SOME PARAMETRIC AND SEMIPARAMETRIC MODELS FOR FINANCIAL TIME SERIES ANALYSIS

XINGFA ZHANG

Ph.D

The Hong Kong Polytechnic University

2012

The Hong Kong Polytechnic University

Department of Applied Mathematics

Some Parametric and Semiparametric Models for Financial Time Series Analysis

by

Xingfa ZHANG

A thesis submitted in partial fulfillment of

the requirements for the Degree of Doctor of Philosophy

June 2011

CERTIFICATION OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

____________________ (Signed)

Xingfa ZHANG (Name of student)

Abstract

Motivated by Ling's (2007) DAR (p) model, in this thesis, we study new classes of GARCH and GARCH-in-mean models which have applications to financial data such as treasury bill rate and stock indices. Unlike the previous models in the literature, the conditional variances in our considered models are specified as functions of the time-lagged observable returns instead of the usual unobservable errors. Such a setting for the conditional variance enables us to give some new insights in the analysis of financial time series.

Under the framework of an alternative specification in the conditional variance, this study considers the following aspects. First, we generalize Ling's (2007) DAR (p) model by considering a piecewise linear conditional mean instead of the single linear conditional mean in the existing models. Issues about parameter estimation and threshold test are discussed. Secondly, for a specific parametric GARCH-M model, we study its ergodicity conditions. Under some regularity assumptions, it can be shown that the quasi maximum likelihood estimator for the model is asymptotically normal. We then attempt to investigate the relationship between risk (conditional variance) and return (conditional mean) based on a class of semiparametric GARCH-M models, in which the conditional mean is specified as an unknown smooth function and the conditional variance is set as a known parametric function of lagged returns. Approaches are given to estimate the unknown function and parameters. Moreover, motivated by the time varying property of the risk aversion and the functional coefficient autoregressive model, we propose a functional coefficient autoregressive GARCH-M model to capture the variation of the risk aversion. By treating the risk aversion as a function of one day lagged return, we are able to study how yesterday's return affects today's risk mag-

nitude. Estimates for the unknown function and parameters are discussed. Finally, we generalize the proposed functional coefficient autoregressive GARCH-M model to functional coefficient GARCH-M model, from which, we can describe the effect of common factors to risk aversion. Improved estimators for the parameters are given and, under some regularity conditions, we can prove that the parametric estimators are consistent.

For all the proposed models, simulations are conducted to assess the performance of the related approaches. Applications to real data are also considered. It is demonstrated that our studied models can have comparable or better fitting performance as compared to other well known models.

Acknowledgements

I wish to express my most sincere gratitude to my chief supervisor, Dr. Heung Wong, for his guidance, encouragement and understanding over the years. I benefited from numerous discussions with him on the ideas and results presented in this thesis.

I am greatly indebted to my co-supervisors, Dr. Wai-Cheung Ip and Prof. Yuan Li, for their enthusiastic guidance and support. Prof. Yuan Li stimulated my interest in Non(Semi)-parametric time series analysis while he was supervising my Master thesis at Guangzhou University and provided many valuable advices.

I am grateful to my senior fellow apprentice Fan Zhang for his kind help in both study and life. I also would like to thank my friends, Yan Liang, Zhangyou Chen and Lin Shu, for their help in the past years.

I shall express my sincere thanks to all the staff in the Department of Applied Mathematics at the Hong Kong Polytechnic University for their kind assistance, and I greatly appreciate the financial support from The Hong Kong Polytechnic University, whose research studentship supported me in the past three years.

Finally and most of all, I would like to thank my parents and my wife, the ones love me and I love. Their care, encouragement and understanding enabled me to finish my study.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The famous ARCH (autoregressive conditional heteroscedastic) model was proposed by Engle (1982) and was then generalized to GARCH (generalized autoregressive conditional heteroscedastic) model by Bollerslev (1986). (G)ARCH models have been successfully used to describe the clustering phenomenon of the stock volatility and hence they are widely applied in practice. Following the publications of (G)ARCH models, there have been numerous extensions which can be summarized in two major directions. The first class of extensions focuses on the purely parametric models such as Nelson's (1991) EGARCH (exponential GARCH), the GJR model of Glosten et al. (1993), the TARCH (threshold ARCH) of Zakoian (1994) and the GARCH-M (GARCH-in-mean) model of Engle et al. (1987). More variants of the parametric GARCH specification can be found in Degiannakis and Xekalaki (2004). Besides the purely parametric extensions, with the rapid development in computing power, nonparametric and semiparametric statistical approaches are also widely adopted to study the (G)ARCH models. For example, Engle and González-Rivera (1991), Linton (1993), Drost and Klaassen (1997) studied the GARCH type models by assuming the error density as some unknown function. Pagan and Hong (1991), Engle and Ng (1993), H*ä*rdle and Tsybakov (1997), Yang (2006) considered nonparametric and semiparametric forms of the volatility function. Linton and Perron (2003), Christensen et al. (2008), Conrad and Mammen (2008) investigated the relationship between mean and variance based on semiparametric and nonparametric methods. For a survey article, one can refer to Linton (2009).

There are three fundamental topics when we study the (G)ARCH models and they are respectively estimation, testing, and stability conditions. Many theoretical results have been obtained for the (G)ARCH models. Some of them are as follows. Nelson (1990), and Bougerol and Picard (1992) established conditions for the stationarity and ergodicity of the GARCH process. Lee and Hansen (1994), and Lumsdaine (1996) proved the consistency and asymptotic normality of the quasi maximum likelihood estimator (QMLE) for the GARCH (1, 1). Jensen and Rahbek (2004) obtained some limiting results on QMLE of GARCH (1, 1) process for the nonstationary case and Berkes et al. (2003) considered the structure and estimation for the general GARCH (p,q) process. Ling and McAleer (2002a, b) derived conditions for the existence of moments in the GARCH (p,q) processes. The formulas for theoretical autocorrelation function in the GARCH (p,q) model were discussed by He and Ter*ä*svirta (1999a, b) and Karanasos (1999).

The GARCH-in-mean (GARCH-M) model proposed by Engle et al. (1987) is also a generalization from the original GARCH models. It is useful to describe relations between the first and second conditional moments of stock returns (French et al., 1987), output growth (Caporale and McKiernan, 1996) and inflation rates (Grier and Perry,

2000) etc. Hong (1991) derived the autocorrelation structure for a GARCH-M process. Motivated by Nelson (1990), Schepper and Goovaerts (1999) studied the probability density of the variance and mean for GARCH-M models. As to the results of QMLE and stability conditions for GARCH-M model, to our knowledge, few theoretical results are available until the recent results of Meitz and Saikkonen (2008) and Christensen et al. (2008). Meitz and Saikkonen (2008) gave a principle to study the stability conditions of GARCH-M model though the article itself mainly focused on applications to the GARCH model and ACD (autoregressive conditional duration) model. Christensen et al. (2008) provided the asymptotic theory of QMLE for a GARCH-M-type model, where a different specification for the conditional variance was adopted as compared to the traditional one. From a technical perspective, we can say that the difficulty for handling the traditional GARCH-M models partly lies in the complicated structure of the score function or the derivatives of the quasi likelihood function with respect to the parameter vector. The fact stems from that the usual error term, say ε_t, in the (G)ARCH equation is unobservable, which causes the perplexing recursion expressions for the related derivatives (Engle et al., 1987, Bera and Ra, 1995). If related derivatives were in simpler forms, then it would be slightly easier to study the GARCH-M-type models.

Ling (2007) proposed a DAR (p) (double autoregressive) model where the conditional variance, say h_t, was set as $h_t = \omega + a_1 y_{t-1}^2 + \cdots + a_p y_{t-p}^2$ instead of the previous $h_t = \omega + a_1\varepsilon_{t-1}^2 + \cdots + a_p\varepsilon_{t-p}^2$. Here, $\{y_{t-s}, s = 1, \cdots, p\}$ are the observable time lagged series and $\{\varepsilon_{t-s}, s = 1, \cdots, p\}$ are the usual unobservable error terms. In Ling (2007), some novel theoretical results (Remark 3.2) were acquired. For the case of $p = 1$, Ling (2004) demonstrated that the DAR(1) model was superior to the usual ARCH models for the considered data. Ling's idea comes from Weiss (1986) where the conditional variance h_t can depend on $\{y_s\}_{t-p}^{t-1}$, $\{\varepsilon_s\}_{t-q}^{t-1}$ and some other exogenous variables. In this regard he has made it more attractive and insightful. For the usual GARCH-M model, if we also substitute the error term ε_s in the (G)ARCH equation by the observable y_s, then the derivatives of the quasi likelihood function with respect to the parameters would be largely simplified. The reason is the observable y_s will not be treated as a function of the unknown parameters. Such a property is very useful and it enables us to find an alternative way to study the (G)ARCH-in-mean model in both parametric and semiparametric forms (Christensen et al., 2008).

By assuming the conditional variance h_t is purely driven by the observable $\{y_s\}_{t-p}^{t-1}$ (p can be ∞), in this thesis, we study several extensions of the ARCH and (G)ARCH-M models. Some results about estimation, testing, and stability conditions are obtained. In addition, two research articles (Zhang et al., 2011a, b) have been written based on Chapter 2 and Chapter 3 respectively.

Based on Ling's (2007) DAR(p) model, in Chapter 2, we study a class of TARCH (threshold autoregressive conditional heteroscedastic) model by considering a piecewise linear mean equation instead of a single linear mean equation. Provided the threshold is given, the asymptotic results for the QMLE of other unknown parameters are acquired. Based on the Lagrange Multiplier principle, a threshold effect test is considered and its asymptotic null distribution is shown to be a functional of a zero-mean Gaussian process. Approximate methods are given to compute the upper percentage points and simulation results show that they perform well. From the empirical studies, we know that the original model can be improved when the threshold effect is considered.

In Chapter 3, we study a special case of the GARCH-in-Mean model proposed by Christensen et al. (2008). The conditions about geometric ergodicity are discussed and under certain regularity assumptions, the asymptotic normality of the QMLE for the model is proved. Simulations demonstrate that the estimation method performs well and the given empirical studies indicate the considered model has comparable performance in data modelling as compared to the standard one. The results indicate that the model of Christensen et al. (2008) can be useful because it provides an alternative way to study the GARCH-in-Mean effect.

Aiming to find the relationship between the excess return (conditional mean) and risk (conditional variance), in Chapter 4, we consider a semiparametric (G)ARCH-M model. We firstly discuss a semiparametric ARCH-M model, which behaves like the restricted single-index model. Following the method of Xia and Tong (2006), we give steps to estimate the model. We then discuss a semiparametric GARCH-M model which generalizes the model in Christensen et al. (2008) by considering a more flexible form of the conditional variance. An improved approach is given to estimate the parameters and some theoretical results are discussed. Through simulations, it is shown that the estimation methods perform well. When applying the models to practical data, they witness nonlinear relationships between the excess return (conditional mean) and the risk (conditional variance), and it seems that a higher risk does not necessarily guarantee a higher excess return. Such results imply that the simple linear relationships or some other commonly adopted monotonically increasing parametric relations could be misspecified.

Chapter 5 proposes a class of functional coefficient autoregressive GARCH-M models to analyze some excess return series which are calculated based on the weighted stock indices. Different from the time-varying parameter GARCH-M model of Chou et al. (1992), we consider the volatility coefficient as an unknown smooth function of time-lagged returns instead of a random walk. Such a setting enables us to study the effect of the previous return to the present risk aversion. An approach is given to estimate the model and simulation results demonstrate that the performance of our method is satisfactory. Through the empirical studies, it is seen that the proposed model can better capture the variation of the excess return series as compared to the purely parametric models. Moreover, some reasonable results and interpretation about risk aversion are presented.

In Chapter 6, we generalize the functional coefficient autoregressive GARCH-M model in Chapter 5 to functional coefficient GARCH-M model, from which, we can describe the effect of common factors to the risk aversion. For the generalized model, the estimation approach is improved. Under some regularity conditions, we can show that the parametric estimators are consistent. Simulations and empirical studies are conducted to show that our method is satisfactory and applicable.

Finally, brief conclusions and prospects for future research are given in Chapter 7. Some key Matlab codes are presented in the Appendix.

Chapter 2

A Class of Threshold Autoregressive Conditional Heteroscedastic Models

2.1 Background

In a recent paper, Ling (2007) considers the double AR (p) or DAR (p) model, which has the form

$$\begin{aligned} y_t = \theta_1 y_{t-1} + \cdots + \theta_i y_{t-i} + \cdots + \theta_p y_{t-p} + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t}, \\ e_t \sim i.i.d(0,1), h_t = \omega + a_1 y_{t-1}^2 + \cdots + a_p y_{t-p}^2, \end{aligned} \tag{2.1}$$

where $\omega, a_i > 0, t \in \mathcal{N} \equiv \{-p, \cdots, 0, 1, 2, \cdots\}$, y_s is independent of $\{e_t\}$ for $t > s$. Let $\mathcal{F}_t$ be the σ-field generated by $\{e_t, \cdots, e_1, y_0, \cdots, y_{-p}\}, t \in \mathcal{N}$, then we have $\mathrm{var}(y_t|\mathcal{F}_{t-1}) = \omega + a_1 y_{t-1}^2 + \cdots + a_p y_{t-p}^2$. As mentioned in the paper, model (2.1) is a special case of the ARMA-ARCH models in Weiss (1986), but it differs from Engle's (1982) ARCH model if at least one $\theta_i \neq 0$. The difference lies in the specifications of the conditional variance: Engle's (1982) conditional variance is driven by the unobserved errors while the conditional variance of model (2.1) depends on the past observations. Such a specification of the conditional variance brings both novelty and difficulty. The novel result acquired by Ling (2007) is that the quasi maximum likelihood estimation can be still consistent and asymptotically normal when $Ey_t^2 = \infty$, which usually does not hold any more for the classical AR(p) model with i.i.d errors. Difficulty lies in finding the conditions under which the series generated from the model is stationary and geometrically ergodic. Though Ling (2007) has gotten a sufficient and necessary condition about stationarity and ergodicity for model (2.1) when $e_t \sim i.i.d.\ N(0, 1)$, it is still a difficult problem for the general case.

When $p = 1$, model (2.1) becomes the DAR(1) model whose theoretical results and practical application have been well discussed by Ling (2004), Ling and Li (2008) for both stationary and nonstationary cases. When DAR(1) model was applied to the US 3-month treasure bill rate series in Ling (2004), it was found that the model was superior to the usual AR(1) model, and seemed to be able to get a more reliable statistical inference when compared to the usual AR(1)-GARCH(1, 1) model. Nevertheless, since financial data usually present some asymmetric effect or nonlinear relationship, it is helpful to take these factors into account. A well-known tool to deal with this is the threshold autoregressive model because of its ability to capture some important characteristics such as jumps and limit cycles (Tong and Lim, 1980, Tong, 1990, Li and Lam, 1995). Consequently, it is worthwhile to consider a generalized DAR(p) model, which is piecewise linear in the mean function.

In this chapter we consider the following threshold autoregressive conditional heteroscedastic (TARCH) model:

$$y_t = \theta_0 + \sum_{i=1}^{p} \theta_i y_{t-i} + I(y_{t-d} \le r)(\phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i}) + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t},$$
$$e_t \sim i.i.d(0,1), h_t = \omega + a_1 y_{t-1}^2 + \cdots + a_m y_{t-m}^2, \tag{2.2}$$

where $\omega, a_i > 0, t \in \mathcal{N} \equiv \{-m, \cdots, 0, 1, 2, \cdots\}$, y_s is independent of $\{e_t\}$ for $t > s$, $I(\cdot)$ is the indicator function and r is the threshold parameter. For simplicity, the nonnegative integers p, d, m are assumed to be known and satisfy $0 \le p \le m, 1 \le d \le m$. The threshold parameter r is assumed to have a known bounded numerical range $\tilde{R}$, usually a finite interval. When $\theta_0 = \phi_0 = \phi_i = 0, p = m$, model (2.2) is reduced to Ling's (2007) DAR(p) model. If $y_{t-i}\,(i = 1, \cdots, m)$ in the conditional variance equation of (2.2) is replaced by $\varepsilon_{t-i}\,(i = 1, \cdots, m)$, then the model would become Li and Lam's (1995) TARCH model. The difference is that: the former belongs to Weiss' ARCH-type model while the latter is an Engle's ARCH-type model. Moreover, we relax the distribution of the process $\{\varepsilon_t\}$ to the general case instead of the original normal distribution.

The chapter is arranged as follows. In Section 2.2, we discuss the QMLE, threshold effect test and some associated asymptotic properties. Simulations and empirical studies are shown in Section 2.3. All proofs are put in Section 2.4 and we summarize the chapter in Section 2.5.

2.2 Estimation and Threshold Effect Test

2.2.1 Parametric estimation

For simplicity, we assume the threshold parameter r in model (2.2) is known. In practice, as that has been done in the subsequent Section 2.3.2, we can adopt the idea of Li and Lam (1995) to estimate r. Let $\psi = (\theta^\tau, \phi^\tau, a^\tau)^\tau, \theta = (\theta_0, \cdots, \theta_p)^\tau, \phi = (\phi_0, \cdots, \phi_p)^\tau$, $a = (\omega, a_1, \cdots, a_m)^\tau$ and $\psi \in \Psi$, which is a bounded parameter space for model (2.2). All throughout this chapter, the superscript τ denotes the transpose of a vector or a matrix. Suppose that the true parameter $\psi^0 = (\theta^{0^\tau}, \phi^{0^\tau}, a^{0^\tau})^\tau$ is an interior point of Ψ. Without loss of generality, we consider Ψ as a neighborhood of ψ^0. We need to estimate ψ based on the observations $\{y_t\}_{t=1}^T$ and initial values $y_0, \cdots, y_{1-m}$.

Consider the following quasi conditional log-likelihood function (apart from a constant term)

$$L_T(\psi) = \sum_{t=1}^{T} l_t(\psi) = \sum_{t=1}^{T} \left[-\frac{1}{2} \log h_t(\psi) - \frac{1}{2} \frac{\varepsilon_t^2(\psi)}{h_t(\psi)} \right]. \tag{2.3}$$

We have

$$\frac{\partial l_t(\psi)}{\partial \psi} = -\frac{1}{2}\left(1 - \frac{\varepsilon_t^2(\psi)}{h_t(\psi)}\right) \frac{1}{h_t(\psi)} \frac{\partial h_t(\psi)}{\partial \psi} - \frac{\varepsilon_t(\psi)}{h_t(\psi)} \frac{\partial \varepsilon_t(\psi)}{\partial \psi}, \tag{2.4}$$

$$\begin{aligned} \frac{\partial^2 l_t(\psi)}{\partial \psi \partial \psi^\tau} &= \frac{1}{2h_t^2(\psi)}\left(1 - \frac{2\varepsilon_t^2(\psi)}{h_t(\psi)}\right) \frac{\partial h_t(\psi)}{\partial \psi} \frac{\partial h_t(\psi)}{\partial \psi^\tau} + \frac{\varepsilon_t(\psi)}{h_t^2(\psi)} \frac{\partial h_t(\psi)}{\partial \psi} \frac{\partial \varepsilon_t(\psi)}{\partial \psi^\tau} \\ &\quad - \frac{1}{h_t(\psi)} \frac{\partial \varepsilon_t(\psi)}{\partial \psi} \frac{\partial \varepsilon_t(\psi)}{\partial \psi^\tau} + \frac{\varepsilon_t(\psi)}{h_t^2(\psi)} \frac{\partial \varepsilon_t(\psi)}{\partial \psi} \frac{\partial h_t(\psi)}{\partial \psi^\tau}, \end{aligned} \tag{2.5}$$

where

$$\frac{\partial \varepsilon_t(\psi)}{\partial \theta} = -(1, y_{t-1}, \cdots, y_{t-p})^\tau,$$
$$\frac{\partial \varepsilon_t(\psi)}{\partial \phi} = -I(y_{t-d} \le r)(1, y_{t-1}, \cdots, y_{t-p})^\tau, \frac{\partial \varepsilon_t(\psi)}{\partial a} = \mathbf{0},$$
$$\frac{\partial h_t(\psi)}{\partial \phi} = \frac{\partial h_t(\psi)}{\partial \theta} = \mathbf{0}, \frac{\partial h_t(\psi)}{\partial a} = (1, y_{t-1}^2, \cdots, y_{t-m}^2)^\tau. \tag{2.6}$$

For the sake of convenience, we put $h_t = h_t(\psi_0)$, $\varepsilon_t = \varepsilon_t(\psi_0)$, $\varsigma = Ee_t^4 - 1, Y_{1t} = (1, y_{t-1}, \cdots, y_{t-p})^\tau$, $Y_{2t} = (1, y_{t-1}^2, \cdots, y_{t-m}^2)^\tau$. Then the following theorem holds under Assumptions 2.1-2.2 in Section 2.4.

Theorem 2.1 *For model (2.2) with known threshold and the considered quasi log-likelihood function $L_T(\psi)$ given by (2.3), under Assumptions 2.1-2.2 in Section 2.4, there exists a fixed open neighborhood $U(\psi_0) \subset \Psi$ such that with probability one, as $T \to \infty$, $L_T(\psi)$ has an unique maximum point $\hat{\psi}_T$ in U. Furthermore, $\sqrt{T}(\hat{\psi}_T - \psi_0) \xrightarrow{L} N(0, \Omega_I^{-1}\Omega_S\Omega_I^{-1})$, where Ω_S, Ω_I are respectively given by*

$$E\begin{pmatrix} \frac{4}{h_t}Y_{1t}Y_{1t}^\tau & \frac{4I(y_{t-d}\le r)}{h_t}Y_{1t}Y_{1t}^\tau & \mathbf{0} \\ \frac{4I(y_{t-d}\le r)}{h_t}Y_{1t}Y_{1t}^\tau & \frac{4I(y_{t-d}\le r)}{h_t}Y_{1t}Y_{1t}^\tau & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{\varsigma}{h_t^2}Y_{2t}Y_{2t}^\tau \end{pmatrix}, \text{ and}$$
$$E\begin{pmatrix} \frac{2}{h_t}Y_{1t}Y_{1t}^\tau & \frac{2I(y_{t-d}\le r)}{h_t}Y_{1t}Y_{1t}^\tau & \mathbf{0} \\ \frac{2I(y_{t-d}\le r)}{h_t}Y_{1t}Y_{1t}^\tau & \frac{2I(y_{t-d}\le r)}{h_t}Y_{1t}Y_{1t}^\tau & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{h_t^2}Y_{2t}Y_{2t}^\tau \end{pmatrix}.$$

Remark 2.1 Through the proof in Section 2.4, it is known that $Ey_t^2 < \infty$ is not required to guarantee the validity of the theorem, which is consistent with Ling (2007). The matrices Ω_I, Ω_S can be calculated by the relevant sample means after the parameters have been estimated.

2.2.2 Threshold effect test

In this section, we consider the test for the threshold effect, i.e., to test

$$H_0 : \phi_0 = \phi_1 = \cdots = \phi_p = 0.$$

Such a test is nonstandard because the threshold parameter r is absent under H_0. From (2.3) and (2.4), we have

$$E\{\partial^2 L_T(\psi)/(\partial\phi\partial a^\tau)\} = E\{\partial^2 L_T(\psi)/(\partial\theta\partial a^\tau)\} = \mathbf{0}.$$

Following Davies (1977, 1987), the LM test statistic for our null hypothesis is

$$S = \sup_{r\in\tilde{R}} \eta_r^\tau (C_r - L_r^\tau C^{-1} L_r)^{-1} \eta_r, \tag{2.7}$$

where

$$\begin{aligned}
\eta_r &= T^{-\frac{1}{2}} \frac{\partial L_T(\psi)}{\partial\phi}\Big|_{\hat{\theta}_T,\, \hat{a}_T,\, \phi=0},\\
C &= -\frac{1}{T} E\left(\frac{\partial^2 L_T(\psi)}{\partial\theta\partial\theta^\tau}\right)\Big|_{\hat{\theta}_T,\, \hat{a}_T,\, \phi=0},\\
C_r &= -\frac{1}{T} E\left(\frac{\partial^2 L_T(\psi)}{\partial\phi\partial\phi^\tau}\right)\Big|_{\hat{\theta}_T,\, \hat{a}_T,\, \phi=0},\\
L_r &= -\frac{1}{T} E\left(\frac{\partial^2 L_T(\psi)}{\partial\theta\partial\phi^\tau}\right)\Big|_{\hat{\theta}_T,\, \hat{a}_T,\, \phi=0}.
\end{aligned}$$

Here $\hat{\theta}_T, \hat{a}_T$ are the QMLEs under the null hypothesis, and the above estimators are consistent due to Theorem 3.1 in Ling (2007). Under the framework of Lagrange Multiplier test (Silvey, 1959), the above quantities η_r, C, C_r, L_r are asymptotically convergent to the ones that are evaluated at the true values for θ and a under H_0. With the abuse of notation, in the rest of this chapter, η_r, C, C_r, L_r stand for the quantities evaluated at the true value of θ and a under H_0. Then we have the following theorem:

Theorem 2.2 *Suppose Assumptions 2.1-2.3 in Section 2.4 hold, then the asymptotic distribution of* $\{\eta_r\}$ *is identical to that of a* $(p+1)$*-dimensional Gaussian process* $\{\xi_r\}$ *indexed by the threshold parameter* $r \in R$*. For* $r, s \in R$*, we have*

$$\xi_r \sim N_{p+1}(0, C_r - L_r^\tau C^{-1} L_r), cov(\xi_r,\ \xi_s) = C_{\min(r,s)} - L_r^\tau C^{-1} L_s.$$

Also, the asymptotic null distribution of the LM test statistic S *in (2.7) is given by the distribution of* $\sup_{r\in\tilde{R}} \xi_r^\tau (C_r - L_r^\tau C^{-1} L_r)^{-1} \xi_r$.

Remark 2.2 Theorem 2.2 is similar to Wong and Li's (1997) Theorem, but concerns different situations. Moreover, our Assumptions 2.1-2.3 are weaker in contrast with theirs (e.g., $E\varepsilon_t^4 < \infty$ is a little stronger than $Ee_t^4 < \infty$). The proof is a generalization of Chan (1990), Wong and Li (1997) , which is given in Section 2.4.

In practice, it is necessary to estimate the upper percentage points of the asymptotic null distribution for S. For model (2.2), note that $C_r = L_r$, C_r and $C - C_r$ are positive definite. Then there exist an invertible matrix Q and a diagonal matrix $D = \text{diag}\{\lambda_1(r), \cdots, \lambda_{p+1}(r)\}$ such that QCQ^τ is an identity matrix and $QC_rQ^\tau = D$, with all $\{\lambda_i(r)\}$ being strictly between 0 and 1. Let $Q\xi_r = (B_{1r}, \cdots, B_{p+1,r})^\tau$. Then B_{ir}'s are independent Gaussian processes with mean zero and

$$\text{cov}(B_{ir}, B_{is}) = \lambda_i\{\min(r, s)\} - \lambda_i(r)\lambda_i(s).$$

As a result,

$$\xi_r^\tau(C_r - L_r^\tau C^{-1}L_r)^{-1}\xi_r = \left\{\frac{B_{1r}^2}{\lambda_1(r) - \lambda_1^2(r)} + \cdots + \frac{B_{p+1,r}^2}{\lambda_{p+1}(r) - \lambda_{p+1}^2(r)}\right\}.$$

When $p = 0$, we need to compute

$$\Pr\left\{\sup_{\beta_1 \le \lambda_1(r) \le \beta_2} \frac{B_{1r}^2}{\lambda_1(r) - \lambda_1^2(r)} > z^2\right\}, (0 < \beta_1 < \beta_2 < 1) \tag{2.8}$$

for a given z, where $\beta_1 = \min\{\lambda_1(r)\}$ and $\beta_2 = \max\{\lambda_1(r)\}$ for $r \in \tilde{R}$. For the $p > 0$ cases, we want to evaluate

$$\Pr\left\{\sup_{r \in \tilde{R}}\left[\frac{B_{1r}^2}{\lambda_1(r) - \lambda_1^2(r)} + \cdots + \frac{B_{p+1,r}^2}{\lambda_{p+1}(r) - \lambda_{p+1}^2(r)}\right] > y\right\}. \tag{2.9}$$

Based on Chan and Tong (1990), and Chan (1991), by using techniques similar to Wong and Li (1997), the probability in (2.8) can be approximated by

$$\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{z^2}{2}\right)\left(\gamma z - \frac{\gamma}{z} + \frac{1}{z}\right), \gamma = \frac{1}{2}\log\left\{\frac{\beta_2(1-\beta_1)}{\beta_1(1-\beta_2)}\right\}, \tag{2.10}$$

and the probability in (2.9) can be approximated by

$$1 - \exp\left\{-2\chi_{p+1}^2(y)\left(\frac{y}{p+1} - 1\right)\sum_{i=1}^{p+1}\int_{\tilde{R}}\frac{dt_i}{dr}dr\right\}, \tag{2.11}$$

where $\chi_{p+1}^2(\cdot)$ denotes the probability density function of the χ^2-distribution with $(p+1)$ degree of freedom and $t_i = \frac{1}{2}\log\{\lambda_i(r)/(1-\lambda_i(r))\}$.

It remains to find the matrix Q or the $\lambda_i(r)$'s. Note that $C, C - C_r$ are positive definite matrices. Let $C^{-\frac{1}{2}}$ be the inverse of the matrix $C^{\frac{1}{2}}$ that satisfies $C^{\frac{1}{2}}C^{\frac{1}{2}} = C$, then $C^{-\frac{1}{2}}(C - C_r)C^{-\frac{1}{2}}$ is also positive definite. Denote the eigenvalues of $C^{-\frac{1}{2}}(C - C_r)C^{-\frac{1}{2}}$ by $(\delta_1(r), \cdots, \delta_{p+1}(r))^\tau$ and accordingly there exists an orthogonal matrix Q_1 satisfying

$Q_1^\tau C^{-\frac{1}{2}}(C - C_r)C^{-\frac{1}{2}}Q_1 = \mathrm{diag}\{\delta_1(r), \cdots, \delta_{p+1}(r)\}$, namely,

$$I - Q_1^\tau C^{-\frac{1}{2}} C_r C^{-\frac{1}{2}} Q_1 = \mathrm{diag}\{\delta_1(r), \cdots, \delta_{p+1}(r)\}.$$

Here, I is an identity matrix of suitable dimension. Define $Q = Q_1^\tau C^{-\frac{1}{2}}$, then we have

$$QCQ^\tau = I, QC_rQ^\tau = \mathrm{diag}\{1 - \delta_1(r), \cdots, 1 - \delta_{p+1}(r)\}.$$

By definition, it is known that $\{1 - \delta_i(r)\}$'s are exactly the eigenvalues of $C^{-\frac{1}{2}} C_r C^{-\frac{1}{2}}$. Hence, to evaluate the probabilities in (2.8-2.9), we firstly need to estimate the eigenvalues of the matrix $C^{-\frac{1}{2}} C_r C^{-\frac{1}{2}}$.

2.3 Simulations and Empirical Studies

2.3.1 Simulations

This section examines the performance of the proposed LM test in finite samples through Monte Carlo simulations. We give examples for $d = 1, p \leq m \leq 2$ as follows.

M1: $y_t = 0.1 + e_t \sqrt{h_t}, h_t = 0.2 + 0.16y_{t-1}^2 + 0.09y_{t-2}^2$.

M2: $y_t = 0.36y_{t-1} + e_t \sqrt{h_t}, h_t = 0.3 + 0.49y_{t-1}^2$.

M3: $y_t = 0.3y_{t-1} + 0.1y_{t-2} + e_t \sqrt{h_t}, h_t = 0.5 + 0.1y_{t-1}^2 + 0.1y_{t-2}^2$.

M4: $y_t = 0.1 + 0.3y_{t-1} + e_t \sqrt{h_t}, h_t = 0.5 + 0.32y_{t-1}^2$.

M5: $y_t = 0.6 - 0.4I(y_{t-1} \leq 0) + e_t \sqrt{h_t}, h_t = 0.3 + 0.15y_{t-1}^2 + 0.1y_{t-2}^2$.

M6: $y_t = 0.1 + 0.1y_{t-1} - I(y_{t-1} \le 0)(0.2 + 0.2y_{t-1}) + e_t\sqrt{h_t}, h_t = 0.05 + 0.36y_{t-1}^2.$

M7: $y_t = 0.4y_{t-1} - 0.3y_{t-1}I(y_{t-1} \le 0) + e_t\sqrt{h_t}, h_t = 0.25 + 0.4y_{t-1}^2.$

M8: $y_t = 0.1y_{t-1} + 0.1y_{t-2} + (0.2y_{t-1} - 0.15y_{t-2})I(y_{t-1} \le 0) + e_t\sqrt{h_t},$

$h_t = 0.5 + 0.16y_{t-1}^2 + 0.1y_{t-2}^2.$

In the above examples, the i.i.d (0, 1) process in (2.2) is set as $e_t \sim i.i.d.\ N(0,1)$. M1-M4 are used to check the empirical size and M5-M8 are adopted to demonstrate the power of the test. We conduct 1000 replications with sample sizes T=100, 300 and 500 for each of the above examples. Following Wong and Li (1997), we choose $\tilde{R}$, the numerical range for the threshold, to be the intervals between the 10th percentile and 90th percentile of y_t. The empirical sizes or powers at the nominal upper 10%, 5%, 2.5% and 1% points are listed in Table 2.1.

Table 2.1 shows that both sizes and powers behave well. The empirical size in each case gets closer to the nominal level (especially at the nominal levels of 2.5% and 1%) and the test gets more powerful with increasing sample size.

2.3.2 Empirical studies

Ling (2004) applied the DAR(1) model to the US 3-month treasury bill rate series from July 1972 to August 2001 and found that the model fitted the data well as compared to the common AR(1) model. For comparison, we also consider the same set of data except for a longer period from January 1951 to October 2008 (totally 694 observations).

We take x_t to be the logarithms of the observed series and $y_t = x_t - x_{t-1}$. Based on Ling (2004), it is reasonable to apply model (2.2) with $p = m = d = 1$ and $\theta_0 = \phi_0 = 0$

Table 2.1: Results of the simulation experiments for assessing the empirical size and power with and without intercept.

			Empirical size and power			
Model	p m d	Sample size	10.0%	5.0%	2.5%	1.0%
M1	0 2 1	$T = 100$	5.9	3.1	1.4	0.3
		$T = 300$	5.4	2.7	1.3	0.5
		$T = 500$	6.0	3.1	1.8	0.7
M2†	1 1 1	$T = 100$	6.7	3.2	1.3	0.8
		$T = 300$	9.8	5.1	2.2	0.7
		$T = 500$	9.9	5.0	2.3	1.2
M3†	2 2 1	$T = 100$	13.8	8.1	6.5	4.6
		$T = 300$	10.7	5.1	3.1	2.1
		$T = 500$	12.9	6.3	2.8	0.9
M4	1 1 1	$T = 100$	17.9	14.6	11.7	10.3
		$T = 300$	7.1	4.3	3.0	2.0
		$T = 500$	7.2	3.7	1.7	1.0
M5	0 2 1	$T = 100$	51.2	37.5	27.2	16.9
		$T = 300$	97.5	94.8	90.8	85.3
		$T = 500$	100.0	99.8	99.8	99.1
M6	1 1 1	$T = 100$	89.8	83.3	77.2	67.5
		$T = 300$	89.3	82.2	75.4	64.5
		$T = 500$	90.3	83.9	76.9	66.8
M7†	1 1 1	$T = 100$	17.1	9.6	5.2	2.4
		$T = 300$	49.7	37.8	28.0	15.3
		$T = 500$	71.1	61.4	50.2	37.8
M8†	2 2 1	$T = 100$	30.1	21.9	17.5	14.9
		$T = 300$	59.6	38.3	27.1	16.7
		$T = 500$	58.36	46.2	35.9	24.5

Notes: (1) †Testing with no intercept; (2) Number of replications=1000.

to the considered data, which has the form

$$y_t = \theta y_{t-1} + \phi y_{t-1} I(y_{t-1} \leq r) + e_t \sqrt{\omega + a_1 y_{t-1}^2}, e_t \sim i.i.d(0, 1). \tag{2.12}$$

Before fitting the data by model (2.12), we first test whether $\phi = 0$ is significant, namely we consider the hypothesis $H_0 : \phi = 0$. With the numerical range $\tilde{R}$ being the interval between the 10th percentile and 90th percentile of y_t, the p value for the considered test is 0.0194, which shows that it is reasonable to introduce the threshold part. To estimate the threshold parameter r, we adopt the idea of Li and Lam (1995). Denote the potential candidates for r by $\mathcal{R} = \{r_1, r_2, \cdots, r_L\}$, the estimation of r is performed by considering

$$\max_{r \in \mathcal{R}} L_T(\hat{\psi}_T(r)),$$

where $\hat{\psi}_T(r)$ is the maximizer of the quasi log-likelihood given by (2.3) with the threshold parameter r being fixed.

For comparison, we use $\{y_t\}_1^{640}$ to estimate model (2.12) and leave $\{y_t\}_{641}^{693}$ for out-of-sample forecasts. Take $\mathcal{R}$ as a series of evenly spaced points in $\tilde{R}$ with step length being 0.001. Then we get $\hat{r} = -0.0422$, based on which, the estimation results of (2.12) are as follows:

$$\begin{aligned} y_t &= \underset{(0.0732)}{0.3149} y_{t-1} + \underset{(0.1337)}{0.2188} y_{t-1} I(y_{t-1} \leq -0.0422) + e_t \sqrt{h_t}, \\ h_t &= \underset{(0.0004)}{0.0022} + \underset{(0.2075)}{0.7656} y_{t-1}^2, e_t \sim i.i.d(0, 1). \end{aligned} \tag{2.13}$$

The values in parentheses are the corresponding standard errors which are calculated based on Theorem 2.1. We also estimate the DAR(1) model based on Ling (2004):

$$y_t = \underset{(0.0614)}{0.4009} y_{t-1} + e_t \sqrt{h_t},$$
$$h_t = \underset{(0.0004)}{0.0021} + \underset{(0.2081)}{0.7740} y_{t-1}^2, e_t \sim i.i.d(0,1). \tag{2.14}$$

Moreover, we have the results below:

For (2.13),

$$E(\log |\ 0.3149 + e_t \sqrt{0.7656}\ |) = -0.8957 < 0,$$
$$E(\log |\ 0.3149 - e_t \sqrt{0.7656}\ |) = -0.9483 < 0.$$

For (2.14),

$$E(\log |\ 0.4009 + e_t \sqrt{0.7740}\ |) = -0.8236 < 0,$$
$$E(\log |\ 0.4009 - e_t \sqrt{0.7740}\ |) = -0.8381 < 0.$$

Note that (2.12) and DAR(1) model can be rewritten respectively as

$$y_t = \theta y_{t-1} + e_t \sqrt{a_1}|y_{t-1}| + \phi y_{t-1} I(y_{t-1} \le r) + e_t\left(\frac{\omega}{\sqrt{\omega + a_1 y_{t-1}^2} + \sqrt{a_1}|y_{t-1}|}\right),$$
$$y_t = \theta y_{t-1} + e_t \sqrt{a_1}|y_{t-1}| + e_t\left(\frac{\omega}{\sqrt{\omega + a_1 y_{t-1}^2} + \sqrt{a_1}|y_{t-1}|}\right).$$

Both $\phi y_{t-1} I(y_{t-1} \le r)$ and $\omega/(\sqrt{\omega + a_1 y_{t-1}^2} + \sqrt{a_1}|y_{t-1}|)$ are $o(|y_{t-1}|)$ as y_{t-1} goes to infinity. We know from Example 4.1 in Cline and Pu (2004) that the estimated parameters for the above models satisfy the geometric ergodicity conditions. The statistic $Q(M)$ in Li and Mak (1994) with $M = 3, 6, 12$ are used for checking the adequacy of the model (2.13) and their values are $Q(3) = 0.8764 < \chi^2_{3,0.95} = 7.815$, $Q(6) = 5.5123 < \chi^2_{6,0.95} = 12.592$, and $Q(12) = 16.8462 < \chi^2_{12,0.95} = 21.026$, which suggests that model (2.13) is adequate for the considered data at the 5% level . The value of the log-likelihood for model (2.13) is 1436 and that for model (2.14) is 1434.4.

Next we apply model (2.13-2.14) to obtain one step ahead forecasts for $\{y_t\}_{642}^{693}$. We get

$$\text{for (2.13)}, \quad RMSE = \sqrt{\frac{1}{52}\sum_{t=642}^{693}(y_t - \hat{y}_t)^2} = 0.1253,$$
$$\text{for (2.14)}, \quad RMSE = \sqrt{\frac{1}{52}\sum_{t=642}^{693}(y_t - \hat{y}_t)^2} = 0.1281.$$

We have also computed the one-step ahead forecast intervals with 95% confidence level for each case. Denote u_{at}, u_{bt} as the upper bound series, which are respectively calculated according to (2.13) and (2.14). Similarly let l_{at} and l_{bt} denote the corresponding lower bounds. We list the percentiles of the difference series between upper and lower bounds in Table 2.2. It can be seen from the table that model (2.13) generates slightly narrower confidence intervals. In term of the log-likelihood values, the RMSEs and the distance between the estimated bounds, we know that model (2.13) is superior to model (2.14) for the considered data.

Table 2.2: Percentiles of difference series between upper and lower bounds.

Difference	Percentiles				
series	10%	25%	50%	75%	90%
$\{u_{at} - l_{at}\}$	0.1873	0.1978	0.2418	0.3487	0.6746
$\{u_{bt} - l_{bt}\}$	0.1875	0.1981	0.2425	0.3502	0.6781

It makes sense to consider Li and Lam's (1995) TARCH model with order p_1 =

$p_2 = 1, d = 1$ for the data. The model is

$$y_t = \theta y_{t-1} + \phi y_{t-1} I(y_{t-1} \le r) + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t},$$
$$h_t = \omega + a_1 \varepsilon_{t-1}^2, e_t \sim i.i.d.\ N(0,1). \tag{2.15}$$

Before fitting the data by model (2.15), we apply the method in the special case (A) of Wong and Li (1997) to test whether ϕ is significantly different from zero, namely $H_0 : \phi = 0$. The p value for the considered test is computed as 0.5591 by choosing $\tilde{R}$ as the interval between the 10th percentile and 90th percentile of y_t, which suggests $\phi = 0$ in (2.15). Hence we switch to estimate the model

$$y_t = \theta y_{t-1} + \varepsilon_t, \varepsilon_t = e_t \sqrt{\omega + a_1 \varepsilon_{t-1}^2}, e_t \sim i.i.d.\ N(0,1).$$

Based on observations $\{y_t\}_1^{640}$, we shall get

$$y_t = \underset{(0.0259)}{0.2603} y_{t-1} + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t},$$
$$h_t = \underset{(0.0001)}{0.0022} + \underset{(0.0704)}{0.8567} \varepsilon_{t-1}^2, e_t \sim i.i.d.\ N(0,1). \tag{2.16}$$

Using (2.16), we obtain one step ahead forecasts for $\{y_t\}_{642}^{693}$ with a RMSE of 0.1330, which is larger than that of model (2.13) and (2.14) respectively. Thus model (2.13) seems to be more reasonable than (2.16) for the considered data.

2.4 Proofs

We make the following assumptions for model (2.2).

Assumption 2.1. The series $\{y_t\}$ generated from model (2.2) is strictly stationary and geometrically ergodic for the considered parameter space Ψ.

Assumption 2.2. The i.i.d (0, 1) process $\{e_t\}$ satisfies $Ee_t^4 < \infty$, and is absolutely continuous with a continuous symmetric probability density function which is positive everywhere.

Assumption 2.3. The process $\{y_t\}$ is ρ-mixing with an exponential decreasing rate, i.e., there exists a μ between 0 and 1 such that $\rho(m) = O(\mu^m), m \in \mathcal{N}$, where $\rho(m) = \sup|\text{corr}(f, g)|$, the supremum being over all square integrable f and g which are measurable with respect to $\{y_t, t \leq 0\}$ and $\{y_t, t \geq m\}$, respectively.

Remark 2.3 To judge the geometric ergodicity required in Assumption 2.1, we can make use of Cline and Pu (2004) (e.g., Corollary 2.2, Theorem 3.5 and Example 4.1). Part of the conditions in Assumptions 2.2-2.3 have been adopted by Chan (1990) to weaken the condition of normality for the error term.

Lemma 2.1 (Lemma 1 of Jensen and Rahbek, 2004) *Denote $L_T(\psi)$ as a function of the observations $y_1, \cdots, y_T$ and the parameter $\psi \in \Psi \subseteq R^k$. Suppose ψ_0 is an interior point of Ψ. Assume $L_T(\cdot) : R^k \to R$ is three times continuously differentiable in ψ and that*

A1: *As* $T \to \infty$, $\sqrt{T}\partial L_T(\psi_0)/\partial\psi \xrightarrow{L} N(0, \Omega_S), \Omega_S > 0.$

A2: *As* $T \to \infty, \partial^2 L_T(\psi_0)/\partial\psi\partial\psi^\tau \xrightarrow{p} \Omega_I > 0.$

A3: $\max_{i,j,k=1,\cdots,p+2} \sup_{\psi \in N(\psi_0)} \left|\partial^3 L_T(\psi)/\partial\psi_i\partial\psi_j\partial\psi_k\right| \leq c_T.$

Here $N(\psi_0)$ is a neighborhood of ψ_0 and $0 \leq c_T \xrightarrow{p} c, 0 < c < \infty$. Then there exists a fixed open neighborhood $U(\psi_0) \subseteq N(\psi_0)$ such that

B1: *As $T \to \infty$, with probability one that there exists a minimum point $\hat{\psi}_T$ of $L_T(\psi)$ in $U(\psi_0)$ and $L_T(\psi)$ is convex in $U(\psi_0)$. Moreover, $\hat{\psi}_T$ is unique and solves*

$\partial L_T(\hat{\psi}_T)/\partial\psi = 0.$

B2: As $T \to \infty, \hat{\psi}_T - \psi_0 \xrightarrow{p} 0,\ \sqrt{T}(\hat{\psi}_T - \psi_0) \xrightarrow{L} N(0, \Omega_I^{-1}\Omega_S\Omega_I^{-1}).$

Proof of Theorem 2.1:

Let

$$L_T^*(\psi) = \frac{1}{T}\sum_{t=1}^{T}[\log h_t(\psi) + \frac{\varepsilon_t^2(\psi)}{h_t(\psi)}] = \frac{1}{T}\sum_{t=1}^{T} l_t^*(\psi), \tag{2.17}$$

and it can be shown:

$$\frac{\partial l_t^*(\psi)}{\partial\psi} = \left(1 - \frac{\varepsilon_t^2(\psi)}{h_t(\psi)}\right)\frac{1}{h_t(\psi)}\frac{\partial h_t(\psi)}{\partial\psi} + \frac{2\varepsilon_t(\psi)}{h_t(\psi)}\frac{\partial\varepsilon_t(\psi)}{\partial\psi}, \tag{2.18}$$

$$\begin{aligned}\frac{\partial^2 l_t^*(\psi)}{\partial\psi\partial\psi^\tau} &= \frac{1}{h_t^2(\psi)}\left(\frac{2\varepsilon_t^2(\psi)}{h_t(\psi)} - 1\right)\frac{\partial h_t(\psi)}{\partial\psi}\frac{\partial h_t(\psi)}{\partial\psi^\tau} - \frac{2\varepsilon_t}{h_t^2(\psi)}\frac{\partial h_t(\psi)}{\partial\psi}\frac{\partial\varepsilon_t(\psi)}{\partial\psi^\tau} \\ &\quad + \frac{2}{h_t(\psi)}\frac{\partial\varepsilon_t(\psi)}{\partial\psi}\frac{\partial\varepsilon_t(\psi)}{\partial\psi^\tau} - \frac{2\varepsilon_t(\psi)}{h_t^2(\psi)}\frac{\partial\varepsilon_t(\psi)}{\partial\psi}\frac{\partial h_t(\psi)}{\partial\psi^\tau},\end{aligned} \tag{2.19}$$

and

$$\begin{aligned}\frac{\partial^3 l_t^*(\psi)}{\partial\psi_i\partial\psi_j\partial\psi_k} &= \left[2\left(1 - \frac{3\varepsilon_t^2(\psi)}{h_t(\psi)}\right)\frac{1}{h_t^3(\psi)}\frac{\partial h_t(\psi)}{\partial\psi_i}\frac{\partial h_t(\psi)}{\partial\psi_j}\frac{\partial h_t(\psi)}{\partial\psi_k}\right] \\ &\quad + \left[\frac{4\varepsilon_t(\psi)}{h_t^3(\psi)}\frac{\partial h_t(\psi)}{\partial\psi_i}\frac{\partial h_t(\psi)}{\partial\psi_j}\frac{\partial\varepsilon_t(\psi)}{\partial\psi_k} + \frac{4\varepsilon_t(\psi)}{h_t^3(\psi)}\frac{\partial h_t(\psi)}{\partial\psi_i}\frac{\partial\varepsilon_t(\psi)}{\partial\psi_j}\frac{\partial h_t(\psi)}{\partial\psi_k}\right. \\ &\quad \left. + \frac{4\varepsilon_t(\psi)}{h_t^3(\psi)}\frac{\partial\varepsilon_t(\psi)}{\partial\psi_i}\frac{\partial h_t(\psi)}{\partial\psi_j}\frac{\partial h_t(\psi)}{\partial\psi_k}\right] \\ &\quad - \left[\frac{2}{h_t^2(\psi)}\frac{\partial h_t(\psi)}{\partial\psi_i}\frac{\partial\varepsilon_t(\psi)}{\partial\psi_j}\frac{\partial\varepsilon_t(\psi)}{\partial\psi_k} + \frac{2}{h_t^2(\psi)}\frac{\partial\varepsilon_t(\psi)}{\partial\psi_i}\frac{\partial\varepsilon_t(\psi)}{\partial\psi_j}\frac{\partial h_t(\psi)}{\partial\psi_k}\right. \\ &\quad \left. + \frac{2}{h_t^2(\psi)}\frac{\partial\varepsilon_t(\psi)}{\partial\psi_i}\frac{\partial h_t(\psi)}{\partial\psi_j}\frac{\partial\varepsilon_t(\psi)}{\partial\psi_k}\right] := l_{1t} + l_{2t} + l_{3t}.\end{aligned} \tag{2.20}$$

Here, $l_{it}, i = 1, 2, 3$ mean the corresponding quantities expressed in the preceding three pairs of square brackets. To prove Theorem 2.1, we just need to verify A1-A3 described in the above Lemma 2.1.

Recall $h_t = h_t(\psi_0), \varepsilon_t = \varepsilon_t(\psi_0), \varsigma = Ee_t^4 - 1$. From the above (2.17-2.20), we know

$$\begin{aligned}\sqrt{T}\frac{\partial L_T^*(\psi_0)}{\partial\psi} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\left(1-\frac{\varepsilon_t^2}{h_t}\right)\frac{1}{h_t}\frac{\partial h_t}{\partial\psi} + \frac{2\varepsilon_t}{h_t}\frac{\partial\varepsilon_t}{\partial\psi}\right] \\ &:= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} S_t.\end{aligned}$$

Consider any non-zero vector $c = (c_1, \cdots, c_q)^\tau, q = 2p + m + 3$, we have

$$\sqrt{T}c^\tau\frac{\partial L_T^*(\psi_0)}{\partial\psi} = \sum_{t=1}^{T}\left(\frac{1}{\sqrt{T}}c^\tau S_t\right) := \sum_{t=1}^{T} W_t.$$

Given the information set up to time $t-1$, $\mathscr{F}_{t-1} = \sigma(e_{t-1}, \cdots, e_1, y_0, \cdots, y_{-m+1})$, then we know $\{W_t\}$ is a martingale difference with respect to the information set, and $E(W_t^2|\mathscr{F}_{t-1}) = c^\tau\frac{1}{T}E(S_tS_t^\tau|\mathscr{F}_{t-1})c$. It is not difficult to get

$$E(S_tS_t^\tau|\mathscr{F}_{t-1}) = \frac{\varsigma}{h_t^2}\frac{\partial h_t}{\partial\psi}\frac{\partial h_t}{\partial\psi^\tau} + \frac{4}{h_t}\frac{\partial\varepsilon_t}{\partial\psi}\frac{\partial\varepsilon_t}{\partial\psi^\tau} := \Omega_{S,t}. \tag{2.21}$$

In fact, we have

$$\begin{aligned}&E(S_tS_t^\tau|\mathscr{F}_{t-1}) \\ &= E\left\{\left[\left(1-\frac{\varepsilon_t^2}{h_t}\right)\frac{1}{h_t}\frac{\partial h_t}{\partial\psi} + \frac{2\varepsilon_t}{h_t}\frac{\partial\varepsilon_t}{\partial\psi}\right]\left[\left(1-\frac{\varepsilon_t^2}{h_t}\right)\frac{1}{h_t}\frac{\partial h_t}{\partial\psi^\tau} + \frac{2\varepsilon_t}{h_t}\frac{\partial\varepsilon_t}{\partial\psi^\tau}\right]|\mathscr{F}_{t-1}\right\} \\ &= E\left\{\left[A_t^2\frac{\partial h_t}{\partial\psi}\frac{\partial h_t}{\partial\psi^\tau} + A_tB_t\frac{\partial h_t}{\partial\psi}\frac{\partial\varepsilon_t}{\partial\psi^\tau} + B_tA_t\frac{\partial\varepsilon_t}{\partial\psi}\frac{\partial h_t}{\partial\psi^\tau} + B_t^2\frac{\partial\varepsilon_t}{\partial\psi}\frac{\partial\varepsilon_t}{\partial\psi^\tau}\right]|\mathscr{F}_{t-1}\right\}.\end{aligned}$$

In the above expressions, A_t, A_t^2, B_t, B_t^2 are given as

$$A_t = \left(1-\frac{\varepsilon_t^2}{h_t}\right)\frac{1}{h_t}, A_t^2 = \left(1-2\frac{\varepsilon_t^2}{h_t}+\frac{\varepsilon_t^4}{h_t^2}\right)\frac{1}{h_t^2}, B_t = \frac{2\varepsilon_t}{h_t}, B_t^2 = \frac{4\varepsilon_t^2}{h_t^2},$$

and we have

$$E(A_tB_t|\mathscr{F}_{t-1}) = 0, E(A_t^2|\mathscr{F}_{t-1}) = \frac{Ee_t^4-1}{h_t^2} = \frac{\varsigma}{h_t^2}, E(B_t^2|\mathscr{F}_{t-1}) = \frac{4}{h_t}.$$

Note $\partial h_t/\partial\psi, \partial\varepsilon_t/\partial\psi$ are part of $\mathscr{F}_{t-1}$, then we get (2.21). Consequently, we have

$$\sum_{t=1}^{T} E(W_t^2|\mathscr{F}_{t-1}) = c^\tau \left(\frac{1}{T}\sum_{t=1}^{T}\Omega_{S,t}\right) c \xrightarrow{p} c^\tau \Omega_S c,$$

where

$$\Omega_S = E(\Omega_{S,t}) = E\left(\frac{\varsigma}{h_t^2}\frac{\partial h_t}{\partial\psi}\frac{\partial h_t}{\partial\psi^\tau} + \frac{4}{h_t}\frac{\partial\varepsilon_t}{\partial\psi}\frac{\partial\varepsilon_t}{\partial\psi^\tau}\right). \tag{2.22}$$

Furthermore, given any $\delta > 0$, we have

$$\begin{aligned}\sum_{t=1}^{T} E\left[W_t^2 I(|W_t| \geq \delta)\right] &= \frac{1}{T}\sum_{t=1}^{T} E\left[c^\tau S_t S_t^\tau c I(|c^\tau S_t S_t^\tau c| \geq \delta^2 T)\right] \\ &= E\left[c^\tau S_1 S_1^\tau c I(|c^\tau S_1 S_1^\tau c| \geq \delta^2 T)\right] \longrightarrow 0.\end{aligned}$$

The above limit can be explained by the fact that $E\Omega_{S,t} < \infty$. By the martingale central limit theorem, see, for example, Theorem 35.12 in Billingsley (1995), we have proved that $\sum_{t=1}^{T} W_t \xrightarrow{L} N(0, c^\tau \Omega_S c)$, which means

$$\sqrt{T}\frac{\partial L_T^*(\psi_0)}{\partial\psi} \xrightarrow{L} N(0, \Omega_S), \tag{2.23}$$

namely condition A1 is satisfied.

Applying the double expectation formula we can get

$$\begin{aligned} E\left(\frac{\partial^2 l_t^*(\psi_0)}{\partial\psi\partial\psi^\tau}\right) &= E\left(\frac{1}{h_t^2}\frac{\partial h_t}{\partial\psi}\frac{\partial h_t}{\partial\psi^\tau} + \frac{2}{h_t}\frac{\partial\varepsilon_t}{\partial\psi}\frac{\partial\varepsilon_t}{\partial\psi^\tau}\right) := \Omega_I, \\ \frac{\partial^2 L_t^*(\psi_0)}{\partial\psi\partial\psi^\tau} &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 l_t^*(\psi_0)}{\partial\psi\partial\psi^\tau} \xrightarrow{p} \Omega_I, \end{aligned} \tag{2.24}$$

which means A2 holds.

We next verify condition A3. For each $\psi \in \Psi$, from (2.6) it is not difficult to show

$$\left|\frac{\partial h_t(\psi)}{\partial \psi_i}\frac{1}{h_t(\psi)}\right|, \left|\frac{\partial \varepsilon_t(\psi)}{\partial \psi_i}\frac{1}{\sqrt{h_t(\psi)}}\right|$$

are bounded by some constants that are independent of ψ. Note

$$2\left(1-\frac{3\varepsilon_t^2(\psi)}{h_t(\psi)}\right) < 6\left(1+\frac{\varepsilon_t^2(\psi)}{h_t(\psi)}\right), \text{ and } \frac{4\varepsilon_t(\psi)}{\sqrt{h_t(\psi)}} < 4(1+\frac{\varepsilon_t^2(\psi)}{h_t(\psi)}),$$

then from (2.20) we know there exist finite positive constants C_1, C_2 and C_3 satisfying the following:

$$|l_{1t}| \leq C_1(1+\frac{\varepsilon_t^2(\psi)}{h_t(\psi)}), |l_{2t}| \leq C_2(1+\frac{\varepsilon_t^2(\psi)}{h_t(\psi)}), |l_{3t}| \leq C_3.$$

Note the true value of the parameter vector is denoted as $\psi^0 = (\theta^{0^\tau}, \phi^{0^\tau}, a^{0^\tau})^\tau$. We have

$$\begin{aligned}
\varepsilon_t(\psi) &= y_t - [\theta_0 + \sum_{i=1}^p \theta_i y_{t-i} + I(y_{t-d} \leq r)(\phi_0 + \sum_{i=1}^p \phi_i y_{t-i})] \\
&= \varepsilon_t + [(\theta_0^0 - \theta_0) + \sum_{i=1}^p (\theta_i^0 - \theta_i) y_{t-i} \\
&\quad + I(y_{t-d} \leq r)((\phi_0^0 - \phi_0) + \sum_{i=1}^p (\phi_i^0 - \phi_i) y_{t-i})] \\
&:= e_t\sqrt{h_t} + [\beta_0 + \sum_{i=1}^p \beta_i y_{t-i}].
\end{aligned}$$

Recall the i.i.d (0, 1) process e_t is independent of y_{t-s} for $s > 0$. Using the formula $(x+y)^2 \leq 2(x^2+y^2)$, then we have

$$\frac{\varepsilon_t^2(\psi)}{h_t(\psi)} \leq 2e_t^2 \cdot \frac{\omega^0 + \Sigma_{i=1}^m a_i^0 y_{t-i}^2}{\omega + \Sigma_{i=1}^m a_i y_{t-i}^2} + 2\frac{(\beta_0 + \sum_{i=1}^p \beta_i y_{t-i})^2}{\omega + \Sigma_{i=1}^p a_i y_{t-i}^2}.$$

Using the Cauchy-Schwarz inequality, we have $(\sum_{i=1}^n x_i)^2 = (\sum_{i=1}^n x_i.1)^2 \leq (\sum_{i=1}^n x_i^2)(\sum_{i=1}^n 1^2)$. As a result, we have $(\beta_0 + \sum_{i=1}^p \beta_i y_{t-i})^2 \leq (p+1)(\beta_0^2 + \sum_{i=1}^p \beta_i^2 y_{t-i}^2)$.

Recall the considered parameter space Ψ is bounded, then it can be shown that

$$\frac{\omega^0 + \Sigma_{i=1}^m a_i^0 y_{t-i}^2}{\omega + \Sigma_{i=1}^m a_i y_{t-i}^2} \leq \frac{\omega^0}{\omega} + \frac{\Sigma_{i=1}^m a_i^0 y_{t-i}^2}{\Sigma_{i=1}^m a_i y_{t-i}^2} \leq \frac{\omega_U}{\omega_L} + \frac{a_U}{a_L} = O(1),$$

where, we assume $\omega_L \leq \omega \leq \omega_U, a_L \leq a \leq a_U$. Analogously,

$$\frac{(\beta_0 + \sum_{i=1}^p \beta_i y_{t-i})^2}{\omega + \Sigma_{i=1}^p a_i y_{t-i}^2} \leq (p+1)\frac{\beta_0^2 + \sum_{i=1}^p \beta_i^2 y_{t-i}^2}{\omega + \Sigma_{i=1}^p a_i y_{t-i}^2} = O(1).$$

Putting $w_t = C(1 + e_t^2)$ with C being a certain positive constant that depends only on the parameter space Ψ, then from the above discussion, we know

$$\max_{i,j,k=1,\cdots,p+2} \sup_{\psi \in N(\psi_0)} \left| \frac{\partial^3 L_T^*(Y,\psi)}{\partial\psi_i \partial\psi_j \partial\psi_k} \right| \leq \frac{1}{T}\sum_{t=1}^T w_t \xrightarrow{p} Ew_t < +\infty.$$

Proof of Theorem 2.2:

Define

$$\eta_\infty = T^{-\frac{1}{2}} \frac{\partial L_T(\psi)}{\partial\theta} = \frac{1}{\sqrt{T}} \sum_{t=1}^T \left[-\frac{\varepsilon_t}{h_t} \frac{\partial\varepsilon_t}{\partial\theta} \right],$$

and let $u_r = (\eta_\infty^\tau, \eta_r^\tau)^\tau, c^* = (c_1^\tau, c_2^\tau)^\tau$. Here, c^* is any nonzero constant vector and $c_1 = (c_{10}, \cdots, c_{1p})^\tau, c_2 = (c_{20}, \cdots, c_{2p})^\tau$. Consider

$$c^{*\tau} u_r = c_1^\tau \eta_\infty + c_2^\tau \eta_r = \sum_{t=1}^T \frac{1}{\sqrt{T}} \left[-\frac{\varepsilon_t}{h_t}(c_1^\tau \frac{\partial\varepsilon_t}{\partial\theta} + c_2^\tau \frac{\partial\varepsilon_t}{\partial\phi}) \right] = \sum_{t=1}^T U_t,$$

then we know $\{U_t\}$ is a martingale difference with respect to $\mathcal{F}_{t-1}$. Using analogous discussion to (2.23), we can show

$$u_r \xrightarrow{L} N\left\{0, \begin{pmatrix} C & L_r \\ L_r^\tau & C_r \end{pmatrix}\right\}.$$

Hence, on the condition that $\eta_\infty = 0$, we have

$$\eta_r \xrightarrow{L} N\left\{0, \left(C_r - L_r^\tau C^{-1} L_r\right)\right\}.$$

For $r \neq s$, let $u_{r,s} = (\eta_\infty^\tau, \eta_r^\tau, \eta_s^\tau)^\tau$, then it can be similarly obtained that

$$u_{r,s} \xrightarrow{L} N\left\{0, \begin{pmatrix} C & L_r & L_s \\ L_r^\tau & C_r & C_{\min(r,s)} \\ L_s^\tau & C_{\min(r,s)}^\tau & C_s \end{pmatrix}\right\},$$

and conditionally, $(\eta_r, \eta_s)'$ converges in distribution to

$$N\left\{0, \begin{pmatrix} C_r & C_{\min(r,s)} \\ C_{\min(r,s)}^\tau & C_s \end{pmatrix} - \begin{pmatrix} L_r^\tau \\ L_s^\tau \end{pmatrix} C^{-1} \begin{pmatrix} L_r & L_s \end{pmatrix}\right\}.$$

Hence, (η_r, η_s) asymptotically follows a joint normal distribution with the covariance being $C_{\min(r,s)} - L_r^\tau C^{-1} L_s$.

Let $b > 0, D_k(-\infty, \infty)(D_k[-b, b])$ denote the function spaces with each element $f : R([-b, b]) \longrightarrow R^k$ being right continuous and having left-hand limit. Equip $D_k(-\infty, \infty)(D_k[-b, b])$ with the topology of uniform convergence over compact sets. Let $C_k(-\infty, \infty)$ be the subspace of $D_k(-\infty, \infty)$ consisting of functions continuous everywhere. More details on these spaces can be found in Pollard (1984). Now, $\{\eta_r, -\infty < r < \infty\}$ lives on $D_{p+1}(-\infty, \infty)$.

Subsequently, we show that η_r converges weakly to $\{\xi_r\}$ in $D_{p+1}(-\infty, \infty)$ and each realization of $\{\xi_r\}$ belongs to $C_{p+1}(-\infty, \infty)$ almost surely. It suffices to verify the tightness of $\{\eta_r, -b \leq r \leq b\}$ componentwise. Without loss of generality, consider the last component of $\{\eta_r, -b \le r \le b\}$. It is tight if and only if

$$g_T(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\frac{\varepsilon_t}{h_t} y_{t-p} I(y_{t-d} \le r) \right]$$

is tight.

Let $-b \le s \le r \le b$ be two arbitrary numbers, M_i, K_i (i=1, 2) be constants independent of T. Then

$$g_T(r) - g_T(s) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\frac{\varepsilon_t}{h_t} y_{t-p} I(s < y_{t-d} \le r) \right].$$

For $i = 1, \cdots, p$, and $\delta = 1, 2, 3, 4$, note $d, p \le m$, and $y_{t-i}^2 / h_t = O(1)$, then we have

$$\begin{aligned}
& E\left\{ \left| \frac{\varepsilon_t}{h_t} y_{t-i} I(s < y_{t-d} \le r) \right|^{\delta} \right\} \\
\le\ & E\left\{ (1 + \frac{y_{t-i}^4}{h_t^2}) I(s < y_{t-d} \le r) E\left((1 + \frac{\varepsilon_t^4}{h_t^2}) | \mathcal{F}_{t-1} \right) \right\} \\
\le\ & M_1 E\left\{ I(s < y_{t-d} \le r) \right\} \le M_2(r - s).
\end{aligned} \tag{2.25}$$

Let $\zeta_t = \frac{1}{\sqrt{T}} \frac{\varepsilon_t}{h_t} y_{t-p} I(s < y_{t-d} \le r)$. Applying Assumption 2.3 and Lemma 3.6 in Peligrad (1982), we have

$$\begin{aligned}
E|g_T(r) - g_T(s)|^4 \ &\le\ K_1 (T^{\frac{1}{4}} \|\zeta_t\|_4 + T^{\frac{1}{2}} \|\zeta_t\|_2)^4 \\
&\le\ K_2[(r - s)/T + (r - s)^2].
\end{aligned} \tag{2.26}$$

Here $\|\cdot\|_\delta$ means the usual L^δ norm. The second line in the above inequalities follows from (2.25). For $[-b, b]$, consider a partition $\{-b = r_0 < r_1 < \cdots < r_L = b\}$ with $u > 0, r_j = r_{j-1} + u, 0 \le j \le L - 1$ and $r_L - r_{L-1} \le u$. Define

$$\kappa_{t,i} = \frac{1}{\sqrt{T}} \frac{|\varepsilon_t y_{t-p}|}{h_t} I(r_{i-1} < y_{t-d} \le r_i),$$

then, $\forall i$, for $r_{i-1} \le r \le r_i$, we have

$$|g_T(r) - g_T(s)| \le \sum_{t=1}^{T} \kappa_{t,i}. \tag{2.27}$$

Based on (2.25), it is not difficult to show

$$\sup_i \sum_{t=1}^{T} \kappa_{t,i} = uO_p(\sqrt{T}). \tag{2.28}$$

In fact, we know that

$$\begin{aligned} \Pr\left\{\frac{1}{\sqrt{T}u} sup_i \sum_{t=1}^{T} \kappa_{t,i} > \delta\right\} &\le \Pr\left\{\sum_{t=1}^{T} sup_i\, \kappa_{t,i} > \sqrt{T}u\delta\right\} \\ \le \frac{TE[sup_i\, \kappa_{t,i}]}{\sqrt{T}u\delta} &= \frac{\sqrt{T}E[sup_i\, \kappa_{t,i}]}{u\delta}. \end{aligned}$$

In the above inequalities, we use the stationarity of $\{sup_i\, \kappa_{t,i}\}_{t=1}^{T}$ (which is easily satisfied according to Assumption 2.1) and the Markov inequality. In terms of (2.25), it is easy to show that

$$\begin{aligned} M_s &:= \sqrt{T}E[sup_i\, \kappa_{t,i}] \\ &= E\left[sup_i \frac{|\varepsilon_t y_{t-p}|}{h_t} I(r_{i-1} < y_{t-d} \le r_i)\right] \\ &\le E\left[\frac{|\varepsilon_t y_{t-p}|}{h_t}\right] < \infty. \end{aligned}$$

For any given $\epsilon > 0$, put $\delta > M_s/(u\epsilon)$, then we have

$$\Pr\left\{\frac{1}{\sqrt{T}u} sup_i \sum_{t=1}^{T} \kappa_{t,i} > \delta\right\} < \epsilon,$$

which means $\frac{1}{\sqrt{T}u} sup_i \sum_{t=1}^{T} \kappa_{t,i} = O_p(1)$, namely $sup_i \sum_{t=1}^{T} \kappa_{t,i} = uO_p(\sqrt{T})$.

In terms of (2.26-2.28), by applying similar discussion used in the proof of Theorem 22.1 of Billingsley (1968), we can show the tightness of $\{g_T(r), -b \le r \le b\}$.

2.5 Summary

This chapter considers a class of threshold ARCH model by adding threshold effect in the mean equation of the DAR (p) model proposed by Ling (2007). Provided the threshold is known, the QMLE of other parameters is shown to be asymptotically normal. A LM test is proposed for testing the threshold effect and approximate methods are given to tabulate the upper percentage points of the asymptotical null distribution. From the simulation results, it is shown that the considered methods perform well. Via the empirical studies, it is seen that the proposed model has improvement over existing models for the considered data.

Chapter 3

An Alternative GARCH-M Model: Structure and Estimation

3.1 Background

GARCH-M models have been widely studied since they were proposed by Engle, et al. (1987), which can be generally described as

$$
\begin{aligned}
y_t &= F_m(h_t) + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t}, \\
e_t &\sim i.i.d(0,1), h_t = F_h(y_{t-1}, h_{t-1}),
\end{aligned} \tag{3.1}
$$

where $F_m(\cdot)$ is the conditional mean function and $F_h(\cdot,\cdot)$ is the conditional variance function. Model (3.1) includes many cases of the existing GARCH-M models by putting $F_m(h_t) = \delta \log h_t, \delta h_t, \delta \sqrt{h_t}$ and $F_h(y_{t-1}, h_{t-1}) = \omega + \alpha(y_{t-1} - F_m(h_{t-1}))^2 + \beta h_{t-1}$. It is well known in the literature that there are two difficult problems to deal with GARCH-M models. The first is under what conditions the model is geometrically ergodic and the second is whether its quasi maximum likelihood estimator (QMLE) is asymptotically normal. Fortunately, some recent works shed insights to the solutions of these problems. When $F_h(y_{t-1}, h_{t-1}) = F_h(y_{t-1} - F_m(h_{t-1}), h_{t-1})$, Meitz and Saikkonen (2008) proposed a principle to study the geometric ergodicity of (3.1) though they mainly focused on applications to the GARCH and ACD (autoregressive conditional duration) models. By assuming $F_h(y_{t-1}, h_{t-1}) = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}$, Christensen et al. (2008) listed several conditions in their Assumption A for generally specified functions $F_m(h_t)$ (except for some unknown parameters), under which the QMLE for (3.1) is asymptotically normal. Christensen et al. (2008) also gave empirical studies to demonstrate that the setting of $F_h(y_{t-1}, h_{t-1}) = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}$ is sensible for analyzing real data. Consequently, for a special case of model (3.1), we can apply the results of Meitz and Saikkonen (2008) and Christensen et al. (2008) to study the ergodicity conditions and the limiting properties of the QMLE.

To be exact, we consider the following model

$$y_t = \delta \sqrt{h_t} + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t}, \tag{3.2}$$

$$e_t \sim i.i.d(0, 1), h_t = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}, \tag{3.3}$$

where $0 < \omega, \alpha, 0 < \beta < 1$, e_t is independent of $y_s, s < t$, and it has a continuous symmetric density function on R. Denote $\theta = (\delta, \omega, \alpha, \beta)^\tau$ as the unknown parameter vector and $\theta \in \Theta$, where Θ is assumed to have the form $\Theta := \{\theta : \delta_L \leq \delta \leq \delta_U, 0 < \omega_L \leq \omega \leq \omega_U, 0 < \alpha_L \leq \alpha \leq \alpha_U, 0 < \beta_L \leq \beta \leq \beta_U < 1\}$. All throughout this chapter, the superscript τ denotes the transpose of a vector or a matrix. The above (3.2-3.3) is a special case of the model in Christensen et al. (2008), by setting $m(h_t) = \delta \sqrt{h_t}$. If y_{t-1}^2

in (3.3) is substituted by ε_{t-1}^2, then h_t becomes the usual GARCH case for which, to the best of our knowledge, few results have been available for the asymptotic normality of QMLE for (3.1). For simplicity, we focus on the case of (3.3), which enables us to study the ergodicity and the QMLE for the considered model. In fact, such a setting for the conditional variance in (3.3) is not new. Ling (2004) and Ling (2007) took advantage of such specifications for the conditional variance with a finite order and some novel results were achieved (See Remark 3.2 in Ling, 2007). Cline (2007a) also adopted an analogous GARCH process when studying the geometric ergodicity of a class of nonlinear AR-GARCH models.

The chapter is arranged as follows. In Section 3.2, we discuss the geometric ergodicity and the asymptotic normality of the QMLE for the considered model. Simulations and empirical studies are given in Section 3.3. Proofs are put in Section 3.4 and we summarize the chapter in Section 3.5.

3.2 Ergodicity and Estimation

3.2.1 Geometric ergodicity

Putting $\sigma_t = \sqrt{h_t}$, we can reformulate (3.2) and (3.3) as

$$y_t = (\delta + e_t)\sigma_t, \tag{3.4}$$

$$\sigma_t = \sqrt{\omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2}. \tag{3.5}$$

Recall e_t is an independent and identically distributed process with mean 0 and variance 1, and e_t is independent of (y_s, σ_s), $s < t$. Define $z_t = (y_t, \sigma_t)$, with y_t being Y_t and σ_t

being X_t, then we can respectively consider (3.4) and (3.5) as special cases of equations (4) and (5) in Meitz and Saikkonen (2008). According to Proposition 1 of Meitz and Saikkonen (2008), if the process σ_t is V_σ geometrically ergodic, then we have z_t to be V_z geometrically ergodic for some function V_z. Hence, it suffices to study the ergodicity of σ_t.

By simple recursion, we have

$$\sigma_t = \sqrt{\omega + [\alpha(\delta + e_{t-1})^2 + \beta]\sigma_{t-1}^2} := F(\sigma_{t-1}, e_{t-1}). \tag{3.6}$$

Hence σ_t can be viewed as a Markov chain of its own and studied in isolation from y_t. Following the notations in Cline (2007a), we can rewrite (3.6) as

$$\sigma_t = B(\sigma_{t-1}, e_t) + C(\sigma_{t-1}, e_t),$$

where,

$$\begin{aligned} B(x, e) &= \sqrt{\alpha(\delta + e)^2 + \beta}x, \\ C(x, e) &= \omega / \left\{ \sqrt{\alpha(\delta + e)^2 + \beta}x + \sqrt{\omega + [\alpha(\delta + e)^2 + \beta]x^2} \right\}. \end{aligned} \tag{3.7}$$

Obviously, $B(x, e)$ is homogeneous in x and satisfies $0 < |B(x/|x|, e)| \leq \bar{b}(1 + |e|)$ for some finite $\bar{b}$, and $|C(x, e)| = O(1) \leq \bar{c}(1 + |e|)$ for some finite $\bar{c}$. Hence, (3.6) belongs to the framework of (1.2) of Cline (2007a) and we may apply Cline's (2007a) approach to study the ergodicity of σ_t. Define a related Markov process as

$$\sigma_t^* = B(\sigma_{t-1}^*, e_t) = \sqrt{\alpha(\delta + e_t)^2 + \beta}\sigma_{t-1}^*. \tag{3.8}$$

Let $W_t^* = |B(\sigma_{t-1}^*/|\sigma_{t-1}^*|, e_t)| = |B(1, e_t)| = \sqrt{\alpha(\delta + e_t)^2 + \beta}$, then we have

$$\gamma := E(\log(W_t^*)) = \frac{1}{2}E\left\{\log[\alpha(\delta + e_t)^2 + \beta]\right\}. \tag{3.9}$$

To study the geometric ergodicity of σ_t, we further define the Lyapounov exponent as

$$\bar{\gamma} := \liminf_{T\to\infty} \limsup_{\sigma\to\infty} \frac{1}{T}E\left(\log\left(\frac{1+\sigma_T}{1+\sigma_0}\right)\middle|\, \sigma_0 = \sigma\right). \tag{3.10}$$

Then we have the following theorem.

Theorem 3.1 *For the considered Θ, suppose Assumption 3.1 in Section 3.4 holds, then $\{\sigma_t\}$ generated from (3.6) and $\{\sigma_t^*\}$ from (3.8) are ϕ-irreducible and aperiodic T chains on $(0, +\infty)$. Furthermore, $\bar{\gamma}$ is equivalently evaluated by γ and geometric ergodicity of $\{\sigma_t\}$ is implied by a negative value of γ, namely $E\left\{\log[\alpha(\delta + e_t)^2 + \beta]\right\} < 0$.*

Proof: It follows from Cline (2007a), and Cline and Pu (1999) that, when $\{\sigma_t\}$ is ϕ-irreducible and aperiodic, geometric ergodicity of $\{\sigma_t\}$ is implied by a negative value of $\bar{\gamma}$. As mentioned before (3.6) is a special case of the recursion model (1.2) of Cline (2007a). If we can show that the listed conditions A.1-A.4 in Cline (2007a) are satisfied for (3.6), then $\bar{\gamma}$ is equivalent to γ. As a result, to prove Theorem 3.1, it suffices to verify the mentioned conditions for (3.6). Referring to Section 5 of Cline (2007a), under Assumption 3.1 in Section 3.4, we can see the conditions A.1, A.2 and A.4 in Cline (2007a) are trivially satisfied for (3.6). Next we are to show that $\{\sigma_t\}$ and $\{\sigma_t^*\}$ are ϕ-irreducible and aperiodic T chains on $(0, +\infty)$, which implies A.3 of Cline (2007a) holds. We just consider the case of $\{\sigma_t^*\}$ and the conclusion for $\{\sigma_t\}$ can be acquired analogously.

Recall σ_t^*=$B(\sigma_{t-1}^*, e_t)$ and $B(\sigma, e) = \sqrt{\alpha(\delta + e)^2 + \beta}\sigma$. We have $\frac{\partial B(\sigma,e)}{\partial e} = \alpha(\delta +$

$e)\sigma/\sqrt{\alpha(\delta+e)^2+\beta}$. Suppose the continuous density function for e_t is $f(e)$ and we define the control set

$$O_e = \{e \in R : f(e) > 0\}.$$

Under Assumption 3.1 in Section 3.4, we know $O_e = R$. Let $\{u_t\} \subset O_e$ be a deterministic control sequence corresponded to $\{e_t\}$. Put $u_t = -\delta + c$ for $t = 1, \cdots, k$, where c is a small positive constant such that $\alpha_U c^2 + \beta_U < 1$ and k satisfies that $(\alpha c^2 + \beta)^k \sigma_0^2 < 1/\beta$ (note $\alpha c^2 + \beta < 1$) for some initial positive value σ_0. Then we have $\frac{\partial B}{\partial e}(\sigma_0, u_1) = \alpha c \sigma_0/\sqrt{\alpha c^2+\beta}$, which is nonzero for any positive initial value σ_0. Applying Proposition 7.1.2 of Meyn and Tweedie (1993), we show that $\{\sigma_t^*\}$ is a T-chain.

Define the control sequence as $\sigma_t^c = B(\sigma_{t-1}^c, u_t)$, and we know $(\sigma_k^c)^2 = (\alpha c^2 + \beta)^k \sigma_0^2 < 1/\beta$. Set $u_{k+1} = -\delta + \sqrt{\frac{1}{\alpha}(\frac{1}{(\sigma_k^c)^2} - \beta)}$ then we shall get $\sigma_{k+1}^c = B(\sigma_k^c, u_{k+1}) = 1$. For $t \geq k+2$, put $u_t = -\delta + \sqrt{\frac{1-\beta}{\alpha}}$ and then we shall get $\sigma_t^c = 1$ for $t \geq k+2$, which means that $\sigma^c = 1$ is a globally attracting state for $\{\sigma_t^c\}$. By using Proposition 7.2.5, Theorem 7.2.6 of Meyn and Tweedie (1993), we know $\{\sigma_t^*\}$ is ψ-irreducible. The above convergence property also shows that any circle must contain the state $\{\sigma^c\}$. From Proposition 7.3.4 of Meyn and Tweedie (1993), aperiodicity follows.

Remark 3.1 In practice, as in Cline (2007b), we can evaluate the expectation γ given in (3.9) by simulation approach after the parameters are estimated, or find the ergodic range for a certain parameter when others are fixed. When $\alpha(\delta^2+1)+\beta < 1$, by Jensen's inequality, we immediately have $\gamma < 0$.

3.2.2 Quasi maximum likelihood estimation

Recall $\theta = (\delta, \omega, \alpha, \beta)^\tau$ and $\theta \in \Theta$, which is a bounded parameter space for model (3.2-3.3). Suppose that the true parameter $\theta_0 = (\delta_0, \omega_0, \alpha_0, \beta_0)^\tau$ is an interior point of the considered parameter space Θ. We need to estimate θ based on the observations $\{y_t\}_{t=1}^T$ and initial values $y_0, y_{-1}, y_{-2}, \cdots$. Following the convention in the literature, we consider the quasi conditional log-likelihood function (apart from a constant term):

$$L_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} l_t(\theta) = \frac{1}{T}\sum_{t=1}^{T}[\log h_t(\theta) + \frac{\varepsilon_t^2(\theta)}{h_t(\theta)}], \tag{3.11}$$

where $\varepsilon_t(\theta) = y_t - \delta\sqrt{h_t(\theta)}$. For the sake of convenience, we put

$$\varsigma = Ee_t^4 - 1, h_t = h_t(\theta_0), \varepsilon_t = y_t - \delta_0\sqrt{h_t},$$

$$H_t = \left[1/(1-\beta_0), \textstyle\sum_{l=1}^{\infty}\beta_0^{l-1}y_{t-l}^2, \sum_{l=1}^{\infty}\beta_0^{l-1}h_{t-l}\right]^\tau,$$

then the following theorem holds under Assumptions 3.1-3.2 in Section 3.4.

Theorem 3.2 *For model (3.2-3.3) and the quasi log-likelihood function $L_T(\theta)$ (3.11), suppose that Assumptions 3.1-3.2 in Section 3.4 hold, then there exists a fixed open neighborhood $U(\theta_0) \subset \Theta$ such that with probability one, as $T \to \infty$, $L_T(\theta)$ has an unique minimum point $\hat{\theta}_T$ in U. Furthermore, $\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{L} N(0, \Omega_I^{-1}\Omega_S\Omega_I^{-1})$, where*

$$\Omega_S = E\begin{pmatrix} 4 & \frac{2\delta_0}{h_t}H_t^\tau \\ \frac{2\delta_0}{h_t}H_t & \frac{\varsigma+\delta_0^2}{h_t^2}H_tH_t^\tau \end{pmatrix} \text{ and } \Omega_I = E\begin{pmatrix} 2 & \frac{\delta_0}{h_t}H_t^\tau \\ \frac{\delta_0}{h_t}H_t & \frac{1+\delta_0^2/2}{h_t^2}H_tH_t^\tau \end{pmatrix}.$$

Remark 3.2 The proof of Theorem 3.2 in Section 3.4 is a generalization of Jensen and Rahbek (2004), through which, it is known that $Ey_t^2 < \infty$ is not required to guarantee the validity of the theorem. Such a result is consistent with Ling (2007). In practice, an initial value h_0 is needed for the calculation of $L_T(\theta), h_t, H_t$. The matrices Ω_I, Ω_S can be approximated by the relevant sample means after the parameters have been estimated.

3.3 Simulations and Empirical Studies

3.3.1 Simulations

This section examines the performance of the (Q)MLE through Monte Carlo experiments. We study the medians and standard deviations (SD) of the estimates. The series y_t is generated through model (3.2-3.3). Recall $\theta = (\delta, \omega, \alpha, \beta)^\tau$, then the following cases are considered

$$
\begin{aligned}
\theta &= (0.1, 0.05, 0.2, 0.5)^\tau, e_t \sim i.i.d.\, N(0,1),\\
\theta &= (1.5, 0.2, 0.1, 0.5)^\tau, e_t \sim i.i.d.\, N(0,1),\\
\theta &= (0.1, 0.05, 0.7, 0.35)^\tau, e_t \sim i.i.d.\, N(0,1),\\
\theta &= (-0.2, 0.1, 0.8, 0.2)^\tau, e_t \sim i.i.d.\, N(0,1),\\
\theta &= (0.6, 0.01, 0.1, 0.85)^\tau, e_t \sim i.i.d.\, t(10),\\
\theta &= (-1.2, 0.5, 0.15, 0.3)^\tau, e_t \sim i.i.d.\, t(6),\\
\theta &= (0.5, 1.2, 0.1, 0.6)^\tau, e_t \sim i.i.d.\, t(4),\\
\theta &= (0.05, 0.1, 0.2, 0.6)^\tau, e_t \sim i.i.d.\, t(3).
\end{aligned}
$$

Here $e_t \sim i.i.d.\, t(k)$ means e_t is the innovation series that follows the distribution $t(k)$ independently. The sample sizes are $T = 300, 600$, and the number of replications is 1000. To run the estimation, we set the initial value for the conditional variance $h_0 =$

$\mathrm{var}(y_t)$ and $\theta = (\delta, \omega, \alpha, \beta)^\tau \in [-10, 10] \times [0.0001, 10] \times [0.0001, 0.99] \times [0.0001, 0.99]$.

The results are summarized in Table 3.1, from which, we know the medians are close to the true values and the standard deviations are relatively small in most cases. Moreover, larger sample sizes witness a convergence trend (smaller SDs) for all cases. The simulation results indicate that the estimation performs well in finite samples.

3.3.2 Empirical studies

In this section, model (3.2-3.3) is applied to some real data sets. We analyze the excess return data on the CRSP value weighted indices, which include the NYSE, the AMEX and NASDAQ. Such data can be regarded as a reasonable proxy for the stock market and it was also studied by Conrad and Mammen (2008) in a different way. The riskless rate used to compute the excess returns is one-month Treasury bill rate (from Ibbotson Associates).

First, we study the monthly data from July 1926 to February 2009 (totally 992 observations). Take the excess return series $\{y_t\}_{t=1}^{992}$ for estimation and use (3.2-3.3) to fit the data. By minimizing (3.11), we get the estimates

$$\begin{aligned} y_t &= \underset{(0.0320)}{0.1244} \sqrt{h_t} + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t}, \\ h_t &= \underset{(0.3172)}{0.6206} + \underset{(0.0296)}{0.1278}\, y_{t-1}^2 + \underset{(0.0299)}{0.8553}\, h_{t-1}. \end{aligned} \tag{3.12}$$

The values in parentheses are the corresponding standard errors which are calculated based on Theorem 3.2. Simple calculation gives $\alpha(\delta^2 + 1) + \beta = 0.9851 < 1$ for (3.12). As mentioned in Remark 3.1, this implies the estimates satisfy the geometric ergodicity conditions. The Ljung-Box statistics of the standardized residuals give

Table 3.1: Medians and standard deviations of (Q)MLEs for model (3.2-3.3)

$\theta = (\delta, \omega, \alpha, \beta)^\tau$		$\hat{\delta}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$
		T=300			
$(0.1, 0.05, 0.2, 0.5)^\tau$	Median	0.1037	0.0538	0.1898	0.4539
	SD	0.0588	0.0287	0.0790	0.2016
$(1.5, 0.2, 0.1, 0.5)^\tau$	Median	1.5131	0.2113	0.0978	0.4827
	SD	0.0835	0.0811	0.0217	0.1060
$(0.1, 0.05, 0.7, 0.35)^\tau$	Median	0.1023	0.0562	0.6310	0.3433
	SD	0.0592	0.0181	0.0811	0.0746
$(-0.2, 0.1, 0.8, 0.2)^\tau$	Median	-0.2038	0.1055	0.7602	0.1927
	SD	0.0594	0.025	0.0980	0.0654
$(0.6, 0.01, 0.1, 0.85)^\tau$	Median	0.6010	0.0161	0.1007	0.8266
	SD	0.0340	0.0774	0.0416	0.1555
$(-1.2, 0.5, 0.15, 0.3)^\tau$	Median	-1.2048	0.5182	0.1459	0.2598
	SD	0.0681	0.3244	0.0499	0.1649
$(0.5, 1.2, 0.1, 0.6)^\tau$	Median	0.5028	1.4814	0.0965	0.5126
	SD	0.0373	1.0076	0.1031	0.2789
$(0.05, 0.1, 0.2, 0.6)^\tau$	Median	0.0488	0.1316	0.1952	0.4525
	SD	0.0318	0.1138	0.2340	0.2909
		T=600			
$(0.1, 0.05, 0.2, 0.5)^\tau$	Median	0.1016	0.0521	0.2029	0.4749
	SD	0.0413	0.0207	0.0579	0.1514
$(1.5, 0.2, 0.1, 0.5)^\tau$	Median	1.5014	0.2068	0.1003	0.4903
	SD	0.0623	0.0512	0.0158	0.0720
$(0.1, 0.05, 0.7, 0.35)^\tau$	Median	0.1025	0.0551	0.6384	0.3458
	SD	0.0408	0.0120	0.0549	0.0510
$(-0.2, 0.1, 0.8, 0.2)^\tau$	Median	-0.2012	0.1039	0.7693	0.1977
	SD	0.0400	0.0179	0.0654	0.0491
$(0.6, 0.01, 0.1, 0.85)^\tau$	Median	0.6008	0.0123	0.1009	0.8398
	SD	0.0181	0.0205	0.0266	0.0487
$(-1.2, 0.5, 0.15, 0.3)^\tau$	Median	-1.2005	0.5106	0.1484	0.2882
	SD	0.0266	0.1436	0.0304	0.1205
$(0.5, 1.2, 0.1, 0.6)^\tau$	Median	0.5016	1.3271	0.0991	0.5602
	SD	0.0231	0.8422	0.0832	0.2326
$(0.05, 0.1, 0.2, 0.6)^\tau$	Median	0.0493	0.1137	0.1906	0.5415
	SD	0.0237	0.0864	0.1947	0.2448

Notes: (1) Number of replications=1000; (2) Different error distributions are used.

$Q(3) = 5.8743\,(0.118)$, $Q(12) = 17.406\,(0.135)$, where the values in the parentheses are the related p-values. The Ljung-Box statistics for the squared standardized residuals show $Q(3) = 1.4144\,(0.702)$, $Q(12) = 6.5556\,(0.886)$. For comparison, we also fit the data by the traditional GARCH-M model:

$$y_t = \underset{(0.03145)}{0.1719}\sqrt{h_t} + \varepsilon_t, \varepsilon_t = e_t\sqrt{h_t},$$
$$h_t = \underset{(0.2201)}{0.7018} + \underset{(0.0197)}{0.1378}\,\varepsilon_{t-1}^2 + \underset{(0.0174)}{0.8543}\,h_{t-1}. \tag{3.13}$$

For (3.13), the Ljung-Box statistics of the standardized residuals give $Q(3) = 5.4967\,(0.139)$, $Q(12) = 17.829\,(0.121)$. The Ljung-Box statistics for the squared standardized residuals show $Q(3) = 1.9389\,(0.585)$, $Q(12) = 6.2185\,(0.905)$. From the computed values of the Ljung-Box statistics, we can see that both (3.12) and (3.13) are adequate at the 5% level [1].

For (3.12), we calculate the RMSE (root mean squared error) and the MAE (mean absolute error) for the in-sample forecasts as 5.4539, 3.8223 and the corresponding ones of (3.13) are 5.4669, 3.8185. Denote h_t^n, h_t^o to be the conditional variances calculated from (3.12) and (3.13) respectively. Correspondingly, we denote f_t^n, f_t^o the in-sample forecasting values. To give clear comparison, we plot $\{h_t^n\}_{t=600}^{990}$ (solid line) and $\{h_t^o\}_{t=600}^{990}$ (circle) in Figure 3.1, $\{f_t^n\}_{t=600}^{990}$ (solid line) and $\{f_t^o\}_{t=600}^{990}$ (circle) in Figure 3.2. From the RMSEs, MAEs, and plots in the figures, we can see that both conditional variances and forecasts generated form (3.12) and (3.13) are quite similar, though $\{f_t^n\}$ is generally a bit smaller than $\{f_t^o\}$. The above results mean that (3.12) has comparable fitting effect to that of (3.13) for the considered data, which can be insightful because a different

[1] The objective of empirical studies is to compare the performance between the considered model and the traditional one, while it should be noted that Christensen et al. (2008) has shown semiparametric GARCH-in-Mean models may be more practical when analyzing the real data.

GARCH process is applied.

Next, we apply the model (3.2-3.3) and the traditional GARCH-M model to the weekly data from 05/07/1963 to 27/02/2009 (totally 2383 observations). Similar to Chou et al. (1992), we choose the weekly data rather than the daily data to avoid the documented anomalies of day-of-the-week effects. Since 30/04/1971, for each quarter, we estimate a value for δ or the Market Price of the Risk (Merton, 1980) based on both (3.2-3.3) and the traditional model. We use the previous 400 observations to estimate the parameter and totally 165 estimators are acquired. For each estimation, we record the corresponding in-sample forecast RMSE and MAE. Let $\{\delta_i^n\}_{i=1}^{165}, \{\delta_i^o\}_{i=1}^{165}$ be the estimated δ values from (3.2-3.3) and the traditional model respectively. Accordingly, denote $\{RE_i^n\}_{i=1}^{165}, \{RE_i^o\}_{i=1}^{165}, \{ME_i^n\}_{i=1}^{165}, \{ME_i^o\}_{i=1}^{165}$ to be the respective RMSE and MAE sequences. To compare, we list the percentiles of the differences between the error sequences in Table 3.2, and plot the $\{\delta_i^n\}_{i=1}^{165}$ (solid line), $\{\delta_i^o\}_{i=1}^{165}$ (dashed line) in Figure 3.3.

Table 3.2: Percentiles of differences between error sequences.

Difference	Percentiles				
series	10%	25%	50%	75%	90%
$\{RE_i^o - RE_i^n\}$	-0.0004	-0.0000	0.0011	0.0035	0.0070
$\{ME_i^o - ME_i^n\}$	-0.0067	-0.0054	-0.0042	-0.0033	-0.0017

Based on Table 3.2, it is shown that the differences between the error sequences recorded from the two models are negligible. In terms of Figure 3.3, we can see the trajectory of $\{\delta_i^n\}$ is analogous to that of $\{\delta_i^o\}$, though the latter one is a bit higher. Consequently, similar to the results obtained from the monthly data, model (3.2-3.3) has comparable fitting performance to that of the traditional one for the considered data.

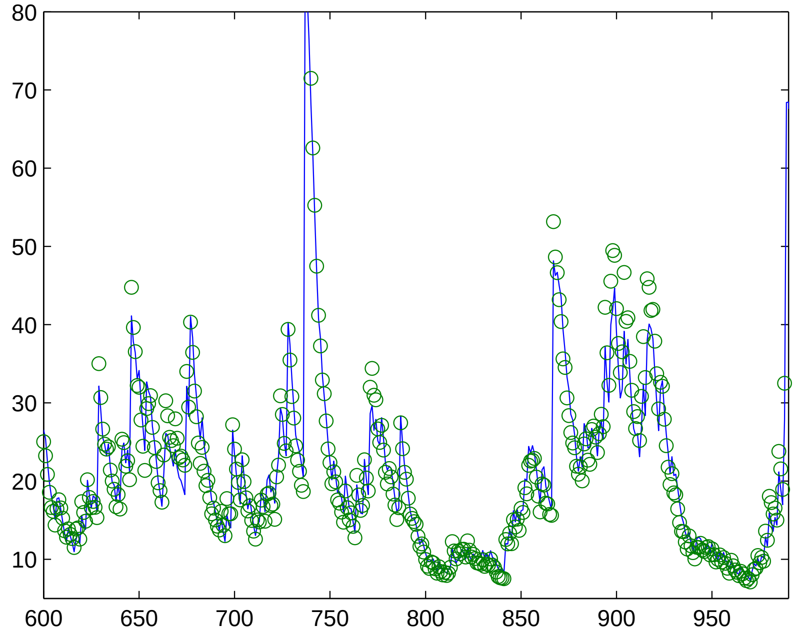


Figure 3.1: Plots of $\{h_t^n\}_{t=600}^{990}$ (solid line) and $\{h_t^o\}_{t=600}^{990}$ (circle).

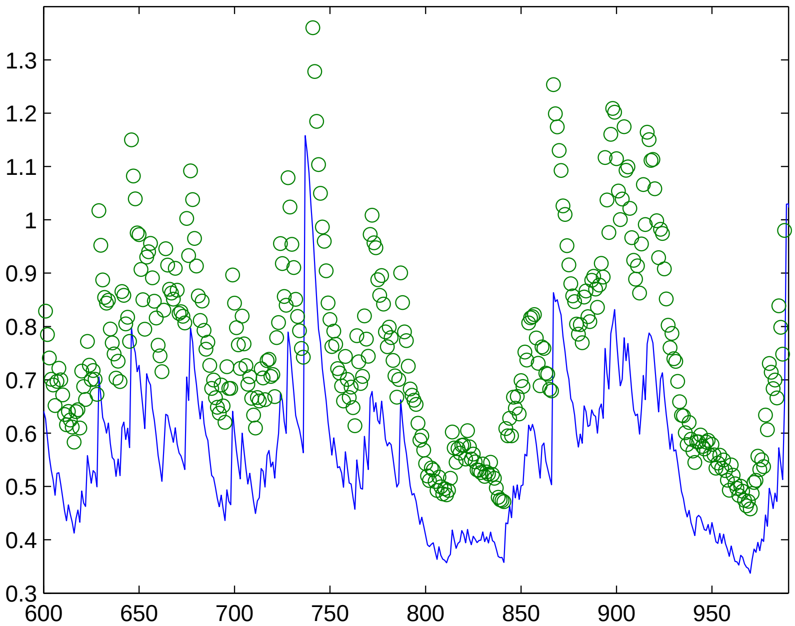


Figure 3.2: Plots of $\{f_t^n\}_{t=600}^{990}$ (solid line) and $\{f_t^o\}_{t=600}^{990}$ (circle).

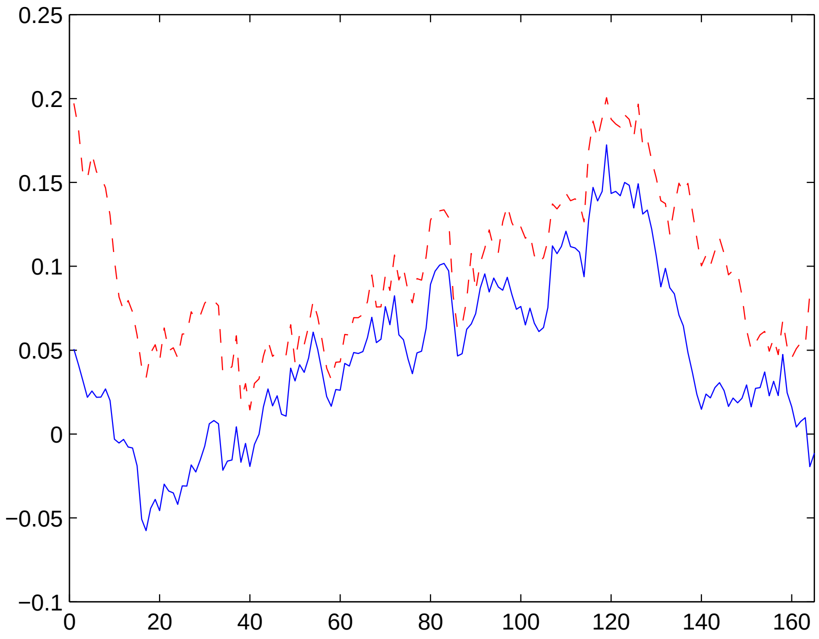


Figure 3.3: Plots of $\{\delta_i^n\}_{i=1}^{165}$ (solid line) and $\{\delta_i^o\}_{i=1}^{165}$ (dashed line).

3.4 Proofs

We make the following assumptions for model (3.2-3.3).

Assumption 3.1. The i.i.d (0, 1) process $\{e_t\}$ satisfies $Ee_t^4 < \infty$, and has a continuous symmetric probability density function which is positive everywhere.

Assumption 3.2. The series $\{y_t, h_t\}$ generated from model (3.2-3.3) are strictly stationary and ergodic for the considered parameter space Θ.

Before giving the proof for Theorem 3.2, we need to show some expressions and state several lemmas. In terms of (3.11), it is not difficult to get the derivatives of the quasi likelihood function with respect to θ:

$$\frac{\partial l_t(\theta)}{\partial \theta} = \left(1 - \frac{\varepsilon_t^2(\theta)}{h_t(\theta)}\right) \frac{1}{h_t(\theta)} \frac{\partial h_t(\theta)}{\partial \theta} + \frac{2\varepsilon_t(\theta)}{h_t(\theta)} \frac{\partial \varepsilon_t(\theta)}{\partial \theta}, \tag{3.14}$$

$$
\begin{aligned}
\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta^\tau} \;=\; & -\frac{1}{h_t^2(\theta)}\left(1-\frac{2\varepsilon_t^2(\theta)}{h_t(\theta)}\right)\frac{\partial h_t(\theta)}{\partial\theta}\frac{\partial h_t(\theta)}{\partial\theta^\tau}-\frac{2\varepsilon_t(\theta)}{h_t^2(\theta)}\frac{\partial h_t(\theta)}{\partial\theta}\frac{\partial\varepsilon_t(\theta)}{\partial\theta^\tau}\\
& +\frac{2}{h_t(\theta)}\frac{\partial\varepsilon_t(\theta)}{\partial\theta}\frac{\partial\varepsilon_t(\theta)}{\partial\theta^\tau}+\frac{2\varepsilon_t(\theta)}{h_t(\theta)}\frac{\partial^2\varepsilon_t(\theta)}{\partial\theta\partial\theta^\tau}\\
& -\frac{2\varepsilon_t(\theta)}{h_t^2(\theta)}\frac{\partial\varepsilon_t(\theta)}{\partial\theta}\frac{\partial h_t(\theta)}{\partial\theta^\tau}+\frac{1}{h_t(\theta)}\left(1-\frac{\varepsilon_t^2(\theta)}{h_t(\theta)}\right)\frac{\partial^2 h_t(\theta)}{\partial\theta\partial\theta^\tau}. \qquad (3.15)
\end{aligned}
$$

Let symbol variables s_1, s_2, s_3 take values from symbol set $\{i, j, k\}$. Then we further have $\frac{\partial^3 l_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k} = d_{1t}(\theta) + d_{2t}(\theta) + d_{3t}(\theta)$ where,

$$
\begin{aligned}
d_{1t}(\theta) \;=\; & \frac{2(h_t(\theta)-3\varepsilon_t^2(\theta))}{h_t^4(\theta)}\frac{\partial h_t(\theta)}{\partial\theta_i}\frac{\partial h_t(\theta)}{\partial\theta_j}\frac{\partial h_t(\theta)}{\partial\theta_k}\\
& +\frac{2\varepsilon_t(\theta)}{h_t^3(\theta)}\sum_{s_1\neq s_2\neq s_3}\frac{\partial h_t(\theta)}{\partial\theta_{s_1}}\frac{\partial h_t(\theta)}{\partial\theta_{s_2}}\frac{\partial\varepsilon_t(\theta)}{\partial\theta_{s_3}}\\
& -\frac{1}{h_t^2(\theta)}\sum_{s_1\neq s_2\neq s_3}\frac{\partial h_t(\theta)}{\partial\theta_{s_1}}\frac{\partial\varepsilon_t(\theta)}{\partial\theta_{s_2}}\frac{\partial\varepsilon_t(\theta)}{\partial\theta_{s_3}}, \qquad (3.16)
\end{aligned}
$$

$$
\begin{aligned}
d_{2t}(\theta) \;=\; & \left(\frac{\varepsilon_t^2(\theta)}{h_t^3(\theta)}-\frac{1}{2h_t^2(\theta)}\right)\sum_{s_1\neq s_2\neq s_3}\frac{\partial h_t(\theta)}{\partial\theta_{s_1}}\frac{\partial^2 h_t(\theta)}{\partial\theta_{s_2}\partial\theta_{s_3}}\\
& -\frac{\varepsilon_t(\theta)}{h_t^2(\theta)}\sum_{s_1\neq s_2\neq s_3}\frac{\partial h_t(\theta)}{\partial\theta_{s_1}}\frac{\partial^2\varepsilon_t(\theta)}{\partial\theta_{s_2}\partial\theta_{s_3}}\\
& -\frac{\varepsilon_t(\theta)}{h_t^2(\theta)}\sum_{s_1\neq s_2\neq s_3}\frac{\partial\varepsilon_t(\theta)}{\partial\theta_{s_1}}\frac{\partial^2 h_t(\theta)}{\partial\theta_{s_2}\partial\theta_{s_3}}\\
& +\frac{1}{h_t(\theta)}\sum_{s_1\neq s_2\neq s_3}\frac{\partial\varepsilon_t(\theta)}{\partial\theta_{s_1}}\frac{\partial^2\varepsilon_t(\theta)}{\partial\theta_{s_2}\partial\theta_{s_3}}, \qquad (3.17)
\end{aligned}
$$

and

$$
d_{3t}(\theta)=\frac{1}{h_t(\theta)}\left(1-\frac{\varepsilon_t^2(\theta)}{h_t(\theta)}\right)\frac{\partial^3 h_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}+\frac{2\varepsilon_t(\theta)}{h_t(\theta)}\frac{\partial^3\varepsilon_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}. \qquad (3.18)
$$

Note that $\varepsilon_t(\theta) = y_t - \delta\sqrt{h_t(\theta)}$, then we can show

$$
\begin{aligned}
\frac{\partial \varepsilon_t(\theta)}{\partial \theta_i} &= -\frac{\partial \delta}{\partial \theta_i}\sqrt{h_t(\theta)} - \frac{\delta}{2\sqrt{h_t(\theta)}}\frac{\partial h_t(\theta)}{\partial \theta_i}, &(3.19)\\
\frac{\partial^2 \varepsilon_t(\theta)}{\partial \theta_i \partial \theta_j} &= \frac{\delta}{4h_t(\theta)\sqrt{h_t(\theta)}}\frac{\partial h_t(\theta)}{\partial \theta_i}\frac{\partial h_t(\theta)}{\partial \theta_j}\\
&\quad -\frac{1}{2\sqrt{h_t(\theta)}}\left[\frac{\partial \delta}{\partial \theta_i}\frac{\partial h_t(\theta)}{\partial \theta_j} + \frac{\partial h_t(\theta)}{\partial \theta_i}\frac{\partial \delta}{\partial \theta_j} + \delta\frac{\partial^2 h_t(\theta)}{\partial \theta_i \partial \theta_j}\right], &(3.20)\\
\frac{\partial^3 \varepsilon_t(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} &= \frac{\delta}{8h_t(\theta)\sqrt{h_t(\theta)}}\sum_{s_1\neq s_2\neq s_3}\frac{\partial^2 h_t(\theta)}{\partial \theta_{s_1}\partial \theta_{s_2}}\frac{\partial h_t(\theta)}{\partial \theta_{s_3}}\\
&\quad +\frac{1}{8h_t(\theta)\sqrt{h_t(\theta)}}\sum_{s_1\neq s_2\neq s_3}\frac{\partial \delta}{\partial \theta_{s_1}}\frac{\partial h_t(\theta)}{\partial \theta_{s_2}}\frac{\partial h_t(\theta)}{\partial \theta_{s_3}}\\
&\quad -\frac{1}{4\sqrt{h_t(\theta)}}\sum_{s_1\neq s_2\neq s_3}\frac{\partial^2 h_t(\theta)}{\partial \theta_{s_1}\partial \theta_{s_2}}\frac{\partial \delta}{\partial \theta_{s_3}}\\
&\quad -\frac{\delta}{2\sqrt{h_t(\theta)}}\frac{\partial^3 h_t(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}\\
&\quad -\frac{3\delta}{8h_t^2(\theta)\sqrt{h_t(\theta)}}\frac{\partial h_t(\theta)}{\partial \theta_i}\frac{\partial h_t(\theta)}{\partial \theta_j}\frac{\partial h_t(\theta)}{\partial \theta_k}. &(3.21)
\end{aligned}
$$

Simple recursion gives

$$
h_t(\theta) = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}(\theta) = \omega/(1-\beta) + \alpha\sum_{l=1}^{\infty}\beta^{l-1}y_{t-l}^2, \tag{3.22}
$$

and hence

$$
\begin{aligned}
\frac{\partial h_t(\theta)}{\partial \delta} = 0,\ \frac{\partial h_t(\theta)}{\partial \omega} = \frac{1}{1-\beta},\\
\frac{\partial h_t(\theta)}{\partial \alpha} = \sum_{l=1}^{\infty}\beta^{l-1}y_{t-l}^2,\ \frac{\partial h_t(\theta)}{\partial \beta} = \sum_{l=1}^{\infty}\beta^{l-1}h_{t-l}(\theta). \qquad (3.23)
\end{aligned}
$$

Define

$$
h_t(\beta) = \frac{\omega_0}{1-\beta} + \alpha_0\sum_{l=1}^{\infty}\beta^{l-1}y_{t-l}^2, \tag{3.24}
$$

then we have the following lemma.

Lemma 3.1 *Let $h_t(\theta), h_t(\beta)$ be given as in (3.22) and (3.24). Note that $h_t(\theta_0) = h_t(\beta_0) =$*

h_t, then we have $h_t(\beta) = h_t + (\beta - \beta_0)\sum_{l=1}^{\infty}\beta^{l-1}h_{t-l}$, $h_t = h_t(\beta) + (\beta_0 - \beta)\sum_{l=1}^{\infty}\beta_0^{l-1}h_{t-l}(\beta)$ and for any t, s

$$\sup_{\theta\in\Theta}\frac{h_t(\theta)}{h_s(\theta)} \leq \kappa_1 \sup_{\beta_L\leq\beta\leq\beta_U}\frac{h_t(\beta)}{h_s(\beta)},$$
$$\sup_{\theta\in\Theta}\frac{1}{h_t(\theta)} \leq \kappa_2 \sup_{\beta_L\leq\beta\leq\beta_U}\frac{1}{h_t(\beta)},$$

where $\kappa_1 = \max\left(\frac{\alpha_U}{\alpha_0}, \frac{\omega_U}{\omega_0}\right)/\min\left(\frac{\alpha_L}{\alpha_0}, \frac{\omega_L}{\omega_0}\right)$ and $\kappa_2 = 1/\min\left(\frac{\alpha_L}{\alpha_0}, \frac{\omega_L}{\omega_0}\right)$.

Proof: Note that

$$h_t(\beta) = \omega_0 + \alpha_0 y_{t-1}^2 + \beta h_{t-1}(\beta) = h_t + \beta h_{t-1}(\beta) - \beta_0 h_{t-1},$$

and hence

$$\begin{aligned}
h_t(\beta) - h_t &= \beta[h_{t-1}(\beta) - h_{t-1}] + (\beta - \beta_0)h_{t-1} \\
&= \beta^2[h_{t-2}(\beta) - h_{t-2}] + (\beta - \beta_0)[h_{t-1} + \beta h_{t-2}] \\
&= \cdots = (\beta - \beta_0)\sum_{l=1}^{\infty}\beta^{l-1}h_{t-l},
\end{aligned}$$

namely the first equality holds. The second equality can be derived analogously. It is known

$$\begin{aligned}
h_t(\theta) &= \omega/(1-\beta) + \alpha\sum_{l=1}^{\infty}\beta^{l-1}y_{t-l}^2 \leq \omega_U/(1-\beta) + \alpha_U\sum_{l=1}^{\infty}\beta^{l-1}y_{t-l}^2 \\
&\leq \frac{\omega_U}{\omega_0}\omega_0/(1-\beta) + \frac{\alpha_U}{\alpha_0}\alpha_0\sum_{l=1}^{\infty}\beta^{l-1}y_{t-l}^2 \leq \max\left(\frac{\alpha_U}{\alpha_0}, \frac{\omega_U}{\omega_0}\right)h_t(\beta). \qquad (3.25)
\end{aligned}$$

Similarly, we can get $h_t(\theta) \geq \min\left(\frac{\alpha_L}{\alpha_0}, \frac{\omega_L}{\omega_0}\right)h_t(\beta)$, which together with (3.25) implies the last two inequalities hold.

Lemma 3.2 *Define the processes*

$$u_{mt}(a,b,c) = m\Sigma_{l=1}^{\infty} a^{l-m}\Pi_{n=1}^{m-1}(l-n)\Pi_{k=1}^{l}\frac{1}{c(\delta_0+e_{t-k})^2+b} \tag{3.26}$$

for $m = 1,\dots,4$ *(where* $\Pi_{n=1}^{0} = 1$*). Then for each* $p \geq 1$ *there exist* β_L *and* β_U *defined in* Θ *such that*

$$E[u_{mt}(\beta_0,\beta_L,\kappa_3\alpha_0)]^p < \infty, E[u_{mt}(\beta_U,\beta_0,\alpha_0)]^p < \infty. \tag{3.27}$$

where, $\kappa_3 = (1-\beta_U)/(2-\beta_0-\beta_U)$.

Proof: The lemma can be established by similar argument to the proof of Lemma 3 in Jensen and Rahbek (2004).

By (3.23), it is not difficult to get

$$h_{1t}(\theta) := \frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial\beta} = \sum_{l=1}^{\infty}\beta^{l-1}\frac{h_{t-l}(\theta)}{h_t(\theta)}, \tag{3.28}$$

$$h_{2t}(\theta) := \frac{1}{h_t(\theta)}\frac{\partial^2 h_t(\theta)}{\partial\beta^2} = 2\sum_{l=1}^{\infty}(l-1)\beta^{l-2}\frac{h_{t-l}(\theta)}{h_t(\theta)}, \tag{3.29}$$

$$h_{3t}(\theta) := \frac{1}{h_t(\theta)}\frac{\partial^3 h_t(\theta)}{\partial\beta^3} = 3\sum_{l=1}^{\infty}(l-1)(l-2)\beta^{l-3}\frac{h_{t-l}(\theta)}{h_t(\theta)}, \tag{3.30}$$

and that

$$\sup_{\theta\in\Theta} h_{it}(\theta) \leq \kappa_1 \sup_{\beta_l\leq\beta\leq\beta_U} h_{it}(\beta) \tag{3.31}$$

with κ_1 being given in Lemma 3.1. In fact we can easily get the above (3.28-3.30) by the following simple but useful equalities: Let $\{Z_k, -\infty \leq k \leq t\}$ be a sequence, then we have

$$\Sigma_{l=1}^{\infty}\beta^{l-1}\Sigma_{k=1}^{\infty}\beta^{k-1}Z_{t-l-k} = \Sigma_{l=1}^{\infty}(l-1)\beta^{l-2}Z_{t-l}, \tag{3.32}$$

$$2\Sigma_{l=1}^{\infty}(l-1)\beta^{l-2}\Sigma_{k=1}^{\infty}\beta^{k-1}Z_{t-l-k} = \Sigma_{l=1}^{\infty}(l-1)(l-2)\beta^{l-3}Z_{t-l}, \tag{3.33}$$

and

$$\begin{aligned} & 3\Sigma_{l=1}^{\infty}(l-1)(l-2)\beta^{l-3}\Sigma_{k=1}^{\infty}\beta^{k-1}Z_{t-l-k} \\ = \ & \Sigma_{l=1}^{\infty}(l-1)(l-2)(l-3)\beta^{l-4}Z_{t-l}. \end{aligned} \tag{3.34}$$

Lemma 3.3 *Let* $0 < \beta_L \le \beta$, *and* $\beta_0 \le \beta_U < 1$, *then*

$$\frac{h_t}{h_t(\beta)} \le \begin{cases} \kappa_4 & \textit{for } \beta_0 \le \beta \\ 1 + (\beta_0 - \beta_L)u_{1t}(\beta_0, \beta_L, \kappa_3\alpha_0) & \textit{for } \beta \le \beta_0 \end{cases},$$

$$h_{1t}(\beta) \le \begin{cases} \kappa_4 u_{1t}(\beta_U, \beta_0, \alpha_0) + \frac{\kappa_4}{2}(\beta_U - \beta_0)u_{2t}(\beta_U, \beta_0, \alpha_0) & \textit{for } \beta_0 \le \beta \\ u_{1t}(\beta_0, \beta_L, \kappa_3\alpha_0) & \textit{for } \beta \le \beta_0 \end{cases},$$

$$h_{2t}(\beta) \le \begin{cases} \kappa_4 u_{2t}(\beta_U, \beta_0, \alpha_0) + \frac{\kappa_4}{3}(\beta_U - \beta_0)u_{3t}(\beta_U, \beta_0, \alpha_0) & \textit{for } \beta_0 \le \beta \\ u_{2t}(\beta_0, \beta_L, \kappa_3\alpha_0) & \textit{for } \beta \le \beta_0 \end{cases},$$

$$h_{3t}(\beta) \le \begin{cases} \kappa_4 u_{3t}(\beta_U, \beta_0, \alpha_0) + \frac{\kappa_4}{4}(\beta_U - \beta_0)u_{4t}(\beta_U, \beta_0, \alpha_0) & \textit{for } \beta_0 \le \beta \\ u_{3t}(\beta_0, \beta_L, \kappa_3\alpha_0) & \textit{for } \beta \le \beta_0 \end{cases},$$

where $\kappa_4 = (2 - \beta_0 - \beta_L)/(1 - \beta_0)$, *and* κ_3 *is given in Lemma 3.2.*

Proof: When $\beta_0 \le \beta$, we know $\Sigma_{l=1}^{\infty}\beta_0^{l-1}y_{t-l}^2 \le \Sigma_{l=1}^{\infty}\beta^{l-1}y_{t-l}^2$. In terms of (3.24), we have

$$\frac{h_t}{h_t(\beta)} \le \frac{1-\beta}{1-\beta_0} + \frac{\Sigma_{l=1}^{\infty}\beta_0^{l-1}y_{t-l}^2}{\Sigma_{l=1}^{\infty}\beta^{l-1}y_{t-l}^2} \le \frac{1-\beta_L}{1-\beta_0} + 1 = \kappa_4. \tag{3.35}$$

When $\beta \le \beta_0$, we know $\Sigma_{l=1}^{\infty}\beta^{l-1}y_{t-l}^2 \le \Sigma_{l=1}^{\infty}\beta_0^{l-1}y_{t-l}^2$. Similar to (3.35), we have

$$\frac{h_t(\beta)}{h_t} \le \frac{1-\beta_0}{1-\beta} + \frac{\Sigma_{l=1}^{\infty}\beta^{l-1}y_{t-l}^2}{\Sigma_{l=1}^{\infty}\beta_0^{l-1}y_{t-l}^2} \le \frac{1-\beta_0}{1-\beta_U} + 1 = \frac{1}{\kappa_3}, \tag{3.36}$$

which implies $h_t/h_t(\beta) \ge \kappa_3$, and hence it can be shown

$$\begin{aligned}\frac{h_{t-l}(\beta)}{h_t(\beta)} &= \prod_{k=1}^{l}\frac{h_{t-k}(\beta)}{h_{t-k+1}(\beta)} = \prod_{k=1}^{l}\frac{h_{t-k}(\beta)}{\omega_0 + \alpha_0 y_{t-k}^2 + \beta h_{t-k}(\beta)} \\ &\le \prod_{k=1}^{l}\frac{1}{\alpha_0(\delta_0 + e_{t-k})^2\frac{h_{t-k}}{h_{t-k}(\beta)} + \beta} \le \prod_{k=1}^{l}\frac{1}{\kappa_3\alpha_0(\delta_0 + e_{t-k})^2 + \beta_L}. \end{aligned} \tag{3.37}$$

Further, in terms of Lemma 3.1 and (3.37), we have

$$\begin{aligned}\frac{h_t}{h_t(\beta)} &= \frac{h_t(\beta) + (\beta_0 - \beta)\sum_{l=1}^{\infty}\beta_0^{l-1}h_{t-l}(\beta)}{h_t(\beta)} \\ &\le 1 + (\beta_0 - \beta_L)\sum_{l=1}^{\infty}\beta_0^{l-1}\prod_{k=1}^{l}\frac{1}{\kappa_3\alpha_0(\delta_0 + e_{t-k})^2 + \beta_L},\end{aligned}$$

which ends the proof of the first inequality. For the other three inequalities, based on (3.32-3.34), they can be shown by analogous argument and we only give the proof for the last one. By definition

$$h_{3t}(\beta) = \frac{1}{h_t(\beta)}\frac{\partial^3 h_t(\beta)}{\partial\beta^3} = 3\sum_{l=1}^{\infty}(l-1)(l-2)\beta^{l-3}\frac{h_{t-l}(\beta)}{h_t(\beta)}.$$

When $\beta \le \beta_0$, the inequality holds by (3.37). Next, for $\beta_0 \le \beta$, Lemma 3.1 yields

$$\begin{aligned}h_{3t}(\beta) &= 3\sum_{l=1}^{\infty}(l-1)(l-2)\beta^{l-3}\frac{h_{t-l} + (\beta - \beta_0)\sum_{s=1}^{\infty}\beta^{s-1}h_{t-l-s}}{h_t(\beta)} \\ &= 3\sum_{l=1}^{\infty}(l-1)(l-2)\beta^{l-3}\frac{h_{t-l}}{h_t(\beta)} \\ &\quad + (\beta - \beta_0)\sum_{l=1}^{\infty}(l-1)(l-2)(l-3)\beta^{l-4}\frac{h_{t-l}}{h_t(\beta)}.\end{aligned}$$

The second equality can be explained by (3.34). Note that, provided $\beta_0 \leq \beta$, (3.35) gives

$$\frac{h_{t-l}}{h_t(\beta)} = \frac{h_t}{h_t(\beta)} \cdot \frac{h_{t-l}}{h_t} \leq \kappa_4 \prod_{k=1}^{l} \frac{1}{\alpha_0(\delta_0 + e_{t-k})^2 + \beta_0}, \tag{3.38}$$

and hence the last inequality follows.

Proof of Theorem 3.2:

According to Lemma 2.1, it suffices to show the conditions A1-A3 in the lemma hold. We consider condition A1 first. Recall $h_t = h_t(\theta_0), \varepsilon_t = \varepsilon_t(\theta_0), \varsigma = Ee_t^4 - 1$. From (3.14-3.15), we know

$$\sqrt{T}\frac{\partial L_T(\theta_0)}{\partial \theta} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\left(1 - \frac{\varepsilon_t^2}{h_t}\right) \frac{1}{h_t} \frac{\partial h_t}{\partial \theta} + \frac{2\varepsilon_t}{h_t} \frac{\partial \varepsilon_t}{\partial \theta} \right] := \frac{1}{\sqrt{T}} \sum_{t=1}^{T} S_t.$$

Consider any non-zero vector $c = (c_1, c_2, c_3, c_4)^\tau$, we have

$$\sqrt{T} c^\tau \frac{\partial L_T(\theta_0)}{\partial \theta} = \sum_{t=1}^{T} \left(\frac{1}{\sqrt{T}} c^\tau S_t \right) := \sum_{t=1}^{T} W_t.$$

Let $\mathscr{F}_{t-1} := \sigma(e_{t-1}, \cdots, e_1, y_0, y_{-1}, \cdots)$ be the information set up to time $t-1$, then we know $\{W_t\}$ is a martingale difference with respect to $\mathscr{F}_{t-1}$, and $E(W_t^2|\mathscr{F}_{t-1}) = c^\tau \frac{1}{T} E(S_t S_t^\tau | \mathscr{F}_{t-1}) c$. Under Assumptions 3.1-3.2, it is not difficult to get

$$E(S_t S_t^\tau | \mathscr{F}_{t-1}) = \frac{\varsigma}{h_t^2} \frac{\partial h_t}{\partial \theta} \frac{\partial h_t}{\partial \theta^\tau} + \frac{4}{h_{0t}} \frac{\partial \varepsilon_t}{\partial \theta} \frac{\partial \varepsilon_t}{\partial \theta^\tau} := \Omega_{S,t}. \tag{3.39}$$

Consequently, we have

$$\sum_{t=1}^{T} E(W_t^2 | \mathscr{F}_{t-1}) = c^\tau \left(\frac{1}{T} \sum_{t=1}^{T} \Omega_{S,t} \right) c \xrightarrow{p} c^\tau \Omega_S c,$$

where

$$\Omega_S = E(\Omega_{S,t}) = E\left(\frac{\varsigma}{h_t^2}\frac{\partial h_t}{\partial\theta}\frac{\partial h_t}{\partial\theta^\tau} + \frac{4}{h_t}\frac{\partial\varepsilon_t}{\partial\theta}\frac{\partial\varepsilon_t}{\partial\theta^\tau}\right). \tag{3.40}$$

Furthermore, given any $\delta > 0$, we have

$$\begin{aligned}\sum_{t=1}^{T} E\left[W_t^2 I(|W_t| \geq \delta)\right] &= \frac{1}{T}\sum_{t=1}^{T} E\left[c^\tau S_t S_t^\tau c I(|c^\tau S_t S_t^\tau c| \geq \delta^2 T)\right] \\ &= E\left[c^\tau S_1 S_1^\tau c I(|c^\tau S_1 S_1^\tau c| \geq \delta^2 T)\right] \longrightarrow 0.\end{aligned}$$

The above limit can be explained by the fact: $E\left[S_1 S_1^\tau\right] = E\Omega_{S,1} < \infty$. By the martingale central limit theorem, see, for example, Theorem 35.12 in Billingsley (1995) we deduce that $\sum_{t=1}^{T} W_t \xrightarrow{L} N(0, c^\tau\Omega_S c)$, which means

$$\sqrt{T}\frac{\partial L_T(\theta_0)}{\partial\theta} \xrightarrow{L} N(0, \Omega_S). \tag{3.41}$$

Applying the double expectation formula we shall get

$$\begin{aligned} E\left(\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta^\tau}\right) &= E\left(\frac{1}{h_t^2}\frac{\partial h_t}{\partial\theta}\frac{\partial h_t}{\partial\theta^\tau} + \frac{2}{h_t}\frac{\partial\varepsilon_t}{\partial\theta}\frac{\partial\varepsilon_t}{\partial\theta^\tau}\right) := \Omega_I, \\ \frac{\partial^2 L_t(\theta_0)}{\partial\theta\partial\theta^\tau} &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta^\tau} \xrightarrow{p} \Omega_I, \end{aligned} \tag{3.42}$$

which means A2 holds.

For condition A3, we just show $\sup_{\theta\in\Theta}\left|\partial^3 L_T(\theta)/\partial\beta^3\right|$ is controlled by a positive ergodic sequence that has desired moments. Other cases can be easily proved by noting the fact that $\partial^i h_t(\theta)/\partial\delta^i = 0$, and $\partial^j h_t(\theta)/\partial\omega^j = \partial^j h_t(\theta)/\partial\alpha^j = 0$ for $i = 1, 2, 3$, $j = 2, 3$.

Moreover,

$$\begin{aligned}
\frac{1}{h_t}\frac{\partial h_t}{\partial \omega} &= \frac{1}{h_t(1-\beta)} \leq \frac{1}{\omega_L(1-\beta_U)} < \infty, \text{and}\\
\frac{1}{h_t}\frac{\partial h_t}{\partial \alpha} &= \frac{\sum_{l=1}^{\infty}\beta^{l-1}y_{t-l}^2}{\omega(\frac{1}{1-\beta}) + \alpha(\Sigma_{l=1}^{\infty}\beta^{l-1}y_{t-l}^2)} < \frac{1}{\alpha_L} < \infty,
\end{aligned}$$

which can be derived from (3.23). Based on (3.19-3.21), we have

$$\begin{aligned}
m_{1t}(\theta) &:= \frac{1}{\sqrt{h_t(\theta)}}\frac{\partial \varepsilon_t(\theta)}{\partial \beta} = -\frac{\delta}{2}h_{1t}(\theta), \qquad (3.43)\\
m_{2t}(\theta) &:= \frac{1}{\sqrt{h_t(\theta)}}\frac{\partial^2 \varepsilon_t(\theta)}{\partial \beta^2} = \frac{\delta}{4}h_{1t}^2(\theta) - \frac{\delta}{2}h_{2t}(\theta), \text{and} \qquad (3.44)\\
m_{3t}(\theta) &:= \frac{1}{\sqrt{h_t(\theta)}}\frac{\partial^3 \varepsilon_t(\theta)}{\partial \beta^3} = \frac{3\delta}{4}h_{1t}(\theta)h_{2t}(\theta)\\
&\quad -\frac{\delta}{2}h_{3t}(\theta) - \frac{3\delta}{8}h_{1t}^3(\theta). \qquad (3.45)
\end{aligned}$$

Then according to (3.16-3.18), it can be calculated that

$$\begin{aligned}
\frac{\partial^3 l_t(\theta)}{\partial \beta^3} &= 2\left(1 - \frac{3\varepsilon_t^2(\theta)}{h_t(\theta)}\right)h_{1t}^3(\theta) + \frac{12\varepsilon_t(\theta)}{\sqrt{h_t(\theta)}}h_{1t}^2(\theta)m_{1t}(\theta) - 6h_{1t}(\theta)m_{1t}^2(\theta)\\
&\quad + \left(\frac{6\varepsilon_t^2(\theta)}{h_t(\theta)} - 3\right)h_{1t}(\theta)h_{2t}(\theta) - \frac{6\varepsilon_t(\theta)}{\sqrt{h_t(\theta)}}h_{1t}(\theta)m_{2t}(\theta)\\
&\quad - \frac{6\varepsilon_t(\theta)}{\sqrt{h_t(\theta)}}h_{2t}(\theta)m_{1t}(\theta) + 6m_{1t}(\theta)m_{2t}(\theta)\\
&\quad + \left(1 - \frac{\varepsilon_t^2(\theta)}{h_t(\theta)}\right)h_{3t}(\theta) + \frac{2\varepsilon_t(\theta)}{\sqrt{h_t(\theta)}}m_{3t}(\theta). \qquad (3.46)
\end{aligned}$$

We also have

$$\begin{aligned}
\frac{\varepsilon_t^2(\theta)}{h_t(\theta)} &= \frac{[y_t - \delta\sqrt{h_t(\theta)}]^2}{h_t(\theta)} \leq 2\left(\frac{y_t^2}{h_t(\theta)} + \delta^2\right)\\
&= 2\left[(\delta_0 + e_t)^2\frac{h_t}{h_t(\theta)} + \delta^2\right]\\
&\leq 2\left[\kappa_2(\delta_0 + e_t)^2\frac{h_t}{h_t(\beta)} + \delta_U^2\right] := V_t(\beta). \qquad (3.47)
\end{aligned}$$

Based on Lemma 3.1 and Lemma 3.3, it can be shown that

$$V_t(\beta) \leq \begin{cases} 2\left[\kappa_2\kappa_4(\delta_0 + e_t)^2 + \delta_U^2\right] & \text{for } \beta_0 \leq \beta \\ 2\left\{\kappa_2(\delta_0 + e_t)^2\left[1 + (\beta_0 - \beta_L)u_{1t}(\beta_0, \beta_L, \kappa_3\alpha_0)\right] + \delta_U^2\right\} & \text{for } \beta \leq \beta_0 \end{cases} . \tag{3.48}$$

Note $\varepsilon_t(\theta)/\sqrt{h_t(\theta)} \leq \varepsilon_t^2(\theta)/h_t(\theta) + 1$. In terms of Lemma 3.1 and (3.43-3.46), then there exists a constant K such that

$$\begin{aligned} \left|\frac{\partial^3 l_t(\theta)}{\partial\beta^3}\right| \leq \ & K(V_t(\beta) + 1)\left[h_{1t}^3(\beta) + h_{1t}^2(\beta)m_{1t}^*(\beta) + h_{1t}(\beta)m_{1t}^{*2}(\beta)\right. \\ & + h_{1t}(\beta)h_{2t}(\beta) + h_{1t}(\beta)m_{2t}^*(\beta) + h_{2t}(\beta)m_{1t}^*(\beta) \\ & \left.+ m_{1t}^*(\beta)m_{2t}^*(\beta) + h_{3t}(\beta) + m_{3t}^*(\beta)\right] := w_t(\beta), \end{aligned} \tag{3.49}$$

where $m_{1t}^*(\beta) = \delta_U\kappa_1 h_{1t}(\beta), m_{2t}^*(\beta) = \delta_U\kappa_1[h_{1t}^2(\beta) + h_{2t}(\beta)]$ and

$$m_{3t}^*(\beta) = 2\delta_U\kappa_1[h_{1t}(\beta)h_{2t}(\beta) + h_{3t}(\beta) + h_{1t}^3(\beta)].$$

From Lemma 3.2, Lemma 3.3 and (3.48), we know $w_t(\beta)$ in (3.49) is bounded by some positive ergodic w_t that has desired moments. Hence we have shown A3 holds for the case of $\partial^3 l_t(\theta)/\partial\beta^3$. Other situations can be proved by similar argument, which ends the proof of Theorem 3.2.

3.5 Summary

In this chapter, we study a special case of the GARCH-M type model proposed by Christensen et al. (2008). Ergodicity conditions are discussed, and by checking the conditions listed in Lemma 1 of Jensen and Rahbek (2004), we show that the QMLE of the model is asymptotically normal. Through simulations and empirical studies, it is seen that the estimation performs well and the considered GARCH-M model has comparable performance in data modeling as compared to the traditional one. Our results indicate that the model of Christensen et al. (2008) can be useful because it gives an alternative way to study the GARCH-in-Mean effect.

Chapter 4

Semiparametric (G)ARCH-M Models

4.1 Background

The relationship between the risk (conditional volatility) and return (conditional mean) is undoubtedly an important topic in finance and many researchers have paid efforts to investigate it. Among the preceding results, the ARCH-M model proposed by Engle, et al. (1987) plays an important role for describing such a relationship. Let y_t denote the excess return of a stock market and h_t denote the conditional volatility at time t. Define $\mathscr{F}_{t-1}$ as the information set up to time $t-1$, then three usual forms are $E(y_t|\mathscr{F}_{t-1}) = \delta h_t, E(y_t|\mathscr{F}_{t-1}) = \delta \log h_t$ and $E(y_t|\mathscr{F}_{t-1}) = \delta\sqrt{h_t}$. Some other generalized forms can be found in Das and Sarkar (2000). Based on the above forms, many empirical studies have been done to analyze real data such as Chou (1988), Chou et al. (1992) and Fama and French (1989). Mixed results about the coefficient δ were obtained (Chou et al., 1992). Such a phenomenon suggests that it can be restrictive to assume the relationship as some parametric form. Recently, Christensen et al. (2008) have proposed a class of semiparametric GARCH-M models to study the conditional volatility and conditional mean. Their model is given by

$$y_t = m(h_t) + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t},$$

$$e_t \sim i.i.d(0, 1), h_t = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}. \tag{4.1}$$

In the above model, the conditional mean (expected excess return) is formulated as an unknown smooth function $m(\cdot)$. Such a description is more flexible compared to that of the usual parametric models. Another novelty of model (4.1) lies in that an adjusted GARCH (1, 1) process is adopted. The specified conditional volatility h_t has a nice property that: with known parameters ω, α and β, h_t is determined by the observable $\{y_s\}_{s=-\infty}^{t-1}$.

Motivated by model (4.1), we are interested in the following two aspects. Firstly it is what the ARCH-in-mean case of (4.1) looks like. Because, on some occasions, considering the finite ordered memory is already adequate. Secondly it is how to deal with (4.1) when we consider a general form of the conditional volatility h_t driven by the observed $\{y_s\}_{s=-\infty}^{t-1}$. Such a generalization enables us to take asymmetric effect into account when describing the conditional volatility (e.g., $h_t = \omega + \alpha[1 + \eta I(y_{t-1} \leq 0)]y_{t-1}^2 + \beta h_{t-1}$). In this chapter, we give some results with respect to these two aspects.

Section 4.2 studies a semiparametric ARCH-M model and the associated estimation method is discussed. In Section 4.3, we consider a generalized semiparametric GARCH-M model from (4.1) and some related issues are investigated. Empirical studies based on the considered two models are displayed in Section 4.4 and a summary is given in Section 4.5.

4.2 A Semiparametric ARCH-M Model

4.2.1 Model and estimation

In this section, for ease of exposition, we study the following semiparametric ARCH-M model

$$y_t = m(h_t) + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t},$$

$$e_t \sim i.i.d(0,1), h_t = \theta_0 + \theta_1 y_{t-1}^2 + \cdots + \theta_L y_{t-L}^2, \tag{4.2}$$

where, $m(\cdot)$ is an unknown smooth function, $\theta_0 > 0, 1 > \theta_1 \geq \theta_2 \geq \cdots \geq \theta_L \geq 0$. Model (4.2) can be considered as a ARCH-in-mean case of model (4.1). We can expect long memory exists if $\theta_L \neq 0$ when L is large. Denote $\theta = (\theta_0, \theta_1, \cdots, \theta_L)^\tau, W_t = (1, y_{t-1}^2, \cdots, y_{t-L}^2)$. All throughout this chapter, the superscript τ denotes the transpose of a vector or a matrix. Then we can rewrite (4.2) as

$$y_t = m(W_t\theta) + \varepsilon_t, \varepsilon_t = e_t \sqrt{W_t\theta}, e_t \sim i.i.d(0,1), \tag{4.3}$$

which is similar to the constrained single-index model studied by Xia and Tong (2006). We can thus apply Xia and Tong's (2006) idea to estimate model (4.3) with some adjustments. In the literature of single-index model, the usual way is to restrict $\| \theta \| = 1$ or $\Sigma_{l=0}^L \theta_l = 1$ for identifiability. Before giving a proper adjustment for identifiability condition for model (4.2), we suppose $\Sigma_{l=0}^L \theta_l = 1$ and give the estimation procedure first.

Following Xia and Tong (2006), we have local linear expansion of $m(W_t\theta)$ at point

w

$$m(W_t\theta) = a_w + b_w(W_t - w)\theta + O\{|(W_t - w)\theta|^2\},$$

where $a_w = m(w\theta), b_w = m'(w\theta)$. Let $K(v)$ be a symmetric kernel function, h be a bandwidth, $K_h(v) = h^{-1}K(v/h)$ and $W_{ij} = W_i - W_j$. By the principle of local linear smoother, we estimate a_w and b_w by minimizing

$$n^{-1}\sum_{i=1}^{n}\{Y_i - a_w - b_w(W_t - w)\theta\}^2 K_h((W_t - w)\theta).$$

Let $a_i = m(W_i\theta), b_i = m'(W_i\theta), i = 1, \cdots, n$, and $\alpha = (a_1, \cdots, a_n, b_1, \cdots, b_n)^\tau$. Following Xia and Tong (2006), the best approximation of θ should minimize the overall departure for all $w = W_j$, $j = 1, \cdots, n$. Thus we can estimate α and θ by

$$\min_{\alpha,\theta} n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{n}\{Y_i - a_j - b_j W_{ij}\theta\}^2 K_h(W_{ij}\theta) \tag{4.4}$$

subject to the constraints: $1 > \theta_0 \geq \theta_1 \geq \cdots \geq \theta_L \geq 0$. The minimization problem in (4.4) can be solved by quadratic programming. Let

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}_{n\times 1}, \quad W = \begin{pmatrix} W_1 \\ \vdots \\ W_n \end{pmatrix}_{n\times(L+1)}, \quad \mathbf{1}_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{n\times 1},$$

$$U_{i,\theta} = \begin{pmatrix} W_{i1}\theta & & \\ & \ddots & \\ & & W_{in}\theta \end{pmatrix}_{n\times n} \quad \text{and } U_\theta = \begin{pmatrix} U_{1,\theta} \\ \vdots \\ U_{n,\theta} \end{pmatrix}_{n^2\times n}.$$

Further, $\mathbb{Y} = Y \otimes \mathbf{1}_n$, $I_n^{\otimes} = \mathbf{1}_n \otimes I_n$ and

$$W_{\theta,h} = \operatorname{diag}\{K_h(W_{11}\theta), \cdots, K_h(W_{1n}\theta), \cdots, K_h(W_{n1}\theta), \cdots, K_h(W_{nn}\theta)\}.$$

Here I_n is the $n \times n$ unit matrix and $\otimes$ denotes the Kronecker product. Following Xia and Tong (2006), the minimization of (4.4) can be achieved by the following algorithm.

Step A: Given an initial estimate $\tilde{\theta}$ of θ (we can set the initial estimate for θ as $\tilde{\theta} = (L, L-1, \cdots, 1, 0)/\sum_{l=0}^{L} l.$), then the minimization of (4.4) becomes

$$\min_{\alpha}\{\mathbb{Y} - (I_n^{\otimes}\vdots U_{\tilde{\theta}})\alpha\}^{\tau} W_{\tilde{\theta},h}\{\mathbb{Y} - (I_n^{\otimes}\vdots U_{\tilde{\theta}})\alpha\}. \tag{4.5}$$

Let

$$C_{\tilde{\theta},h} = (I_n^{\otimes}\vdots U_{\tilde{\theta}})^{\tau} W_{\tilde{\theta},h}\mathbb{Y},$$

$$D_{\tilde{\theta},h} = (I_n^{\otimes}\vdots U_{\tilde{\theta}})^{\tau} W_{\tilde{\theta},h}(I_n^{\otimes}\vdots U_{\tilde{\theta}}).$$

Minimization of (4.5) is equivalent to

$$\min_{\alpha}\{\alpha^{\tau} D_{\tilde{\theta},h}\alpha - 2C_{\tilde{\theta},h}^{\tau}\alpha\}. \tag{4.6}$$

Since $D_{\tilde{\theta},h}$ and $C_{\tilde{\theta},h}$ are independent of α, minimization of (4.6) is a typical quadratic programming problem for α.

Step B: Put $\gamma = (\theta_0 - \theta_1, \theta_1 - \theta_2, \cdots, \theta_{L-1} - \theta_L, \theta_L)^{\tau}$. Let B be such that $B\gamma = \theta$. Define

$$Q_{\tilde{\theta},h} = B^{\tau}\sum_{j=1}^{n} b_j^2 \sum_{i=1}^{n} K_h(W_{ij}\tilde{\theta})W_{ij}^{\tau}W_{ij}B,$$

$$P_{\tilde{\theta},h} = B^{\tau}\sum_{j=1}^{n} b_j \sum_{i=1}^{n} K_h(W_{ij}\tilde{\theta})W_{ij}^{\tau}(Y_i - a_j).$$

Given α, minimization of (4.4) is equivalent to

$$\min_{\gamma}\{\gamma^{\tau}Q_{\tilde{\theta},h}\gamma - 2P^{\tau}_{\tilde{\theta},h}\gamma\} \tag{4.7}$$

subject to $\gamma > 0$. Denote $\tilde{\gamma}$ to be the solution of the above quadratic programming. Let $\tilde{\theta} = B\tilde{\gamma}, \tilde{\theta} = \tilde{\theta}/\mathrm{sum}(\tilde{\theta})$.

Repeat steps A and B until convergence. Denote the final estimates for θ and $m(v)$ by $\hat{\theta}$, $\hat{m}(v)$.

4.2.2 Some adjustments

Some adjustments are needed for (4.2) before Xia and Tong's (2006) method may be adopted. Recall $W_t = (1, y^2_{t-1}, \cdots, y^2_{t-L})$, whose first element is a fixed constant rather than some random variable. Hence the first element of W_{ij} is zero. Incorporated with the characteristic of the matrix B, it is easy to see that either the first row or the first column of $Q_{\tilde{\theta},h}$ is composed of zeros and the first element of $P_{\tilde{\theta},h}$ is also zero. Consequently, no matter what value the first element in γ takes, it does not affect the value of $\{\gamma^{\tau}Q_{\tilde{\theta},h}\gamma - 2P^{\tau}_{\tilde{\theta},h}\gamma\}$ in (4.7). The above property makes it hard to directly estimate $\gamma(1)$, namely $\theta_0 - \theta_1$, from (4.4).

To avoid the above difficulty in estimation, a possible approach we can apply is to assume that θ_0 in (4.2) is known. As a result, we only need to estimate $\theta_1, \cdots, \theta_L$. With an abuse of notation, we still put $\theta = (\theta_1, \cdots, \theta_L)^{\tau}$. To make the model identifiable, we can assume $\sum_{l=1}^{L}\theta_l = \lambda, 0 < \lambda \leq 1$. When $\lambda < 1$ and the function $m(h)$ is $O(1)$ or $o(h)$, according to Theorem 1 in Lu (1998), we know that the series generated from model (4.2) is geometrically ergodic. Hence it is sensible to consider the identifiability condition like $\sum_{l=1}^{L} \theta_l = \lambda, 0 < \lambda < 1$, because for time series models, ergodicity is usually a necessary condition for the study of the limiting properties of parametric estimators.

Based on the assumption that θ_0 is known, we can estimate model (4.2) with slight revision of the method given in Section 4.2.1. To tackle (4.6), we add the known θ_0 into θ to get the estimation for α. To deal with (4.7), we delete the first row and first column of $Q_{\tilde{\theta},h}$ and the first element of $P_{\tilde{\theta},h}$ to estimate θ. The original $\tilde{\theta} = \tilde{\theta}/\text{sum}(\tilde{\theta})$ should be changed to $\tilde{\theta} = \lambda\tilde{\theta}/\text{sum}(\tilde{\theta})$.

4.2.3 Simulations

In this section, simulation experiments are conducted to show that the proposed method works satisfactorily. Because the quadratic programming in (4.7) is relatively time-consuming when the sample size or the lag L is large, hence we just give two examples. Some improvement (faster method) may be achieved by adopting other method in approximating $m(\cdot)$ (e.g., spline method of Wang and Yang, 2009). The data is generated from the model

$$y_t = m(h_t) + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t},$$

$$e_t \sim i.i.N(0,1), h_t = \theta_0 + \theta_1 y_{t-1}^2 + \cdots + \theta_L y_{t-L}^2.$$

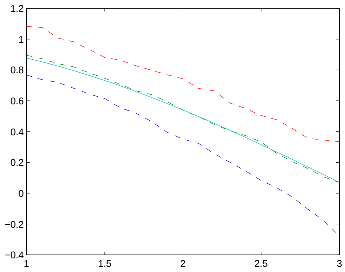
Ex 4.1: $\theta_0 = 0.18, L = 20, \theta_i = 1/(0.5i + 1), \theta_i = 0.9 * \theta_i/\text{sum}(\theta_i), i = 1, \cdots, 20$ and $m(v) = cos(1.5v)$.

Ex 4.2: $\theta_0 = 0.03, L = 100, \theta_i = 1/(0.5i + 1), \theta_i = 0.9 * \theta_i/\text{sum}(\theta_i), i = 1, \cdots, 100$ and $m(v) = cos(0.5v)$.

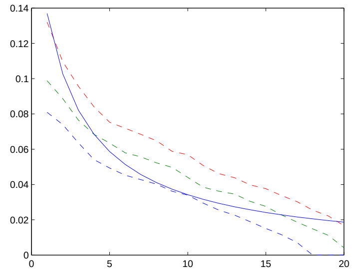
In the above two examples, we set λ in the preceding section as $\lambda = 0.9$. When applying steps A and B to estimate Ex 4.1-4.2, the kernel function $K(\cdot)$ in (4.4) is chosen as $K(x) = 0.75(1 - x^2)I(|x| \leq 1)$. For the choice of the bandwidth h, we follow Xia and Tong (2006) by taking $h = c_n n^{-1/5}$, where $c_n = 1.06 \cdot \mathrm{std}(W_t \theta)$. The convergence criterion is set as $||\hat{\theta}^{(i+1)} - \hat{\theta}^{(i)}|| \leq 0.001$.

We conduct 100 replications with sample size n=500 for Ex 4.1-4.2. The estimated results are shown in Array 4.1. In each subplot of Array 4.1, the three dashed lines are the 10%, 50% and 90% percentile lines obtained from 100 replications and the solid lines are the plots of true values of $m(\cdot)$ or θ. From Array 4.1, it is seen that the true values of θ and $m(\cdot)$ are close to the corresponding medians (Q50) and lie within the interval of $[Q10, Q90]$ for most cases. The results suggest that the proposed estimators for either the parametric part or functional part are satisfactory.

100 simulations with $n = 500$ and $L = 20$ for Ex 4.1.

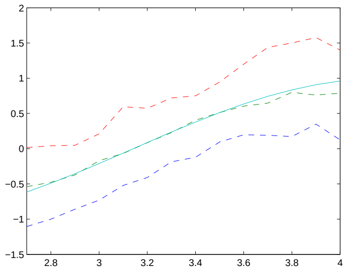


$\hat{m}(\cdot)$

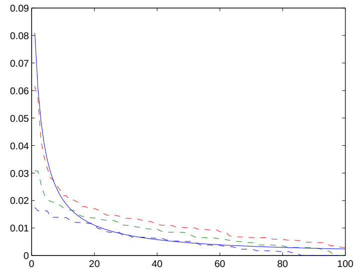


$\hat{\theta}$

100 simulations with $n = 500$ and $L = 100$ for Ex 4.2.



$\hat{m}(\cdot)$



$\hat{\theta}$

Array 4.1: Results of the simulation experiments for Ex 4.1-4.2.

4.3 A Semiparametric GARCH-M Model

4.3.1 Model and estimation

Motivated by Yang (2006) and Engle and Ng (1993), in this section we study a general case of (4.1) which has the form

$$y_t = m(h_t) + \varepsilon_t, \varepsilon_t = e_t\sqrt{h_t},$$

$$e_t \sim i.i.d(0,1), h_t = v(y_{t-1}, \phi) + \beta h_{t-1}, \tag{4.8}$$

where $m(\cdot)$ is an unknown smooth function, $v(y, \phi)$ is a known function (except for the unknown parameters) and it can be called the "news impact curve" as in Engle and Ng (1993). When $\phi = (\omega, \alpha)^\tau, v(y, \phi) = \omega + \alpha y^2$, (4.8) is reduced to (4.1). When $\phi = (\omega, \alpha, \eta)^\tau, v(y, \phi) = \omega + \alpha[1 + \eta I(y \leq 0)]y^2$, asymmetric factor for the conditional volatility is considered. For model (4.8), there are two unknown parts: $m(\cdot)$ and $\theta = (\phi^\tau, \beta)^\tau$. We describe the estimation in two parts. The first one is estimating the functions when the parameters are known and the second is to estimate the parameters consistently.

Suppose that the true values of the parameters ϕ, β are known, say ϕ_0, β_0, then

$$h_t = v(y_{t-1}, \phi_0) + \beta_0 h_{t-1} = \sum_{l=1}^{\infty} \beta_0^{l-1} v(y_{t-l}, \phi_0) \tag{4.9}$$

can be considered as an observable quantity provided that y_0, h_0 are given. We shall have

$$E(y_t|h_t = h) = m(h), \mathrm{var}\,(y_t|h_t = h) = h. \tag{4.10}$$

Equation (4.10) gives a basis for the following estimation procedure. For any fixed $h \in A$, a set as described in Assumption 4.3 in Section 4.3.3, define the estimator

$$\hat{m}^{(\lambda)}(h) = \lambda! b^{-\lambda} E_\lambda^\tau (Z^\tau W Z)^{-1} Z^\tau W Y, 0 \leq \lambda \leq p \tag{4.11}$$

where

$$Z = \begin{pmatrix} 1 & \frac{h_1 - h}{b} & \cdots & (\frac{h_1 - h}{b})^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \frac{h_n - h}{b} & \cdots & (\frac{h_n - h}{b})^p \end{pmatrix}, W = \mathrm{diag} \begin{pmatrix} \frac{1}{n} K_b(h_1 - h) \\ \vdots \\ \frac{1}{n} K_b(h_n - h) \end{pmatrix}, \text{ and}$$

$Y = (y_1, \cdots, y_n)^\tau$. In addition, all elements of the $(p+1)$ vector E_λ are zeros except the $(\lambda+1)$th element is 1, $p > 0$ is an odd integer, $b > 0$ is a bandwidth varying with the sample size n, and $K_b(\cdot) = \frac{1}{b}K(\frac{\cdot}{b})$ with $K(\cdot)$ being a compactly supported and symmetric kernel function. From Yang (2006) or Fan and Gijbels (1996), under Assumptions 4.1-4.4 in Section 4.3.3, we know (4.11) behaves like the standard univariate local polynomial estimator. Let $||K||_2^2 = \int K^2(h)dh$ and $K_\lambda^*(h)$ be defined as in (4.22), then we have the following theorem. The proof is quite standard and is omitted.

Theorem 4.1 *Under Assumptions 4.1-4.4 in Section 4.3.3, for any fixed $h \in A$ and $\lambda \geq 0$ such that $p - \lambda$ is odd, when $nb^{2\lambda+1} \to \infty, nb^{2p+3} = O(1)$, the estimator $\hat{m}^{(\lambda)}(h)$ in (4.11) satisfies*

$$\sqrt{nb^{2\lambda+1}}\{\hat{m}^{(\lambda)}(h) - m^{(\lambda)}(h) - b^{p+1-\lambda}b_\lambda(h)\} \xrightarrow{D} N(0, v_\lambda(h)), \tag{4.12}$$

where $b_\lambda(h) = \lambda!\Lambda_{\lambda,p+1}m^{(p+1)}(h)/(p+1)!, v_\lambda(h) = (\lambda!)^2||K_\lambda^*||_2^2 h\varphi^{-1}(h)$, $\varphi(\cdot)$ is the density of h_t and $\Lambda_{\lambda,p+1}$ is defined in (4.23).

In practice, the parameters in (4.8) are unknown. What follows we shall give a method to estimate them. Let $\theta = (\phi^\tau, \beta)^\tau$, $\phi = (\phi_1, \cdots, \phi_q)^\tau$. For simplicity, suppose that θ lies in the interior of $\Theta = [\phi_{1L}, \phi_{1U}]\times\cdots\times[\phi_{qL}, \phi_{qU}]\times[\beta_L, \beta_U]$, where $0 < \beta_L < \beta_U < 1$. For $\theta \in \Theta$, define

$$h_t(\theta) = v(y_{t-1}, \phi) + \beta h_{t-1}(\theta) = \sum_{l=1}^{\infty} \beta^{l-1} v(y_{t-l}, \phi), \tag{4.13}$$

$$m(\theta, h) = E(y_t | h_t(\theta) = h), \tag{4.14}$$

$$L(\theta) = E\left\{\left[\log h_t(\theta) + \frac{[y_t - m(\theta, h_t(\theta))]^2}{h_t(\theta)}\right]\pi(\tilde{h}_t)\right\}, \tag{4.15}$$

where $\pi(\cdot)$ is a nonnegative weight function whose compact support is contained in A.

The series $\tilde{h}_t$ satisfies $h_t(\theta) \leq \tilde{h}_t, \theta \in \Theta, t = 1, 2, \cdots$. Let $\phi_U = (\phi_{1U}, \cdots, \phi_{qU})^\tau$, then, without loss of generality, we can put

$$\tilde{h}_t = v(y_{t-1}, \phi_U) + \beta_U \tilde{h}_{t-1}. \tag{4.16}$$

When $n \to \infty$, it is known that $L(\theta)$ can be consistently estimated by

$$L_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left\{ \log h_t(\theta) + \frac{[y_t - m(\theta, h_t(\theta))]^2}{h_t(\theta)} \right\} \pi(\tilde{h}_t). \tag{4.17}$$

Under Assumption 4.6, $L(\theta)$ has an unique minimum point at θ_0 and is locally convex. Hence, under some regularity conditions, the minimizer of $L_n(\theta)$ is expected to locate the true parameter θ_0 consistently. However, in practice, we still have no idea about $m(\theta, h_t(\theta))$ and hence the minimizer of $L_n(\theta)$ is not practicable. To obtain a feasible estimator, for each $h \in A, \theta \in \Theta$, define the estimator of $m(\theta, h)$ in (4.14) as

$$\hat{m}(\theta, h) = E_0^\tau (Z_\theta^\tau W_\theta Z_\theta)^{-1} Z_\theta^\tau W_\theta Y, \tag{4.18}$$

where

$$Z_\theta = \begin{pmatrix} 1 & \frac{h_1(\theta)-h}{b} & \cdots & (\frac{h_1(\theta)-h}{b})^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \frac{h_n(\theta)-h}{b} & \cdots & (\frac{h_n(\theta)-h}{b})^p \end{pmatrix} \text{ and } W_\theta = \operatorname{diag} \begin{pmatrix} \frac{1}{n} K_b(h_1(\theta) - h) \\ \vdots \\ \frac{1}{n} K_b(h_n(\theta) - h) \end{pmatrix}.$$

Define next

$$\hat{L}_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left\{ \log h_t(\theta) + \frac{[y_t - \hat{m}(\theta, h_t(\theta))]^2}{h_t(\theta)} \right\} \pi(\tilde{h}_t), \tag{4.19}$$

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} \hat{L}_n(\theta). \tag{4.20}$$

The above estimator $\hat{\theta}_n$ is reduced to that of Christensen et al. (2008) if we set $\pi(\tilde{h}_t) = 1$ for $t = 1, \cdots, n$. However, the added weight function $\pi(\cdot)$ is not redundant. It is known, for boundary point of $\{h_t(\theta)\}$, the estimator in (4.18) may be seriously biased because of inadequate observations. By appropriate choice for $\pi(\cdot)$, we can avoid to use the estimates of $m(\theta, \cdot)$ for boundary points to compute (4.19). The added $\pi(\cdot)$ is also useful to establish the following theorem, from which, it is known that $\hat{\theta}_n$ converges to the true value in probability.

Theorem 4.2 *Under Assumptions 4.1-4.7 in Section 4.3.3, as* $n \to \infty$ *and* $\left(\sqrt{nb}\right)^{-1} \log n = o(1)$, *we have* $\hat{\theta}_n \xrightarrow{P} \theta_0$.

Proof: See Section 4.3.3.

4.3.2 Simulations

In this section, we examine the estimation performance by simulation studies. The data is generated from (4.8) with specifying $\lambda = (\omega, \alpha, \eta)^\tau$, $v(y, \lambda) = \omega + \alpha[1 + \eta I(y \leq 0)]y^2$. Put $\theta = (\omega, \alpha, \eta, \beta)^\tau$ and the examples are given as follows.

Ex 4.3: $\theta = (0.01, 0.2, 0.1, 0.5)^\tau$, $m(h) = 0.5 \sin(3 + 30h)$.

Ex 4.4: $\theta = (0.02, 0.2, 0.3, 0.7)^\tau$, $m(h) = 0.4 - \exp(-3h^2)$.

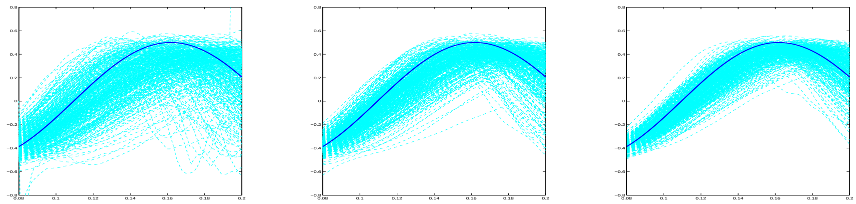
For each of the above examples, we conduct 500 replications with sample sizes $n = 400, 800, 1200$ respectively. To estimate θ and $m(\cdot)$ for Ex 4.3-4.4, we set $\Theta = [0.001, 0.1] \times [0.05, 0.6] \times [0.01, 0.6] \times [0.1, 0.9]$. The p in expression (4.11) is set equal to 1, namely we apply the local linear regression. For the choice of the bandwidth b, like Xia and Tong (2006), we simply put $b = 1.06 \cdot \mathrm{std}(h_t(\theta)) \cdot n^{-1/5}$. As to the weight function, following Yang (2002), we put $\pi(h) = I(h \in A)$, where A is set as the interval between the 10th and 90th percentiles of the explanatory variable.

We study the medians and standard deviations (SD) of the estimators for the parameters and the results are displayed in Table 4.1. For the functional part, we overlay the 500 function estimates $\hat{m}(h)$ with the true function $m(h)$ on the same scale for all the considered sample sizes. The plots are displayed in Array 4.2 (dashed lines are the estimated curves and the real line is the true curve). From Table 4.1, we can see the medians are close to the true values in most cases and the standard deviation becomes smaller when the sample size gets larger. In terms of the plots in Array 4.2, it is shown that the estimated functions capture the trend of the true function and it also witnesses better fitting performance when the sample size gets larger. All the above implies the considered estimates are satisfactory.

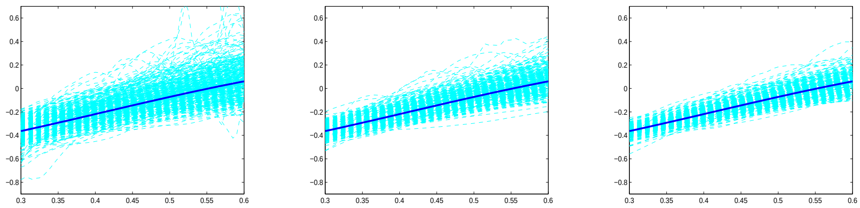
Table 4.1: Medians and standard deviations of parameter estimates for Ex 4.3-4.4

$\theta = (\omega, \alpha, \eta, \beta)^\tau$		$\hat{\omega}$	$\hat{\alpha}$	$\hat{\eta}$	$\hat{\beta}$
		T=400			
$(0.01, 0.2, 0.1, 0.5)^\tau$	Median	0.0010	0.2048	0.0930	0.5266
	SD	0.0093	0.0538	0.2066	0.1087
$(0.02, 0.2, 0.3, 0.7)^\tau$	Median	0.0171	0.1985	0.3427	0.6912
	SD	0.0292	0.0613	0.2577	0.0836
		T=800			
$(0.01, 0.2, 0.1, 0.5)^\tau$	Median	0.0048	0.2084	0.1122	0.5152
	SD	0.0072	0.0375	0.1664	0.0817
$(0.02, 0.2, 0.3, 0.7)^\tau$	Median	0.0185	0.1923	0.3438	0.7000
	SD	0.0224	0.0468	0.2340	0.0620
		T=1200			
$(0.01, 0.2, 0.1, 0.5)^\tau$	Median	0.0059	0.2071	0.1186	0.5151
	SD	0.0069	0.0288	0.1393	0.0703
$(0.02, 0.2, 0.3, 0.7)^\tau$	Median	0.0189	0.1967	0.3471	0.7011
	SD	0.0179	0.0385	0.2158	0.0494

Note: Number of replications=500.



Ex 4.3: n=400 Ex 4.3: n=800 Ex 4.3: n=1200



Ex 4.4: n=400 Ex 4.4: n=800 Ex 4.4: n=1200

Array 4.2: Results of the function estimates for Ex 4.3-4.4.

4.3.3 Proofs

In this part, we give a brief proof for Theorem 4.2 based on the results of Yang (2006). An alternative proof with weaker conditions can be obtained by referring to the ideas in Chapter 6. Firstly, we make the following assumptions for model (4.8).

Assumption 4.1. The i.i.d (0, 1) process $\{e_t\}$ satisfies $Ee_t^4 < \infty$, and has a continuous symmetric probability density function which is positive everywhere.

Assumption 4.2. The function $m(\cdot)$ has Lipschitz continuous $(p+1)th$ derivative.

Assumption 4.3. The process $\{h_t\}$ has a stationary density $\varphi(\cdot)$ which is Lipschitz continuous and satisfy $\inf_{h\in A}\varphi(h) > 0$, where A is a compact subset of R with nonempty interior.

Assumption 4.4. The processes $\{(h_t, h_t(\theta), \tilde{h}_t)\}_{t\geq 1}, \theta \in \Theta$ are uniformly geometrically ergodic and ϕ-mixing. Further the processes have stationary densities $\varphi(h, h_\theta, \tilde{h})$ and there are two constants m and M such that $0 < m \leq \varphi_\theta(h) \leq M < \infty, h \in A, \theta \in \Theta$ where

$\varphi_\theta(\cdot)$ is the marginal stationary density of $h_t(\theta)$.

Assumption 4.5. The function $m(\theta, h), \theta \in \Theta$ defined in (4.14) satisfies

$$\sup_{\theta\in\Theta}\sup_{h\in A}|\frac{m^{(i+j)}(\theta, h)}{\partial\theta^i\partial h^j}| < \infty, 0 \le i \le 1, 0 \le j \le p+1$$

and the process $\{y_t\}$ satisfies $E\exp\{a|y_t|^r\} < +\infty$ for some constants $a > 0$ and $r > 0$.

Assumption 4.6. The function $L(\theta)$ in (4.15) is continuous and has a positive definite Hessian matrix at its unique minimum θ_0.

Assumption 4.7. There exists a stationary positive processes $\{w_t\}$ depending only on Θ such that for any $\theta_1, \theta_2 \in \Theta$

$$\left\|\frac{1}{h_t(\theta_1)}\frac{\partial h_t(\theta_2)}{\partial\theta}\right\| \le w_t, Ew_t < +\infty.$$

Remark 4.1 Assumption 4.4 is similar to the conclusion of Lemma A.1 in Yang (2006), and it is useful to show the uniform convergence of some variables as in Lemma A.2 of Yang (2006). Assumption 4.7 is helpful to establish the uniform laws of large numbers and when $m(h) = \gamma\sqrt{h}$, it can be shown that the conditions in Assumption 4.7 are satisfied (See Chapter 3). Other assumptions have also been analogously applied by Yang (2006).

Following the notations in Yang (2006), denote $\mu_r(K) = \int u^r K(u)du$ and define the matrix S as

$$S = \begin{pmatrix} \mu_0(K) & 0 & \mu_2(K) & \cdots & 0 \\ 0 & \mu_2(K) & 0 & \cdots & \mu_{p+1}(K) \\ \mu_2(K) & 0 & \ddots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \mu_{p+1}(K) & 0 & \cdots & \mu_{2p}(K) \end{pmatrix}. \tag{4.21}$$

Let $S^{-1} = (s_{\lambda\lambda'})_{0\le\lambda,\lambda'\le p}$. As in Fan and Gijbels (1996) page 64 or Yang (2006), we further define the equivalent kernel

$$K^*_\lambda(u) = \sum_{\lambda'=0}^{p} s_{\lambda\lambda'} u^{\lambda'} K(u), \lambda = 0, 1, \ldots, p. \tag{4.22}$$

It is not difficult to have

$$\int K^*_\lambda(u) u^{\lambda''} du = \begin{cases} 1 & \lambda'' = \lambda, \\ 0 & 0 \le \lambda'' \le p, \lambda'' \ne \lambda, \\ \Lambda_{\lambda,p+1} & \lambda'' = p+1, \end{cases} \tag{4.23}$$

where $\Lambda_{\lambda,p+1}$ is a nonzero constant.

Proof of Theorem 4.2:

Based on Assumptions 4.1-4.5, using similar argument to the proof of lemma A.3 in Yang (2006), we can get

$$\sup_{\theta\in\Theta} \sup_{h\in A} |m(\theta, h) - \hat{m}(\theta, h)| = o_p(1). \tag{4.24}$$

Simple calculation gives

$$
\begin{aligned}
[y_t - \hat{m}(\theta, h_t(\theta))]^2 &= [y_t - m(\theta, h_t(\theta))]^2 + [m(\theta, h_t(\theta)) - \hat{m}(\theta, h_t(\theta))]^2 \\
&\quad + 2[y_t - m(\theta, h_t(\theta))][m(\theta, h_t(\theta)) - \hat{m}(\theta, h_t(\theta))]. \qquad (4.25)
\end{aligned}
$$

Then, according to (4.17) and (4.19), we have

$$
\begin{aligned}
\hat{L}_n(\theta) - L_n(\theta) &= \frac{1}{n}\sum_{t=1}^{n}\left\{\frac{[m(\theta, h_t(\theta)) - \hat{m}(\theta, h_t(\theta))]^2}{h_t(\theta)}\right\}\pi(\tilde{h}_t) \\
&\quad + \frac{2}{n}\sum_{t=1}^{n}\left\{\frac{[y_t - m(\theta, h_t(\theta))][m(\theta, h_t(\theta)) - \hat{m}(\theta, h_t(\theta))]}{h_t(\theta)}\right\}\pi(\tilde{h}_t). \qquad (4.26)
\end{aligned}
$$

In conjunction with (4.25) and Assumption 4.5, we know

$$
\sup_{\theta\in\Theta} |\hat{L}_n(\theta) - L_n(\theta)| = o_p(1). \qquad (4.27)
$$

From (4.17), for any $\theta_1, \theta_2 \in \Theta$, it can be derived that

$$
L_n(\theta_1) - L_n(\theta_2) = \frac{1}{n}\sum_{t=1}^{n}[I_{1t} + I_{2t} + I_{3t}]\,\pi(\tilde{h}_t), \qquad (4.28)
$$

where

$$
\begin{aligned}
I_{1t} &= \log h_t(\theta_1) - \log h_t(\theta_2) = \frac{1}{h_t(\theta^*_{1,t})}\frac{\partial h_t(\theta^*_{1,t})}{\partial \theta^\tau}(\theta_1 - \theta_2), \\
I_{2t} &= [y_t - m(\theta_1, h_t(\theta_1))]^2\left(\frac{1}{h_t(\theta_1)} - \frac{1}{h_t(\theta_2)}\right) \\
&= [y_t - m(\theta_1, h_t(\theta_1))]^2\frac{1}{h_t^2(\theta^*_{2,t})}\frac{\partial h_t(\theta^*_{2,t})}{\partial \theta^\tau}(\theta_1 - \theta_2),
\end{aligned}
$$

and

$$
\begin{aligned}
I_{3t} &= \frac{1}{h_t(\theta_2)} \left\{ [y_t - m(\theta_1, h_t(\theta_1))]^2 - [y_t - m(\theta_2, h_t(\theta_2))]^2 \right\} \\
&= -\left[2y_t - m(\theta_1, h_t(\theta_1)) - m(\theta_2, h_t(\theta_2))\right] \\
&\quad \times \left[\frac{\partial m(\theta_1, h_t(\theta_{3,t}^*))}{\partial h} \frac{1}{h_t(\theta_2)} \frac{\partial h_t(\theta_{3,t}^*)}{\partial \theta^\tau} + \frac{\partial m(\theta_{4,t}^*, h_t(\theta_2))}{\partial \theta^\tau} \frac{1}{h_t(\theta_2)} \right] \\
&\quad \times (\theta_1 - \theta_2).
\end{aligned}
$$

In the above expressions, $\theta_{i,t}^*, i = 1, 2, 3, 4, t = 1, 2, \cdots, n$ are parameter vectors between θ_1 and θ_2. According to Assumption 4.5 and Assumption 4.7, we know there exist B_{1t}, B_{2t}, B_{3t} (independent of θ_1, θ_2) such that

$$
|I_{1t}| \le B_{1t} ||\theta_1 - \theta_2||, |I_{2t}| \le B_{2t} ||\theta_1 - \theta_2|| \text{ and } |I_{3t}| \le B_{3t} ||\theta_1 - \theta_2||. \tag{4.29}
$$

Consequently, we get

$$
\begin{aligned}
|L_n(\theta_1) - L_n(\theta_2)| &\le \left\{ \frac{1}{n} \sum_{t=1}^{n} [B_{1t} + B_{2t} + B_{3t}] \, \pi(\tilde{h}_t) \right\} ||\theta_1 - \theta_2|| \\
&:= B_n ||\theta_1 - \theta_2||,
\end{aligned}
$$

with $B_n = O_p(1)$. In terms of Lemma 1 and Theorem 1 in Andrews (1992), we know

$$
\sup_{\theta \in \Theta} |L_n(\theta) - L(\theta)| = o_p(1). \tag{4.30}
$$

By the triangle inequality, (4.27) and (4.30) imply that

$$
\sup_{\theta \in \Theta} |\hat{L}_n(\theta) - L(\theta)| = o_p(1). \tag{4.31}
$$

Under Assumption 4.6, applying Lemma 14.3 (page 258) and Theorem 2.12 (page 28) in Kosorok (2006), we then have $\hat{\theta}_n - \theta_0 = o_p(1)$, which ends the proof of the theorem.

4.4 Empirical Studies

In this section, based on model (4.2) and (4.8), we study the monthly excess return data on the CRSP value weighted indices, including the NYSE, the AMEX and NASDAQ. The riskless rate used to compute the excess returns is the one-month Treasury bill rate (from Ibbotson Associates). The range of the considered data is from July 1926 to February 2009 (totally 992 observations).

First, we model the data by (4.2). Before fitting the data , we need to get a reasonable estimate for θ_0 in model (4.2), which is assumed to be known. As a compromise, we approximate θ_0 by $\hat{\omega}$ which is estimated from the parametric model below:

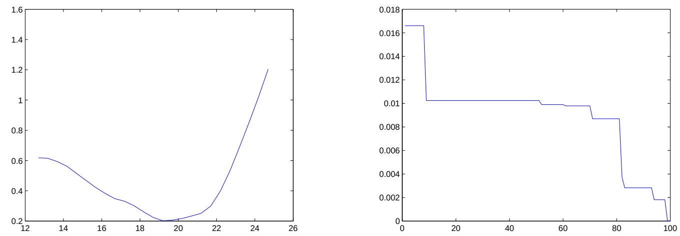
$$y_t = m(h_t, \delta) + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t},$$

$$e_t \sim i.i.d(0, 1), h_t = \omega + \alpha y_{t-1}^2 + \beta h_{t-1},$$

where, $m(h_t, \delta)$ can be $\delta h_t, \delta \sqrt{h_t}, \delta \log h_t$ or other forms. In our study, we put $m(h_t, \delta) = \delta h_t$, which is commonly applied in the literature. For the choices of L, the lagged time, in our study, we try two cases $L = 100$, $L = 200$ with $\lambda = 0.9$. The results are displayed in Array 4.3.

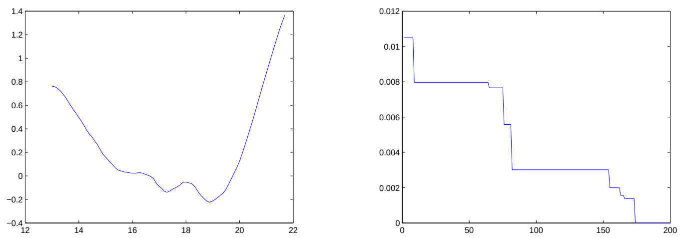
From the plots in Array 4.3, we can see that the estimated results are similar for the two candidates of L. There is no obvious cut ($\theta_i = 0$ when i is bigger than some j) for θ in each case, implying that there may be long-memory cumulative effect for the conditional variance. According to the estimated mean function $\hat{m}(\cdot)$, it is suggested that the nonlinear relationship between risk and return could be more reasonable and it seems that a higher h_t (risk) does not necessarily correspond to higher $m(h_t)$ (return).

Estimation with $L = 100$.



$\hat{m}(\cdot)$ $\hat{\theta}$

Estimation with $L = 200$.



$\hat{m}(\cdot)$ $\hat{\theta}$

Array 4.3: Results of the empirical studies based on (4.2).

Next, we study the data by (4.8). Take $\{y_t\}_{t=1}^{992}$ to be the considered excess return series. For comparison, we use $\{y_t\}_{t=1}^{900}$ to estimate model (4.8) and leave $\{y_t\}_{t=901}^{992}$ for out-of-sample forecasts. Before minimizing (4.20) to get the estimation for the parameters, we need to set a proper scope Θ for the parameters. To get a reasonable scope, we first estimate three parametric models by putting $m(h_t) = \delta h_t$, $m(h_t) = \delta \log h_t$, $m(h_t) = \delta\sqrt{h_t}$ and the conditional variance h_t in (4.8) is specified by $h_t = \omega + \alpha[1+\eta I(y_{t-1} \le 0)]y_{t-1}^2 + \beta h_{t-1}$. Recall $\theta = (\omega, \alpha, \eta, \beta)^\tau$, then the estimates of θ in the above three parametric models are respectively $[0.9108, 0.1075, 0.7227, 0.8328]^\tau$, $[0.7980, 0.1102, 0.7497, 0.8337]^\tau$ and $[0.8147, 0.1104, 0.7336, 0.8332]^\tau$. According to the above results, we set $\Theta =$

$[0.05, 1] \times [0.01, 0.5] \times [0.5, 10] \times [0.1, 0.95]$. Then we get:

$$y_t = m(h_t) + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t},$$

$$h_t = 0.9567 + 0.0351[1 + 2.6854 I(y_{t-1} \leq 0)] y_{t-1}^2 + 0.8798 h_{t-1}. \tag{4.32}$$

For comparison, we also estimate the model of Christensen et al. (2008), namely (4.1), based on the same data set. They are

$$y_t = m(h_t) + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t},$$

$$h_t = 0.8301 + 0.0865 y_{t-1}^2 + 0.8744 h_{t-1}. \tag{4.33}$$

We tabulate the RMSEs and the MAEs for both in-sample forecasts and out-of-sample forecasts in Table 4.2 for (4.32-4.33). According to tabulated fitting errors, it is seen that model (4.32) has better performance than that of model (4.33), which empirically justifies the generalization of the conditional variance. Thus it makes sense to take the asymmetric factors into account for the considered data.

Table 4.2: In-sample and out-of-sample forecast performance.

Model	RMSE (in)	MAE (in)	RMSE (out)	MAE (out)
(4.32)	4.9760	3.6824	4.5249	3.3013
(4.33)	5.0960	3.7137	4.6163	3.3486

Set a grid vector for h_t as $h = [12 : 0.25 : 30]$ which is contained in the intervals between 10th and 90th percentiles of the estimated h_t for both (4.32) and (4.33). Based on the observations $\{y_t\}_{t=1}^{900}$, we can estimate the value of $m(h(i))$ $(i = 1, \cdots, 73)$ according to (4.11) in Section 4.3.1 and (7) of Christensen et al. (2008). We plot the estimated $\hat{m}(\cdot)$ from (4.32) (solid line) and (4.33) (dashed line) in Figure 4.1. Similar to the plots in Array 4.3, we can see that both the estimated curves are nonlinear and they are not monotonically increasing either. The above results indicate the traditional parametric specifications for the conditional mean such as $m(h_t) = \delta h_t, \delta \log h_t, \delta\sqrt{h_t}$ may not be appropriate. From Figure 4.1, it is worthwhile to further consider the following two questions. Firstly, how to interpret the behavior of $m(h_t)$ in the figure? Partial answer can be found in Rossi and Timmermann (2010), where non-monotonic relation between conditional volatility and expected stock market returns has been evidenced. The second question is whether we can obtain similarly shaped curves when a different data set is applied. We leave the detailed interpretations for future study.

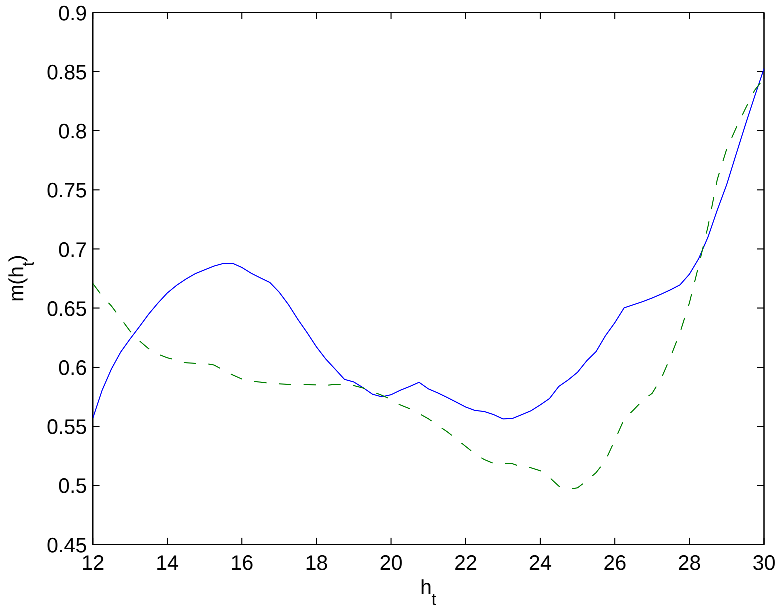


Figure 4.1: Plots of estimated $\hat{m}(h_t)$ for model (4.32) (solid line), (4.33) (dashed line).

4.5 Summary

In this chapter, we study the relationship between risk (conditional variance) and return (conditional mean) by semiparametric (G)ARCH-M models. By adopting the idea of the constrained single-index model of Xia and Tong (2006), we give a method to estimate the considered semiparametric ARCH-M model, which enables us to check the long memory of the conditional volatility. Motivated by Christensen et al. (2008), we further study their model by considering a general "news impact curve", which allows to take the asymmetric factor into account. An improved estimation method is proposed for the generalized model. Through the simulations, it is shown that the proposed estimates perform well. From the empirical studies, we find that the curve between h_t (conditional variance) and $m(h_t)$ (conditional mean) is neither linear nor monotonically increasing for the considered data. Such a phenomenon implies that the traditional parametric forms such as $m(h_t) = \delta h_t, \delta \log h_t, \delta \sqrt{h_t}$ may not be appropriate.

Chapter 5

A Functional Coefficient Autoregressive GARCH-M Model

5.1 Background

ARCH-M models have been considered by many researchers since they were proposed by Engle, et al. (1987). One of its forms is

$$y_t = \delta h_t + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t},$$

$$e_t \sim i.i.d \ \ N(0, 1), h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}. \tag{5.1}$$

There are different appellations for δ such as: "price of volatility" in Chou et al. (1992) and "relative risk aversion parameter" in Das and Sarkar (2010). In this chapter, we simply address it as volatility coefficient. Many empirical studies have been done based on (5.1), but mixed results were reported. For example, Glosten et al. (1993) obtained a negative value for δ and Harvey (1989) found δ nonconstant and counter-cyclical.

Backus and Gregory (1988) argued that the relationship between the conditional mean and the conditional variance was not necessarily linear. To explain the above phenomena, Chou et al. (1992) proposed a time-varying parameter GARCH-M (henceforth TVP-GARCH-M) model to capture the variation of the volatility coefficient. The TVP-GARCH-M model has the form

$$
\begin{aligned}
&y_t = \delta_t h_t + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t}, \delta_t = \delta_{t-1} + v_t, \\
&e_t \sim i.i.d\ \ N(0,1), h_t = \omega + \alpha\eta_{t-1}^2 + \beta h_{t-1}.
\end{aligned} \tag{5.2}
$$

Here, the GARCH surprise variable $\eta_t := y_t - E_{t-1}(y_t)$ with $E_{t-1}(y_t)$ is the optimal forecast of y_t given all information up to time $t-1$. The errors ε_t, v_t are assumed to be uncorrelated Gaussian variates with zero means and variances h_t and Q, respectively. The coefficient δ_t in (5.2) is assumed to follow a random walk, which together with the system parameters, can be estimated by the Kalman filter and maximum likelihood methods.

Motivated by the TVP-GARCH-M model (5.2), it is sensible to study the GARCH-M model (5.1) with a time-varying volatility coefficient. When explaining why the δ_t was time-varying, Chou et al. (1992) suggested some macroeconomic variables such as inflation rate and interest rate could have impact on it. Therefore it is worthwhile to consider δ as a function of some explanatory variables. Because of the availability of data and complexity in computation, it is hard to cover all the related factors. An alternative method is to assume δ as a function of the time-lagged returns, say $\delta = \delta(y_{t-1})$, which is a standard approach in time series analysis. A further argument about setting $\delta = \delta(y_{t-1})$ can be given as follows. As in Chou et al. (1992) and Das and Sarkar

(2010), we can consider the volatility coefficient δ to be a measure of risk aversion. It is generally accepted that yesterday's return has impact on today's risk attitude or the risk aversion. Hence it is reasonable to assume $\delta = \delta(y_{t-1})$.

In this chapter, we shall study the following functional coefficient autoregressive GARCH-M (henceforth FCA-GARCH-M) model:

$$y_t = \delta(y_{t-1})h_t + \varepsilon_t, \varepsilon_t = e_t\sqrt{h_t},$$

$$e_t \sim i.i.d(0,1), h_t = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}. \tag{5.3}$$

Here $\omega, \alpha, \beta > 0$, y_s is independent of $\{e_t\}$ for $t > s$, and $\delta(\cdot)$ is an unknown smooth function. Also, h_t is assumed to be driven by an alternative GARCH (1, 1) process, where the original ε_{t-1}^2 is replaced by the observable y_{t-1}^2. Similar to (5.2), such modification for h_t is helpful to estimating the model and it has been adopted by Christensen et al. (2008). Model (5.3) belongs to the scope of semiparametric models because it includes a nonparametric part $\delta(\cdot)$ and a parametric part h_t. Considering $\delta(\cdot)$ in (5.3) as a measure of risk aversion as in Chou et al. (1992), the improvement of (5.3) lies in that it enables us to understand the impact of previous return to the present risk aversion. When h_t is observable, model (5.3) is reduced to the common functional coefficient autoregressive model which has been well discussed by Chen and Tsay (1993), Chen and Liu (2001), among others.

This chapter is arranged as follows. In Section 5.2, we explain the idea about estimation. Practical procedure for testing goodness of fit is also given. Simulation and empirical studies are respectively shown in Section 5.3 and Section 5.4. We give a summary in Section 5.5.

5.2 Estimation and Goodness of Fit Test

In this section, we give an approach to estimate the parametric part and functional part of the model. The method is based on the ideas of functional coefficient model of Cai et al. (2000) and the quasi maximum likelihood estimation. Following the principle of bootstrap test of Cai et al. (2000), we also give practical procedure to construct a goodness of fit test.

5.2.1 Estimation

For model (5.3), we need to estimate the system parameters ω, α, β and the value $\delta(u_0)$ for a given u_0 based on the observations $\{y_t\}_{t=1}^T$. Let $\theta = (\omega, \alpha, \beta)^\tau$ and $\theta \in \Theta$, which is the considered parameter space including the true value, say θ_0, of θ. All throughout this chapter, the superscript τ denotes the transpose of a vector or a matrix. We apply two-step method to estimate the model. The first step is to estimate the function $\delta(\cdot)$ provided θ is given. The second step is to estimate θ based on the obtained function estimates in the first step. The details are as followed.

Given θ, then $h_t(\theta)$ is determined according to (5.3) and hence it may be considered as an observable quantity. Suppose the function $\delta(\cdot)$ has second derivative at point u_0. Then for x in a small neighborhood $[u_0 - b, u_0 + b]$, we have

$$\delta(x) \approx \delta(u_0) + \delta'(u_0)(x - u_0) := a_0 + a_1(x - u_0).$$

Based on $\{y_t, h_t(\theta)\}_{t=1}^T$, we can define the estimators for (a_0, a_1) by

$$(\hat{a}_0(u_0, \theta), \hat{a}_1(u_0, \theta)) = \min_{a_0, a_1} \sum_{t=2}^{T} \{y_t - [a_0 + a_1(y_{t-1} - u_0)]h_t(\theta)\}^2 k_b(y_{t-1} - u_0).$$

Here, $k_b(\cdot) = b^{-1}k(\cdot/b)$ with $k(\cdot)$ being a kernel function and $b > 0$ a bandwidth. Following Cai et al. (2000), we have

$$\hat{\delta}(u_0, \theta) := \hat{a}_0(u_0, \theta) = \sum_{t=2}^{T} K_T(y_{t-1} - u_0, h_t(\theta))y_t, \tag{5.4}$$

where

$$K_T(u, x) = (1, 0)(H^\tau WH)^{-1}(x, ux)^\tau k_b(u),$$

H being a $(T-1)\times 2$ matrix with $(h_i(\theta), h_i(\theta)(y_{i-1}-u_0))$ as its $(i-1)$th row $(i = 2, \cdots, T)$, and $W = \mathrm{diag}\{k_b(y_1 - u_0), \cdots, k_b(y_{T-1} - u_0)\}$. Let u_0 take values in $\{y_{t-1}\}_{t=2}^T$ one by one, then we can get a series of estimators $\{\hat{\delta}(y_{t-1}, \theta)\}_{t=2}^T$ based on (5.4).

Define $\hat{\Delta}_T(\theta) = (\hat{\delta}(y_1, \theta), \cdots, \hat{\delta}(y_{T-1}, \theta))^\tau$, then we get following approximate quasi log likelihood function

$$L_T(\theta, \hat{\Delta}_T(\theta)) = \sum_{t=2}^{T} l_t(\theta, \hat{\delta}(y_{t-1}, \theta)) = \sum_{t=2}^{T}[-\frac{1}{2}\log h_t(\theta) - \frac{1}{2}\frac{\hat{\varepsilon}_t^2(\theta)}{h_t(\theta)}], \tag{5.5}$$

where $\hat{\varepsilon}_t(\theta) = y_t - \hat{\delta}(y_{t-1}, \theta)h_t(\theta)$. Let $\hat{\theta}_T = \max_{\theta\in\Theta} L_T(\theta, \hat{\Delta}_T(\theta))$ which may be considered as an estimator for θ. Based on (5.4), $\delta(u_0)$ is then approximated by $\hat{\delta}(u_0, \hat{\theta}_T)$.

In practice, because of the complicated form of $\hat{\delta}(y_{t-1}, \theta)$, it is not easy to directly calculate the maximizer $\hat{\theta}_T = \max_{\theta\in\Theta} L_T(\theta, \hat{\Delta}_T(\theta))$. Subsequently, we give an algorithm to obtain $\hat{\theta}_T$. The procedure can be seen a generalization of Christensen et al. (2008).

Step 1: Assign a set of initial parameters $\hat{\theta}^{(i)}(i = 0)$ and compute $h_t(\hat{\theta}^{(i)})$ for $t = 2, \cdots, T$, according to model (5.3)[1].

[1]Initial value for θ can be acquired by estimating the model (5.3) with $\delta(\cdot)$ being a constant δ. The sample variance of $\{y_t\}_{t=1}^T$ can be used as an initial value for h_t.

Step 2: Based on the sequence $\{y_t, h_t(\hat{\theta}^{(i)})\}_{t=1}^{T}$, we can get $\{\hat{\delta}(y_1, \theta^{(i)}), \cdots, \hat{\delta}(y_{T-1}, \theta^{(i)})\}$ by (5.4).

Step 3: Update $\hat{\theta}^{(i)}$ (i.e., find $\hat{\theta}^{(i+1)}$) by performing quasi maximum likelihood estimation on the GARCH(1, 1) model

$$\hat{\varepsilon}_t^{(i)} = \sqrt{h_t} e_t,$$

$$e_t \sim i.i.d(0, 1), h_t = \omega + \alpha y_{t-1}^2 + \beta h_{t-1},$$

where, $\hat{\varepsilon}_t^{(i)} = y_t - \hat{\delta}(y_{t-1}, \theta^{(i)}) h_t(\hat{\theta}^{(i)}), t = 2, \cdots, T$.

Step 4: Repeat steps 1-3 until convergence.

5.2.2 Goodness of fit test

After estimating the FCA-GARCH-M model (5.3), a possible question is whether the proposed model performs better than the GARCH-M model with a constant volatility coefficient. Equivalently, we want to test the hypothesis for model (5.3):

$$H_0 : \delta(\cdot) = \delta \text{ vs } H_1 : \delta(\cdot) = \delta(y_{t-1}).$$

By applying the method of Cai et al. (2000), we give the following procedure to test the above hypothesis approximately. Let $\hat{\delta}_0$ (a constant), $\hat{\delta}_1(\cdot)$ (a function), $\hat{\theta}_0$ and $\hat{\theta}_1$ be the estimators of $\delta(\cdot), \theta$ under H_0 and H_1 respectively. Calculate the RSS (residual sum of squares) under H_0 as

$$RSS_0 = (T-1)^{-1}\Sigma_{t=2}^{T}\{y_t - \hat{\delta}_0 h_t(\hat{\theta}_0)\}^2$$

and the RSS under H_1 as

$$RSS_1 = (T-1)^{-1}\Sigma_{t=2}^{T}\{y_t - \hat{\delta}_1(y_{t-1}, \hat{\theta}_1)h_t(\hat{\theta}_1)\}^2.$$

The test statistic is defined to be

$$\Lambda_T = RSS_0/RSS_1 - 1. \tag{5.6}$$

We reject the null hypothesis for large values of Λ_T. The following steps are used to evaluate the p-value of the test.

Step I: Generate the bootstrap residuals $\{\varepsilon_t^*\}_{t=2}^T$ from the centered residuals $\{\hat{\varepsilon}_t - \bar{\hat{\varepsilon}}\}_{t=2}^T$, where

$$\hat{\varepsilon}_t = y_t - \hat{\delta}_1(y_{t-1}, \hat{\theta}_1)h_t(\hat{\theta}_1), \bar{\hat{\varepsilon}} = \frac{1}{T-1}\Sigma_{t=2}^{T}\hat{\varepsilon}_t$$

and define $y_t^* = \hat{\delta}_0 h_t(\hat{\theta}_0) + \varepsilon_t^*$.

Step II: Let $y_1^* = y_1$. According to (5.6), calculate the test statistic Λ_T^* based on $\{y_t^*\}_{t=1}^T$.

Step III: Replicate steps I-II K times to get a sequence $\{\Lambda_{T,i}^*\}_{i=1}^K$. The null hypothesis will be rejected when Λ_T is greater than the upper-α point of $\{\Lambda_{T,i}^*\}_{i=1}^K$.

As mentioned in Cai et al. (2000) that the p-value of the above test is the relative frequency of the event $\{\Lambda_T^* \geq \Lambda_T\}$ in the above replications. For more details about the test, one may refer to Cai et al. (2000), Lee and Ullah (2000) and Kreiss et al. (2008).

5.3 Simulations

In this section, we give three simulation examples to examine the performance of the considered estimators. Series are simulated from the model

$$y_t = \delta(y_{t-1})h_t + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t},$$

$$e_t \sim i.i.d \;\; N(0,1), h_t = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}.$$

The cases considered are:

Ex 5.1: $\delta(x) = 0.4 - (0.659 + 1.26x)e^{-3.1x^2}, (\omega, \alpha, \beta)^\tau = (0.1, 0.1, 0.55)^\tau.$

Ex 5.2: $\delta(x) = 0.2sin^2(\pi x)/x, (\omega, \alpha, \beta)^\tau = (0.5, 0.1, 0.75)^\tau.$

Ex 5.3: $\delta(x) = 0.3I(x > 0) - 0.3I(x \leq 0), (\omega, \alpha, \beta)^\tau = (0.2, 0.2, 0.3)^\tau.$

When applying steps 1-4 to estimate Ex 5.1-5.3, the kernel function $k(\cdot)$ in (5.4) is chosen as $k(x) = 0.75(1 - x^2)I(|x| \leq 1)$. We adopt $AMS(b)$-minimized method of Cai et al. (2000) for choice of the bandwidth b. Let m and Q be two positive integers such that $T > mQ$. The idea is first to use the sub-series of lengths $T - qm\,(q = 1, \cdots, Q)$ to estimate the coefficient function $\delta(\cdot)$ for the next segment of the time series of length m and then to compute the one-step ahead forecast errors. The bandwidth b is chosen to minimize the average of the mean square forecast errors

$$AMS(b) = \Sigma_{q=1}^{Q} AMS_q(b),$$

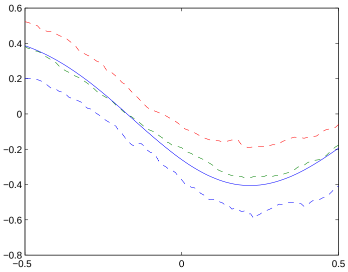
where

$$AMS_q(b) = \frac{1}{m} \sum_{t=T-qm+1}^{t=T-qm+m} [y_t - \hat{\delta}(y_{t-1}, \theta) h_t(\theta)]^2$$

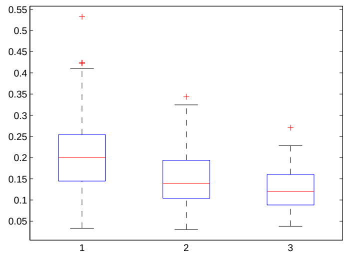
and $\hat{\delta}(\cdot)$ are computed from the sample $\{y_{t-1}, h_t\}_{t=1}^{T-qm}$.

In our simulations, we apply $m = [0.1T]$, $Q = 4$. To run the estimation, we assume the parameters satisfying $\omega \in [0.005, 1], \alpha \in [0.005, 0.9]$ and $\beta \in [0.1, 0.99]$. The convergence criterion is set as $\|\hat{\theta}_T^{(i+1)} - \hat{\theta}_T^{(i)}\| \leq 0.0001$. We conduct 500 replications with sample size T=400, 800, 1200 respectively for each of the above examples. The estimation results for the parameters are given in Table 5.1 and those for the function $\delta(\cdot)$ restricted on the grid vector $u_0 = [-0.5 : 0.01 : 0.5]$ are displayed in Array 5.1. In Table 5.1, Q10, Q50 and Q90 denote respectively the 10%, 50% and 90% percentiles of the estimators among 500 replications. SD means the standard deviation. In each row of Array 5.1, for each grid point $u_0(i)\,(i = 1, \cdots, 101)$, we plot the true value $\delta(u_0(i))$ (solid line) and the 10%, 50% and 90% percentiles of $\{\hat{\delta}_j(u_0(i))\}_{j=1}^{500}$ (dashed line) in the left figure. For clear comparison, we just plot the percentile lines that are estimated under the sample size 1200. The right figure is the box plot of the RMSE sequences for $T = 400, 800, 1200$ (from left to the right). Ex 5.1-5.3 corresponds to (a), (b) and (c) respectively.

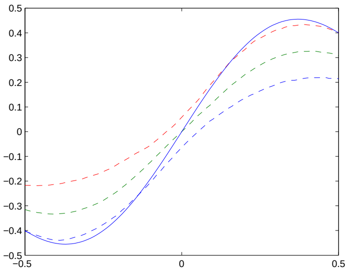
From Table 5.1, it is shown that, in most cases, the true values of parameters are close to the corresponding medians (Q50) and are contained in the interval $[Q10, Q90]$. From Array 5.1, we can see that most of the estimated values for the functions are close to the true ones. From Table 5.1 and the box plots in Array 5.1, both the SDs and RMSEs get smaller gradually when the sample size becomes larger. Simulation results suggest that the considered estimators for either the parametric part or the functional part are asymptotically convergent to the true ones.



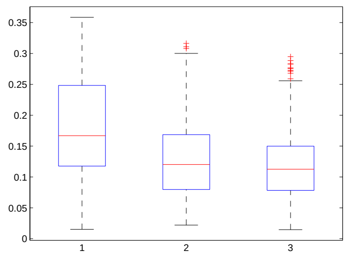
(*a*1)



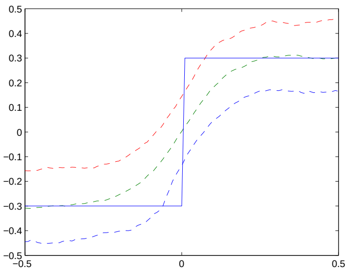
(*a*2)



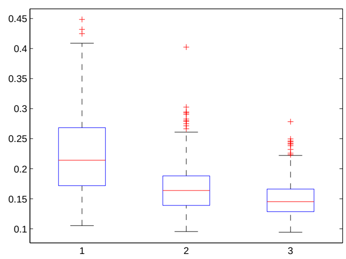
(*b*1)



(*b*2)



(*c*1)



(*c*2)

Array 5.1: Results of the simulation experiments for function estimation.

Table 5.1: Results of the simulation experiments for parameter estimation.

Example	True value	Q10	Q50	Q90	SD
Ex 5.1	$\omega = 0.1$	0.0236	0.0999	0.2314	0.0753
(T=400)	$\alpha = 0.1$	0.0050	0.0500	0.1295	0.0511
	$\beta = 0.55$	0.1000	0.5563	0.8832	0.2845
Ex 5.1	$\omega = 0.1$	0.0418	0.1002	0.2168	0.0615
(T=800)	$\alpha = 0.1$	0.0299	0.0722	0.1235	0.0380
	$\beta = 0.55$	0.1000	0.5573	0.7866	0.2295
Ex 5.1	$\omega = 0.1$	0.0485	0.0996	0.1892	0.0541
(T=1200)	$\alpha = 0.1$	0.0336	0.0748	0.1229	0.0344
	$\beta = 0.55$	0.2334	0.5661	0.7660	0.2002
Ex 5.2	$\omega = 0.5$	0.1898	0.5038	1.0000	0.2902
(T=400)	$\alpha = 0.1$	0.0080	0.0526	0.1162	0.0398
	$\beta = 0.75$	0.6225	0.7821	0.9081	0.1074
Ex 5.2	$\omega = 0.5$	0.2314	0.4678	0.8290	0.2222
(T=800)	$\alpha = 0.1$	0.0348	0.0702	0.1050	0.0277
	$\beta = 0.75$	0.6595	0.7819	0.8734	0.0835
Ex 5.2	$\omega = 0.5$	0.2722	0.4790	0.7897	0.2005
(T=1200)	$\alpha = 0.1$	0.0459	0.0772	0.1076	0.0246
	$\beta = 0.75$	0.6639	0.7747	0.8557	0.0742
Ex 5.3	$\omega = 0.2$	0.0825	0.1966	0.2888	0.0772
(T=400)	$\alpha = 0.2$	0.0374	0.1278	0.2359	0.0754
	$\beta = 0.3$	0.1000	0.3230	0.6846	0.2263
Ex 5.3	$\omega = 0.2$	0.1203	0.2044	0.2757	0.0586
(T=800)	$\alpha = 0.2$	0.0923	0.1542	0.2264	0.0520
	$\beta = 0.3$	0.1000	0.2986	0.5551	0.1667
Ex 5.3	$\omega = 0.2$	0.1367	0.1991	0.2673	0.0493
(T=1200)	$\alpha = 0.2$	0.1109	0.1633	0.2217	0.0432
	$\beta = 0.3$	0.1174	0.3225	0.4941	0.1420

5.4 Empirical Studies

In this section, we apply (5.3) to model real data sets. We analyze the monthly and weekly excess returns on the CRSP value weighted indices, which include the NYSE, the AMEX and NASDAQ. These data can be regarded as reasonable proxies for the stock market and they were also studied by Conrad and Mammen (2008) in a different perspective. The riskless rate used to compute the excess returns is one-month Treasury bill rate (from Ibbotson Associates).

5.4.1 Analysis for monthly excess return

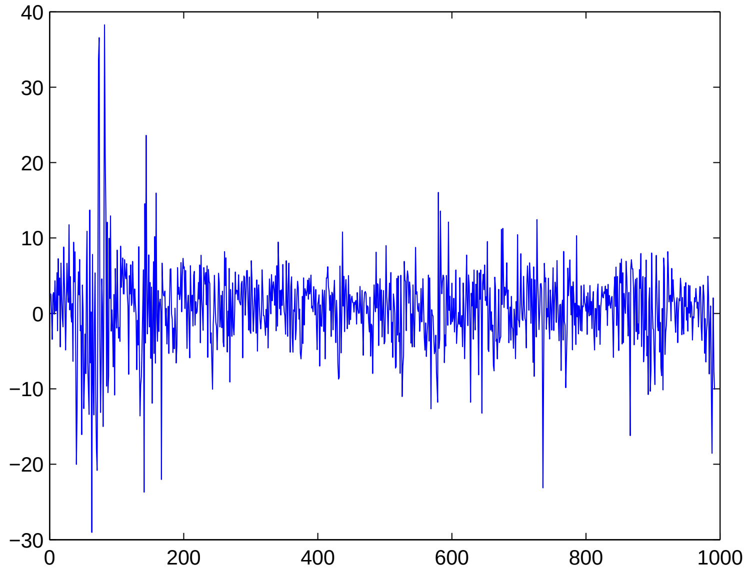


Figure 5.1: Plots of y_t (the monthly excess return).

The range of the data considered is from July 1926 to February 2009 (totally 992 observations). The return series $\{y_t\}_{t=1}^{992}$ is plotted in Figure 5.1. For comparison, we use $\{y_t\}_{t=1}^{900}$ to estimate model (5.3) and leave $\{y_t\}_{t=901}^{992}$ for out-of-sample forecasts. To get a reasonable parameter space, we first fit the data $\{y_t\}_{t=1}^{900}$ by (5.1) and (5.3) with $\delta(\cdot)$ being a constant. Their results are:

$$\begin{aligned} y_t &= 0.0291h_t + \varepsilon_t, \varepsilon_t = e_t\sqrt{h_t}, \\ e_t &\sim i.i.d(0,1), h_t = 0.9307 + 0.1161\varepsilon_{t-1}^2 + 0.8516h_{t-1}. \end{aligned} \tag{5.7}$$

$$\begin{aligned} y_t &= 0.0207h_t + \varepsilon_t, \varepsilon_t = e_t\sqrt{h_t}, \\ e_t &\sim i.i.d(0,1), h_t = 0.7288 + 0.1523y_{t-1}^2 + 0.8306h_{t-1}. \end{aligned} \tag{5.8}$$

Based on the above estimates, we may reasonably set Θ as $\omega \in [0.001, 10], \alpha \in [0.001, 0.9]$ and $\beta \in [0.001, 0.99]$. Applying steps 1-4 in Section 5.2, we get the estimation:

$$\begin{aligned} y_t &= \delta(y_{t-1})h_t + \varepsilon_t, \varepsilon_t = e_t\sqrt{h_t}, \\ e_t &\sim i.i.d(0,1), h_t = 0.8192 + 0.0977y_{t-1}^2 + 0.8673h_{t-1}. \end{aligned} \tag{5.9}$$

We also get the test statistic $\Lambda_{900} = 0.0987$ and the 90%, 95%, 97.5%, 99% percentiles of the bootstrap test statistic sequence $\{\Lambda_{900,i}^*\}_{i=1}^{1000}$ (based on 1000 re-samplings) to be 0.0459, 0.0568, 0.0640 and 0.0822. It is found the calculated Λ_{900} is greater than all the above percentiles, leading to the rejection of $H_0 : \delta(\cdot) = \delta$.

The RMSEs and MAEs for both in-sample and out-of-sample forecasts are presented in Table 5.2. It is found that model (5.8) has similar fitting performance to that of

Table 5.2: In-sample and out-of-sample forecast performance.

Model	RMSE (in)	MAE (in)	RMSE (out)	MAE (out)
(5.7)	5.5502	3.8918	4.7471	3.4076
(5.8)	5.5271	3.8825	4.6851	3.4014
(5.9)	5.2886	3.7397	4.5573	3.3579

model (5.7) for the parametric case. Our results empirically justify the modification for the conditional variance of (5.3). On the whole, the FC-GARCH-M model (5.9) has the best forecasting performance. To illustrate the results graphically, we plot part of the in-sample forecasts in Figure 5.2 (to have a better graph, we do not plot the forecasted values from (5.7) whose performance is comparable to that of (5.8)). From the figure, it is shown that model (5.9) can better capture the variation of the excess return series compared to the model (5.8).

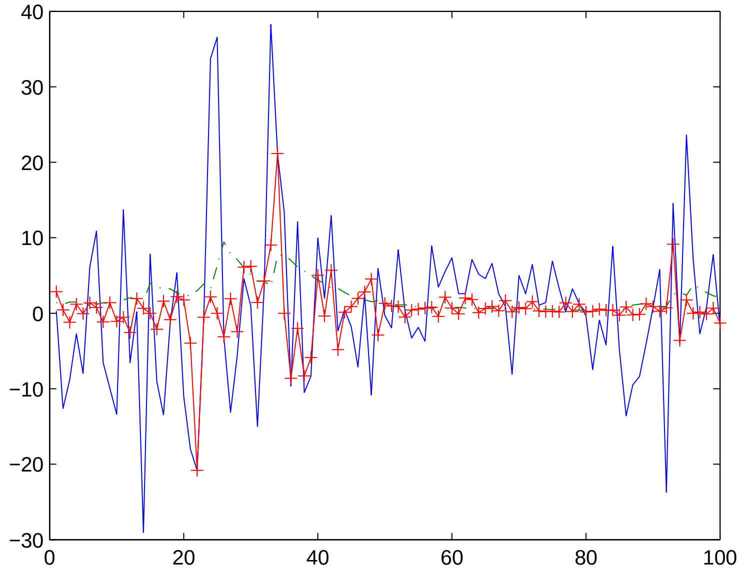


Figure 5.2: Plots of $y_{t=50}^{151}$ $(-)$ and its corresponding in-sample forecasts from model (5.8) $(-\cdot)$, (5.9) $(-+)$.

It is of some interest to study the relationship between the volatility coefficient δ and the time-lagged excess return y_{t-1}. Set a grid vector for y_{t-1} as $u_0 = (-3 : 0.1 : 4)$. Based on $\{y_t\}_{t=1}^{900}$ and model (5.9), the estimate of each $\delta(u_0(i))\,(i = 1, \cdots, 71)$ can be gotten via (5.4). We plot the estimated $\hat{\delta}(\cdot)$ in Figure 5.3 (solid line). It is shown that the obtained $\hat{\delta}(\cdot)$ is nonconstant and can take positive or negative values. Such a result is consistent with those of Backus and Gregory (1993) and Rossi and Timmermann (2010). Namely, the relation between equity premium and conditional volatility is unrestricted with increasing, decreasing, flat or non-monotonic patterns. To compare, we also plot the estimated value of $\delta = 0.0207$ (dotted line) in model (5.8) in the same figure. For

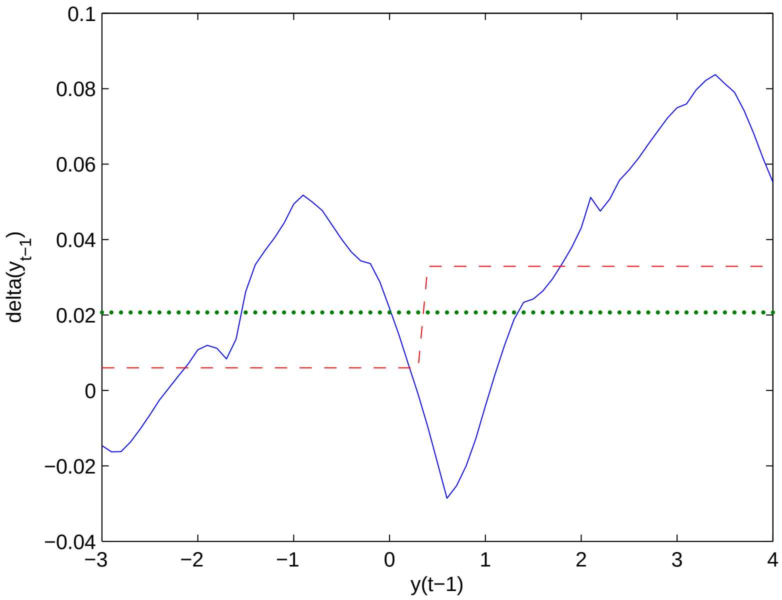


Figure 5.3: Plots of estimated volatility coefficient for model (5.8) (dotted line), (5.9) (solid line) and (5.10) (dashed line).

certain $r \in (0, 1)$, it is shown that when $y_{t-1} < r$, about two thirds of the values $\hat{\delta}(\cdot)$ are under the dotted line of $\delta = 0.0207$. On the contrary, when $y_{t-1} > r$, more than two thirds of the values $\hat{\delta}(\cdot)$ are above the dotted line of $\delta = 0.0207$. The results suggest there may exist some effect of asymmetry.

To give a further understanding, we suppose the unknown function $\delta(\cdot)$ has a form of $\delta(y_{t-1}) = \delta_1 I(y_{t-1} > r) + \delta_2 I(y_{t-1} \leq r)$, where we apply the idea of Tong's (1990) SETAR model to capture the asymmetric effect. Following the idea of Li and Lam (1995), we can estimate the threshold parameter r and other parameters in the GARCH equation. Assuming the threshold r belongs to the interval between 25% and 75% percentiles of y_{t-1}, based on $\{y_t\}_{t=1}^{900}$, the estimates are

$$y_t = [0.0329I(y_{t-1} > 0.3950) + 0.0060I(y_{t-1} \leq 0.3950)]\, h_t + \varepsilon_t,$$
$$\varepsilon_t = e_t\sqrt{h_t}, e_t \sim i.i.d(0, 1), h_t = 1.3445 + 0.1347y_{t-1}^2 + 0.8159h_{t-1}. \tag{5.10}$$

We see $\hat{r} = 0.3950 \in (0, 1), \hat{\delta}_1 = 0.0329 > \hat{\delta}_2 = 0.0060$. This seems reasonable according to Figure 5.3. Applying (5.10) to do the in-sample and out-of-sample forecasts, the computed RMSEs are 5.5172, 4.6074 respectively and the corresponding MAEs are 3.8792, 3.3585. From results presented in Table 5.2, the forecasting performance of (5.10) is seen more satisfactory than (5.7-5.8) but not as good as (5.9). The calculated RMSEs and MAEs for (5.10) give empirical evidences for its usefulness in data modelling. However, here we only take advantage of model (5.10) to justify the proposed FC-GARCH-M model. The properties of estimators and threshold effect test about model (5.10) can be interesting topics for future research.

We have also added the estimated $\delta_1 = 0.0329, \delta_2 = 0.006$ of (5.10) in Figure 5.3 (dashed line). By considering the volatility coefficient δ as a measure of risk aversion, the above asymmetric effect can be explained by a common phenomenon in psychology: When the acquired return is small (less than 0.3950 for example, which usually happens when the stock is at a low price), people are not that risk averse and they tend to take risk to get higher returns. Once a high return (larger than 0.3950 for example, which usually means the stock is at a high price already) has been obtained, they are easy to become conservative and it will require higher premium for per unit of risk to attract them to invest on the risky assets.

5.4.2 Rolling estimation for weekly excess return

Next, we apply model (5.3) to the weekly excess return from 05/07/1963 to 27/02/2009 (totally 2383 observations). Since 30/04/1971, for each quarter, we firstly estimate values of δ based on (5.3) with $\delta(\cdot) = \delta$ and (5.1). The previous 400 observations are used to estimate the parameters and totally 165 estimators are obtained. Let $\{\delta_i^n\}_{i=1}^{165}$, $\{\delta_i^o\}_{i=1}^{165}$ be the estimated δ values for each quarter based on (5.3) with $\delta(\cdot) = \delta$ and (5.1) respectively. These estimated results are plotted in Figure 5.4. We can see both the estimated sequences are time-varying rather than constant. These results are consistent with those of Chou et al. (1992) and Das and Sarkar (2010). Moreover it is shown that $\{\delta_i^n\}$ has similar trajectory to that of $\{\delta_i^o\}$ though the latter one is a bit higher.

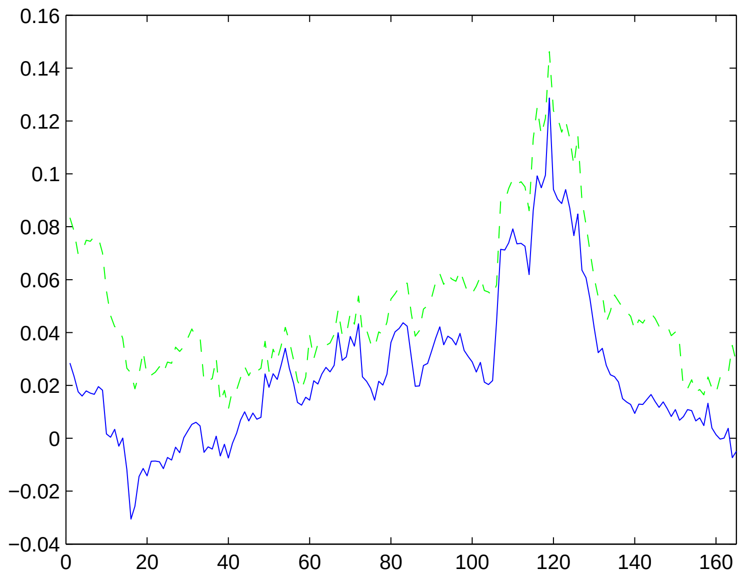


Figure 5.4: Plots of $\{\delta_i^n\}_{i=1}^{165}$ (solid line) and $\{\delta_i^o\}_{i=1}^{165}$ (dashed line).

To show the superiority of model (5.3), we also fit each group of data that is used to estimate the constant δ. The corresponding fitting errors, RMSEs and MAEs, are also recorded. Denote $\{RE_i^n\}_{i=1}^{165}$, $\{RE_i^o\}_{i=1}^{165}$, $\{RE_i^f\}_{i=1}^{165}$, $\{ME_i^n\}_{i=1}^{165}$, $\{ME_i^o\}_{i=1}^{165}$, $\{ME_i^f\}_{i=1}^{165}$ to be the

RMSE and MAE sequences, which are respectively acquired from (5.3) with $\delta(\cdot) = \delta$, (5.1) and (5.3). For comparison, we list the percentiles of the error sequences in Table 5.3. It is seen that the performance of considered two parametric models are similar, and model (5.3) appears to perform most satisfactorily as the fitting errors are much smaller in most cases.

Table 5.3: Percentiles of error sequences.

Error series	Percentiles 10%	25%	50%	75%	90%
$\{RE^n\}$	1.6311	1.9439	2.0694	2.3211	2.5376
$\{RE^o\}$	1.6371	1.9602	2.0836	2.3238	2.5427
$\{RE^f\}$	0.5264	0.6717	1.0186	1.2529	1.3780
$\{ME^n\}$	1.2201	1.4414	1.5891	1.7243	1.8407
$\{ME^o\}$	1.2099	1.4420	1.5874	1.7255	1.8696
$\{ME^f\}$	0.2282	0.3048	0.4984	0.6403	0.8251

5.5 Summary

Motivated by the time-varying risk aversion and the functional coefficient autoregressive model, a functional coefficient autoregressive GARCH-M model is studied. We consider the volatility coefficient in a modified GARCH-M model as an unknown function of the time lagged excess return. Such a setting is useful to seek for the relationship between present risk aversion and the previous return. An approach is given to estimate the parameters and the unknown function. From the simulation studies, it seems that the considered method works effectively. Empirical studies show that the proposed model can capture the variation of the excess return series well. Moreover, the model can also shed insight on the choice of some potential parametric models.

Chapter 6

A Functional Coefficient GARCH-M Model

6.1 Background

Motivated by the FCA-GARCH-M model (5.3), in this chapter, we further consider the following functional coefficient GARCH-M (FC-GARCH-M) model of the form

$$y_t = m(U_t)h_t + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t},$$
$$e_t \sim i.i.d(0, 1), h_t = \upsilon(y_{t-1}, \phi) + \beta h_{t-1}. \tag{6.1}$$

Here $\{y_t, U_t\}$ are observable series and (y_s, U_s) is independent of $\{e_t\}$ for $t > s$. $\theta = (\phi^\tau, \beta)^\tau$ is the unknown parameter vector and $m(\cdot)$ is an unknown smooth function. All throughout this chapter, the superscript τ denotes the transpose of a vector or a matrix. $\upsilon(\cdot, \phi)$ is a known positive function and, as Yang (2006) or Engle and Ng (1993), it can be considered as an "impact news curve". For the specified conditional variance h_t of

(6.1), the original ε_{t-1} is replaced by the observable y_{t-1}. Similar to model (5.3), such a modification for h_t is helpful to estimating the model because h_t is deterministic once the parameters ϕ and β are known. If we set $U_t = y_{t-1}, \phi = (\omega, \alpha)^\tau, \upsilon(y, \phi) = \omega + \alpha y^2$, then (6.1) is reduced to (5.3). When h_t is observable, model (6.1) becomes the common functional coefficient model which has been well discussed by Cai et al. (2000) and Cai et al. (2009).

This chapter is arranged as follows. In Section 6.2, we explain the idea about estimation and some related asymptotic results are given. Simulation and empirical studies are shown in Section 6.3. Detailed proof for Theorem 6.2 is presented in Section 6.4 and we summarize the chapter in Section 6.5.

6.2 Estimation

For the considered model (6.1), there are two unknown parts: $m(\cdot)$ and $\theta = (\phi^\tau, \beta)^\tau$. In this section, we describe the estimation in two subsections: the first is estimating the function provided the parameters are known and the second is to estimate the parameters consistently.

6.2.1 Estimating the function

Suppose that the true values of the parameters ϕ, β are known, say ϕ_0, β_0, then

$$h_t := v(y_{t-1}, \phi_0) + \beta_0 h_{t-1} = \sum_{l=1}^{\infty} \beta_0^{l-1} v(y_{t-l}, \phi_0) \tag{6.2}$$

can be considered as an observable quantity provided that y_0, h_0 are given. We know

$$E(y_t|U_t = u, h_t = h) = m(u)h, \mathrm{var}(y_t|U_t = u, h_t = h) = h. \tag{6.3}$$

As mentioned before, when h_t is observable, model (6.1) becomes the common functional coefficient model of Cai et al. (2000). Hence, as in Cai et al. (2000), we can apply the local linear smoothing technique to estimate the unknown function $m(\cdot)$. Assume the process $\{U_t\}$ to have a continuous pdf $f(u)$ satisfying $\inf_{u\in A} f(u) > 0$, where A is a compact subset of R with nonempty interior. For any fixed $u \in A$, based on (10-11) of Cai et al. (2000), define the estimator of $m(u)$ in (6.3) by

$$\hat{m}(u) = E_0^\tau (Z^\tau W Z)^{-1} Z^\tau W Y, \tag{6.4}$$

where

$$Z = \begin{pmatrix} h_1 & h_1(\frac{U_1-u}{b}) \\ \vdots & \vdots \\ h_n & h_n(\frac{U_n-u}{b}) \end{pmatrix}, W = \mathrm{diag}\begin{pmatrix} \frac{1}{n}k_b(U_1 - u) \\ \vdots \\ \frac{1}{n}k_b(U_n - u) \end{pmatrix},$$

$Y = (y_1, \cdots, y_n)^\tau$ and $E_0 = (1, 0)^\tau$. $b > 0$ is a bandwidth varying with the sample size n, $k_b(\cdot) = \frac{1}{b}k(\frac{\cdot}{b})$ with $k(\cdot)$ being a compactly supported and symmetric kernel function. For the sake of convenience, we put

$$\mu_j = \int_{-\infty}^{+\infty} x^j k(x)dx,\ \upsilon_j = \int_{-\infty}^{+\infty} x^j k^2(x)dx,\ \sigma_j(u) = E(h_t^j|U_t = u),$$
$$c_0 = \mu_2/(\mu_2 - \mu_1^2) \text{ and } c_1 = -\mu_1/(\mu_2 - \mu_1^2). \tag{6.5}$$

Based on Cai et al. (2000), under conditions A and B in Section 6.4, we know (6.4) behaves like the standard functional coefficient estimator. Hence, we state the following theorem without proof.

Theorem 6.1 *Under conditions A and B in Section 6.4, for any fixed $u \in A$, the estimator $\hat{m}(u)$ in (6.4) satisfies*

$$\sqrt{nb}\{\hat{m}(u) - m(u) - \frac{b^2}{2}\frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2} m''(u)\} \xrightarrow{D} N(0, \Sigma(u)), \tag{6.6}$$

where $\Sigma(u) = \frac{c_0^2 \upsilon_0 + 2c_0c_1\upsilon_1 + c_1^2\upsilon_2}{f(u)} \times \frac{\sigma_3(u)}{\sigma_2^2(u)}$ *and* $f(u)$ *is given in condition A.2 in Section 6.4.*

Remark 6.1 One can find detailed proof in Cai et al. (2000). The expression of $\Sigma(u)$ can be easily obtained by incorporating Theorem 2 of Cai et al. (2000) and the equality $\mathrm{var}\,(y_t|U_t = u, h_t = h) = h$ in (6.3).

6.2.2 Estimating the parameters

In practice, the parameters in (6.1) are unknown. Now we give a method to estimate them. Let $\theta = (\phi^\tau, \beta)^\tau$, $\phi = (\phi_1, \cdots, \phi_r)^\tau$. For simplicity, suppose that θ lies in the interior of $\Theta = [\phi_{1L}, \phi_{1U}] \times \cdots \times [\phi_{rL}, \phi_{rU}] \times [\beta_L, \beta_U]$, where $0 < \beta_L < \beta_U < 1$. For $\theta \in \Theta$, define

$$h_t(\theta) = v(y_{t-1}, \phi) + \beta h_{t-1}(\theta) = \textstyle\sum_{l=1}^{\infty} \beta^{l-1} v(y_{t-l}, \phi), \tag{6.7}$$

and we assume the conditional mean of y_t given $U_t = u, h_t(\theta) = h$ has the following form

$$E(y_t|U_t = u, h_t(\theta) = h) = m(\theta, u)h. \tag{6.8}$$

Treating $\{y_t, U_t, h_t(\theta)\}$ as observable processes, (6.8) enables us to estimate $m(\theta, u)$ in the framework of (1-2) of Cai et al. (2000). Define

$$L(\theta) := E\left\{\left[\log h_t(\theta) + \frac{[y_t - m(\theta, U_t)h_t(\theta)]^2}{h_t(\theta)}\right]\pi(U_t)\right\}, \tag{6.9}$$

where $\pi(\cdot)$ is a nonnegative weight function whose compact support is contained in A. When $n \to \infty$, it is known that $L(\theta)$ can be consistently estimated by

$$L_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\left\{\log h_t(\theta) + \frac{[y_t - m(\theta, U_t)h_t(\theta)]^2}{h_t(\theta)}\right\}\pi(U_t). \tag{6.10}$$

Under condition C.6 in Section 6.4, $L(\theta)$ has an unique minimum point at θ_0 and is locally convex. Hence the minimizer of $L_n(\theta)$ should locate the true parameter θ_0 consistently under some regularity conditions. However, in practice, we still have no idea about $m(\theta, U_t)$ and hence the minimizer of $L_n(\theta)$ is not practicable. To obtain a feasible estimator, for each $u \in A, \theta \in \Theta$, define the estimator of $m(\theta, u)$ in (6.8) by

$$\hat{m}(\theta, u) = E_0^{\tau}(Z_\theta^{\tau} W Z_\theta)^{-1} Z_\theta^{\tau} W Y, \tag{6.11}$$

where

$$Z_\theta = \begin{pmatrix} h_1(\theta) & h_1(\theta)(\frac{U_1-u}{b}) \\ \vdots & \vdots \\ h_n(\theta) & h_n(\theta)(\frac{U_n-u}{b}) \end{pmatrix}.$$

Recall that the estimator $\hat{m}(\theta, u)$ in (6.11) is based on the assumption (6.8). Define next

$$\hat{L}_n(\theta) := \frac{1}{n}\sum_{t=1}^{n}\left\{\log h_t(\theta) + \frac{[y_t - \hat{m}(\theta, U_t)h_t(\theta)]^2}{h_t(\theta)}\right\}\pi(U_t), \tag{6.12}$$

$$\hat{\theta}_n := \arg\min_{\theta\in\Theta} \hat{L}_n(\theta). \tag{6.13}$$

Based on $\{y_t,\ U_t\}_{t=1}^n$ and given y_0, $h_0(\theta)$, we can get the minimizer $\hat{\theta}_n$. The above estimator $\hat{\theta}_n$ is similar to that adopted in Chapter 5 if we set $\pi(U_t) = 1$ for $t = 1, \cdots, n$. However, the added weight function $\pi(\cdot)$ is not redundant. It is known, for boundary point of $\{U_t\}$, the estimator in (6.11) may be seriously biased because of inadequate observations. By appropriate choice for $\pi(\cdot)$, we can avoid to use the estimates of $m(\theta, \cdot)$ for boundary points in computing (6.12). The added $\pi(U_t)$ is also useful to establish the following theorem, from which, we know $\hat{\theta}_n$ converges to the true value in probability under some regularity conditions.

Theorem 6.2 *Under conditions A and C in Section 6.4, as* $n \to \infty, b \to 0$, *we have* $\hat{\theta}_n \xrightarrow{P} \theta_0$.

Proof: See Section 6.4.

After we have obtained $\hat{\theta}_n$, the value of $m(u)$ can be approximated by $\hat{m}(\hat{\theta}_n, u)$. The related confidence intervals can be computed based on Theorem 6.1 by treating $\hat{\theta}_n$ as the true value θ_0. In practice, to get the estimator of (6.13), we need to minimize (6.12) with respect to $\theta \in \Theta$. By generalizing the idea of Linton and Perron (2003), the following procedure can be used to obtain the minimizer:

Step 1: Provide a set of initial parameters $\hat{\theta}^{(i)}(i = 0)$, compute $h_t(\hat{\theta}^{(i)})$ for $t = 1, \cdots, n$ according to (6.7)[1].

[1]Initial value for θ can be obtained by estimating model (6.1) with $m(\cdot)$ being a constant δ and the sample variance of $\{y_t\}_{t=1}^T$ can be used as an initial value for h_t or $h_t(\theta)$.

Step 2: Based on the sequence $\{y_t, U_t, h_t(\hat{\theta}^{(i)})\}_{t=1}^n$, calculate

$$\left\{\hat{m}(\hat{\theta}^{(i)}, U_1), \cdots, \hat{m}(\hat{\theta}^{(i)}, U_n)\right\}$$

in terms of (6.11).

Step 3: Update $\hat{\theta}^{(i)}$ to $\hat{\theta}^{(i+1)}$ by performing weighted quasi maximum likelihood estimation (WQMLE) on a GARCH (1, 1) model

$$\begin{aligned}\hat{\varepsilon}_t^{(i)} &= \sqrt{h_t} e_t, \\ e_t \sim i.i.d(0, 1), h_t(\theta) &= \nu(y_{t-1}, \phi) + \beta h_{t-1}(\theta),\end{aligned}$$

where, $\hat{\varepsilon}_t^{(i)} = y_t - \hat{m}(\hat{\theta}^{(i)}, U_t) h_t(\hat{\theta}^{(i)}),\ t = 1, \cdots, n$. Namely, consider $\{\hat{\varepsilon}_t^{(i)}\}$ as an observable series and acquire $\hat{\theta}^{(i+1)}$ by minimizing

$$\hat{L}_n^{(i)}(\theta) = \frac{1}{n} \sum_{t=1}^n \left[\log h_t(\theta) + \frac{(\hat{\varepsilon}_t^{(i)})^2}{h_t(\theta)}\right] \pi(U_t)$$

with respect to $\theta \in \Theta$.

Step 4: Repeat steps 1-3 until convergence.

6.3 Simulations and Empirical Studies

6.3.1 Simulations

In this section, we give four simulation examples to demonstrate that the proposed estimators for the parameters and functions are satisfactory. The series are generated from the model

$$y_t = m(U_t)h_t + \varepsilon_t, \varepsilon_t = e_t\sqrt{h_t},$$

$$e_t \sim i.i.d\ \ N(0,1), h_t = \upsilon(y_{t-1},\phi) + \beta h_{t-1}.$$

The cases considered are:

Ex 6.1: $m(x) = 0.4 - (0.659 + 1.26x)e^{-3.1x^2},\ \upsilon(y,\phi) = \omega + \alpha[1 + \eta I(y \le 0)]y^2,$

$U_t = y_{t-1},\ (\omega,\alpha,\eta,\beta)^\tau = (0.1, 0.12, -0.3, 0.55)^\tau.$

Ex 6.2: $m(x) = 0.2sin^2(\pi x),\ \upsilon(y,\phi) = \omega + \alpha y^2,$

$U_t \sim i.i.d\ \ U(-1,1),\ (\omega,\alpha,\beta)^\tau = (0.5, 0.1, 0.75)^\tau.$

Ex 6.3: $m(x) = 0.3I(x > 0) - 0.3I(x \le 0),\ \upsilon(y,\phi) = \omega + \alpha y^2, U_t = 0.8U_{t-1} + v_t,$

$v_t \sim i.i.d\ \ N(0, 0.36),\ (\omega,\alpha,\beta)^\tau = (0.2, 0.2, 0.3)^\tau.$

Ex 6.4: $m(x) = 0.7x,\ \upsilon(y,\phi) = \omega + \alpha|y|, U_t = 0.6U_{t-1} + 0.3U_{t-2} + v_t,$

$v_t \sim i.i.d\ \ N(0, 0.64),\ (\omega,\alpha,\beta)^\tau = (0.08, 0.1, 0.8)^\tau.$

When applying steps 1-4 of Section 6.2 to estimate Ex 6.1-6.4, the kernel function $k(\cdot)$ in (6.11) is chosen as $k(x) = 0.75(1 - x^2)I(|x| \le 1)$. For the choice of the bandwidth b, as described in Chapter 5, we adopt $AMS(b)$-minimized method of Cai et al. (2000) to save the cost of computation. For the weight function $\pi(\cdot)$ in (6.12), following Yang (2002), we put $\pi(u) = I(u \in A)$, where A is set to be the interval between the 10th and 90th percentiles of the explanatory variable U_t. For the estimation of the system parameters, we set $\omega \in [0.005, 1], \alpha \in [0.005, 0.9], \eta \in [-1, 1]$ and $\beta \in [0.1, 0.99]$. The convergence criterion is $||\hat{\theta}_n^{(i+1)} - \hat{\theta}_n^{(i)}|| \le 0.0001$. We conduct 500 replications with sample size $n = 400, 800, 1200$ respectively for each of the above examples.

For the part of parameter estimates, we calculate their percentiles and standard deviations. The results are displayed in Table 6.1 and Table 6.2. In the tables, Q10, Q50 and Q90 denote respectively the 10%, 50% and 90% percentiles of the estimators, and SD the standard deviation. For the part of function estimation, under sample size $n = 1200$, we plot the estimated curve $\hat{m}(u)$ with median performance (whose RMSE is close to the median among 500 replications) and the true function $m(u)$ in the same figure. The plots are displayed in the left column of Array 6.1 (dashed line is the estimated curve and solid line is the true curve). For comparison, in the right column of Array 6.1, we give box plots for the RMSE sequences for $T = 400, 800, 1200$ (from left to the right).

From Table 6.1 and Table 6.2, we can see the medians are close to the true values in most cases. The standard deviation becomes smaller when the sample size gets larger. From the estimated curves in the left column of Array 6.1, it is seen that the estimated functions with median performance can capture the true trends. The box plots of RMSEs in the right column witness a better fitting performance when the sample size gets larger. These suggest that the considered estimators perform well.

Table 6.1: Results of the parameter estimation for Ex 6.1-6.2.

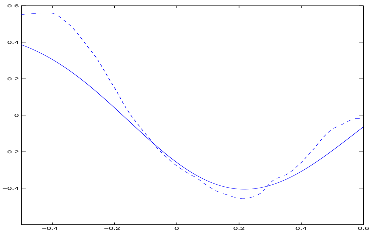
Example	True value	Q10	Q50	Q90	SD
	$\omega = 0.1$	0.0245	0.1146	0.2140	0.0700
Ex 6.1	$\alpha = 0.12$	0.0239	0.1264	0.3862	0.1515
(N=400)	$\eta = -0.3$	-1.0000	-0.3645	1.0000	0.8193
	$\beta = 0.55$	0.1000	0.4556	0.8515	0.2835
	$\omega = 0.1$	0.0381	0.1131	0.2040	0.0600
Ex 6.1	$\alpha = 0.12$	0.0350	0.1233	0.3145	0.1092
(N=800)	$\eta = -0.3$	-1.0000	-0.3014	1.0000	0.7240
	$\beta = 0.55$	0.1000	0.4940	0.8091	0.2451
	$\omega = 0.1$	0.0361	0.1021	0.1918	0.0562
Ex 6.1	$\alpha = 0.12$	0.0393	0.1189	0.2529	0.0864
(N=1200)	$\eta = -0.3$	-0.9749	-0.2819	1.0000	0.6242
	$\beta = 0.55$	0.1537	0.5372	0.8197	0.2289
Ex 6.2	$\omega = 0.5$	0.1703	0.5713	1.0000	0.3112
(N=400)	$\alpha = 0.1$	0.0302	0.0819	0.1539	0.0492
	$\beta = 0.75$	0.5683	0.7304	0.8855	0.1227
Ex 6.2	$\omega = 0.5$	0.2658	0.5464	1.0000	0.2580
(N=800)	$\alpha = 0.1$	0.0545	0.0954	0.1450	0.0370
	$\beta = 0.75$	0.5855	0.7300	0.8465	0.0978
Ex 6.2	$\omega = 0.5$	0.2957	0.5462	0.9548	0.2312
(N=1200)	$\alpha = 0.1$	0.0596	0.0952	0.1384	0.0303
	$\beta = 0.75$	0.6045	0.7371	0.8349	0.0860

Note: Number of replications=500.

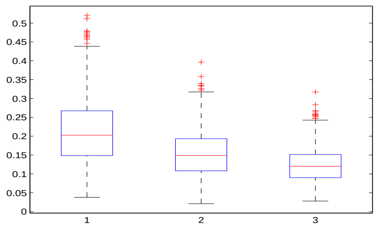
Table 6.2: Results of the parameter estimation for Ex 6.3-6.4.

Example	True value	Q10	Q50	Q90	SD
Ex 6.3	$\omega = 0.2$	0.0973	0.2020	0.2811	0.0723
(N=400)	$\alpha = 0.2$	0.0828	0.1743	0.2814	0.0780
	$\beta = 0.3$	0.1000	0.2724	0.6053	0.2061
Ex 6.3	$\omega = 0.2$	0.1239	0.1978	0.2763	0.0590
(N=800)	$\alpha = 0.2$	0.1074	0.1869	0.2649	0.0603
	$\beta = 0.3$	0.1000	0.3063	0.5305	0.1698
Ex 6.3	$\omega = 0.2$	0.1321	0.1996	0.2677	0.0510
(N=1200)	$\alpha = 0.2$	0.1267	0.1903	0.2558	0.0493
	$\beta = 0.3$	0.1000	0.3037	0.5073	0.1460
Ex 6.4	$\omega = 0.08$	0.0438	0.1445	0.5407	0.1961
(N=400)	$\alpha = 0.1$	0.0493	0.1349	0.2744	0.0886
	$\beta = 0.8$	0.1005	0.6827	0.8600	0.2305
Ex 6.4	$\omega = 0.08$	0.0489	0.1036	0.2056	0.0879
(N=800)	$\alpha = 0.1$	0.0697	0.1192	0.1961	0.0557
	$\beta = 0.8$	0.5919	0.7455	0.8422	0.1219
Ex 6.4	$\omega = 0.08$	0.0547	0.0983	0.1682	0.0576
(N=1200)	$\alpha = 0.1$	0.0729	0.1149	0.1706	0.0423
	$\beta = 0.8$	0.6599	0.7599	0.8377	0.0817

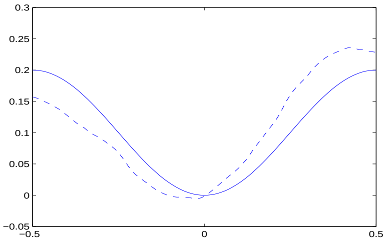
Note: Number of replications=500.



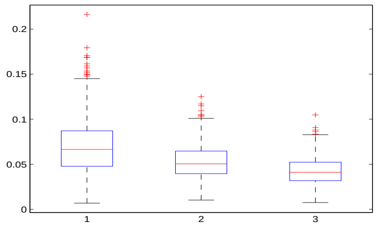
Ex 6.1: $\hat{m}(u)$(- -) and $m(u)$(-)



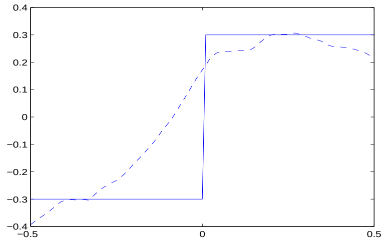
Ex 6.1: Box plot of RMSE



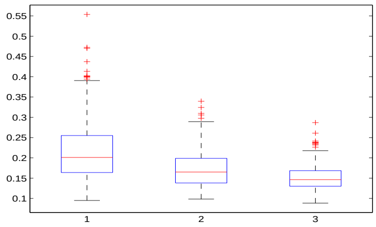
Ex 6.2: $\hat{m}(u)$(- -) and $m(u)$(-)



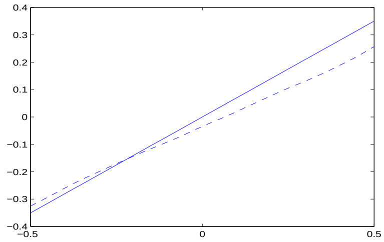
Ex 6.2: Box plot of RMSE



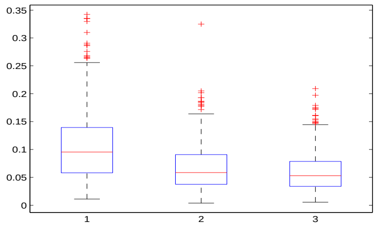
Ex 6.3: $\hat{m}(u)$(- -) and $m(u)$(-)



Ex 6.3: Box plot of RMSE



Ex 6.4: $\hat{m}(u)$(- -) and $m(u)$(-)



Ex 6.4: Box plot of RMSE

Array 6.1: Results of the function estimation for Ex 6.1-6.4.

6.3.2 Empirical studies

In this section, we apply model (6.1) to study some real data sets. We analyze the monthly and weekly excess returns on the CRSP value weighted indices, including NYSE, AMEX and NASDAQ. These data can be regarded as reasonable proxies for the stock market and they have been studied by other researchers. The riskless rate used to compute the excess returns is one-month Treasury bill rate (from Ibbotson Associates).

To apply model (6.1), we set $U_t = y_{t-1}, \upsilon(y, \phi) = \omega + \alpha[1 + \eta I(y \le 0)]y^2$. As in Chapter 5, this setting for U_t enables us to understand the impact of previous return on the present risk aversion. The specified conditional variance takes the asymmetric effect into account, namely positive and negative returns will cause asymmetric fluctuations. Firstly, we analyze the monthly excess returns from July 1926 to February 2009 (totally 992 observations). Take $\{y_t\}_{t=1}^{992}$ to be the considered excess return series. For comparison, we use $\{y_t\}_{t=1}^{900}$ to estimate model (6.1) and leave $\{y_t\}_{t=901}^{992}$ for out-of-sample forecasts. Before estimating the model (6.1), we display the following two fitted models based on $\{y_t\}_{t=1}^{900}$:

$$y_t = 0.0291h_t + \varepsilon_t, \varepsilon_t = e_t\sqrt{h_t},$$

$$e_t \sim i.i.d(0,1), h_t = 0.9307 + 0.1161\varepsilon_{t-1}^2 + 0.8516h_{t-1}, \tag{6.14}$$

$$y_t = 0.0207h_t + \varepsilon_t, \varepsilon_t = e_t\sqrt{h_t},$$

$$e_t \sim i.i.d(0,1), h_t = 0.7288 + 0.1523y_{t-1}^2 + 0.8306h_{t-1}. \tag{6.15}$$

Here (6.14) is the traditional GARCH-M model, (6.15) is the FC-GARCH-M model with $m(\cdot) = m$, a constant, and $\upsilon(y, \phi) = \omega + \alpha y^2$. Based on the above two results, for estimating (6.1) with $U_t = y_{t-1}, \upsilon(y, \phi) = \omega + \alpha[1 + \eta I(y \le 0)]y^2$, we set the parametric space Θ as $\omega \in [0.001, 10], \alpha \in [0.001, 0.9], \eta \in [0.001, 1000], \beta \in [0.1, 0.99]$. Applying steps 1-4 in Section 6.2, we shall get the estimates with the basis of $\{y_t\}_{t=1}^{900}$:

$$y_t = m(y_{t-1})h_t + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t}, e_t \sim i.i.d(0, 1),$$

$$h_t = 0.8442 + 0.0532[1 + 2.4782I(y_{t-1} \leq 0)]y_{t-1}^2 + 0.8663h_{t-1}. \tag{6.16}$$

To compare, we also estimate the TVP-GARCH-M model (5.2) based on $\{y_t\}_{t=1}^{900}$. The result is

$$y_t = \delta_t h_t + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t}, \delta_t = \delta_{t-1} + v_t,$$

$$e_t \sim i.i.d \;\; N(0, 1), h_t = 0.7472 + 0.1176\eta_{t-1}^2 + 0.8592h_{t-1}. \tag{6.17}$$

The estimated value of Q (the variance of v_t in (6.17)) is 2.2×10^{-6} which is quite small. This implies that (6.17) is close to the original model (6.14) because model (5.2) converges to (5.1) when Q becomes small (Chou et al. 1992). The mean of the estimated δ_t is 0.024 which is near the fixed value 0.029 in (6.14).

We also tabulate the RMSEs and MAEs for both the in-sample and out-of-sample forecasts in Table 6.3. For coefficient-fixed cases, it is found that models (6.14-6.15)

Table 6.3: In-sample and out-of-sample forecast performance.

Model	RMSE (in)	MAE (in)	RMSE (out)	MAE (out)
(6.14)	5.5524	3.8926	4.7356	3.4048
(6.15)	5.5295	3.8837	4.6692	3.3955
(6.16)	5.2130	3.7257	4.5455	3.3648
(6.17)	5.5275	3.8360	4.6518	3.3936

have comparable performance which empirically justifies the modification of the conditional variance that is driven by the past returns. For the time-varying cases, (6.16) has better fitting performance than that of model (6.17). On the whole, model (6.16) has the best forecasting performance for the considered data.

It is of some interest to study the relationship between the volatility coefficient and the time-lagged excess return y_{t-1}. Set a grid vector for y_{t-1} as $u_0 = (-3 : 0.1 : 4)$. Based on $\{y_t\}_{t=1}^{900}$ and (6.16), we can estimate the value of each $m(u_0(i))\,(i = 1, \cdots, 71)$ according to (6.4). The estimated $\hat{m}(\cdot)$ (solid line) and the related 95% confidence band (dashed line) are plotted in Figure 6.1. From the plots, it is seen that the estimated $\hat{m}(\cdot)$ is similarly shaped to that of Figure 5.3 in Chapter 5. Hence, we can adopt analogous interpretation for the plots.

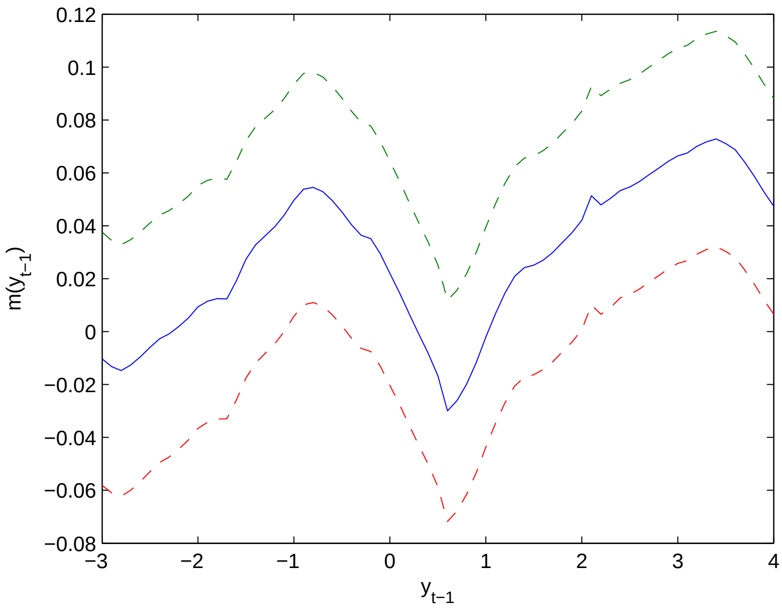


Figure 6.1: Plots of $\hat{m}(y_{t-1})$ from (6.16) (solid line) and the related confidence band (dashed lines).

Next, we apply model (6.1) to the weekly excess returns from 05/07/1963 to 27/02/2009 (totally 2383 observations). Starting from 30/04/1971, for each quarter, we firstly estimate constant volatility coefficient δ based on (6.1) with $m(\cdot) = \delta, \upsilon(y, \phi) =$

$\omega + \alpha y^2$ and (5.1). The previous 400 observations are used to estimate the parameter and totally 165 estimators are obtained. Let $\{\delta_i^n\}_{i=1}^{165}, \{\delta_i^o\}_{i=1}^{165}$ be the respective estimated coefficient values which are plotted in Figure 6.2. We can see from the plots that the estimated two sequences are time-varying rather than constant. These results are consistent with those in Chou et al. (1992) and Das and Sarkar (2010). Moreover, $\{\delta_i^n\}$ has similar trajectory to that of $\{\delta_i^o\}$ though the latter one is a bit higher.

For each group of the data used to estimate the constant volatility coefficient in the above, we also fit them by (6.1) (with $U_t = y_{t-1}, \upsilon(y, \phi) = \omega + \alpha[1 + \eta I(y \leq 0)]y^2$) and the TVP-GARCH-M model (5.2). The corresponding fitting errors are calculated. Denote $\{RE_i^n\}_{i=1}^{165}$, $\{RE_i^o\}_{i=1}^{165}$, $\{RE_i^f\}_{i=1}^{165}$, $\{RE_i^v\}_{i=1}^{165}$, $\{ME_i^n\}_{i=1}^{165}$, $\{ME_i^o\}_{i=1}^{165}$, $\{ME_i^f\}_{i=1}^{165}$, $\{ME_i^v\}_{i=1}^{165}$ to be the RMSE and MAE sequences, which are respectively recorded from the (6.1) with $m(\cdot) = \delta, \upsilon(y, \phi) = \omega + \alpha y^2$, (5.1), (6.1) with $U_t = y_{t-1}, \upsilon(y, \phi) = \omega + \alpha[1 + \eta I(y \leq 0)]y^2$ and (5.2). For comparison, we list the percentiles of the error sequences in Table 6.4.

Table 6.4: Percentiles of error sequences.

Error series	Percentiles 10%	25%	50%	75%	90%
$\{RE^n\}$	1.5497	1.9206	2.0948	2.3285	2.5447
$\{RE^o\}$	1.5483	1.9243	2.1051	2.3314	2.5481
$\{RE^f\}$	0.6140	0.8271	1.1946	1.4306	1.7550
$\{RE^v\}$	1.5582	1.9220	2.0975	2.3309	2.5551
$\{ME^n\}$	1.1938	1.4320	1.5796	1.7312	1.8669
$\{ME^o\}$	1.1816	1.4311	1.5712	1.7312	1.8786
$\{ME^f\}$	0.2644	0.3833	0.6321	0.8758	1.1117
$\{ME^v\}$	1.2040	1.4373	1.5857	1.7312	1.8835

From the table, similar to the case of the monthly data, it is seen that the two constant-coefficient models have comparable fitting performance. According to the values in the fourth and eighth rows, we can see that the performance of TVP-GARCH-M model is

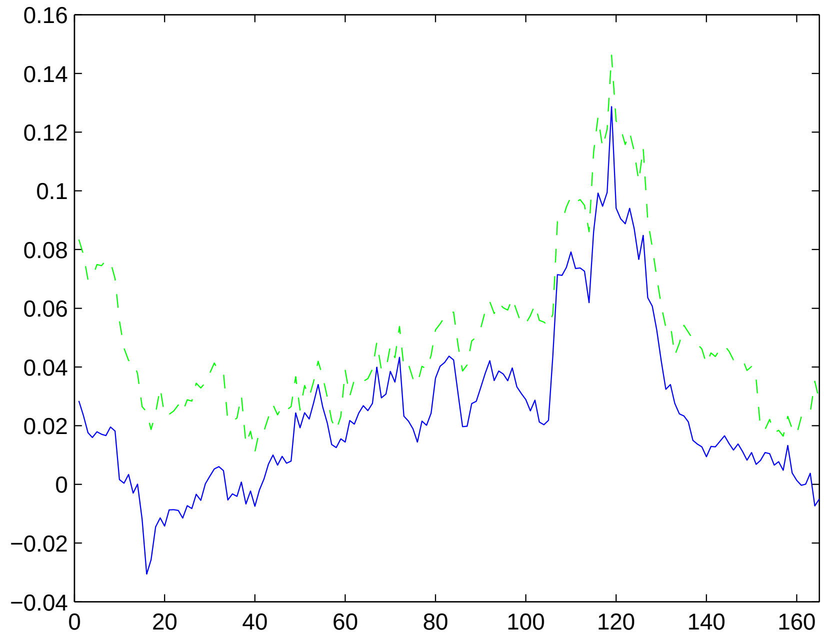


Figure 6.2: Plots of $\{\delta_i^n\}_{i=1}^{165}$ (solid line) and $\{\delta_i^o\}_{i=1}^{165}$ (dashed line).

similar to the two coefficient-fixed models. This implies the volatility coefficient does not necessarily vary in the form of a random walk and can take some other forms. On the whole, the FC-GARCH-M model (6.1) seems to be superior to other models (the fitting errors are much smaller in most cases) for the considered data.

6.4 Proofs

We first state some conditions for model (6.1). Throughout this section, we let M, m denote certain positive constants, which may take different values at different places.

Condition A

A.1 The kernel function $k(\cdot)$ in (6.4) is a bounded density with a bounded support $[-1, 1]$.

A.2 The process $\{U_t\}$ has a continuous pdf $f(u)$ satisfying $\inf_{u\in A} f(u) > 0$, where A is a compact subset of R with nonempty interior. For each $u \in A$, $|f'(x)|$ is bounded when x takes values in a neighborhood of u.

Condition B

B.1 The α-mixing processes $\{(y_t, U_t, h_t)\}$ satisfies $\sum l^c[\alpha(l)]^{1-2/\delta} < \infty$ for some $\delta > 2$, $c > 1 - 2/\delta$, where h_t is given in (6.2).

B.2 $E\{y_0^2 + y_l^2 | U_0 = u_0, h_0 = w, U_l = u_l, h_l = v\} \leq M < \infty$ for all $l \geq 1, w, v \in R, u_0$ and u_l in a neighborhood of u.

B.3 $b \to 0$ and $nb \to \infty$. Further, there exists a sequence of positive integers s_n such that $s_n \to \infty$, $s_n = o(\sqrt{nb})$, $\sqrt{n/b}\alpha(s_n) \to 0$, as $n \to \infty$.

B.4 There exists $\delta^* > \delta$, where δ is given in B.1, such that $E\{|y|^{\delta^*}|U = v, h = w\} \leq M < \infty$ for all $w \in R$ and v in a neighborhood of u, and $\alpha(n) = O(n^{-\delta^{**}})$ where $\delta^{**} \geq \delta\delta^*/\{2(\delta^* - \delta)\}$.

B.5 $Eh_t^{2\delta^*} < \infty$, and $n^{1/2-\delta/4} b^{\delta/\delta^* - 1/2 - \delta/4} = O(1)$.

B.6 Let conditional density of (U_0, U_l) given (h_0, h_l), say $f(u_0, u_l | h_0, h_l)$, satisfies that $|f(u_0, u_l | h_0, h_l)| \leq M < \infty$ for all $l \leq 1$ and $f(u_t | h_t) \leq M < \infty$, where $f(u|h)$ is the conditional density of U given $h_t = h$.

Condition C

C.1 For each $\theta \in \Theta$, the process $\{(y_t, U_t, h_t(\theta))\}$ generated from (6.1) is strictly stationary and ergodic.

C.2 For each $\theta \in \Theta, u \in A$, functions $m(\theta, x)$ in (6.8), $\sigma_2(\theta, x)$ in (6.21), $\|\frac{\partial m(\theta,u)}{\partial \theta}\|, |\frac{\partial m^2(\theta,x)}{\partial x^2}|$ and $|\frac{\partial \sigma_2(\theta,x)}{\partial x}|$ are uniformly bounded when x takes values in a neighborhood of u.

C.3 For any $\theta_1, \theta_2 \in \Theta$, there exist two positive processes $\{w_{1t}\}, \{w_{2t}\}$ (independent of θ_1, θ_2) such that $\|\frac{1}{h_t(\theta_1)} \frac{\partial h_t(\theta_2)}{\partial \theta}\| \leq w_{1t}$ and $h_t(\theta_1)/h_t(\theta_2) \leq w_{2t}$.

C.4 $Ew_{1t} < \infty, Ew_{2t} < \infty, E|y_t| < \infty, Eh_t(\theta_U) < \infty$, where w_{1t}, w_{2t} are given in C.3 and $h_t(\theta_U)$ is described in (6.23).

C.5 The function $\upsilon(y, \phi) \geq m > 0$ holds uniformly.

C.6 The function $L(\theta)$ defined in (6.9) is continuous and has a positive definite Hessian matrix at its unique minimum θ_0.

Remark 6.2 Conditions A and B are basically the same as those in Cai et al. (2000). Conditions C.1, C.2, C.5 and C.6 have been analogously adopted by Yang (2006). C.3 and C.4 are useful for proving uniform convergence in probability for some processes. And they hold for the case of the usual GARCH (1, 1) process (Jensen and Rahbek, 2004), which gives some basis for the assumptions.

Before giving the proof of Theorem 6.2, we firstly state three useful lemmas.

Lemma 6.1 *Let $G_n(\theta) = Q_n(\theta) - Q(\theta)$, where $Q(\theta)$ is a nonrandom function that is continuous in $\theta \in \Theta$, and Θ is a bounded metric space. Suppose for $\theta, \theta_1, \theta_2 \in \Theta$, $Q_n(\theta) \xrightarrow{P} Q(\theta)$, $|Q_n(\theta_1) - Q_n(\theta_2)| \leq B_n\|\theta_1 - \theta_2\|$ and $B_n = O_p(1)$, then we have* $\sup_{\theta\in\Theta} |G_n(\theta)| = o_p(1)$.

Proof: The lemma is a direct result based on Lemma 1 and Theorem 1 in Andrews (1992).

Lemma 6.2 *Suppose $Q_n(\theta)$ is an approximation of nonrandom continuous function $Q(\theta)$, which has an unique minimum θ_0 in Θ. Let*

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} Q_n(\theta),$$

if $\sup_{\theta\in\Theta} |Q_n(\theta) - Q(\theta)| = o_p(1)$, *then we have* $\hat{\theta}_n \xrightarrow{P} \theta_0$.

Proof: The lemma is a direct result following Lemma 14.3 (page 258) and Theorem 2.12 (page 28) in Kosorok (2006).

Lemma 6.3 *Let $\theta \in \Theta$,*

$$A_n(\theta) = \begin{pmatrix} A_{11}(\theta) + a_n^{11}(\theta) & A_{12}(\theta) + a_n^{12}(\theta) \\ A_{21}(\theta) + a_n^{21}(\theta) & A_{22}(\theta) + a_n^{22}(\theta) \end{pmatrix} \textit{ and } A(\theta) = \begin{pmatrix} A_{11}(\theta) & A_{12}(\theta) \\ A_{21}(\theta) & A_{22}(\theta) \end{pmatrix}.$$

For $1 \le i, j \le 2$, *if* $\sup_{\theta\in\Theta} |a_n^{i,j}(\theta)| = o_p(1)$ *and the element of nonsingular matrix* $A(\theta)$ *satisfies* $\sup_{\theta\in\Theta} |A_{i,j}(\theta)| = O_p(1)$, *then we have* $A_n(\theta)^{-1} = A(\theta)^{-1} + \rho_n(\theta)$. *Here* $\rho_n(\theta) = (\rho_n^{i,j}(\theta))_{1\le i,j\le 2}$ *and* $\sup_{\theta\in\Theta} |\rho_n^{i,j}(\theta)| = o_p(1)$.

Proof: Simple algebra yields $|A_n(\theta)| = |A(\theta)| + a_n(\theta)$, $\sup_{\theta\in\Theta} |a_n(\theta)| = o_p(1)$. By definition, we know the elements in the first row and the first column of $A_n(\theta)^{-1}$ and $A^{-1}(\theta)$ are respectively $|A_n(\theta)|^{-1}[A_{22}(\theta) + a_n^{22}(\theta)]$ and $|A(\theta)|^{-1}A_{22}(\theta)$. Let

$$\begin{aligned} \rho_n^{1,1}(\theta) &= |A_n(\theta)|^{-1}[A_{22}(\theta) + a_n^{22}(\theta)] - |A(\theta)|^{-1}A_{22}(\theta) \\ &= \frac{a_n^{2,2}(\theta)|A(\theta)| - A_{2,2}(\theta)a_n(\theta)}{|A(\theta)|^2 + a_n(\theta)|A(\theta)|}, \end{aligned}$$

then we know $\sup_{\theta\in\Theta} |\rho_n^{1,1}(\theta)| = o_p(1)$ when $0 \neq |A(\theta)| = O(1)$. Using similar argument to other elements of $A_n(\theta)^{-1}$ and $A^{-1}(\theta)$, then the proof can be completed.

Proof of Theorem 6.2:

Define

$$S_n(\theta) = Z_\theta^\tau W Z_\theta = \begin{pmatrix} S_{n,0}(\theta) & S_{n,1}(\theta) \\ S_{n,1}(\theta) & S_{n,2}(\theta) \end{pmatrix}, \tag{6.18}$$

where,

$$S_{n,j}(\theta) = \frac{1}{n}\sum_{t=1}^{n} h_t^2(\theta)\left(\frac{U_t - u}{b}\right)^j k_b(U_t - u),\ j = 0, 1, 2. \tag{6.19}$$

For each $\theta \in \Theta$, under condition C.1, we know that $S_{n,j}(\theta) = ES_{n,j}(\theta) + o_p(1)$. Here

$$\begin{aligned} ES_{n,j}(\theta) &= E[h_t^2(\theta)(\frac{U_t - u}{b})^j k_b(U_t - u)] \\ &= E[\sigma_2(\theta, U_t)(\frac{U_t - u}{b})^j k_b(U_t - u)] \\ &= \int_{-\infty}^{+\infty} f(U_t)\sigma_2(\theta, U_t)(\frac{U_t - u}{b})^j k_b(U_t - u)dU_t \\ &= \int_{-1}^{+1} f(bx + u)\sigma_2(\theta, bx + u)x^j k(x)dx \\ &= \mu_j \cdot f(u)\sigma_2(\theta, u) + R(\theta). \end{aligned} \tag{6.20}$$

In the above expressions, $\sigma_i(\theta, u)$ is defined as

$$\sigma_i(\theta, u) = E[h_t^i(\theta)|U_t = u]. \tag{6.21}$$

$R(\theta)$ is given by

$$\begin{aligned} R(\theta) &= bf(u)\int_{-1}^{+1} \frac{\partial\sigma_2(\theta, b\lambda_1 x + u)}{\partial u} x^{j+1}k(x)dx \\ &\quad + b\sigma_2(\theta, u)\int_{-1}^{+1} f'(b\lambda_2 x + u)x^{j+1}k(x)dx \\ &\quad + b^2\int_{-1}^{+1} \frac{\partial\sigma_2(\theta, b\lambda_1 x + u)}{\partial u} f'(b\lambda_2 x + u)x^{j+2}k(x)dx, \end{aligned}$$

where $\lambda_1, \lambda_2 \in [0,1]$. Under conditions A.2 and C.2, we know that

$$f(x),\ f'(x),\ \sigma_2(\theta, x) \text{ and } \partial\sigma_2(\theta, x)/\partial x$$

are uniformly bounded in a neighborhood of u. As a result, it is easy to have

$$\sup_{\theta\in\Theta} |R(\theta)| = o(1), \tag{6.22}$$

as $b \to 0$. From (6.19), we know that

$$\begin{aligned} & S_{n,j}(\theta_1) - S_{n,j}(\theta_2) \\ = \ & \frac{1}{n}\sum_{t=1}^{n}(\frac{U_t-u}{b})^j k_b(U_t-u)[h_t^2(\theta_1) - h_t^2(\theta_2)] \\ = \ & \frac{2}{n}\sum_{t=1}^{n}(\frac{U_t-u}{b})^j k_b(U_t-u)h_t^2(\theta^*_{1,t})\frac{1}{h_t(\theta^*_{1,t})}\frac{\partial h_t(\theta^*_{1,t})}{\partial\theta^\tau}(\theta_1-\theta_2). \end{aligned}$$

Here $\theta^*_{1,t}$ and the subsequent $\theta^*_{i,t}, i = 2, \cdots, 8, t = 1, \cdots, n$ are parameter vectors between θ_1 and θ_2. Without loss of generality, we can suppose there exists a $\theta_U \in \Theta$ such that

$$h_t(\theta_U) \geq h_t(\theta), \theta \in \Theta, t = 1, 2, \cdots. \tag{6.23}$$

Put $B_{1n} = \frac{2}{n}\sum_{t=1}^{n}|\frac{U_t-u}{b}|^j k_b(U_t-u)h_t^2(\theta_U)w_{1t}$ and we know $|S_{n,j}(\theta_1) - S_{n,j}(\theta_2)| \leq B_{1n}\|\theta_1 - \theta_2\|$. Based on C.2 and C.4, it is easy to show $B_{1n} = O_p(1)$. According to Lemma 6.1, we have

$$\sup_{\theta\in\Theta} |S_{n,j}(\theta) - ES_{n,j}(\theta)| = o_p(1). \tag{6.24}$$

(6.20), (6.22) and (6.24) imply that

$$S_{n,j}(\theta) = \mu_j f(u)\sigma_2(\theta, u) + r_{n,j}(\theta), \sup_{\theta\in\Theta} |r_{n,j}(\theta)| = o_p(1). \tag{6.25}$$

Consequently, we have

$$S_n(\theta) = f(u)\sigma_2(\theta, u)S + r_n(\theta), \tag{6.26}$$

as $n \to \infty, b \to 0$, where

$$S = \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix}, r_n(\theta) = \begin{pmatrix} r_{n,0}(\theta) & r_{n,1}(\theta) \\ r_{n,1}(\theta) & r_{n,2}(\theta) \end{pmatrix}. \tag{6.27}$$

Applying Lemma 6.3, we know that there is a matrix $R_n(\theta)$ of the form

$$R_n(\theta) = \begin{pmatrix} R_{n,0}(\theta) & R_{n,1}(\theta) \\ R_{n,1}(\theta) & R_{n,2}(\theta) \end{pmatrix}, \ \sup_{\theta\in\Theta} |R_{n,j}(\theta)| = o_p(1) \tag{6.28}$$

such that

$$S_n(\theta)^{-1} = \frac{S^{-1}}{f(u)\sigma_2(\theta, u)} + R_n(\theta). \tag{6.29}$$

Suppose $S^{-1} = (S^{i,j})_{0\leq i,j\leq 1}$, then it can be calculated that

$$S_n^{-1}(\theta)Z_\theta^\tau W = \tfrac{1}{f(u)\sigma_2(\theta,u)} S^{-1}Z_\theta^\tau W + R_n(\theta)Z_\theta^\tau W,$$

$$S^{-1}Z_\theta^\tau W = \tfrac{1}{n} \begin{pmatrix} K_{0,b}(U_1 - u)h_1(\theta) & \cdots & K_{0,b}(U_n - u)h_n(\theta) \\ K_{1,b}(U_1 - u)h_1(\theta) & \cdots & K_{1,b}(U_n - u)h_n(\theta) \end{pmatrix},$$

$$R_n(\theta)Z_\theta^\tau W = \tfrac{1}{n} \begin{pmatrix} C_{11}(\theta) & \cdots & C_{1n}(\theta) \\ C_{21}(\theta) & \cdots & C_{2n}(\theta) \end{pmatrix},$$

where

$$K_j(x) = \sum_{i=0}^{1} S^{ji} x^i k(x),\ j = 0, 1, K_{j,b}(x) = \frac{1}{b} K_j(x/b), \tag{6.30}$$

$$C_{1,j}(\theta) = \sum_{l=0}^{1} R_{n,l}(\theta) k_b(U_j - u) h_j(\theta) (\tfrac{U_j - u}{b})^l,\ 1 \le j \le n,$$

$$C_{2,j}(\theta) = \sum_{l=1}^{2} R_{n,l}(\theta) k_b(U_j - u) h_j(\theta) (\tfrac{U_j - u}{b})^{l-1},\ 1 \le j \le n. \tag{6.31}$$

It is not difficult to prove $\sup_{\theta \in \Theta} |C_{i,j}(\theta)| = o_p(1)$ in terms of (6.28) for $i = 1, 2$, $j = 1, \cdots, n$. Let $E_1 = (0, 1)^\tau$, by definition, we know

$$E_0^\tau (Z_\theta^\tau W Z_\theta)^{-1} Z_\theta^\tau W Z_\theta E_0 = 1, E_0^\tau (Z_\theta^\tau W Z_\theta)^{-1} Z_\theta^\tau W Z_\theta E_1 = 0.$$

Then we have

$$\begin{aligned}
& \hat{m}(\theta, u) - m(\theta, u) \\
= \ & E_0^\tau (Z_\theta^\tau W Z_\theta)^{-1} Z_\theta^\tau W Y - m(\theta, u) E_0^\tau (Z_\theta^\tau W Z_\theta)^{-1} Z_\theta^\tau W Z_\theta E_0 \\
& - \frac{\partial m(\theta, u)}{\partial u} b E_0^\tau (Z_\theta^\tau W Z_\theta)^{-1} Z_\theta^\tau W Z_\theta E_1 \\
= \ & \{E_0^\tau S_n(\theta)^{-1} Z_\theta^\tau W [Y - M_\theta]\} \\
& + \left\{ E_0^\tau S_n(\theta)^{-1} Z_\theta^\tau W \left[M_\theta - m(\theta, u) Z_\theta E_0 - \frac{\partial m(\theta, u)}{\partial u} b Z_\theta E_1 \right] \right\} \\
:= \ & I_1(\theta) + I_2(\theta),
\end{aligned} \tag{6.32}$$

where $M_\theta = (m(\theta, U_1) h_1(\theta), \cdots, m(\theta, U_n) h_n(\theta))^\tau$, $I_1(\theta)$ and $I_2(\theta)$ mean the corresponding items in the preceding two pairs of braces. Via some algebraic calculations, we further get

$$I_1(\theta) = I_{11}(\theta) + I_{12}(\theta), \tag{6.33}$$

where

$$I_{11}(\theta) = \frac{1}{nf(u)\sigma_2(\theta,u)}\sum_{t=1}^{n} K_{0,b}(U_t-u)h_t(\theta)[y_t - m(\theta,U_t)h_t(\theta)], \tag{6.34}$$

$$I_{12}(\theta) = \frac{1}{n}\sum_{t=1}^{n} C_{1,t}(\theta)[y_t - m(\theta,U_t)h_t(\theta)]. \tag{6.35}$$

It is not difficult to obtain that (here we omit the $O(1)$ term $\frac{1}{f(u)\sigma_2(\theta,u)}$ for simplifying the deductions)

$$\begin{aligned}
& I_{11}(\theta_1) - I_{11}(\theta_2) \\
= & \frac{1}{n}\sum_{t=1}^{n} K_{0,b}(U_t-u)y_t[h_t(\theta_1)-h_t(\theta_2)] \\
& + \frac{-1}{n}\sum_{t=1}^{n} K_{0,b}(U_t-u)m(\theta_1,U_t)[h_t^2(\theta_1)-h_t^2(\theta_2)] \\
& + \frac{-1}{n}\sum_{t=1}^{n} K_{0,b}(U_t-u)h_t^2(\theta_2)[m(\theta_1,U_t)-m(\theta_2,U_t)] \\
= & \frac{1}{n}\sum_{t=1}^{n} K_{0,b}(U_t-u)y_t h_t(\theta_{2,t}^*)\frac{1}{h_t(\theta_{2,t}^*)}\frac{\partial h_t(\theta_{2,t}^*)}{\partial\theta^\tau}(\theta_1-\theta_2) \\
& + \frac{-2}{n}\sum_{t=1}^{n} K_{0,b}(U_t-u)m(\theta_1,U_t)h_t^2(\theta_{3,t}^*)\frac{1}{h_t(\theta_{3,t}^*)}\frac{\partial h_t(\theta_{3,t}^*)}{\partial\theta^\tau}(\theta_1-\theta_2) \\
& + \frac{-1}{n}\sum_{t=1}^{n} K_{0,b}(U_t-u)h_t^2(\theta_2)\frac{\partial m(\theta_{4,t}^*,U_t)}{\partial\theta^\tau}(\theta_1-\theta_2).
\end{aligned} \tag{6.36}$$

Define

$$\begin{aligned}
B_{2n} = & \frac{1}{n}\sum_{t=1}^{n} K_{0,b}^*(U_t-u)|y_t|h_t(\theta_U)w_{1t} \\
& + \frac{2M}{n}\sum_{t=1}^{n} K_{0,b}^*(U_t-u)h_t^2(\theta_U)w_{1t} \\
& + \frac{M}{n}\sum_{t=1}^{n} K_{0,b}^*(U_t-u)h_t^2(\theta_U),
\end{aligned}$$

where

$$K_j^*(x) = \sum_{i=0}^{1} |S^{ji}||x|^i k(x),\ j = 0, 1, K_{j,b}^*(x) = \frac{1}{b} K_j^*(x/b). \tag{6.37}$$

Then, in terms of C.2 and C.3, for certain M, we have

$$|I_{11}(\theta_1) - I_{11}(\theta_2)| \leq B_{2n}\|\theta_1 - \theta_2\|. \tag{6.38}$$

It is easy to show $B_{2n} = O_p(1)$ based on C.2 and C.4. Applying Lemma 6.1 to $I_{11}(\theta)$, we shall get

$$\sup_{\theta\in\Theta} |I_{11}(\theta) - E[I_{11}(\theta)]| = \sup_{\theta\in\Theta} |I_{11}(\theta)| = o_p(1) \tag{6.39}$$

by noting that $m(\theta, U_t)h_t(\theta) = E[y_t|U_t, h_t(\theta)]$ in terms of (6.8), which implies that $E[I_{11}(\theta)] = 0$. In fact, (apart from the $O(1)$ term $\frac{1}{f(u)\sigma_2(\theta,u)}$), based on (6.34), we have

$$\begin{aligned}
E[I_{11}(\theta)] &= E\{K_{0,b}(U_t - u)h_t(\theta)[y_t - m(\theta, U_t)h_t(\theta)]\} \\
&= E\{K_{0,b}(U_t - u)h_t(\theta)[y_t - E[y_t|U_t, h_t(\theta)]]\} \\
&= E\{K_{0,b}(U_t - u)h_t(\theta)y_t - K_{0,b}(U_t - u)h_t(\theta)E[y_t|U_t, h_t(\theta)]\} \\
&= E\{K_{0,b}(U_t - u)h_t(\theta)y_t - E[K_{0,b}(U_t - u)h_t(\theta)y_t|U_t, h_t(\theta)]\} \\
&= 0.
\end{aligned}$$

One can also easily see $\sup_{\theta\in\Theta} |I_{12}(\theta)| = o_p(1)$ based on (6.31) and (6.35). Incorporated with (6.39), it follows that

$$\sup_{\theta\in\Theta} |I_1(\theta)| = o_p(1). \tag{6.40}$$

We next show that $\sup_{\theta\in\Theta}|I_2(\theta)| = o_p(1)$. Recall that

$$\begin{aligned} I_2(\theta) &= E_0^\tau S_n(\theta)^{-1} Z_\theta^\tau W\left[M_\theta - m(\theta,u)Z_\theta E_0 - \frac{\partial m(\theta,u)}{\partial u} b Z_\theta E_1\right] \\ &= I_{21}(\theta) + I_{22}(\theta), \end{aligned}$$

where

$$\begin{aligned} I_{21}(\theta) &= \frac{1}{nf(u)\sigma_2(\theta,u)}\sum_{t=1}^n K_{0,b}(U_t-u)h_t^2(\theta)[m(\theta,U_t) \\ &\quad -m(\theta,u) - \frac{\partial m(\theta,u)}{\partial u}(U_t-u)] \\ &= \frac{1}{2nf(u)\sigma_2(\theta,u)}\sum_{t=1}^n K_{0,b}(U_t-u)h_t^2(\theta)\frac{\partial^2 m(\theta,u_{1,t}^*)}{\partial u^2}(U_t-u)^2, \end{aligned} \tag{6.41}$$

$$\begin{aligned} I_{22}(\theta) &= \frac{1}{n}\sum_{t=1}^n C_{1,t}(\theta)\left[m(\theta,U_t) - m(\theta,u) - \frac{\partial m(\theta,u)}{\partial u}(U_t-u)\right]h_t(\theta) \\ &= \frac{1}{2n}\sum_{t=1}^n C_{1,t}(\theta)h_t(\theta)\frac{\partial^2 m(\theta,u_{1,t}^*)}{\partial u^2}(U_t-u)^2. \end{aligned} \tag{6.42}$$

Here, for each $t = 1,\cdots,n$, $u_{1,t}^*$ takes value between u and U_t. For $u\in A$, under C.2, it can be derived that

$$\begin{aligned} \sup_{\theta\in\Theta}|I_{21}(\theta)| &\leq \frac{Mb^2}{nf(u)}\sum_{t=1}^n K_{0,b}^*(U_t-u)h_t^2(\theta_U)(\frac{U_t-u}{b})^2 \\ &= b^2 O_p(1) = o_p(1), b\to 0. \end{aligned} \tag{6.43}$$

From (6.31), $\sup_{\theta\in\Theta}|I_{22}(\theta)| = o_p(1)$, in conjunction with (6.43), which gives

$$\sup_{\theta\in\Theta}|I_2(\theta)| = o_p(1). \tag{6.44}$$

(6.32), (6.40) and (6.44) imply that

$$\sup_{\theta\in\Theta} |\hat{m}(\theta, u) - m(\theta, u)| = o_p(1). \tag{6.45}$$

From (6.10), it is not difficult to get

$$\begin{aligned}
& L_n(\theta_1) - L_n(\theta_2) \\
= & \frac{1}{n}\sum_{t=1}^{n}[\log h_t(\theta_1) - \log h_t(\theta_2)]\pi(U_t) \\
& + \frac{1}{n}\sum_{t=1}^{n}[\frac{1}{h_t(\theta_1)} - \frac{1}{h_t(\theta_2)}][y_t - m(\theta_1, U_t)h_t(\theta_1)]^2\pi(U_t) \\
& + \frac{1}{n}\sum_{t=1}^{n}\frac{\pi(U_t)}{h_t(\theta_2)}\left\{[y_t - m(\theta_1, U_t)h_t(\theta_1)]^2 - [y_t - m(\theta_2, U_t)h_t(\theta_2)]^2\right\} \\
= & \frac{1}{n}\sum_{t=1}^{n}\pi(U_t)\frac{1}{h_t(\theta^*_{5,t})}\frac{\partial h_t(\theta^*_{5,t})}{\partial\theta^\tau}(\theta_1 - \theta_2) \\
& + \frac{1}{n}\sum_{t=1}^{n}[y_t - m(\theta_1, U_t)h_t(\theta_1)]^2\pi(U_t)\frac{1}{h^2_t(\theta^*_{6,t})}\frac{\partial h_t(\theta^*_{6,t})}{\partial\theta^\tau}(\theta_1 - \theta_2) \\
& + \frac{-1}{n}\sum_{t=1}^{n}\frac{\pi(U_t)}{h_t(\theta_2)}\{[2y_t - m(\theta_1, U_t)h_t(\theta_1) - m(\theta_2, U_t)h_t(\theta_2)] \\
& \times[\frac{\partial m(\theta^*_{7,t}, U_t)}{\partial\theta^\tau}h_t(\theta_1) + m(\theta_2, U_t)\frac{\partial h_t(\theta^*_{8,t})}{\partial\theta^\tau}]\}(\theta_1 - \theta_2).
\end{aligned} \tag{6.46}$$

Note that $0 < \pi(U_t)$ implies $U_t \in A$ and $h_t(\theta) \geq m > 0$ for all $\theta \in \Theta$ under C.5. Define

$$\begin{aligned}
B_{3n} = & \frac{1}{n}\sum_{t=1}^{n}\pi(U_t)w_{1t} + \frac{M}{n}\sum_{t=1}^{n}[|y_t| + h_t(\theta_U)]^2\pi(U_t)w_{1t} \\
& + \frac{M}{n}\sum_{t=1}^{n}[|y_t| + h_t(\theta_U)]\pi(U_t)(w_{1,t} + w_{2,t}).
\end{aligned}$$

Then we know

$$|L_n(\theta_1) - L_n(\theta_2)| \leq B_{3n}||\theta_1 - \theta_2|| \tag{6.47}$$

for some M and it is easy to show $B_{3n} = O_p(1)$. Applying Lemma 6.1 to $L_n(\theta)$, then it follows that

$$\sup_{\theta\in\Theta}|L_n(\theta) - L(\theta)| = o_p(1). \tag{6.48}$$

It is not difficult to derive that

$$\begin{aligned}
[y_t - \hat{m}(\theta, U_t)h_t(\theta)]^2 &= [y_t - m(\theta, U_t)h_t(\theta)]^2 + h_t^2(\theta)[m(\theta, U_t) - \hat{m}(\theta, U_t)]^2 \\
&\quad +2[y_t - m(\theta, U_t)h_t(\theta)][m(\theta, U_t) - \hat{m}(\theta, U_t)]h_t(\theta).
\end{aligned}$$

From (6.10) and (6.12), it follows that

$$\begin{aligned}
\hat{L}_n(\theta) - L_n(\theta) &= \frac{1}{n}\sum_{t=1}^{n}\pi(U_t)h_t(\theta)[m(\theta, U_t) - \hat{m}(\theta, U_t)]^2 \\
&\quad +\frac{2}{n}\sum_{t=1}^{n}\pi(U_t)[m(\theta, U_t) - \hat{m}(\theta, U_t)][y_t - m(\theta, U_t)h_t(\theta)].
\end{aligned}$$

Further,

$$\begin{aligned}
\sup_{\theta\in\Theta}|L_n(\theta) - \hat{L}_n(\theta)| &\le \frac{1}{n}\sum_{t=1}^{n}\pi(U_t)h_t(\theta_U)\sup_{\theta\in\Theta}[m(\theta, U_t) - \hat{m}(\theta, U_t)]^2 \\
&\quad +\frac{2}{n}\sum_{t=1}^{n}\pi(U_t)[|y_t| + Mh_t(\theta_U)]\sup_{\theta\in\Theta}|m(\theta, U_t) - \hat{m}(\theta, U_t)| \\
&= o_p(1).
\end{aligned} \tag{6.49}$$

(6.48) and (6.49) imply that

$$\sup_{\theta\in\Theta}|L(\theta) - \hat{L}_n(\theta)| = o_p(1). \tag{6.50}$$

Under condition C.6, and applying Lemma 6.2, we obtain the consistency from (6.50).

6.5 Summary

In this chapter, motivated by the FCA-GARCH-M model in Chapter 5, we further study a functional coefficient GARCH-M model, where the volatility coefficient is treated as an unknown function of a certain variable. Such a setting enables us to study the relationship between risk aversion and some related variable (e.g., the time-lagged return). An improved approach is given to estimate the parameters in the GARCH equation. Under some regularity conditions, the parametric estimators are shown to be consistent. Simulation studies have shown the method performs well. Through the empirical studies, the proposed FC-GARCH-M model seems to be superior to the usual parametric models for the considered data.

Chapter 7

Conclusions

In this thesis, based on previous work in the literature (e.g., Ling, 2004, Christensen et al., 2008, Cai et al., 2000), we have studied some parametric and semiparametric models for financial time series. Besides the TARCH model in Chapter 2, we mainly consider the GARCH-in-Mean models by assuming the conditional variance to be driven by the past returns. Such a setting for the conditional variance makes it slightly easier to study related issues in estimation and inference. Some theoretical results have been obtained such as: asymptotic null distribution (Theorem 2.2) in Chapter 2, geometric ergodicity condition (Theorem 3.1) and asymptotic normality (Theorem 3.2) in Chapter 3, consistency (Theorem 6.2) in Chapter 6. The conducted simulation studies for the considered models suggest that the proposed methods work satisfactorily. From the empirical studies, for the considered data, it is seen that the proposed models have comparable or better fitting performance as compared to the traditional ones. Moreover, some interesting results have been gotten like the relationship between the conditional mean and variance (Chapter 4), and the relations between volatility coefficient and time lagged return (Chapters 5-6).

Our study reveals that there are several areas that may be worthwhile to be explored further. First, when we study the asymptotic properties for the QMLE of the considered TARCH model in Chapter 2, assuming the threshold parameter is known is rather restrictive. The results would be more general if one can study the estimation of all the parameters (including the threshold parameter) jointly. Second, for the specific GARCH-in-Mean model in Chapter 3, it is worthwhile to further study the QMLE of the model for non-stationary cases (namely the true parameters are located in a non-stationary region.). Conditions for the existence of moments can also be an interesting topic. It would be a significant contribution if one can study the ergodicity and QMLE for other popular choices of the mean function, such as $m(h_t) = \delta h_t, m(h_t) = \delta \log h_t$. Third, for the semiparametric models discussed in Chapters 4 and 6, though parametric estimators are shown to be consistent, we have not established the asymptotic normality of the estimators. Also, few theoretical results about the functional estimation in Chapters 4 and 6 have been given. Future studies are expected to fill this gap. Finally, for empirical studies in Chapters 5-6, besides the time lagged returns, further study of the relationship between volatility coefficient and other explanatory variables is of practical value.

Appendix

In this part, we present the Matlab codes for estimating the models (6.15-6.17) in Chapter 6. The codes for previous chapters can be analogously developed based on the presented ones and hence we omit them. The codes for (6.14) are similar to those of (6.15) and hence we do not present them here either.

A.1 Matlab Code for the estimation of model (6.15)

To get (6.15), given observations $\{y_t\}_{t=1}^{T}$, we equivalently need to estimate $\theta = (\delta, a_0, a_1, a_2)$ for the model below:

$$y_t = \delta h_t + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t},$$
$$e_t \sim i.i.d\,(0,1), h_t = a_0 + a_1 y_{t-1}^2 + a_2 h_{t-1}.$$

The codes are as follows:

```
function theta=garchmn(y)
format compact;
data=y;
% initial values
ss=std(data);
h0=ss^2;
delta=0.2;
a0=0.2;
a1=0.5;
a2=0.1;
theta=[delta a0 a1 a2];
% constraints
thetaL=[-2 0.0001 0.0001 0.0001];
thetaU=[2 50 0.99 0.99];
% optimization options
optopt=optimset('MaxIter',1000);
[x, fval]=fmincon(@garchlik,theta,[], [],[],[],...
thetaL, thetaU,[],optopt,data,h0);
theta=x; lik=-fval;
[lik,V]=garchlik0(theta, data, h0);


function lik=garchlik(theta,data,h0)
 [lik,V]=garchlik0(theta,data,h0);

function [lik,V] = garchlik0(theta,data,h0)
delta=theta(1);
a0=theta(2);
a1=theta(3);
a2=theta(4);
N=size(data);
```

```
V=zeros(N+1);
VV=h0;
V(1)=VV;
 lik=0.; % negative likelihood
  for indx=1:N
     error=data(indx)-delta* VV;
     lik=lik+(error^2)/VV+log(VV);
    VV=a0+a1*data(indx)^2+a2*VV;
    V(indx+1)=VV;
  end
```

A.2 Matlab Code for the estimation of model (6.16)

To get (6.16), given observations $\{y_t,\}_{t=1}^{T}$, we equivalently need to estimate $\theta = (\omega, \alpha, \eta, \beta)$ for the model below:

$$y_t = m(y_{t-1})h_t + \varepsilon_t, \varepsilon_t = e_t\sqrt{h_t}, e_t \sim i.i.d(0,1),$$
$$h_t = \omega + \alpha[1 + \eta I(y_{t-1} \leq 0)]y_{t-1}^2 + \beta h_{t-1}.$$

The codes are followed:

```
function y=FCE(Y,U,X,U0,b)
N=length(Y);
l=length(U0);
I=ones(N,1);
y=[];
for i=1:l
      Ui=U-U0(i)*I;
      KI=(0.75/b)*(I-Ui.*Ui/(b^2)).*(abs(Ui)/b<0.999);
      S11=sum(KI.*X.*X);
      S12=sum(KI.*X.*X.*Ui);
      S21=sum(KI.*X.*X.*Ui);
      S22=sum(KI.*X.*X.*Ui.*Ui);
      T1=sum(KI.*X.*Y);
      T2=sum(KI.*X.*Y.*Ui);
      fenmu=S22*S11-S12*S21;
      fenzi=S22*T1-S12*T2;
         if fenmu==0
             temp=fenzi/(fenmu+0.0001);
         else
             temp=fenzi/fenmu;
         end
         y=[y;temp];
end
```

The above function "FCM "is used to estimate the common functional coefficient model proposed by Cai et al. (2000). The inputs Y, U, X are observation vectors. U_0 is the grid point vector which has the same length to that of output y. b is the given bandwidth.

```
function theta= WQMLE(Y,error,U,theta0)
format compact;
data1=Y;
data2=error;
ss=std(data1);
Vbar=ss^2;
omega=theta0(1);
```

```
alpha=theta0(2);
eta=theta0(3);
beta=theta0(4);
par=[omega alpha eta beta];
% constraints
parL=[0.001 0.001 0.001 0.1];
parU=[ 10 0.9 1000 0.99];
% optimization options
optopt=optimset('MaxIter',1000,'LargeScale','off');
[x, fval, hess]= fmincon(@garchlik, par, [], [], [], [], ...
parL, parU, [], optopt, data1,data2,Vbar,U);
par=x;
theta=x;
stderrs=diag(sqrt(inv(hess)))'
lik=-fval;
[lik,V]=garchlik0(par, data1,data2,Vbar,U);

function lik=garchlik(par, data1, data2,Vbar,U)
[lik,V]=garchlik0(par, data1, data2, Vbar,U);

function [lik,V]=garchlik0(par, data1, data2, Vbar,U)
omega=par(1);
alpha=par(2);
eta=par(3);
beta=par(4);
N=size(data1);
V=zeros(N+1);
VV=Vbar;
V(1)=VV;
P10=prctile(U,10); P90=prctile(U,90);
lik=0.; % negative likelihood
for indx=1:N-1
    lik=lik+((data2(indx)^2)/VV+log(VV))*(P10<=U(indx))*(U(indx)<=P90);
    VV=omega + alpha*(1+eta*(data1(indx)<=0))*data1(indx)^2+beta*VV ;
    V(indx+1)=VV;
end
```

The above function “WQMLE ”is for the weighted QMLE in Step 3 in Section 6.2.2. The elements of inputs Y, error are respectively $y_t, \hat{\varepsilon}_t^{(i)}$ in Step 3. The elements of input U are the lagged returns and the input “theta0 ”is the initial value for the output “theta”. Based on the above two functions, “FCM ”and “WQMLE ”, we give the following function “estimate ”to estimate the model. For the function “estimate ”, the input “theta0 ”is an initial estimator for the parameter vector and Y is the observation vector.

```
function [theta_e wucha]=estimate(Y, theta0)
N=length(Y);
k=50;
temp=theta0;
wuchas=[];
for j=1:k
V=ones(N,1);
V(1,1)=std(Y)^2;
for i=2:N
 V(i,1)=theta0(1)+theta0(2)*(1+theta0(3)*(Y(i-1)<=0))*Y(i-1)^2...
 +theta0(4)*V(i-1,1);
end
X=V;
Y1=Y(50:N);
```

```
U1=Y(49:N-1);
X1=X(50:N);
h=1.06*std(U1)*(N-51)^(-1/5);
M=FCE(Y1,U1,X1,U1,h);
error1=Y1-M.*X1;
theta1=WQMLE(Y1,error1,U1,theta0)
clc
wucha=norm(theta1-theta0);
if wucha<=0.0001
    break
else
    theta0=theta1;
    temp=[temp;theta1];
    wuchas=[wuchas;wucha];
end
end
theta_e=theta1;
thetaes=temp;
```

A.3 Matlab Code for the estimation of model (6.17)

To estimate model (6.17), namely we are to estimate

$$y_t = \delta_t h_t + \varepsilon_t, \varepsilon_t = e_t \sqrt{h_t}, \delta_t = \delta_{t-1} + v_t,$$
$$e_t \sim i.i.d \quad N(0,1), h_t = a_0 + a_1\eta_{t-1}^2 + a_2 h_{t-1}.$$

Here, the GARCH surprise variable $\eta_t := y_t - E_{t-1}(y_t)$ with $E_{t-1}(y_t)$ being the optimal forecast of y_t given all information up to time $t-1$. The errors ε_t, v_t are assumed to be uncorrelated Gaussians with zero means and variances h_t and Q, respectively. In above model, the coefficient δ_t is assumed to follow a random walk, which together with the system parameters, can be estimated by the Kalman filter (see page 399-400 in Section 13.8 in Hamilton 1994) and maximum likelihood methods. The codes are presented as follows:

```
function [theta lik h f] = kalmanest(y)
format long;
% initial values
ini=[10 2 var(y) 0.001]; %ini=[b10, p10, h0, eta0];
a0=0.9;
a1=0.1;
a2=0.6;
Q=50;
par=[a0 a1 a2 Q];
% constraints
parL=[0.000001 0.000001 0.000001 0.000001];
parU=[ 10000    0.999    0.999    10000];
% optimization options
optopt=optimset('MaxFunEvals', 5000, 'MaxIter',5000,...
'TolFun',10^-10,'TolX',10^-10,...
'TolCon',10^-10,'Display','iter','LargeScale','off')

[x, fval, hess]=fmincon(@garchlik, par, [], [], [], [],...
parL, parU, [], optopt, y, ini);
par=x;
theta=par;
stderrs=diag(sqrt(inv(hess)))';
lik=-fval
```

```
[lik, h, f]=garchlik0(par, y, ini);


function lik=garchlik(par,y,ini)
 [lik,h,f]=garchlik0(par, y, ini);

function [lik,h,f]=garchlik0(par, y, ini)
a0=par(1);
a1=par(2);
a2=par(3);
Q=par(4);
N=length(y);
h=zeros(N,1); % h=(h1, h2, ht,..., hN)
H=zeros(N,1); % H=(H1, H2, Ht,..., HN), var(eta_t)=H_t
b=zeros(N,1); % b=(b1|0, ..., bt|t-1, .., bN|N-1)
p=zeros(N,1); % p=(p1|0, ..., pt|t-1, .., pN|N-1)
e=zeros(N,1); % e means eta=(eta1, etat, ..., etaN)
f=zeros(N,1);
b10=ini(1);p10=ini(2);h0=ini(3);eta0=ini(4);
h(1)=a0+a1*eta0^2+a2*h0;
f(1)=b(1)*h(1);
e(1)=y(1)-b(1)*h(1);
H(1)=h(1)+h(1)^2*p(1);
lik = 0.; % negative likelihood
for t=1:N-1
    temp1=b(t)+p(t)*(y(t)-b(t)*h(t))/(h(t)*p(t)+1);
    temp2=p(t)-p(t)^2*h(t)/(h(t)*p(t)+1);
    b(t+1)=temp1;
    p(t+1)=temp2+Q;
    h(t+1)=a0+a1*e(t)^2+a2*h(t);
    f(t+1)=b(t+1)*h(t+1);
    e(t+1)=y(t+1)-b(t+1)*h(t+1);
    H(t+1)=h(t+1)+h(t+1)^2*p(t+1);
    lik=lik+(e(t+1)^2)/H(t+1)+log(H(t+1));
end
```

References

Andrews, D. W. K. (1992). Generic uniform convergence. *Econometric Theory* **8** 241-257.

Backus, D. and Gregory, A. (1993). Theoretical relations between risk premiums and conditional variances. *Journal of Business and Economic Statistics* **11** 177-185.

Bera, A. K. and Ra, S. (1995). A test for the presence of conditional heteroskedasticity within ARCH-M framework. *Econometric Reviews* **14(4)** 473-485.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Billingsley, P. (1995). *Probability and Measure*, 3rd edition. Wiley, New York.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31** 307-327.

Bougerol, P. and Picard, N. (1992). Stationarity of GARCH Processes and of some non-negative time series. *Journal of Econometrics* **52** 115-128.

Cai, Z. , Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association* **95** 941-956.

Cai, Z. , Li, Q. and Park, Y. J. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics* **148** 101-113.

Caporale, T. and McKiernan, B. (1996). The relationship between output variability and growth: evidence from post war UK data. *Scottish Journal of Political Economy* **43** 229-236.

Chan, K. S. (1990). Testing for threshold autoregression. *The Annals of Statistics* **18** 1886-1894.

Chan, K. S. (1991). Percentage points of likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society B* **53** 691-696.

Chan, K. S. and Tong, H. (1990). On likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society B* **52** 469-476.

Chen, R. and Liu, L. (2001). Functional Coefficient Autoregressive Models: Estimation and Tests of Hypotheses. *Journal of the Time Series Analysis* **22** 151-173.

Chen, R. and Tsay, R. S. (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association* **88** 298-308.

Chou, R. (1988). Volatility persistence and stock valuations: some empirical evidence using GARCH. *Journal of Applied Econometrics* **3** 279-294.

Chou, R., Engle, R. F. and Kane, A. (1992). Measuring risk aversion from excess returns on a Stock Index. *Journal of Econometrics* **52** 201-224.

Christensen, B. J., Dahl, C. M. and Iglesias, E. M. (2008). Semiparametric inference in a GARCH-in-Mean model. *Manuscript, Michigan State University*.

Cline, D. B. H. (2007a). Stability of nonlinear stochastic recursions with application to nonlinear AR-GARCH models. *Advances in Applied Probability* **39** 462-491.

Cline, D. B. H. (2007b). Evaluating the Lyapounov exponent and existence of moments for threshold AR-ARCH models. *Journal of Time Series Analysis* **28** 241-260.

Cline, D. B. H. and Pu, H. H. (1999). Geometric ergodicity of nonlinear time series. *Statistica Sinica* **9** 1103-1118.

Cline, D. B. H. and Pu, H. H. (2004). Stability and the Lyaponouv exponent of threshold AR-ARCH models. *The Annals of Applied Probability* **14** 1920-1949.

Conrad, C. and Mammen, E. (2008). Nonparametric Regression on Latent Covariates with an Application to Semi-parametric GARCH-in-Mean Models. *Manuscript, University of Mannheim*.

Das, S. and Sarkar, N. (2000). An ARCH in the nonlinear mean(ARCH-NM) Model. *Sankhya Ser B* **62** 327-344.

Das, S. and Sarkar, N. (2010). Is the relative risk aversion parameter constant over time? A multi-country study. *Empirical Economics* **38** 605-617.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64** 247-254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74** 33-43.

Degiannakis, S. and Xekalaki, E. (2004). ARCH models: a review. *Quality Technology and Quantitative Management* **1** 271-324.

Drost, F. C. and Klaassen, C. A. J. (1997). Efficient estimation in semiparametric GARCH models. *Journal of Econometrics* **81** 193-221.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* **50(4)** 987-1007.

Engle, R. F. and González-Rivera, G. (1991). Semiparametric ARCH models. *Journal of Business and Economic Statistics* **9** 345-359.

Engle, R. F., Lilien, D. M. and Robins, R. P. (1987). Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica* **55** 391-407.

Engle, R. F. and Ng, V. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance* **48** 1749-1778.

Fama, E. F. and French, K. R. (1989). Business conditions and expected returns on stock and bonds. *Journal of Financial Economics* **25** 23-49.

Fan, J. and Gijbels, I. (1996). *Local polynomial modeling and its applications.* Chapman & Hall, London.

French, K. R., Schwert, G. W. and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics* **19** 3-29.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993). On the relationship between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance* **XLVIII** 1779-1801.

Grier, K. B. and Perry, M. J. (2000). The effects of real and nominal uncertainty on inflation and output growth: some GARCH-M evidence. *Journal of Applied Econometrics* **15** 45-58.

H*ä*rdle, W. and Tsybakov, A. B. (1997). Local polynomial estimators of the volatility function. *Journal of Econometrics* **81** 223-242.

Harvey, C. (1989). Is the expected compensation for market volatility constant through time. *Manuscript, Duke University*.

Hamilton, J. D. (1994). Time Series Analysis. Princeton University Press, Princeton.

He, C. and Ter*ä*svirta, T. (1999a). Fourth moment structure of the GARCH(p, q) processes. *Econometric Theory* **15** 824-846.

He, C. and Ter*ä*svirta, T. (1999b). Properties of moments of a family of GARCH processes. *Journal of Econometrics* **92** 173-192.

Hong, E. P. (1991). The autocorrelation structure for the GARCH-M process. *Economics letters* **37** 129-132.

Jensen, S. T and Rahbek, A. (2004). Asymptotic inference for non-stationary GARCH. *Econometric theory* **20** 1203-1226.

Karanasos, M. (1999). The second moment and the autocovariance function of the squared errors of the GARCH model. *Journal of Econometrics* **90** 63-76.

Kosorok, M. R. (2006). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York.

Kreiss, J. P., Neumann, M. and Yao, Q. (2008). Bootstrap tests for simple structures in nonparametric time series regression. *Statistics and its interface* **1(2)** 367-380.

Lee, S-W. and Hansen, B. E. (1994). Asymptotic theory for the GARCH (1, 1) quasi-maximum likelihood estimator. *Econometric Theory* **10** 29-52.

Lee, T-H. and Ullah, A. (2000). Nonparametric bootstrap tests for neglected nonlinearity in time series regression models. *Manuscript, University of California*.

Li, W. K. and LAM, K. (1995). Modelling the asymmetry in stock returns using threshold ARCH model. *Statistician* **44** 333-41.

Li, W. K. and Mak, T. K. (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of time series analysis* **15** 627-636.

Ling, S. (2004). Estimation and testing of stationarity for double autoregressive models. *Journal of the Royal Statistical Society B* **66** 63-78.

Ling, S. (2007). A double AR (p) model: structure and estimation. *Statistica Sinica* **17** 161-175.

Ling, S. and Li, D. (2008). Asymptotic inference for a non-stationary double AR(1) model. *Biometrika* **95** 257-263.

Ling, S. and McAleer, M. (2002a). Necessary and sufficient moment conditions for the GARCH (p,q) and asymmetric power GARCH (p,q) models. *Econometric Theory* **18** 722-729.

Ling, S. and McAleer, M. (2002b). Stationarity and the existence of moments of a family of GARCH processes. *Journal of Econometrics* **106** 109-117.

Linton, O. B. (1993). Adaptive estimation in ARCH models. *Econometric Theory* **9** 539-569.

Linton, O. B. (2009). Semiparametric and nonparametric ARCH modeling, in Andersen, T. G., Davis, R. A., Kreiss, J. P. and Mikosch, Th. (eds.), *Handbook of Financial Time Series*, Springer.

Linton, O. B. and Perron, B. (2003). The shape of the risk premium: evidence from a semiparametric generalized autoregressive condtional heteroskedasticity model. *Journal of Business and Economic Statistics* **21(3)** 354-367.

Lu, Z. D. (1998). On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term. *Statistica Sinica* **8** 1205-1217.

Lumsdaine, R. L. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH (1, 1) and covariance stationary GARCH (1, 1) models. *Econometrica* **64** 575-596.

Meitz, M. and Saikkonen, P. (2008). Ergodicity, mixing, and existence of moments of a class of markov models with applications to GARCH and ACD models. *Econometric Theory* **24** 1291-1320.

Merton, R. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* **8** 323-361.

Meyn, S.P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer-Verlag, London.

Nelson, D. B. (1990). ARCH models as diffusion approximations. *Journal of Econometrics* **45** 7-38.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59** 347-370.

Pagan, A. R. and Hong, Y. S. (1991). Nonparametric Estimation and the Risk Premium, in Barnett, W., Powell, J. and Tauchen, G. E. (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge University Press.

Peligrad, M. (1982). Invariance principles for mixing sequence of random variables. *The Annual of Probability* **10** 968-981.

Pollard, D. (1984). *Convergence of Stochastic Process*. Springer, New York.

Rossi, A. and Timmermann, A. (2010). What is the shape of the risk-return relation? *Manuscript, University of California San Diego, CREATES*.

Schepper, A. D. and Goovaerts, M. J. (1999). The GARCH (1, 1)-M model: results for the densities of the variance and the mean. *Insurance: Mathematics and Economics* **24** 83-94.

Silvey, S. D. (1959). The Lagrangian Multiplier test. *The Annuals of Mathematical Statistics* **30** 389-407.

Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach.* Oxford University Press, Oxford.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data (with Discussion). *Journal of the Royal Statistical Society B* **42** 245-92.

Wang, L. and Yang, L. (2009). Spline estimation of single-index models. *Statistic Sinica* **19** 765-783.

Weiss, A. A. (1986). Asymptotic theory for ARCH models: estimation and testing. *Econometric theory* **2** 101-131.

Wong, C. S. and Li, W. K. (1997). Testing for threshold autoregression with conditional heteroscedasticity. *Biometrika* **84** 407-418.

Xia, Y. and Tong, H. (2006). Cumulative effects of air pollution on public health. *Statistics in Medicine* **25** 3548-3559.

Xia, Y., Tong, H. and Li, W. K. (1999). On extended partially linear single-index models. *Biometrika* **86** 831-842.

Xia, Y., Tong, H. and Li, W. K. (2002). Single-index volatility models and estimation. *Statistica Sinica* **12** 785-799.

Yang, L. (2002). Direct estimation in an additive model when the components are proportional. *Statistica Sinica* **12** 801-821.

Yang, L. (2006). A semiparametric GARCH model for foreign exchange volatility. *Journal of Econometrics* **130** 365-384.

Zakoian, J. M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* **18** 931-995.

Zhang, X., Wong, H., Li, Y. and Ip, W. (2011a). A class of threshold autoregressive conditional heteroscedastic models. *Statistics and Its Interface* **4 (2)** 149-158.

Zhang, X., Wong, H., Li, Y. and Ip, W. (2011b). An alternative GARCH-in-Mean model: structure and estimation. (Accepted by *Communications in Statistics - Theory and Methods*).