as below:

$$E(|u_t^2|^m) < \infty \text{ for some integer } m \leq 1.$$

$$\rho(\Omega) < 1.$$

(Francq et al., 2001) proves that the maximum likelihood procedure for RS-ARCH process is consistent. (Francq et al., 2001) also discusses necessary and sufficient stationarity conditions for strictly stationary solution and brings up misspecification situation using simulation experiments.

(Francq et al., 2001) claims fit a standard GARCH model to simulated data sets with at least two very different regimes, typically estimates close to the nonstationary region might be obtained (i.e. $\sum_i \alpha_i + \beta_i \simeq 1$). Such parameter estimates have been observed for many financial time series by earlier researchers. Without the linear constraint $\sum_i \alpha_i + \beta_i \leq 1$, the parameter estimates are generally outside the stationary region. This leads us to think that GARCH parameter estimates close to the nonstationary region could indicate that an alternative model, such as RS-GARCH and structural break process. Consequently it should be carefully considered if GARCH parameters close to non-stationary region. Note that due to the constraints out of stationary conditions, practically parameter estimates of turbulent regime are quite often insignificant or even 0. Another interesting statement is that, some estimated GARCH regimes are exactly the same, which can be a implication that an overparameterization occurs. although, which does not affect the forecasting performance of the model. More details can be found in (Francq and Roussignol, 1997) and (Francq et al., 2001).

2.4 Test with null as single regime against multiple regimes

The test against multiple regimes under certain model (or without specified model). Even though there has been a lot of research covering this topic, it still stands out as a quite difficult problem out of regime switching literature.

The major difficulty of regime test lies in the fact that the likelihood function of regime switching model can be substantially flat, the nuisance parameters arise when the true parameters in the dynamics of single regime is not identified in null, see (Hansen, 1996) for more details. Second, the information matrix is also singular under the null (Carrasco et al., 2014), as the transition probabilities p and q is unobserved. As a consequence, regular model specification test like standard likelihood ratio test is not applicable under regime switching test. To solve the nuisance parameters problem, (Hansen, 1996) presents a supere LR test statistic combined with grid search technique, but the steps can be quite cumbersome. (Cho and White, 2007) provides quasi-likelihood ratio(QLR) test. Following penalized quasi likelihood function from Chen et al.(2001), (Kasahara et al., 2014) propose a modified quasi-likelihood function by introducing a penalty term to the original QLR statistics from (Cho and White, 2007). The asymptotic distribution of the quasi-likelihood ratio(QLR) test is Gaussian, proved by (Cho and White, 2007).

To statistically describe the regime test, we start from a broad regime switching model specification:

$$\epsilon_t | F_{t-1} \sim \begin{cases} F(|\bar{\epsilon}_{t-1}|; \theta_0^*, \theta_1^*), & \text{if } S_t = 1 \\ F(|\bar{\epsilon}_{t-1}|; \theta_0^*, \theta_2^*), & \text{if } S_t = 2 \end{cases}$$

where S_t follows first order Markov Chain, same as the assumptions in section 2.

And if we denote

$$\pi^* = \text{Pr } ob(S_t = 1)$$

as the stationary or steady distribution of S_t , then the quasi-likelihood function can be written as:

$$L_n^*(\pi, \theta) = \sum_{t=1}^n l_t(\pi, \theta), l_t(\pi, \theta) = \log(\pi f_t(\theta^1) + (1 - \pi) f_t(\theta^2)), \quad (12)$$

with the conditional density term $f_t(\theta^i)$, as

$$f_t(\theta^i) = f(X_t | \overrightarrow{X_{t-1}}; \overrightarrow{\theta^i}), \quad \overrightarrow{\theta^i} = (\theta_0, \theta_i), \quad i = 1, 2.$$

Take $\overrightarrow{\theta_i^*}$, $i = 1, 2$ as true value of $\overrightarrow{\theta^i}$ under null of a single regime, following the notations from Cho and White(2007) and Kasahara et al.(2008), the null hypothesis can be described as follows:

$$H_0 : \{\pi^* = 0 \cap \overrightarrow{\theta^1} = \overrightarrow{\theta^*}\} \cup \{\overrightarrow{\theta^1} = \overrightarrow{\theta^2} = \overrightarrow{\theta^*}\} \cup \{\pi^* = 1 \cap \overrightarrow{\theta^2} = \overrightarrow{\theta^*}\}.$$

It is obvious any of the three cases above violates assumptions of null as a single regime. In order to derive the asymptotic distribution of Equation 12, we can consider the first two partitions together and the last one individually (it is symmetrical the other way around), and the limiting distribution of both can be found in (Cho and White, 2007). The QLR statistic is model-dependent and one has to calculate critical values for each alternative, which is quite burdensome. Even though Cho and White provide lower bound on the tail distribution of QLR, it is still not quite practical. citeKasahara2013 propose the new test statistic based on penalized likelihood function as below:

$$L_n(\pi, \theta) = L_n^*(\pi, \theta) + nc_n \log(\pi(1 - \pi)) = \sum_{t=1}^n l_t(\pi, \theta) + nc_n \log(\pi(1 - \pi)) \quad (13)$$

where c_n is a non-random sequence that goes to 0 as $n \rightarrow \infty$. And since $\log(\pi(1-\pi)) = -\infty$ while $\pi = 0, 1$, the penalty term always bounds away from both 0 and 1.

In addition, (Hu and Shin, 2008) and (Carrasco et al., 2014) give an Information matrix approach in which they prove the test statistic is asymptotically optimal, and meanwhile the testing procedure does not require a full specification of the coefficients in alternatives but only covariance structure.

2.5 Autocorrelation function of regime switching model

Several criteria have been defined to detect the existence of long memory behavior inside real data. Here we adopt asymptotic decay of the autocorrelation function to define the long memory behavior of a process. Furthermore, this ACF definition is consistent with what we are going to take into consideration in Chapter 2.

The moments and autocovariance function of regime switching models are discussed in (Timmermann, 2000) and (Francq et al., 2001). (Guegan and Stéphanie, 2005) shows how to generate a long memory process generated only by a regime switching model. From (Guegan and Stéphanie, 2005), we can see how regime switching effect generates slower decay in autocorrelation function. Let X_t defines a mean-switching model:

$$X_t = \begin{cases} u_1 + \varepsilon_t & S_t = 1 \\ u_2 + \varepsilon_t & S_t = 2 \end{cases} \quad (14)$$

The process X_t is composed of a switching deterministic part plus a white noise process. The process switching with respect to the unobserved ergodic

Markov chain characterized as Equation 1 with $k = 2$, which is independent of the white noise process.

According to (Guegan and Stéphanie, 2005), the autocorrelation function of Equation 14 is defined by:

$$\Gamma(h) = C_{u_i, p_{ii}} \delta^h, i = 1, 2 \quad (15)$$

where $\delta := p_{11} + p_{22} - 1$, and

$$C_{u_i, p_{ii}} = \frac{(u_1 - u_2)^2 (1 - p_{11})(1 - p_{22})}{(1 - \delta)^2 [\pi_1 u_1^2 + \pi_2 u_2^2 + 1 - (\pi_1 u_1 + \pi_2 u_2)^2]},$$

$\pi_1 = (1 - p_{22}) / (2 - p_{11} - p_{22})$ and $\pi_2 = (1 - p_{11}) / (2 - p_{11} - p_{22})$,

The autocorrelation function Eq. 15 convergence rate depends on $\delta \in [-1, 1]$. When δ is close to 1, both regime has high persistence level, the decay of autocorrelation function (Equation 15) would be rather slow. However, as δ approaches to 0, the decay of Equation 15 gets quicker.

3 Normal Tempered Stable distribution introduction

As stylized facts in financial time series suggest, since the distribution of innovations of financial time series often appears fat-tailed behavior, the general stable class of distribution is introduced. Stable distributions are a class of distributions that is able to describe both skewness and fat tails. The stable family of distributions was first characterized by Levy (1924) in his study of normalized sums of i.i.d. terms. The stable distributions play an important role not only because they exhibit closure properties under convolution, but also that a stable distribution can be the limiting distribution of sums of i.i.d. variables (DuMouchel, 1973). The general stable distribution is described by four parameters: an index of tailness or stability $\alpha \in (0, 2)$, a skewness parameter $\beta \in [-1, 1]$, a scale parameter γ and a location parameter σ . There is no general closed form for the class of stable distributions. Here we provide the definition of stable distribution given by Levy (1924). If $X \sim S_1(\alpha, \beta, \gamma, \delta_1)$, then the characteristic function of X is as follows:

$$E[\exp(iuX)] = \left\{ \begin{array}{ll} \exp(-\gamma^\alpha |u|^\alpha [1 - i\beta \operatorname{sgn}(u) (\tan \frac{\pi\alpha}{2})] + i\delta_1 u), & \alpha \neq 1 \\ \exp(-\gamma |u| [1 + i\beta \frac{2}{\pi} \operatorname{sign}(u) \ln |u|] + i\delta_1 u), & \alpha = 1 \end{array} \right\}, \quad (16)$$

where $\sigma_1 \in R$ and $\operatorname{sign}(u)$ stands for sign function of u . When $\alpha = 2$, the distribution is Gaussian distribution, but for $\alpha < 2$ the variance of stable distribution does not exist, and mean of stable distribution does not converge for $\alpha < 1$. We don't have a parameter represents standard deviation for stable distribution, either. In addition, this representation of stable distribution does not continue at $\alpha = 1$.

An alternative parametrization from (Zolotarev, 1986) overcomes discontinu-

ity at $\alpha = 1$, if $X \sim S_2(\alpha, \beta, \gamma, \delta_1)$:

$$E[\exp(iuX)] = \left\{ \begin{array}{ll} \exp(-\gamma^\alpha |u|^\alpha [1 + i\beta \operatorname{sgn}(u) (\tan \frac{\pi\alpha}{2}) (|\gamma u|^{1-\alpha} - 1)] + i\delta_2 u), & \alpha \neq 1 \\ \exp(-\gamma |u| [1 + i\beta \frac{2}{\pi} \operatorname{sgn}(u) \ln |u|] + i\delta_2 u), & \alpha = 1 \end{array} \right\} \quad (17)$$

The new formulation of characteristic function has less intuitive meaning yet much more useful for statistical purposes. The substitution of δ_2 is:

$$\delta_2 = \left\{ \begin{array}{ll} \delta_1 + \beta \gamma \tan \frac{\pi\alpha}{2} \text{ if } \alpha \neq 1 \\ \delta_1 + \beta \frac{2}{\pi} \gamma \ln \gamma \text{ if } \alpha = 1 \end{array} \right\} \quad (18)$$

Beside the stable distributions which have infinite moments of all orders, (Koponen, 1995) introduced the idea of tempered Stable distribution applying an exponentially decreasing term times the original density function on each half of real axis. Because of the exponential tempering, the tempered stable distributions are bounded to have finite moments. Furthermore, a wide range of different types of Tempered Stable distributions and generalize the tempered class of distributions, see more about the class of tempered stable distributions in (Kim et al., 2008), (Kim et al., 2009) and (Rachev et al., 2011).

The Normal Tempered Stable (NTS) distribution is defined with parameters $(\alpha, \lambda, \gamma, \sigma^2, \mu)$, where $\alpha \in (0, 2)$, $\lambda, \sigma^2 > 0$ and $\gamma, \mu \in R$.

The subordination scheme can be used to simplify the representation of this distribution by:

$$X_t = \mu t + \beta(T_t - t) + \gamma B_{T_t} \quad (19)$$

where B_{T_t} is standard Brownian motion, and the characteristic function of X_t is given by:

$$\phi_{X_t}(u) = e^{iu(\mu-\beta)t} \phi_{T_t}(\psi_{N(\beta, \gamma^2)})$$

$$= \exp(iu(\mu - \beta)t - \frac{2\theta^{1-\frac{\alpha}{2}}}{\alpha}((\theta - i\beta u + \frac{\gamma^2 u^2}{2})^{\frac{\alpha}{2}} - \theta^{\frac{\alpha}{2}})t).$$

Thus we can define a univariate NTS random variable: X_1 follows a Normal Tempered Stable distribution if and only if X_t is a NTS process with parameters $(\alpha, \lambda, \gamma, \sigma^2, \mu)$, where $\alpha \in (0, 2)$, $\theta, \gamma > 0$ and $\beta, \mu \in R$.

An equivalent representation from (Rachev et al., 2011):

$$\begin{aligned} \phi_{NTS}(u; \tilde{\alpha}, C, \lambda, b, m) = \\ \exp(ium - iu2^{-\frac{\tilde{\alpha}-1}{2}}\sqrt{\pi}C\Gamma(1 - \frac{\tilde{\alpha}}{2})b(\lambda^2 - b^2)^{\frac{\tilde{\alpha}}{2}-1} \dots \\ \dots + 2^{-\frac{\tilde{\alpha}+1}{2}}\sqrt{\pi}C\Gamma(-\frac{\tilde{\alpha}}{2})((\lambda^2 - (b + iu)^2)^{\frac{\tilde{\alpha}}{2}} - (\lambda^2 - b^2)^{\frac{\tilde{\alpha}}{2}})), \end{aligned}$$

with $\tilde{\alpha} \in (0, 2)$, $C, \lambda > 0$, $|b| < \lambda$, and $m \in R$. This representation is equivalent to Equation 19

The representation accommodates the equalities as below:

$$c\tilde{\alpha} = \alpha\lambda = \sqrt{\frac{2\theta}{\gamma^2} + b^2} \quad (20)$$

$$b = \frac{\beta}{\gamma^2} \quad (21)$$

$$C = \frac{\sqrt{2}\gamma^\alpha}{\sqrt{\pi}\Gamma(1 - \frac{\alpha}{2})\theta^{\frac{\alpha}{2}-1}} \quad (22)$$

$$m = \mu \quad (23)$$

We apply (Rachev et al., 2011) representation in this thesis: given two independent random variables $X \sim NTS(\tilde{\alpha}, \lambda, b, C_1, m_1)$ and $Y \sim NTS(\tilde{\alpha}, \lambda, b, C_2, m_2)$, then their sum will follow the NTS distribution with the following parameters:

$$X + Y \sim NTS(\tilde{\alpha}, \lambda, b, C_1 + C_2, m_1 + m_2). \quad (24)$$

We then obtain the standardized NTS distribution after a few parameter manipulation steps in order to have zero mean and unit variance by setting $m = 0$, and via linear transformation:

$$C = 2^{1-\frac{\tilde{\alpha}}{2}}(\sqrt{\pi}C\Gamma(-\frac{\tilde{\alpha}}{2})\tilde{\alpha}(\lambda^2 - b^2)^{\frac{\tilde{\alpha}}{2}-2}(\tilde{\alpha}b^2 - \lambda^2 - b^2))^{-1}.$$

Thus, the standard NTS can be defined as:

$$stdNTS(\alpha, \theta, \beta) \equiv NTS(\alpha, \theta, \beta, 1 - \beta^2(\frac{2-\alpha}{2\theta}), 0), \quad (25)$$

with characteristic function:

$$\phi_X(u) = \exp(-\beta ui - \frac{2\theta^{1-\frac{\alpha}{2}}}{\alpha}((\theta - i\beta u + (1 - \beta^2(\frac{2-\alpha}{2\theta}))\frac{u^2}{2})^{\frac{\alpha}{2}} - \theta^{\frac{\alpha}{2}})) \quad (26)$$

Within standard NTS distribution, α stands for tail index as the same in original stable distributions, θ captures scale, while β is the skewness parameter. Figure 3 shows illustratively how the three parameters would character the stdNTS distribution.

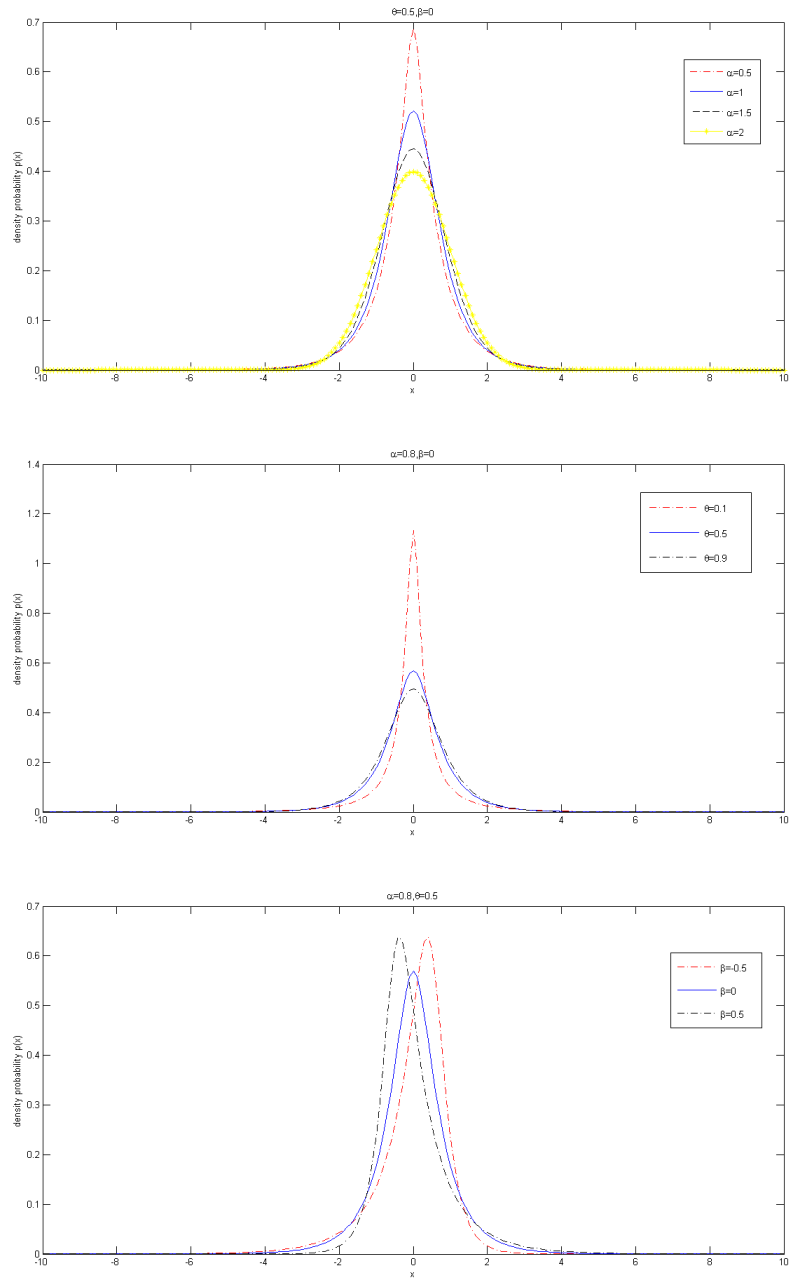


Figure 3: Illustration of stdNTS under different parameters

Chapter 2

4 FIGARCH introduction

If the AR polynomial of the GARCH representation has a unit root, then we have an Integrated GARCH model (IGARCH), which was first introduced by (Engle and Bollerslev, 1986). A key feature of IGARCH model is that the impact of the past squared shock is persistent and the pricing of risky securities, including long-term options and future contracts, may show extreme dependence on the initial conditions. Several studies report the presence of apparent long-memory in the autocorrelations of squared or absolute returns of various financial assets. Motivated by these observations, (Baillie et al., 1996) introduced the Fractionally Integrated Generalized Autoregressive Conditional Heteroscedastic (FIGARCH) process. The primary purpose of introducing FIGARCH model was to develop a more flexible class of processes for the conditional variance, that are capable of explaining and representing the observed temporal dependencies in financial market volatility. In particular, the FIGARCH model allows only a slow hyperbolic rate of decay for the lagged squared or absolute innovations in the autocorrelation function, while weakly stationary process like EGARCH has short memory and its autocorrelation function (ACF) is geometrically bounded. This model can accommodate the time dependence of the variance and a leptokurtic unconditional distribution for the returns with a long memory behaviour for the conditional variances.

Let ϵ_t denote a real-valued discrete-time stochastic process, and \mathcal{F}_t be the information set of all information up to time t , i.e., $\mathcal{F}_t = \sigma\{\dots, \epsilon_{t-2}, \epsilon_{t-1}, \epsilon_t\}$,

where ϵ_t follows ARCH(q) process, when

$$\epsilon_t = h_t^{1/2} z_t,$$

with

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2, \quad z_t \sim N(0, 1)$$

and $\alpha_i \geq 0, i = 1, \dots, q$. The conditional variance can be generally expressed as

$$h_t = h(\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-q}, \vec{\alpha}),$$

where $h(\cdot)$ is a nonnegative function of its arguments and $\vec{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_q)'$

is a vector of unknown parameters. An alternative expression of ARCH model is shown below:

$$E(\epsilon_t | \Phi_{t-1}) = 0 \quad \text{and} \quad \text{var}(\epsilon_t | \Phi_{t-1}) = h_t$$

According to (Caporin, 2002), and GARCH process cannot fully capture long memory effect even with particular characteristics such as asymmetric behavior, regime switching effect. Here we refer long memory to slow decaying of autocorrelation function (abbrev. ACF) of a time series: in long memory process, the ACF is hyperbolically decaying, while in regular GARCH model it is exponentially decaying. The difference is illustrated as below:

Suppose dynamic process x_t has spectral density $f(\lambda)$ at frequency λ . Then the process has long memory can be formally constructed if:

$$f(\lambda) \simeq C \lambda^{-2d},$$

as $\lambda \rightarrow 0+$ with $d \in (-0.5, 0.5)$ and $G \in (0, \infty)$, and C is a positive constant term. Granger and Joyeux(1980) and Hosking (1981) first introduce the

FARIMA(p,d,q) processes as:

$$A(L)(1 - L)^d x_t = B(L)\epsilon_t, \quad (27)$$

where $A(L)$ and $B(L)$ are lag operators such that

$$A(L) = 1 - a_1L - a_2L^2 - \dots - a_pL^p, \quad B(L) = 1 + b_1L + b_2L^2 - \dots + b_qL^q,$$

and ϵ_t here is a white noise process.

For $-0.5 < d < 0$, x_t is characterized as a stationary short memory series and addressed as anti-persistent. For $0.5 < d < 1$, the series is nonstationary long memory but mean reverting. For $d = 0$, we have an $I(0)$ series and when $d = 1$, we have an $I(1)$ series. To familiarize the transformation steps applied in FIGARCH, the GARCH(p, q) process can be expressed similar as FARIMA(p,d,q) in Equation 27:

$$\begin{aligned} h_t &= \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j} \\ &= \alpha_0 + \alpha(L)\epsilon_t^2 + \beta(L)h_t, \end{aligned}$$

where $p > 0, q > 0, \alpha_i \geq 0, i = 1, \dots, q, \beta_j \geq 0, j = 1, \dots, p$, and $\alpha(L) = \alpha_1L + \alpha_2L^2 + \dots + \alpha_qL^q, \beta(L) = \beta_1L + \beta_2L^2 + \dots + \beta_pL^p$, and in order to have a stationary GARCH process, $\sum \alpha_i + \sum \beta_j < 1$. For $p = 0$, the process reduces to an ARCH(q) and for $p = q = 0, \epsilon_t$ turns to a white noise process.

(Baillie et al., 1996) proposes FIGARCH first using FARIMA(p,d,q) representation 27, yet (Chung, 1999) argues that the method of parameterization of the FIGARCH model of (Baillie et al., 1996) may have a specification problem, and points out some drawbacks in the (Baillie et al., 1996) model. A structural problem in the (Baillie et al., 1996) specification with respect to the FARIMA

framework of the conditional mean equation, thereby leading to difficult interpretations of the estimated parameters. Indeed, the fractional differencing operator applies to the constant term in the mean equation of representation 27, while it is not satisfied in the variance equation of FIGARCH. Therefore, (Chung, 1999) redefines the FIGARCH model.

Denote $v_t = \epsilon_t^2 - h_t$, then the fractionally integrated GARCH or FIGARCH class of model can be obtained by using similar technique, with $(1 - L)$ replaced by $(1 - L)^d$, where d is fractional integration index with $0 < d < 1$. Thus, the FIGARCH class of models can be obtained by considering

$$[1 - \alpha(L) - \beta(L)](1 - L)^d \epsilon_t^2 = \alpha_0 + [1 - \beta(L)]v_t$$

The fractional differencing operator $(1 - L)^d$ can be written in terms of hypergeometric function,

$$(1 - L)^d = \sum_{k=0}^{\infty} \Gamma(k - d)\Gamma(k + 1)^{-1}\Gamma(-d)^{-1}L^k.$$

Rearranging the terms in the last equation, an alternative representation for the FIGARCH(p, d, q) model may be obtained as

$$[1 - \beta(L)]h_t = \alpha_0 + [1 - \beta(L) - \phi(L)(1 - L)^d]\epsilon_t^2.$$

From the above, the conditional variance h_t is given by

$$h_t = \alpha_0[1 - \beta(1)]^{-1} + \{1 - [1 - \beta(L)]^{-1}\phi(L)(1 - L)^d\}\epsilon_t^2 = \alpha_0[1 - \beta(1)]^{-1} + \lambda(L)\epsilon_t^2, \quad (28)$$

where $\lambda(L) = \lambda_1 L + \lambda_2 L^2 + \dots$. To define the FIGARCH(p, d, q) of 28 properly, the conditional variance in the ARCH(∞) representation must be non-negative, i.e., $\lambda_k \geq 0$ for $k = 1, 2, \dots$

The condition for the nonnegativity of the conditional variance of the FIGARCH $(1, d, 1)$ model is covered in (Baillie et al., 1996). Following the discussion above,

$$h_t = \alpha_0[1 - \beta(1)]^{-1} + [1 - [1 - \beta_1 L]^{-1} \phi_1 L(1 - L)^d] \epsilon_t^2,$$

and

$$\lambda(L) = \lambda_1 L + \lambda_2 L^2 + \dots = 1 - [1 - \beta_1 L]^{-1} \phi_1 L(1 - L)^d.$$

Then from Taylor expansion we have the equations for the coefficients

$$\lambda_1 = \phi_1 - \beta_1 + d$$

,

$$\lambda_2 = (d - \beta_1)(\beta_1 - \phi_1) + \frac{d(1 - d)}{2}$$

,

$$\lambda_3 = \beta_1[d\beta_1 - d\phi_1 - \beta_1^2 + \beta_1\phi_1 + \frac{d(1 - d)}{2}] + d\frac{1 - d}{2}\left(\frac{2 - d}{2}\left(\frac{2 - d}{3} - \phi_1\right)\right)$$

,

$$\lambda_k = \beta_1 \lambda_{k-1} + \left(\frac{k - 1 - d}{k} - \phi_1\right) \delta_{d, k-1}, \quad k = 2, 3, \dots,$$

,

where $\delta_{d, k} = \delta_{d, k-1}(k - 1 - d)k^{-1}$ refer to the coefficients in the series expansion of $(1 - L)^d$ for $k = 2, 3, \dots$ That is,

$$\delta_d(L) = \sum_{k=1}^{\infty} \delta_{d, k} L^k,$$

with $\delta_{d, 0} = 1$. Using the non-negativity of λ_k 's, it is possible to derive inequalities which are sufficient for all conditional variances h_t to be non-negative:

$$\beta_1 - d \leq \phi_1 \leq \frac{2 - d}{3} \quad \text{and} \quad d\left(\phi_1 - \frac{1 - d}{2}\right) \leq \beta_1(d - \beta_1 + \phi_1).$$

Practically FIGARCH(1, d , 1) is enough to analyze financial volatility process, however, it is necessary to mention that non-negativity restriction conditions of conditional variances for higher order models cannot be derived easily (Caporin, 2002), so as to stationary conditions.

4.1 Estimation of FIGARCH

The estimation of parameters of FIGARCH model is generally carried out using the maximum likelihood method (which is most efficient) with normality assumption for z_t . But as the same to GARCH, the normality assumption can be questioned with some empirical evidence and therefore the use of quasi-maximum likelihood estimator is preferred. The likelihood of a FIGARCH(p, d, q) process based on the sample $\{\epsilon_1, \epsilon_2, \dots, \epsilon_T\}$ may be written as

$$\log L(\theta, \epsilon_1, \epsilon_2, \dots, \epsilon_T) \simeq -0.5T \log(2\pi) - 0.5 \sum_{t=1}^T T [\log(h_t) + \epsilon_t^2 h_t^{-1}],$$

where $\theta' \equiv (\alpha_0, d, \beta_1, \dots, \beta_p, \phi_1, \dots, \phi_q)$. The likelihood function is maximized conditional on pre-sample values of ϵ_t^2 for $t = 0, -1, -2, \dots$ in the truncated ARCH $_{\infty}$ representation at the unconditional sample variance.

(Baillie et al., 1996) claimed the asymptotic normality of the quasi-maximum likelihood estimator $\hat{\theta}_T$, when $(\epsilon_1, \dots, \epsilon_T)$ form a sample from FIGARCH(1, d , 0) through extending a similar result available for IGARCH(1, 1). (Baillie et al., 1996) used an upper bound for the infinite sequence of coefficients of the ARCH(*infty*) representation of an IGARCH model. A similar argument was also used in claiming the asymptotic properties of the QMLE for the FIGARCH. (Mikosch and Starica, 2003) cast doubts on Baillie's claims by pointing out it is not possible to bound a hyperbolically decaying sequence (the ARCH(∞) coefficients of

any FIGARCH process) by an exponentially decaying sequence, which deficiets the proof of the claim from (Caporin, 2002), too. However, the reported Monte Carlo experiment result in (Baillie et al., 1996) and (Bollerslev and Mikkelsen, 1996) is still valid and can be used to conclude the consistency and asymptotic normality of the QMLE estimators empirically for FIGARCH(1, d , 0). However, the proofs of consistency and asymptotic normality of the QMLE estimator for the general FIGARCH(p , d , q) model are still open to be resolved. The estimation of FIGARCH parameters using QMLE can be implemented in MFE toolbox of MATLAB (Sheppard, 2009).

To estimate the parameters a FIGARCH model individually, we can employ the methods listed as below. For fractionally integrated index d , we can use GPH estimator proposed in (Geweke and Porter-Hudak, 1983). To obtain the estimation of other parameters see (Härdle and Mungo, 2008). There are several ways to estimate fractionally integrated index and (Robinson, 1995) is some of them.

When it comes down to forecasting, the one-step ahead forecast of h_t can be given by

$$h_t(1) = \alpha_0[1 - \beta(1)]^{-1} + \lambda_1\epsilon_t^2 + \lambda_2\epsilon_{t-1}^2 + \dots$$

And the l -step ahead forecast is

$$h_t(l) = \alpha_0[1 - \beta(1)]^{-1} + \lambda_1 h_t(l-1) + \lambda_2 h_t(l-2) + \dots + \lambda_{l-1} h_t(1) + \lambda_l \epsilon_t^2 + \lambda_{l+1} \epsilon_{t-1}^2 + \dots$$

When applying FIGARCH to forecast volatility, we can see from Figure ?? and ?? that FIGARCH introduces oscillations in volatility forecasts, also implies it may be over-sensitive to one point replacement or outliers. Thus, to include persistence effect into volatility forecasting results, we are going to introduce regime switching into FIGARCH model.

4.2 regime switching FIGARCH

Even though regime switching has been applied in a lot long memory processes, there are quite few discussing the incorporation between regime switching and long memory in volatility processes except some recent research from (Dark,), (Xie, 2009). Even though FIGARCH has not been well developed in terms of stationary conditions in higher order cases, FIGARCH(1,d,1) is somehow enough to capture the stylized fact that squared returns of financial time series exhibit some long memory behavior. and for parsimonious purpose, we do not consider more than 2 regimes in RS-FIGARCH. The estimation scheme directly follows what we discussed in RS-GARCH given Hamilton filter and QMLE of FIGARCH introduced in last section.

For RS-FIGARCH model, the specification is defined as: residual terms ϵ_t can be defined as

$$\epsilon_t = \sqrt{h_t}\eta_t,$$

where conditional variance depends on regime at time t

$$h_{t,S_t} = \alpha_{0,S_t}[1 - \beta_{1,S_t}]^{-1} + \lambda_{S_t}(L)\epsilon_t^2,$$

where λ_{S_t} is the sequence derived using regime specific FIGARCH(1,d,1) parameters β_i, ϕ_i, d_i for $S_t = i$.

To present a general feeling of the estimation performance, we include a sample of histograms of estimation results from RS-FIGARCH DGP, see Figure 4.2. The number of data points simulated is 3000, and the step is replicated 100 times. QMLE optimization method is usually starting from randomly grid search of different starting points, and then seek optimization for each starting point individually, then aggregate to find semi-global QMLE. We might that this estimation

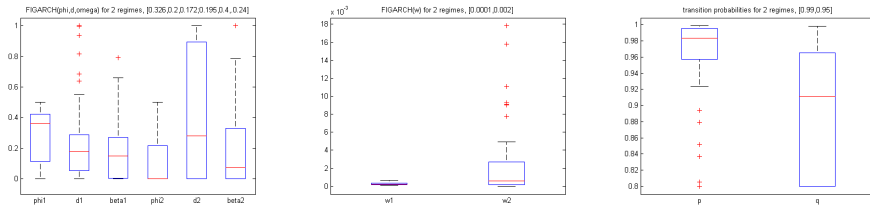


Figure 4: box plot of estimation results for RS-FIGARCH
100 replications, 3000 data points used for estimation.

results do not seem to be robust, especially for the turbulent regime (see Figure 4.2). One reason is that the likelihood function of RS-FIGARCH is rather flat since there are ten parameters need to be estimated in a row, the second reason is that the turbulent or high volatility regime has small unconditional probabilities compared to the stable regime, thus acquires very little information compared to stable regime. Thus we can give another approach to the estimation step based on the empirical finding. It is safe to say that the stable regime has more robust estimation, and we can constrain the parameters in the regime for the constrained optimization step (can be adjusted in `fmincon` linear constraints) and then we can proceed our estimation so forth. Typically, we can start with d_1 and ϕ_1 first, and bounded β_1 based on the restriction conditions, then shrink the likelihood function. Same as what we discussed in RS-GARCH, RS-FIGARCH can also have intrigue turbulent regime, but not the same as in RS-GARCH the second regime of RS-FIGARCH usually estimated as IGARCH instead of GARCH with parameters equal to 0.

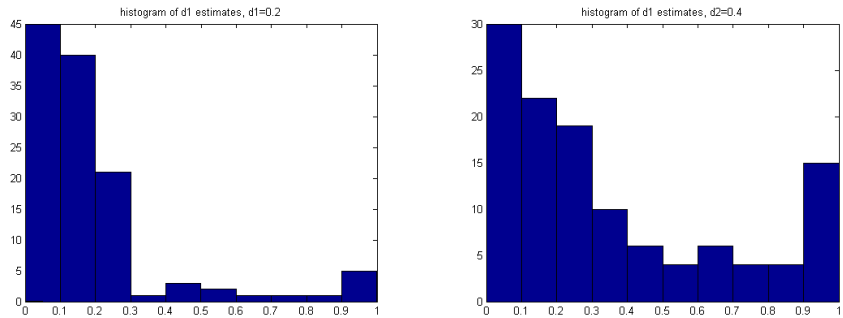


Figure 5: box plot of estimation results for RS-FIGARCH
 100 replications, 3000 data points used for estimation, parameters used for simulation are
 (0.0295,0.1255,0.2000,0.1669,0.0055,0.0309,0.4000,0.0510,0.9900,0.9500).

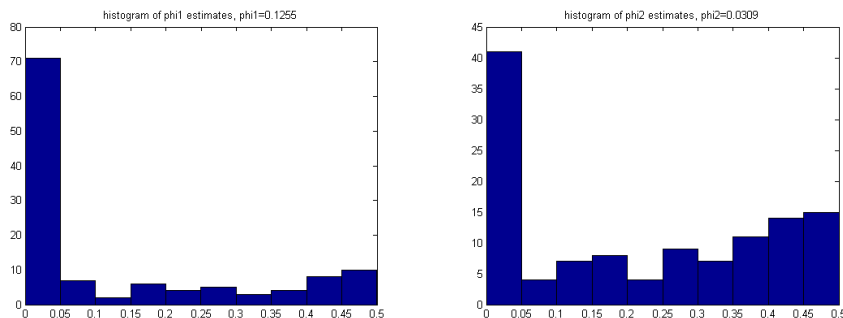


Figure 6: box plot of estimation results for RS-FIGARCH
 100 replications, 3000 data points used for estimation, parameters used for simulation are
 (0.0295,0.1255,0.2000,0.1669,0.0055,0.0309,0.4000,0.0510,0.9900,0.9500).

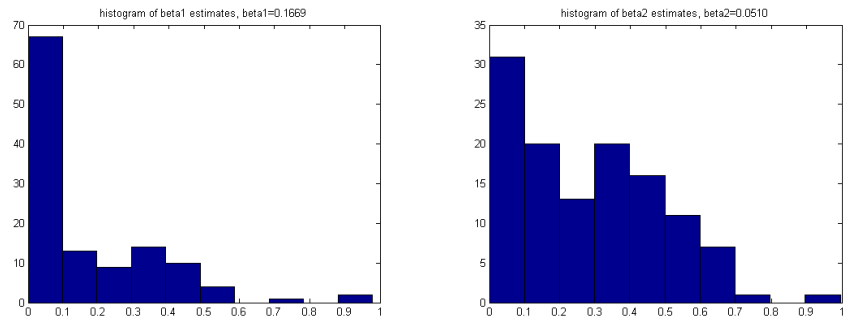


Figure 7: box plot of estimation results for RS-FIGARCH
 100 replications, 3000 data points used for estimation, parameters used for simulation are
 (0.0295,0.1255,0.2000,0.1669,0.0055,0.0309,0.4000,0.0510,0.9900,0.9500).

Chapter 3

5 Backtesting

On a practical perspective, to test model validity risk management plays an important role. While there are many aspects of risks, we would concentrate on market risk in this thesis, and the highly standard risk measure that has been used in market risk modeling is Value-at-risk(VaR). The regulators nowadays released a whole package of VaR-based criterions to the requirements in terms of regulation and supervision of Banking system to evaluate their risks. VaR can be also used to evaluate the risks of individual assets and a portfolio as well. Technically VaR for significance level α can be defined as:

$$VaR_{\alpha}(X) = -inf(x \in R | P(X \leq x)) > \alpha \quad (29)$$

where $\alpha \in (0, 1)$ and usally taken as 95% or 99% in practical use, and $F^{-1}(\alpha)$ stands for the inverse function of cumulative probability density of log returns. We are going to focus on VaR backtests in this section. Some knowledge regarding both Unconditional Coverage tests and Independence Tests will be covered and performed to test VaR performance of our RS-GARCH and RS-FIGARCH models.

Backtesting aims to take ex ante var forecasts from a particular model and compare them with ex post realized returns. whenever losses exceed var, a VaR violation is said to have occurred. There are a few methods to backtest models. We basiely will discuss binomial and berkowitz tests here. (At each time point, we use previous historical data to fit model, and make forecasts for the next period(in

the context of one-step forecast). The first statistical test is kupiec test (Kupiec, 1995), also named as unconditional coverage test. Roughly speaking it tests if the violation of sequence occurs with probability p or not. We can give a 0/1 indicator sequence by following the criteria:

$$\begin{cases} 0 & \text{if } VaR_{\alpha}(r_t) > -r_t, \\ 1 & \text{if } VaR_{\alpha}(r_t) \leq -r_t \end{cases} \quad (30)$$

where r_t denotes realized return series. Given independence of sequence of I_t , the null of hypothesis is $E(I_t) = p$ while the alternative is $E(I_t) \neq p$. The likelihood under the null hypothesis is :

$$L(p; I_1, \dots, I_T) = (1 - p)^{n_0} p^{n_1}, \quad (31)$$

where n_1 is the number of violations, i.e. $\sum_0^T I_t$, and $n_0 + n_1 = T$. Under the alternative

$$L(\pi; I_1, \dots, I_T) = (1 - \pi)^{n_0} \pi^{n_1} \quad (32)$$

Then the likelihood ratio test statistic can be calculated as:

$$LR_{uc} = -2 \log \frac{L(\pi; I_1, \dots, I_T)}{L(\hat{\pi}; I_1, \dots, I_T)} \quad (33)$$

where $\hat{\pi}$ is maximum likelihood estimator of π from Equation 32, and $\hat{\pi} = \frac{n_1}{n_0 + n_1}$.

To evaluate the non-reject regions for Kupiec test we propose a table with different number of forecasts as Table 1 and 2 with significance level of 0.05 and 0.01 respectively.

(Christoffersen, 1998) also gives a test with respect to both independence and unconditional coverage simultaneously, whose test statistic is simply the summation of Equation 33 and likelihood ratio test statistic of Christoffersen independence test, and its asymptotically distributed as chi-square distribution with degree of freedom as 2. See details in (Christoffersen, 1998). Both Kupiec test and

Table 1: Non-rejection region for Kupiec test of different confidence levels, significance level of the test is 0.05

VaR confidence levels	T=100	T=200	T=400
99.9%	$x = 1$	$x = 1$	$0 < x < 3$
99.5%	$0 < x < 3$	$0 < x < 4$	$0 < x < 6$
99%	$0 < x < 4$	$0 < x < 6$	$0 < x < 9$

This table shows non-rejection regions of 99%, 99.5%, 99.9% VaR in 100, 200, 400 forecast regions, significance level of test is 0.01.

Table 2: Non-rejection region for Kupiec test of different confidence levels, significance level of the test is 0.01

VaR confidence levels	T=100	T=200	T=400
99.9%	$x = 1$	$0 < x < 3$	$0 < x < 3$
99.5%	$0 < x < 4$	$0 < x < 5$	$0 < x < 7$
99%	$0 < x < 5$	$0 < x < 7$	$0 < x < 12$

This table shows non-rejection regions of 99%, 99.5%, 99.9% VaR in 100, 200, 400 forecast regions, significance level of test is 0.01.

Christoffersen tests are corresponding to violation indicator sequence in Equation 30, which does not acquire much information out of return series. To obtain a test with great power compared with the tests above, Berkowitz(2001) propose a density evaluation method to be able to extract more information from sample data. To begin with, the return series r_t is transformed to uniform random variable with:

$$x_t = \int_{-\infty}^{r_t} f(u)du,$$

where $f(u)$ is the probability density function of r_t determined by arbitrary model specification. In our backtesting part of the thesis, we do it both in the context of standard NTS distribution and normal distribution. After x_t being transformed to standard normal random variable as $z_t = F^{-1}(x_t)$, we perform the test against the alternative that a first order autoregressive mechanism by:

$$z_t - \mu = \rho(z_{t-1} - \mu) + \varepsilon_t$$

Then the log-likelihood function can be written as:

$$\begin{aligned} L(\mu, \sigma^2, \rho) &= -0.5 \log 2\pi - 0.5 \log\left(\frac{\sigma^2}{1 - \sigma^2}\right) - \frac{(z_t - \mu / (1 - \rho))^2}{2\sigma^2 / (1 - \rho^2)} \\ &= -\frac{T-1}{2} \log 2\pi - \frac{T-1}{2} \log(\sigma^2) - \sum_{t=2}^T \frac{(z_t - \mu - \rho z_{t-1})^2}{2\sigma^2}, \end{aligned}$$

where σ^2 is the variance of ε_t . The test statistic of null as independence is:

$$LR_{ind} = -2(L(\hat{\mu}, \hat{\sigma}^2, 0) - L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho})) \sim \chi^2(1), \quad (34)$$

where in practice $\hat{\mu}, \hat{\sigma}^2$ can be approximated by mean and variance of return series respectively, and $\hat{\rho}$ can be calculated by sample autocorrelation with lag 1. Check out Berkowitz(2001) to see details of the test statistic, With the use of

the transformed z_t , Berkowitz also gives a tail test with respect to large losses in returns. The new variable z_t^* is defined:

$$z_t^* = \begin{cases} \Phi^{-1}(\alpha) & \text{if } z_t \geq \Phi^{-1}(\alpha) \\ z_t & \text{if } z_t < \Phi^{-1}(\alpha) \end{cases}$$

Then the conditional likelihood function is:

$$\begin{aligned} L(\mu, \sigma^2 | z^*) &= \sum_{z^* < \Phi^{-1}(\alpha)} \left(-0.5 \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (z^* - \mu)^2 \right) \dots \\ &\dots + \sum_{z^* = \Phi^{-1}(\alpha)} \log\left(1 - \Phi\left(\frac{\Phi^{-1}(\alpha) - \mu}{\sigma}\right)\right) \end{aligned}$$

Similarly, the asymptotic distribution of the test statistic $-2(L(0, 1) - L(\hat{\mu}, \hat{\sigma}^2))$ follows chi-square with degree of freedom as 2. To perform standard backtesting procedure, we are going to perform all these five likelihood-ratio based tests in the empirical exhibition section.

6 Empirical results in high frequency

Data used in the thesis is extracted from tick-by-tick archive through trade and quote (TAQ) database of New York Stock Exchange. Data is aggregated for 1 minute, 5minute respectively and cleaned as well. Missing values are mid-point interpolated already, and data sets used in the thesis are close price for each aggregation period. Dates of each data set are adjusted corresponding to Eastern Standard Time Zone, especially for data of Shanghai Shenzhen 300 Index data.

To give an overview of this section, we are going to introduce the steps in model estimation and forecasting steps. First of all, the models used in this section include ARMA(1,1)-GARCH(1,1) with standard Normal Tempered Stable innovations (GARCH-stdNTS henceforth), ARMA(1,1)-GARCH(1,1)-Gaussian, ARMA(1,1)-FIGARCH(1,d,1)-Gaussian, ARMA(1,1)-FIGARCH(1,d,1)-stdNTS, ARMA(1,1)-2-regime switching-GARCH(1,1)-Gaussian, ARMA(1,1)-2-regime switching-FIGARCH. To get results as robust as possible, we filter residuals from ARMA(1,1) first, and then estimate parameters of dynamic volatility processes accordingly. Sometimes this sequential estimation method is called back-out method. It is worth mentioning here, to obtain better fit results of three parameters of standard NTS distribution, we apply the steps following (Mittnik et al., 1999). We need to first estimate QMLE under GARCH or FIGARCH with student-T innovation, and then apply MLE-FFT method to fit standard NTS parameters using extracted residuals. QQ plots of residuals fits for both GARCH-stdNTS and FIGARCH-stdNTS are shown in Figure 13. QQ plot of GARCH-Gaussian residuals versus fitted Gaussian distribution is shown in Figure 10, and this result is consistent with the stylized fact that GARCH innovation is fat-tailed, especially for high frequency data.

All backtesting procedure in this section are performed with sliding window size as 2000 and forecasting horizon as 400, one step forecast is used in models of interest.

6.1 Goodness of fit test

To find out how the specified model is fit to the data, it is critical to verify the model's descriptive power. Statistically a goodness of fit (Gof) examination for the employed distribution of the innovation is used. Several statistical tests can do the work, and within each test p-value has to be compared to some predetermined confidence level, usually chosen as 95% or 99%, and if the resulting p-value falls below some arbitrary values, 0.05 and 0.01 respectively, null hypothesis should be rejected in favor of alternative hypothesis.

One most frequently used goodness of fit test is Kolmogorov-Smirnov test. It aims to assess if two cumulative probability density functions come from the same distributions with some predetermined confidence. Given two distributions $F_1(x)$ and $F_2(x)$, the test statistics is constructed as:

$$KS = \sqrt{n} \sup_{x_i} |F_1(x_i) - F_2(x_i)|$$

, and furthermore, this test is also applicable to the cases where either of $F_1(x)$ or $F_2(x)$ can be empirical distribution of sample data, or some theoretical cumulative density distribution.

Another test often referred to as goodness of fit test is Anderson-Darling test (AD test), which is carried out in very similar procedure but has accentuates more discrepancies in tails of desired distributions, and empirically relatively good

power. The A-D test is computed by:

$$AD = \sup_{x \in R} \frac{|F_1(x_i) - F_2(x_i)|}{\sqrt{F_1(x_i)(1 - F_2(x_i))}}$$

Specially for Gof of Gaussian distribution, (Shapiro and Wilk, 1965) test and J(Jarque and Bera, 1980) test are also frequently used.

Residuals goodness of fit tests of both ARMA(1,1)-GARCH(1,d,1)-NTS and ARMA(1,1)-FIGARCH(1,d,1)-NTS versus estimated distributions are performed in Table 3 and Table 4. It is evident that RS-GARCH, even RS-FIGARCH cannot fully capture the fat-tailed behavior of residuals, indicating that we might still require further incorporation between RS-FIGARCH and standard NTS innovations.

6.2 Some VaR forecasts versus log return series

In this subsection we will show standard backtesting results under both significance level 0.01, 0.05 and also using different frequency data for both S&P 500 data and ShanghaiShenzhen 300 Index(SHSZ 300) data. Before doing backtesting formally, the S&P 500 price from 2008-09-02 to 2008-09-15 is shown as Figure 6.2: After filtering the residual terms, we use them to forecast next period (2008-09-12 to early hour of 2008-09-15). First we take forecasts of volatilities into consideration to illustratively describe how long memory effect and regime switching effect together impact the forecast performance. The characteristics of GARCH, FIGARCH and RS-FIGARCH forecasting results during financial crisis are depicted in Figure 16: GARCH model barely moves when the huge plunge occurs, and FIGARCH have lots of small spike, while RS-FIGARCH only have

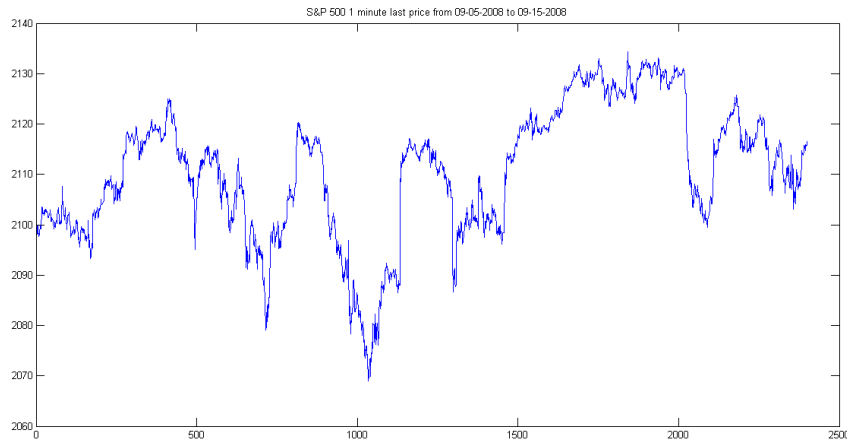


Figure 8: Data is between 2008-09-02 and 2008-09-15, in the middle of Lehmann Brother bankruptcy event (at time index 350)

very few but larger spikes than FIGARCH. Mostly RS-FIGARCH is smoothing and stable, but it is sensitive to large events. It reacts earlier than FIGARCH forecasts, also. For 99% VaR here, we can see GARCH has flat results as usual, and FIGARCH has lots of small spikes. All models but GARCH-Gaussian and stdNTS-RS-GARCH are able to capture that huge spike. The violation numbers for different forecasting regions are in 6

Now we compare the results among all candidates, which are already introduced at the beginning of the section. Figures in this subsection are all shown in Appendix. Figure 18, 19, 20 show VaR forecasts of the same recent data. At the very beginning of the whole period, 99% VaR of RS-FIGARCH stays at the very bottom among all the models, and after a while, market somehow turns into calm state and the forecasts of RS-FIGARCH elevated slowly, which is consistent with

our assumption that RS-FIGARCH tends to be less conservative during tranquil period. For 99.5% and 99.9% VaRs, the candidate models except for GARCH-NTS exhibit bandwidth and do not drop very much, but since GARCH-NTS has fatter tail index in distribution of innovations, the forecasts drops dramatically, no matter at the first peak or in a general sense.

The data of the second group do not have very huge spikes but overall its scale is also large compared to last group of data. We can observe that RS-FIGARCH, under this case, have largest VaR values most of the time, but when spikes of log returns occur, it still responds very quickly. All candidates exhibit the same behavior except GARCH-Gaussian. And there is no surprise that GARCH-NTS is again very conservative at 99.9% VaR. In both second and third group GARCH-NTS has quite close 99.5% VaR values while in first group it does not have. And in 99.9% VaR it just dominates.

Since Christoffersen independence test are not rejected in any data sets out of this thesis, we would perform the first three tests in an easier way. For that we already have the non-rejection region table of Kupiec test in both significance levels, we are going to simply display violation numbers. The violation numbers of Kupiec test (significance level is 0.01) for 99% VaR in different horizons are shown in Table 6. More violation results combined with descriptions and explanations from different data and horizons are shown in Appendix. Table 8, 9, 10 are results for 99% VaR violations in 100, 200 and 400 forecasting regions, and Table 11, 12, 13 are results for 99% VaR violations. From these tables we can see our model RS-FIGARCH has relatively low rejection rates and thus can be helpful in risk management.

7 Conclusion

Regime switching model is well known as a model able to capture both persistence and abrupt changes, and is widely used in economics and financial time series. We incorporate regime switching model with Fractionally Integrated GARCH (FIGARCH) and observe that it has very impressive performance as shown in Table 8-13. If we take closer view at its forecasts in Figure 13-24, RS-FIGARCH also tends to have conservative volatility forecasts compared to alternative models in long term, but in short term, RS-FIGARCH is able to capture volatility spikes and can react fast to large spikes compared to FIGARCH alone which is consistent with the properties: persistence and flexibility. We also find that FIGARCH without regimes has forecasting results more oscillated and meanwhile sensitive to outliers.

We introduce standard Normal Tempered stable distribution, and model fat-tailed and skewed innovations in this thesis too. Even though FIGARCH, RS-GARCH and RS-FIGARCH can be quite flexible, none of them is able to fully interchange fat-tailed innovations according to normality test results. However when performing Goodness of fit test on Standard NTS with GARCH, FIGARCH, RS-GARCH and RS-FIGARCH residuals, and none of them were rejected. Introducing Standard NTS also improves backtesting results of GARCH, FIGARCH, and RS-GARCH, as suggested in 6. In this thesis We include insights in strength and weakness of semi-global QMLE estimation method and propose the outline of an alternative sequential estimation method.

According to violation results in the context of Kupiec test, RS-FIGARCH overall has RS-FIGARCH also has potential to predict if we have entered into high volatility market, determined by whether cross-over between RS-FIGARCH

forecasts and forecasts of other model, especially RS-GARCH. And the validation of this indicator will be included in my future research. In future work, I am also going to apply RS-FIGARCH and GARCH-stdNTS in option pricing and portfolio optimization. I will discover a broad extension to this model, such as time varying regime switching model, and multifractal regime switching model, etc.

8 Appendix

Table 3: normality test of residuals extracted from RS-GARCH

KS test	AD test	Jarque-Bera test	Shapiro-Wilk test
0.0784	0	0	3.83e-06

Notes: AD, JB, SW tests are rejected at significance level 0.05 and 0.01, while only KS test is not rejected, at both significance levels.

Table 4: normality test of residuals extracted from RS-FIGARCH

KS test	AD test	JB test	Shapiro-Wilk test
0.0735	0	0	8.4040e-07

Notes: AD, JB, SW tests are rejected at significance level 0.05 and 0.01, while only KS test is not rejected, at both significance levels. Same conclusion as in RS-FIGARCH

Table 5: Goodness of fit test statistics

Models/statistics	p-value of KS test	p-value of AD test
GARCH-stdNTS	0.1969	0.06
FIGARCH-stdNTS	0.2378	0.12

The data is S&P 500 with 5 minute frequency, extracted from TAQ database, ranged from 2015-4-20 9:35 to 2015-6-1 13:25. Residuals from both ARMA-GARCH and ARMA-FIGARCH are not rejected by KS test and AD test of null as Gof of standard NTS, thus we can not reject Gof of standardized residuals to standard NTS of both models. The qq plots are shown in Figure ??.

Table 6: violation numbers for 99% VaR in different horizons, significance level is 0.01

Forecasting Horizon	100	200	400
GARCH-Gaussian	0	0	1
FIGARCH-Gaussian	1	1	3
GARCH-stdNTS	1	1	5
stdNTS-RS-GARCH	0	0	1
FIGARCH-stdNTS	0	0	4
RS-FIGARCH-Gaussian	0	0	1

The table presents violation numbers of 99% VaRs in all models. Data is from 2008-09-02 to 2008-09-12. The notorious event 'lehman brothers filed bankruptcy' happens approximately at 350 index of forecasting horizon.

Table 7: Backtesting results (p-values) of SP 1min with 200 forecasting horizon, significance level is 0.05

data	KupUc	ChInd	ChCc	BerkInd	BerkTail
GARCH-Gaussian	0.1622	0.1545	0.1366	0.3447	0.0276
FIGARCH-Gaussian	0.0737	0.6117	0.1766	0.8084	0.1357
GARCH-stdNTS	0.7493	0.6228	0.8420	0.4466	0.8858
stdNTS-RS-GARCH	0.1622	0.5413	0.3124	0.4317	0.2669
FIGARCH-stdNTS	0.5020	0.3083	0.4750	0.3051	0.7994
RS-FIGARCH	0.0080	0.7619	0.0284	0.3382	0.0183

Note: This table shows p values of the tests: Kupiec coverage test (KupUc), Christoffersen independence test (ChInd) and combined test (ChCc). Berkowitz independence test (BerkInd) and tail test (BerkTail). Performed under 2000 sliding window size. Data still comes from pre-crisis period. RS-FIGARCH is rejected by both Kupiec test and Berkowitz Tail test at 0.05.

Table 8: violation numbers for 99% VaR in 100 forecasting horizon, significance level is 0.05

data	SP1min	SP5min	SHSZ1min	SHSZ5min
GARCH-Gaussian	3	1	1	0
FIGARCH-Gaussian	5	1	3	0
GARCH-stdNTS	2	1	0	1
RS-GARCH	2	2	1	0
RS-FIGARCH	2	1	1	2

SP1min: 2015-05-08 to 2015-06-03 , SP5min: 2015-4-20 to 2015-6-1, and SHSZ1min: 2015-05-28 to 2015-06-09, SHSZ5min: 2014-12-19 to 2015-03-03. Looking up Table 1, none of these models is rejected by Kupiec test at 0.05 significance level, yet FIGARCH-Gaussian in SP1min, and all VaRs in SHSZ5min except RS-GARCH are rejected at 0.01 significance level.

Table 9: violation numbers for 99% VaR in 200 forecasting horizon, significance level is 0.05

data	SP1min	SP5min	SHSZ1min	SHSZ5min
GARCH-Gaussian	6	3	3	2
FIGARCH-Gaussian	8	3	3	2
GARCH-stdNTS	4	2	1	6
RS-GARCH	3	2	1	0
RS-FIGARCH	5	3	1	1

Data range: SP1min is 2015-05-08 to 2015-06-03 , SP5min is 2015-4-20 to 2015-6-1, and SHSZ1min is 2015-05-28 to 2015-06-09, SHSZ5min is 2014-12-19 to 2015-03-03.

In the column of SP1min, both GARCH-Gaussian and FIGARCH are rejected, and in other panels of data, every violation is in the range of 1 to 5, except GARCH-stdNTS and RS-GARCH in SHSZ5min.

Table 10: violation numbers for 99% VaR in 400 forecasting horizon, significance level is 0.05

data	SP1min	SP5min	SHSZ1min	SHSZ5min
GARCH-Gaussian	7	8	3	1
FIGARCH-Gaussian	9	8	3	1
GARCH-stdNTS	5	5	1	9
RS-GARCH	3	5	1	0
RS-FIGARCH	8	7	1	5

SP1min: 2015-05-08 to 2015-06-03 , SP5min: 2015-4-20 to 2015-6-1, and SHSZ1min: 2015-05-28 to 2015-06-09, SHSZ5min: 2014-12-19 to 2015-03-03.

These models are not rejected by Kupiec test at 0.05 significance level overall except FIGARCH-Gaussian in SP1min, and reasonable violations from Table 1 are in the range of 1 to 8.

Table 11: violation numbers for 99.5% VaR in 100 forecasting horizon, significance level is 0.01

data	SP1min	SP5min	SHSZ1min	SHSZ5min
GARCH-Gaussian	3	1	1	1
FIGARCH-Gaussian	4	1	3	3
GARCH-stdNTS	2	1	1	5
RS-GARCH	2	2	0	1
RS-FIGARCH	2	1	3	3

SP1min: 2015-05-08 to 2015-06-03 , SP5min: 2015-4-20 to 2015-6-1, and SHSZ1min: 2015-05-28 to 2015-06-09, SHSZ5min: 2014-12-19 to 2015-03-03.

FIGARCH-Gaussian, GARCH-stdNTS are jected according to Table 2.

Table 12: violation numbers for 99.5% VaR in 200 forecasting horizon, significance level is 0.01

data	SP1min	SP5min	SHSZ1min	SHSZ5min
GARCH-Gaussian	6	2	0	1
FIGARCH-Gaussian	7	3	2	1
GARCH-stdNTS	3	1	0	6
RS-GARCH	3	2	0	1
RS-FIGARCH	5	2	1	3

SP1min: 2015-05-08 to 2015-06-03 , SP5min: 2015-4-20 to 2015-6-1, and SHSZ1min: 2015-05-28 to 2015-06-09, SHSZ5min: 2014-12-19 to 2015-03-03.

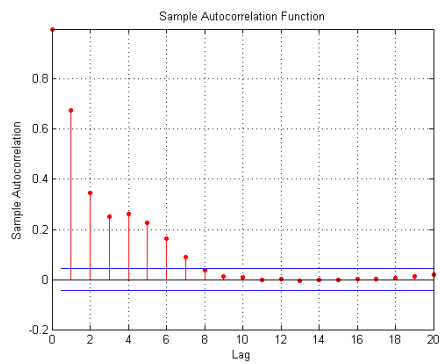
FIGARCH, GARCH and RS-FIGARCH are rejected in SP1min, while GARCH-stdNTS are rejected in SHSZ 5min.

Table 13: violation numbers for 99.5% VaR in 400 forecasting horizon, significance level is 0.01

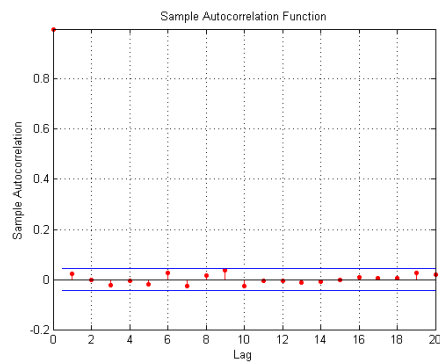
data	SP1min	SP5min	SHSZ1min	SHSZ5min
GARCH-Gaussian	7	4	1	1
FIGARCH-Gaussian	8	7	3	1
GARCH-stdNTS	3	1	0	10
RS-GARCH	3	3	1	0
RS-FIGARCH	6	6	1	3

SP1min: 2015-05-08 to 2015-06-03 , SP5min: 2015-4-20 to 2015-6-1, and SHSZ1min: 2015-05-28 to 2015-06-09, SHSZ5min: 2014-12-19 to 2015-03-03.

GARCH-Gaussian is rejected in SP1min FIGARCH-Gaussian is rejected in SP 1min and 5min, GARCH-stdNTS is rejected in SHSZ 5min. RS-GARCH and RS-FIGARCH are not rejected by in any data sets according to Table 2.



(a) ACF of long memory process



(b) ACF of short memory process

Figure 9: ACF of long memory process and short memory process

The data from left panel is absolute values generated from IGARCH(1,1), and the right panel is generated from regular GARCH(1,1) model.

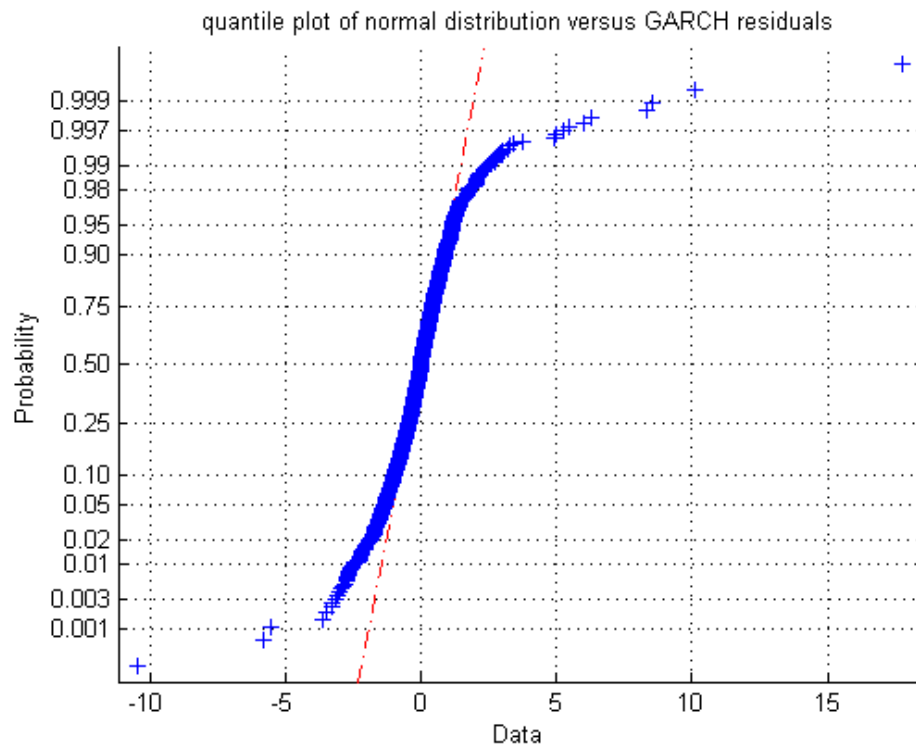


Figure 10: QQ plot of residuals versus Gaussian distribution

The residuals is filtered from ARMA(1,1)-GARCH(1,1) model, and then compared versus the Gaussian distribution in terms of quantiles.

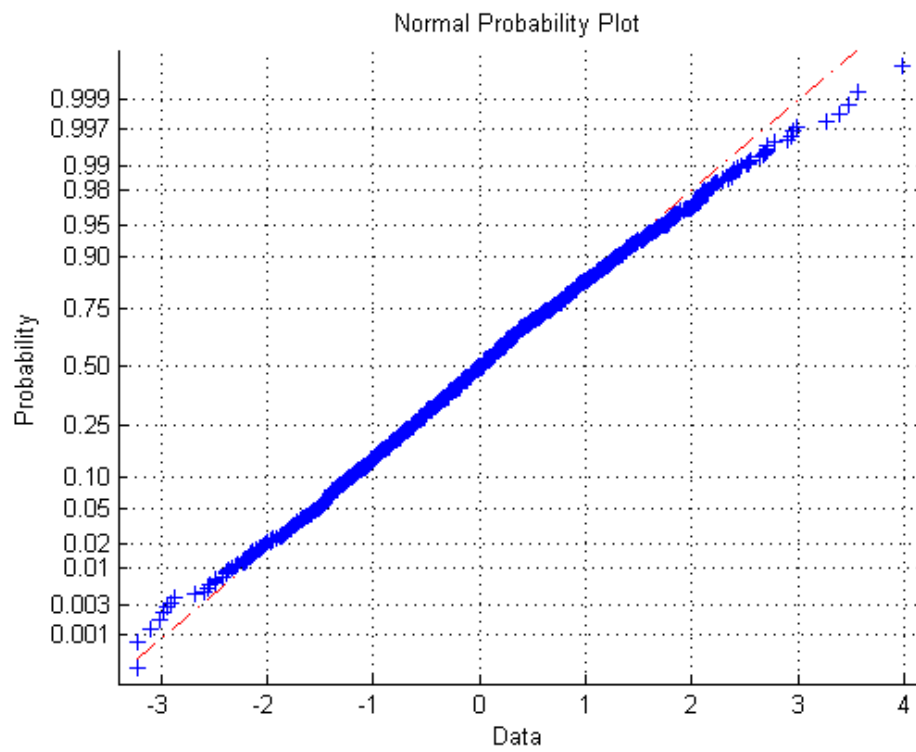


Figure 11: QQ plot of residuals versus Gaussian distribution

The residuals is filtered from ARMA(1,1)-RS-GARCH(1,1) model, and then compared versus the Gaussian distribution in terms of quantiles.

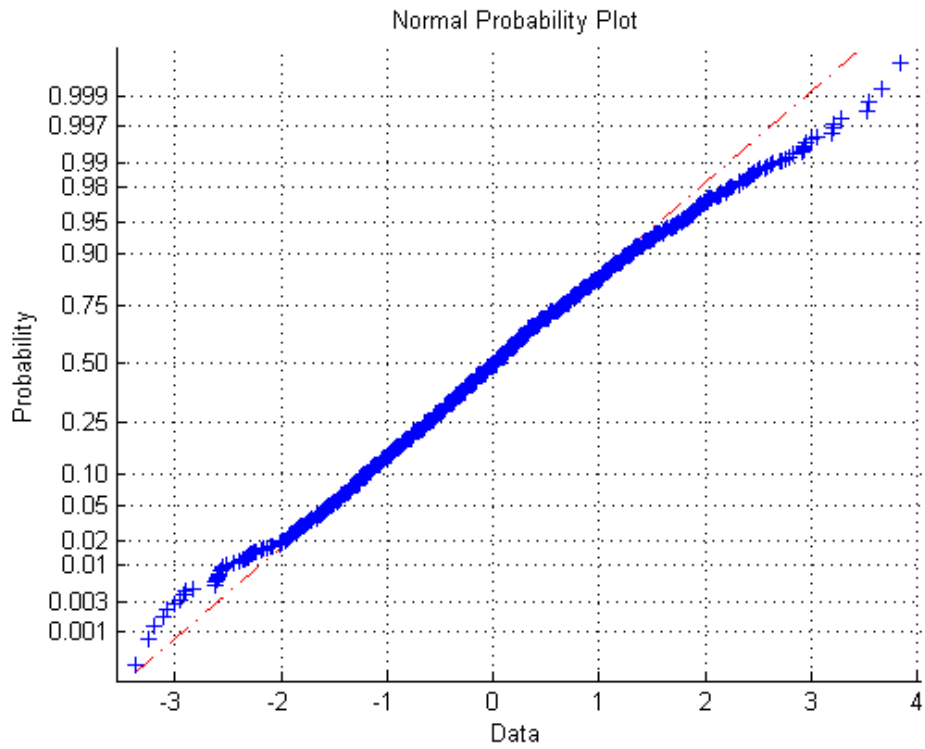
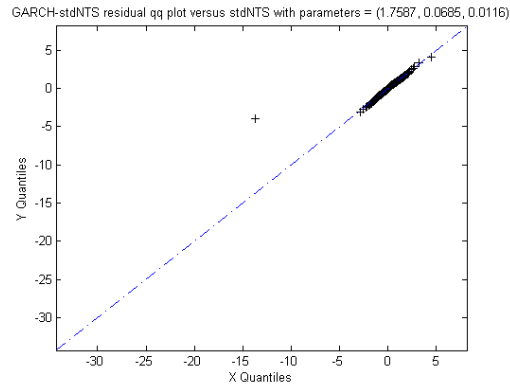
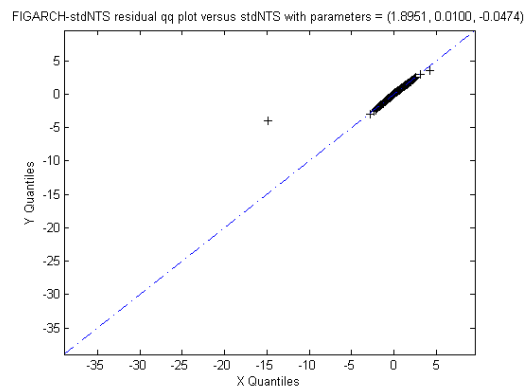


Figure 12: Standardized residuals extracted from RS-GARCH versus fitted standard normal distribution

The residuals are filtered from ARMA(1,1)-RS-GARCH(1,1) model, and then compared versus the Gaussian distribution in terms of quantiles.

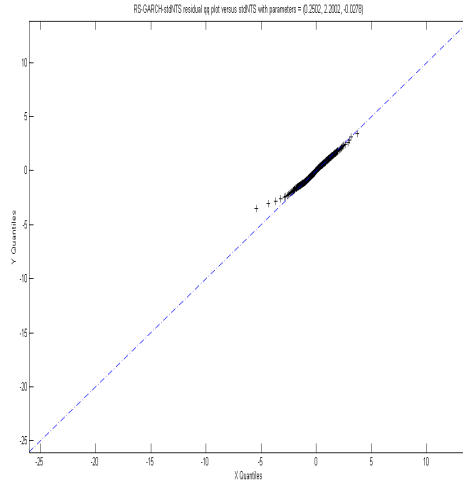


(a) GARCH-stdNTS

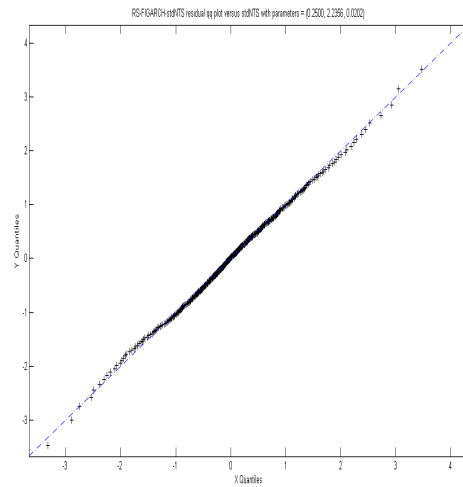


(b) FIGARCH-stdNTS

Figure 13: Standardized residuals versus fitted stdNTS



(a) RS-GARCH-stdNTS



(b) RS-FIGARCH-stdNTS

Figure 14: QQ plots of residuals versus stdNTS, estimate SHSZ 1minute data, from 2015-5-28 to 2015-6-9

The left panel is QQ plots of innovations of ARMA(1,1)-RS-GARCH(1,1) model versus fitted standard NTS, and the right panel is for QQ plot of ARMA(1,1)-RS-FIGARCH(1,1) innovations.

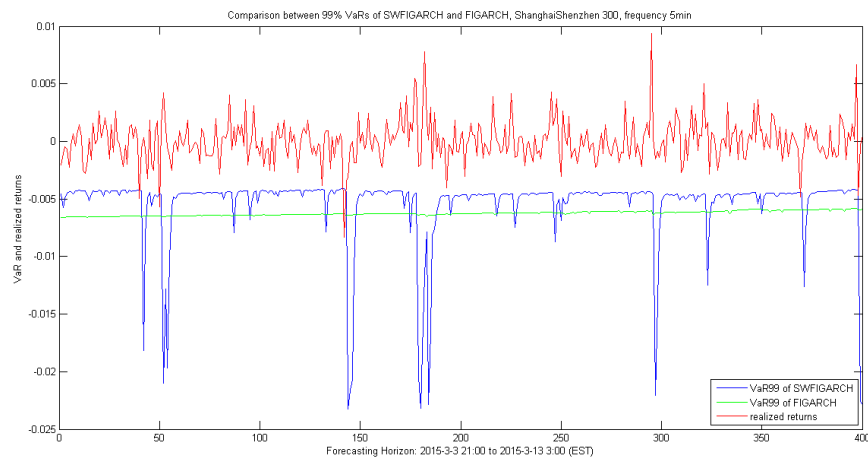


Figure 15: VaR forecasting results of RS-FIGARCH and FIGARCH

This figure shows the comparison result between RS-FIGARCH forecasts and FIGARCH forecasts. It is noteworthy that in long terms FIGARCH alone can be a good tool in risk management compared to RS-FIGARCH, yet RS-FIGARCH performs much better as a portfolio optimization tool of short terms.

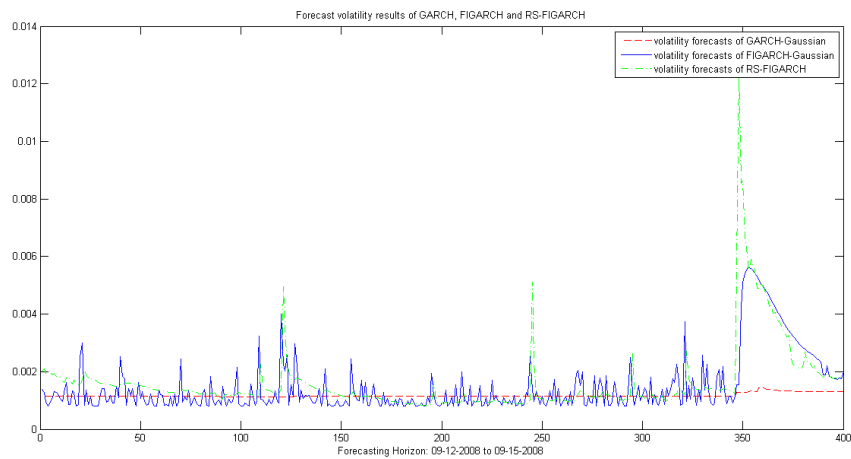


Figure 16: volatility forecasting results of GARCH, FIGARCH and RS-FIGARCH

This figure shows the comparison of volatility forecasts among three models, GARCH, FIGARCH and RS-FIGARCH. GARCH has flat results in the whole region, and FIGARCH has lots of small spikes. However, RS-FIGARCH forecasts stay stable mostly and much less but larger spikes are featured.

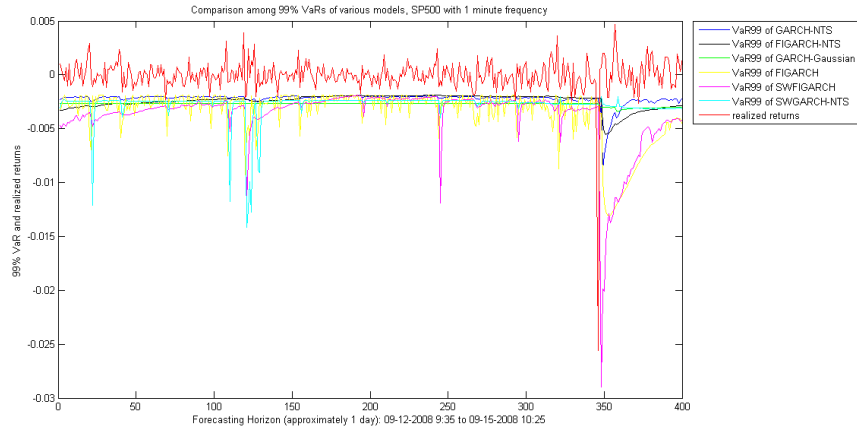


Figure 17: 99% VaR forecasting results of GARCH, FIGARCH, RS-FIGARCH, GARCH-NTS, FIGARCH-NTS, stdNTS-RS-GARCH

This figure shows the comparison of 99% VaR forecasts among all candidate models. Data is from 2008-09-02 to 2008-09-15

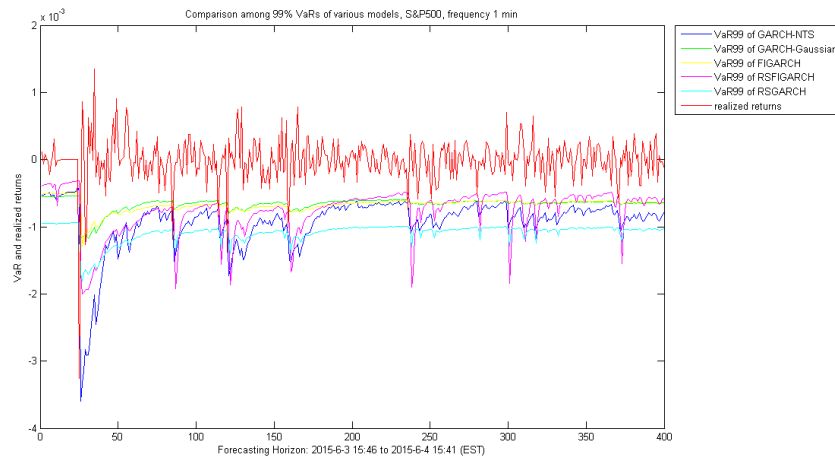


Figure 18: 99% VaRs of different models versus realized returns

The data used for initial estimation is S&P 500 with 1 minute frequency, from 2015-05-28 to 2015-06-03.

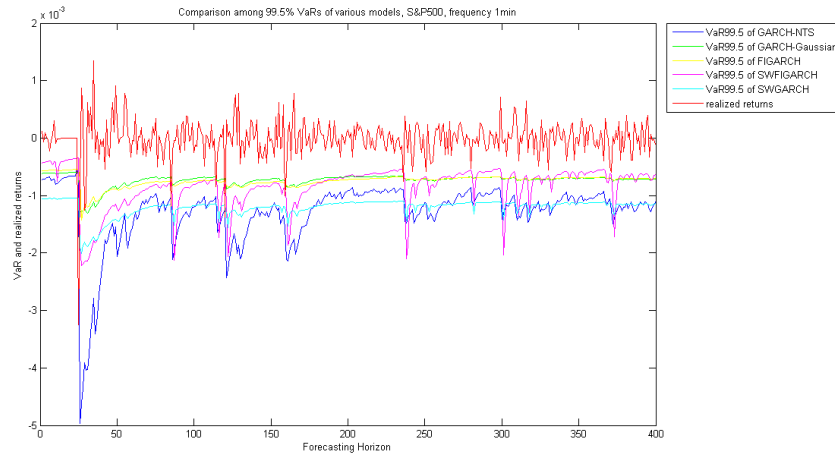


Figure 19: 99.5% VaRs of different models versus realized returns

The data used for initial estimation is S&P 500 with 1 minute frequency, from 2015-05-28 to 2015-06-03.

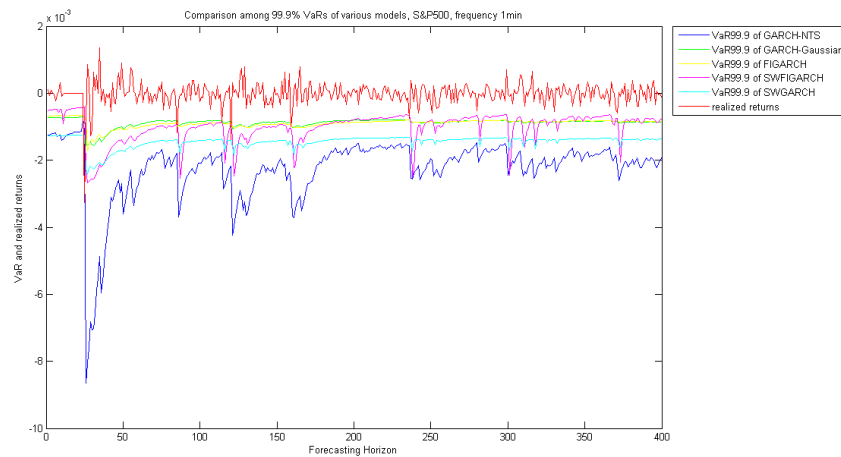


Figure 20: 99.9% VaRs of different models versus realized returns

The data used for initial estimation is S&P 500 with 1 minute frequency, from 2015-05-28 to 2015-06-03.

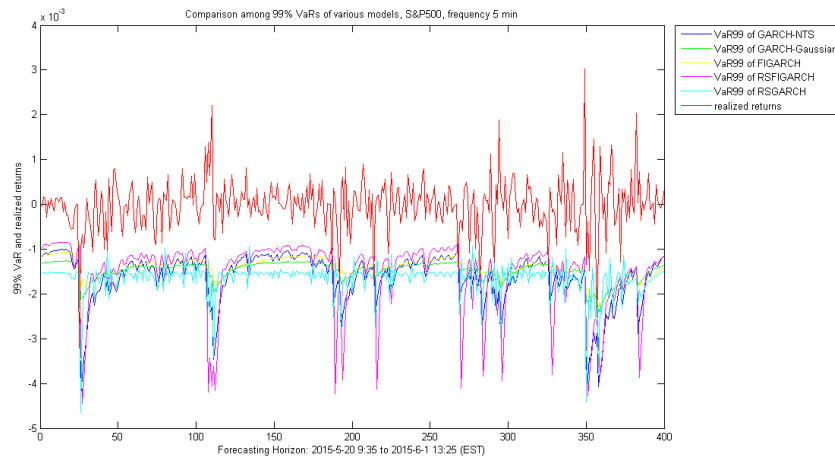


Figure 21: 99% VaRs of different models versus realized returns

The data used for initial estimation is S&P 500 with 5 minute frequency, from 2015-04-20 to 2015-05-20, forecasting horizon: 2015-5-20 9:35 to 2015-6-1 13:25.

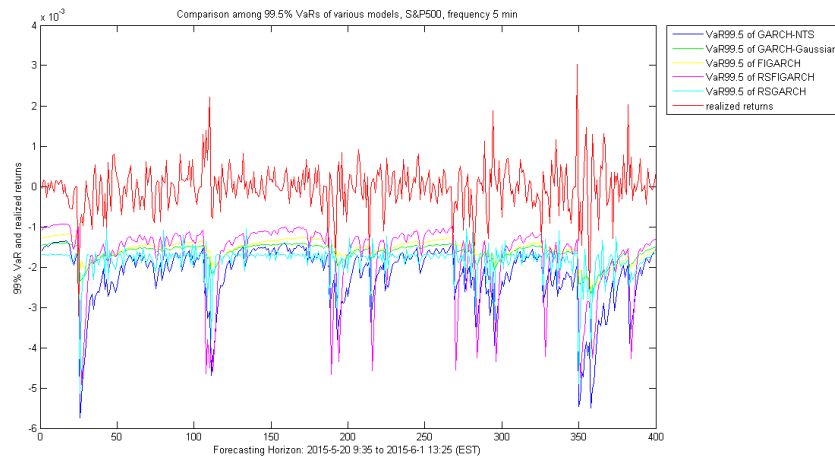


Figure 22: 99.5% VaRs of different models versus realized returns

The data used for initial estimation is S&P 500 with 5 minute frequency, from 2015-04-20 to 2015-05-20, forecasting horizon: 2015-5-20 9:35 to 2015-6-1 13:25.

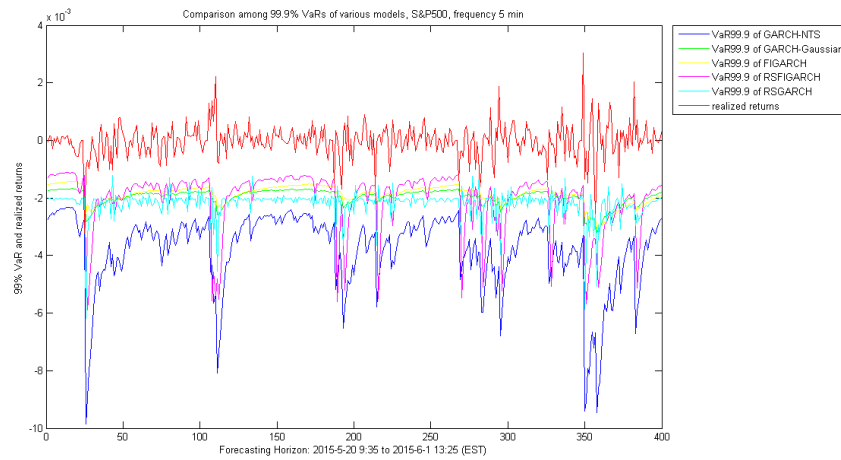


Figure 23: 99.9% VaRs of different models versus realized returns

The data used for initial estimation is S&P 500 with 5 minute frequency, from 2015-04-20 to 2015-05-20, forecasting horizon: 2015-5-20 9:35 to 2015-6-1 13:25.

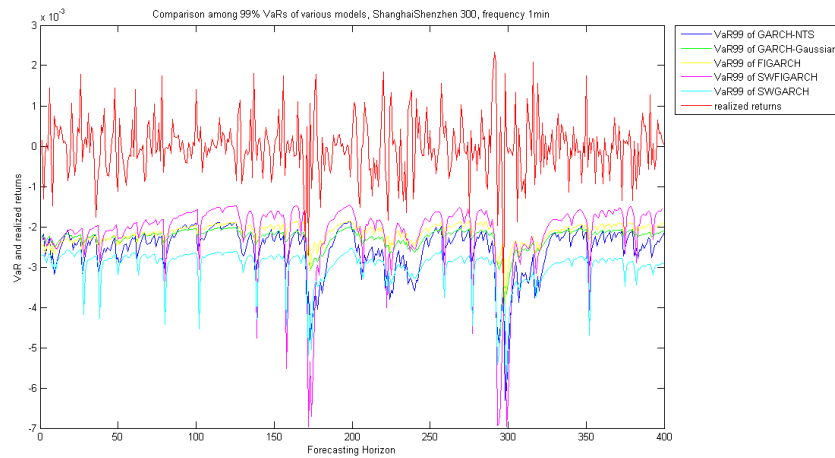


Figure 24: 99% VaRs of different models versus realized returns

The data used for initial estimation is SHSZ 300 with 1 minute frequency, from 2015-06-01 to 2015-06-09, forecasting horizon: 2015-6-9 21:00 to 2015-6-11 3:00.

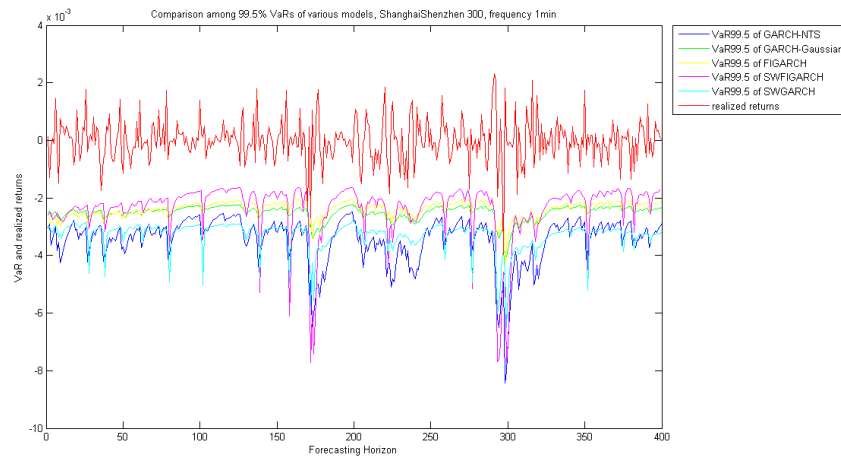


Figure 25: 99.5% VaRs of different models versus realized returns

The data used for initial estimation is SHSZ 300 with 1 minute frequency, from 2015-06-01 to 2015-06-09, forecasting horizon: 2015-6-9 21:00 to 2015-6-11 3:00.

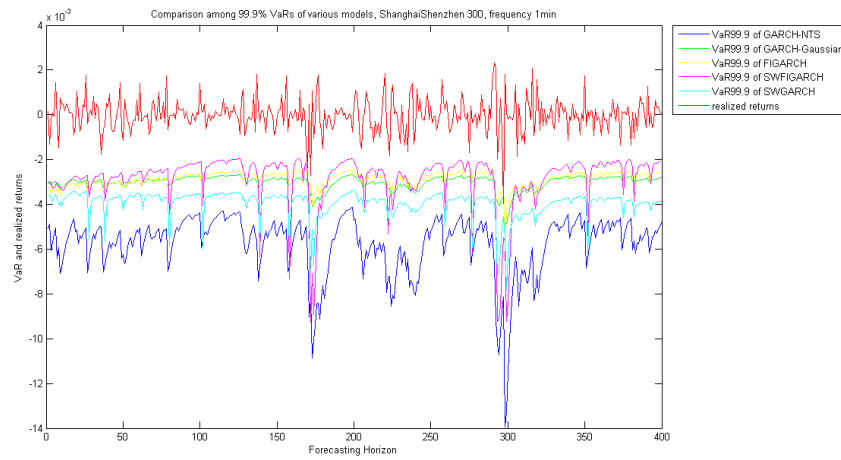


Figure 26: 99.9% VaRs of different models versus realized returns

The data used for initial estimation is SHSZ 300 with 1 minute frequency, from 2015-06-01 to 2015-06-09, forecasting horizon: 2015-6-9 21:00 to 2015-6-11 3:00.

References

- Ardia, D. (2009). Bayesian estimation of a markov-switching threshold asymmetric garch model with student-t innovations. *The Econometrics Journal*, 12(1):105–126.
- Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3–30.
- Bauwens, L., Hafner, C. M., and Laurent, S. (2012). *Handbook of volatility models and their applications*, volume 3. John Wiley & Sons.
- Bauwens, L., Preminger, A., and Rombouts, J. V. (2007). Theory and Inference for a Markov-Switching GARCH Model. *Cahiers de recherche*.
- Bauwens, L., Preminger, A., and Rombouts, J. V. (2010). Theory and inference for a markov switching garch model. *The Econometrics Journal*, 13(2):218–244.
- Bollerslev, T. (1986a). Generalized AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY. 31:307–327.
- Bollerslev, T. (1986b). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.
- Bollerslev, T. and Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility. *Journal of econometrics*, 73(1):151–184.
- Caporin, M. (2002). Figarch models: stationarity, estimation methods and the identification problem. *Università de Ca’Foscari à Venice*, (02.02).

- Carrasco, M., Hu, L., and Ploberger, W. (2014). Optimal Test for Markov Switching Parameters. *Econometrica*, 82(2):765–784.
- Cho, J. S. and White, H. (2007). Testing for regime switching. *Econometrica*, 75(6):1671–1720.
- Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, pages 841–862.
- Chung, C.-F. (1999). Estimating the fractionally integrated garch model. *National Taiwan University*, 1:20.
- Dark, J. Dynamic Hedging using a Bivariate Markov Switching FIGARCH model.
- DuMouchel, W. H. (1973). On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *The Annals of Statistics*, pages 948–957.
- Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric reviews*, 5(1):1–50.
- Francq, C. and Roussignol, M. (1997). On white noises driven by hidden markov chains. *Journal of Time Series Analysis*, 18(6):553–578.
- Francq, C., Roussignol, M., and Zakoian, J.-M. (2001). Conditional Heteroskedasticity Driven by Hidden Markov Chains. *Journal of Time Series Analysis*, 22(2):197–220.
- Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long-memory times series models.

- Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1):27–62.
- Guegan, D. and Stéphanie, R. (2005). Regime switching models: real or spurious long memory? *IDHE-MORA Note of Research*, (02-2005).
- Haas, M., Mittnik, S., and Paolella, M. S. (2004). A new approach to markov-switching garch models. *Journal of Financial Econometrics*, 2(4):493–530.
- Hamilton, D. (1994). and Changes in Regime. 64:307–333.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series.
- Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica: Journal of the econometric society*, pages 413–430.
- Härdle, W. and Mungo, J. (2008). Value-at-risk and expected shortfall when there is long range dependence.
- Henneke, J. S., Rachev, S. T., Fabozzi, F. J., and Nikolov, M. (2011). MCMC-based estimation of Markov Switching ARMA- $\hat{\text{G}}$ GARCH models. *Applied Economics*, 43(3):259–271.
- Hu, L. and Shin, Y. (2008). Optimal test for markov switching garch models. *Studies in Nonlinear Dynamics & Econometrics*, 12(3).
- Jarque, C. M. and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, 6(3):255–259.

- Kasahara, H. (2013). Modified Quasi-Likelihood Ratio Test for. (23530249):1–23.
- Kasahara, H., Okimoto, T., and Shimotsu, K. (2014). Modified quasi-likelihood ratio test for regime switching. *Japanese Economic Review*, 65(1):25–41.
- Kim, Y. S., Rachev, S. T., Chung, D. M., and Bianchi, M. L. (2008). A modified tempered stable distribution with volatility clustering. *New Developments in Financial Modelling*, pages 344–365.
- Kim, Y. S., Rachev, S. T., Chung, D. M., and Bianchi, M. L. (2009). The modified tempered stable distribution, garch-models and option pricing. *Probability and Mathematical statistics*, 29(1):91–117.
- Klaassen, F. (2002). EMPIRICAL Improving GARCH volatility forecasts with regime- switching GARCH. pages 363–394.
- Koponen, I. (1995). Analytic approach to the problem of convergence of truncated lévy flights towards the gaussian stochastic process. *Physical Review E*, 52(1):1197.
- Kupiec, P. (1995). Techniques for verifying the accuracy of risk management models. *Journal of Derviatives*.
- Marcucci, J. (2005). Forecasting Stock Market Volatility with Regime-Switching GARCH Models. *Studies in Nonlinear Dynamics & Econometrics*, 9(4).
- Meyn, S. P. and Tweedie, R. L. (1993). Stability of markovian processes ii: continuous-time processes and sampled chains. *Advances in Applied Probability*, pages 487–517.

- Mikosch, T. and Starica, C. (2003). Long-range dependence effects and arch modeling. *Long-Range Dependence: Theory and Applications*. Birkhauser, Boston, pages 439–459.
- Mittnik, S., Doganoglu, T., Chenyao, D., et al. (1999). Maximum likelihood estimation of stable paretian models. *Mathematical and Computer Modelling*, 29(10):275–293.
- Piplack, J. et al. (2009). Estimating and forecasting asset volatility and its volatility: a markov-switching range model. *Discussion Paper Series/Tjalling C. Koopmans Research Institute*, 9(08).
- Rachev, S. T., Kim, Y. S., Bianchi, M. L., Fabozzi, F. J., et al. (2011). *Financial models with Lévy processes and volatility clustering*, volume 187. John Wiley & Sons.
- Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of statistics*, pages 1630–1661.
- Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, pages 591–611.
- Sheppard, K. (2009). Mfe matlab function reference financial econometrics.
- Timmermann, A. (2000). Moments of Markov switching models. *Journal of Econometrics*, 96(1):75–111.
- Xie, Y. (2009). Consistency of maximum likelihood estimators for the regime-switching GARCH model. *Statistics*, 43(2):153–165.

Zolotarev, V. M. (1986). *One-dimensional stable distributions*, volume 65. American Mathematical Soc.