March 2012

Essays on Financial Return and Volatility Modeling

Jing Wu
The University of Western Ontario

Supervisor
John Knight
The University of Western Ontario

Graduate Program in Economics

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

ESSAYS ON FINANCIAL RETURN AND VOLATILITY MODELING

(Spine Title: Financial Return and Volatility)

(Thesis Format: Integrated Article)

by

Jing Wu

Department of Economics

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES
THE UNIVERSITY OF WESTERN ONTARIO
LONDON, CANADA
2012

THE UNIVERSITY OF WESTERN ONTARIO
SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES

CERTIFICATE OF EXAMINATION

Supervisor	Examiners
Dr. John Knight	Dr. Youngki Shin
Supervisory Committee	Dr. Maria Ponomareva
Dr. Youngki Shin	Dr. Reg Kulperger
Dr. Martijn van Hasselt	Dr. Dinghai Xu

The thesis by

Jing Wu

entitled

Essays on Financial Return and Volatility Modeling

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Date ______________________

____
Chair of the Thesis Examining Board

Abstract

My dissertation consists of three essays focusing on modeling financial asset return and volatility.

The first essay proposes a threshold GARCH model to describe the regime-switching in volatility dynamics of financial asset returns. In the threshold model the switching of regimes is triggered by an observable threshold variable, while volatility follows a GARCH process within each regime. This model can be viewed as a special case of the random coefficient GARCH model. We establish theoretical conditions, which ensure that the return process in the threshold model is strictly stationary, as well as conditions for the existence of finite variance and fourth moment. A simulation study is further conducted to examine the finite sample properties of the maximum likelihood estimator.

The second essay extends our study of the threshold GARCH model to an empirical application. The empirical results support the use of the threshold variable to identify the regime shifts in the volatility process. Especially when VIX is used as the threshold, we are able to separate the clustering of volatile periods corresponding to various financial crises. According to 5 common measures on forecasting evaluation, the threshold GARCH model provides better volatility forecasts for stocks as well as currency exchange rates.

The third essay examines the effect of time structure on the estimation performance of independent component analysis (ICA) models and provides guidance in applying the ICA model to time series data. We compare the performance of the basic ICA model to the time series ICA model in which the cross-autocovariances are used as a measure to achieve independence. We conduct a simulation study to evaluate the time series ICA model under different time structure assumptions about the underlying components that generate financial time series. Moreover, the empirical results support the use of the time series ICA model.

Keywords: Regime-Switching Volatility Model, Threshold GARCH Model, Non-Linear Time Series, Volatility Forecasting, Quasi-Maximum Likelihood, Realized Volatility, Independent Component Analysis.

To my family

Table of Contents

List of Tables

List of Figures

Acknowledgements

This work would not have been possible without the support of many people.

I am deeply indebted to my supervisor, Professor John Knight, for his invaluable guidance and continuous support. Without his knowledge and assistance this work would not have been possible. I would also like to thank other members of my supervisory committee, Professor Youngki Shin and Professor Martijn van Hasselt, for their helpful comments and constant encouragement.

I wish to express my deep gratitude to Professor Dinghai Xu for sharing the high frequency data and providing valuable suggestions. Special thanks to Yvonne Adams and Sharon Phillips for their help through the duration of my studies.

Finally, I would like to thank all the faculty, staff, and fellow students in the Economics Department at the University of Western Ontario, all of whom made my experience at Western unforgettable.

London, Canada Jing Wu
March, 2012

Chapter 1

Introduction

My doctoral thesis consists of three essays focusing on financial return and volatility modeling. The first two essays theoretically and empirically examine the use of a threshold GARCH model to describe the regime shifts in the volatility process of financial asset returns. The forecasting performance is also investigated based on various measures. The third essay studies the effect of time structure on the estimation performance of ICA models and provides guidance in applying the ICA model to time series data.

Modeling the temporal dependencies in the volatility of financial time series has drawn the attention of many econometricians and financial analysts. For decades, researchers have implemented ARCH and GARCH models with a regime switching framework to capture non-linearity in the volatility process. Hidden Markov models, which assume that states of the world are unknown, are widely used regime switching models. While estimation is not difficult, these models often fail to generate accurate predictions due to the unknown state in the future. In the first essay, aiming to incorporate the non-linearity and the additional information provided by trading intensity variables in regime shifts, we model volatility dynamics as a threshold model with an observable trigger, while volatility follows a GARCH process within each regime. This model can be viewed as a special case of the random coefficient GARCH model. We establish theoretical conditions, which ensure that the return process in the threshold model is strictly stationary, as well as conditions for the existence of various moments. A simulation study is further conducted to examine the finite sample properties of the maximum likelihood estimator. Simulation results reveal that the maximum likelihood estimator performs well for modest sample sizes when the stationarity conditions hold and the variance of the return series exists. The endogeneity between volatility and threshold variables has no effect on the estimation performance, may even improve it.

The second essay empirically investigates the performance of the threshold GARCH model in identifying regime shifts in the volatility process of asset returns. Since volatility forecast is a key component in pricing derivative securities and risk management, the success of a volatility model is determined crucially by its out-of-sample predicting power. We focus on the forecasting performance of the threshold model. Using VIX as the threshold variable we apply the threshold GARCH model to 20 stocks from MMI and evaluate the performance of 250 out-of-sample daily volatility forecasts. The results reveal an improvement in forecasting accuracy especially when the financial crisis in 2008 is included in the out-of-sample period. The forecasting performance is improved further when we update the true volatility proxy from the squared return to a realized volatility using high frequency data of IBM and GE. We employ stocks from NASDAQ to study the effect of an endogenous threshold variable on the forecasting performance. Two trading activity variables volume turnover and number of trades demonstrate their use in identifying the regime shifts in the period other than financial crises. The forecasting performance of three currency exchange rates also support the use of the threshold model in volatility forecasting.

Closely related to principal component analysis, independent component analysis (ICA) is a statistical and computational method for revealing hidden factors that underlie a set of random variables. The increasing demand of multivariate modeling in economics and finance as well as the strength of ICA analysis to extract statistically independent components from multivariate signals have attracted economists and financial analyst's attention to ICA. In standard applications of the ICA model, the components are assumed to be uncorrelated over time. In contrast to this random components assumption, the autocovariances of independent components in time series are well-defined statistics and can be used in estimating the ICA model. Therefore, in the last essay, we examine the effect of time structure on the estimation performance of ICA models and provide guidance in applying the ICA model to time series data. We compare the performance of the basic ICA model to the time series ICA model in which the cross-autocovariance is used as a measure to achieve independence. We conduct a simulation study to evaluate the time series ICA model under different time structure assumptions about the underlying components that generate financial time series. Moreover, the empirical study supports the use of the time series ICA model.

1.1 Bibliography

ANDERSEN, T. G. (1996): "Return volatility and trading volume: An information flow interpretation of stochastic volatility," *Journal of Finance*, 51, 169–204.

BACK, A. D., AND A. S. WEIGEND (1997): "A first application of independent component analysis to extracting structure from stock returns," *International Journal of Neural Systems*, 8, 473–84.

ENGLE, R. F., AND T. BOLLERSLEV (1986): "Modeling the persistence of conditional variances," *Econometric Reviews*, 5, 1–50.

GRAY, S. F. (1996): "Modeling the conditional distribution of interest rates as a regime-switching process," *Journal of Financial Economics*, 42, 27–62.

KNIGHT, J., AND S. E. SATCHELL (2011): "Some new results for threshold AR(1) models," *Journal of Time Series Econometrics*, 3(2), Article1.

Chapter 2

Threshold GARCH Model: Theory and Simulation

2.1 Introduction

In finance, volatility refers to the variation of financial asset returns over time. It is used to quantify the risk associated with a financial instrument. In many cases, volatility itself is a risk measure. Therefore volatility is a key concept in financial economics, as noted by Campbell, Lo and MacKinlay (1997):

> "...what distinguishes financial economics is the central role that uncertainty plays in both financial theory and its empirical implementation...Indeed in the absence of uncertainty, the problems of financial economics reduce to exercises in basic microeconomics".

The well recognized positive trade-off between risk and expected return makes volatility modeling and forecasting among the most important pursuits in empirical finance and risk management. Volatility modeling and forecasting are also very important in making monetary policy.

The volatility of asset returns is defined as the standard deviation of the return series, or square root of variance. For convenience we usually refer modeling of volatility to modeling of conditional variances. Volatility of financial asset return series appears to be time varying and serially correlated. Therefore modeling the temporal dependencies in the conditional variance of financial time series has been the interest of many economists and financial analysts. The most popular approaches are the ARCH model introduced by Engle (1982) and its extension GARCH model by Bollerslev (1986). ARCH and GARCH models can well explain the persistence in the conditional variance process, however they fail to account for the stylized fact that the conditional variance is often clustering with high volatility in one period and with low volatility in another period.

To capture the striking feature that asset prices move more rapidly during some periods than others, a regime switching framework has been brought into ARCH and GARCH models. A widely used class of regime switching models is the hidden Markov model, which assumes that states of the world are unknown. While estimation is not difficult, these models often fail to generate accurate predictions due to the unknown state in the future. In this chapter we employ a different type of regime switching models – the threshold model – to describe the conditional variance process. In this threshold model, the state of the world is determined by an observable threshold variable and therefore known, while conditional variance follows a GARCH process within each state. This model can be viewed as a special case of the random coefficient GARCH model. First, we examine the theoretical properties of the threshold model with an exogenous threshold variable. We establish theoretical conditions, which ensure that the return process in the threshold model is strictly stationary, as well as conditions for the existence of various moments. A simulation study is then conducted to examine the finite sample properties of the maximum likelihood estimator. The simulation results reveal that the accuracy of maximum likelihood estimation increases as sample size increases[1] when the stationarity conditions hold. We also explore the properties of the threshold GARCH model when the threshold variable is endogenous through simulation studies.

The outline of the rest of this chapter is as follows. In Section 2.2, we briefly review the literature on development of time varying volatility models and volatility volume relationships. In Section 2.3, we introduce the setup of the threshold GARCH model and derive the theoretical conditions for the existence of various moments in the return series. Section 2.4 runs various Monte Carlo experiments to examine the finite sample properties of the maximum likelihood estimator as well as the forecasting performance of the threshold GARCH model. A brief conclusion is contained in Section 2.5.

2.2 Literature Review

2.2.1 Regime Shifts in Conditional Variance

In most widely used GARCH models the conditional variance is defined as a linear function of lagged conditional variances and squared past returns. Formally, let r_t be a sequence of returns, ε_t be a series of innovations that are usually assumed to be independent identically distributed (i.i.d.) zero-mean random variable, σ_t^2 be the variance of r_t given information at time t, the GARCH(p, q) model for returns r_t is defined as follows:

[1] We have estimated return data according to sample sizes ranging from 500, 1000, to 2000.

$$r_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \alpha_j r_{t-j}^2$$

where $p,\ q = 0,\ 1...$ are integers, $i = 1, ...p$, $j = 1, ...q$, $\varepsilon_t \sim i.i.d.N(0,1)$. The parameters must satisfy $\alpha_0 > 0, \alpha_j \geq 0, \beta_i \geq 0$, and $\sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i < 1$ to ensure that the conditional variance is positive and that the asset return series $\{r_t\}$ is covariance stationary. Obviously in this process, the variance is a deterministic function of past variance and squared returns.

Though these models have been proved to be adequate for explaining the dependence structure in conditional variances, they have several important limitations, one of which is that they fail to capture the stylized fact that conditional variance tends to be higher after a decrease in return than after an equal increase. In order to account for this asymmetry many alternative models have been proposed. The exponential GARCH (EGARCH) introduced by Nelson (1991) specifies the conditional variance in logarithmic form[2]:

$$ln\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \beta_i ln\sigma_{t-i}^2 + \sum_{k=1}^{q} \alpha_k [\theta Z_{t-k} + \gamma(|Z_{t-k}| - (\frac{2}{\pi})^{\frac{1}{2}})]$$

$$Z_t = \epsilon_t / \sigma_t$$

The threshold GARCH (TGARCH) model proposed by Zakoian (1994) and GJR GARCH model studied by Glosten, Jagannathan, and Runkle (1993) define the conditional variance as a linear piecewise function. In TGARCH(1,1),

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \delta D_t r_{t-1}^2 + \beta \sigma_{t-1}^2$$

[2]The model takes the asymmetry into account while keeping the linear function form of conditional variance.

$$D_t = \begin{cases} 1 & r_{t-1} < 0 \\ 0 & r_{t-1} \geq 0 \end{cases}$$

More details of such alternative models can be found in the survey of GARCH models by Bollerslev, Chou, and Kroner (1992). The above alternative models are able to characterize some stylized facts better than the GARCH model. However there is no evidence that any alternative model consistently outperforms the GARCH model, for example Hansen and Lunde (2005) claim that nothing beats a GARCH (1,1) in the analysis of the exchange rate data.

The TGARCH and GJR-GARCH models also relax the linear restriction on the conditional variance dynamics. Questioning the common finding of a high degree of persistence to the conditional variance in GARCH model, Lamoureux and Lastrapes (1990) suggest that such high persistence may be spurious if there are regime shifts in the volatility process. From then both ARCH and GARCH models have been implemented with regime switching (RS) framework. The early RS applications, such as Hamilton and Susmel (1994), only allow a Markov-switching ARCH model to describe the conditional variances. Gray (1996) and Klaassen (2002), on the other hand, develop a generalized Markov-switching model, in which a GARCH process in conditional variance is permitted in each regime.

In comparison to the popular Markov-switching models, threshold models have clear conceptual advantages while receiving less attention. The Markov-switching models often fail to generate accurate predictions due to the unknown state in the future. While in the threshold model, the state is governed by an observable threshold variable, therefore is predictable. Knight and Satchell (2011) derive theoretical conditions for the existence of stationary distributions for the threshold models. Based on their work it is now convenient to apply the threshold model to financial time series. To account for the possible structural changes in the conditional variances, we use a threshold model to describe regime switches in the conditional variance process. We simply assume 2 regimes for the conditional variance, which follows a GARCH process within each regime. Different from Markov-switching models, regimes are predictable because their shifts are triggered by an observable variable. We just need to estimate the threshold value which in turn determines the change of the state. The model is more complex since the parameters controlling the conditional variances are changing over time, however it is still flexible in the sense that single regime is a possible outcome in the estimation procedure [3].

2.2.2 Exogenous and Endogenous Threshold Variables

In addition to incorporating the nonlinearity in the threshold GARCH model, the threshold or trigger variable takes into account the effect of correlation between conditional variance and other observable variables that represent trading activities. The use of the threshold model is particularly motivated by the volatility-volume relationship.

At the time of advancing the volatility modeling, an extensive study on stock return volatility-volume relation has been developed. Here the term volume includes any function of raw trading volume. As mentioned in Poon and Granger (2003) the volatility-volume research may lead to a new and better way for modeling the return distribution. The early works on the relationship between stock returns and trading volume summarized in Karpoff (1987) show that volume is positively related to the absolute value of the price change.

[3] If the estimated threshold value is the minimum or maximum of the trigger.

Later works further identify the positive contemporaneous correlation between return volatility and volume (Gallant, Rossi, and Tauchen (1992), and Lamoureux and Lastrapes (1990)).

The empirical works establish various relationships between stock returns and trading volume, yet there is no consensus on how to model the underlying generating process theoretically. The favored theoretical explanation of positive price-volume correlation is the mixture of distribution hypothesis (MDH), which states that the stock returns and trading volume are driven by the same underlying latent information variable (Clark (1973), Epps and Epps (1976), Tauchen and Pitts (1983), Andersen (1996), and Bollerslev and Jubinski (1999)). One encouraging attempt is Andersen's (1996) MDH model in which the joint dynamics of returns and volume is generalized and estimated with a result of significant reduction in the volatility persistence.

More interestingly, recent findings suggest that the size of the trading volume, more specifically the above average volume has significant effect on conditional variance (Wagner and Marsh 2004). Intuitively, the price changes in a stock market can be regarded as a response to arrivals of information, while the volume of shares traded reflects the arrival rate of information. As mentioned in many studies stock prices experience volatile periods with high intensity of information arrivals and tranquil periods accompanied by moderate trading activities. If we assume that the volatility follows different processes in different regimes, obviously volume provides information about which regime the volatility is in.

The established volatility-volume relation motivates the use of volume as the trigger variable in our threshold GARCH model. Since volume and volatility are highly correlated, volume must be treated as an endogenous threshold variable. Nevertheless, other variables that reflect trading activities can also be accommodated. In this chapter we first choose the Chicago Board Options Exchange (CBOE) Volatility Index (VIX) as an exogenous threshold variable since it is a measure of market expectations of near-term volatility and therefore has almost no correlation with current volatility but provides information on the state of the current volatility. The VIX is calculated and disseminated in real-time by CBOE since 1993. It is a weighted blend of prices for a range of options on the S&P 500 index. The formula uses a kernel-smoothed estimator that takes as inputs the current market prices for all near-term and next-term out-of-the-money calls and puts with at least 8 days left to expiration. The goal is to estimate the implied volatility of the S&P 500 index over the next 30 days.

Even though the theoretical conditions for an endogenous threshold variable model cannot be derived, we simulate data according to endogenous threshold variables and examine the performance of the maximum likelihood estimators (MLE) based on different endogeneity levels between the threshold variable and volatility. The simulation results reveal that the increase of the endogeneity coefficient does not affect the performance of MLE.

2.3 Threshold GARCH Model

The setup of the threshold GARCH model is very similar to that of a threshold AR (TAR) model. Knight and Satchell (2011) derive the stationary conditions for TAR models, we follow their approach in deriving the stationary conditions for the return series accordingly.

2.3.1 Introduction

The threshold GARCH model we study in this chapter is defined as follows:

$$r_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \omega_{s_t} + \alpha_{s_t} r_{t-1}^2 + \beta_{s_t} \sigma_{t-1}^2$$

where r_t is the series of demeaned returns and σ_t^2 is the conditional variance of returns given time t information. We assume that the sequence of innovations ε_t follows an independent and identical distribution with mean 0 and variance 1: $\varepsilon_t \sim iidD(0,1)$. The parameters $\{\omega_{S_t}, \alpha_{S_t}, \beta_{S_t}\}$ in the conditional variance equation depend on a threshold variable y_t:

$$\sigma_t^2 = \omega_0 + \alpha_0 r_{t-1}^2 + \beta_0 \sigma_{t-1}^2 \qquad \text{if } S_t = I(y_{t-1} \leq y^*) = 0$$

$$\sigma_t^2 = \omega_1 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \qquad \text{if } S_t = I(y_{t-1} > y^*) = 1$$

where the state or regime of the world S_t is determined by the threshold variable y_{t-1} which can be treated as exogenous or endogenous and the threshold value y^* determines the probability $p(S_t = 1) = p(y_{t-1} > y^*) = \pi$. To simplify the theoretical derivation, we assume the threshold variable is independent of σ_t^2.

As in the standard GARCH(1,1) model we impose the non-negative constraints on all parameters to ensure the conditional variance to be non-negative. However, the conventional stationary conditions for GARCH model may not apply here. Since the conditional variance can fall into 2 different regimes, it is possible that conditional variance is not stationary in one regime but stationary in the other.

For the threshold variable y, we assume that it is a stationary process. This assumption is not critical. We just want to ensure that given a threshold value, if we leave one regime, it is possible that we will return to that regime in the future. If the threshold variable is not stationary, then it is possible that after a point in time, we will only observe one state of the world. This is not a case of interest in this study.

The conditional variance dynamics in the threshold GARCH model we define above is similar to a threshold AR (TAR) model. Knight and Satchell (2011) derive the stationarity conditions for TAR model following the work of Quinn (1982). We follow Knight and Satchell (2011) in deriving the stationarity conditions for the conditional variance and the return series accordingly. Proposition 1 gives conditions for the existence of stationary solution of return process as well as the existence of the mean in the threshold GARCH model. We also examine the conditions for the existence of higher order moments. Proposition 2 provides conditions for the return process to have a stationary variance and Proposition 3 presents conditions for the existence of the fourth moment. Since the return processes experience low autocorrelation but squared returns are highly correlated, we are also interested in examining the theoretical autocorrelation structure of the squared return. Proposition 4 expresses the formulas for the squared return autocovariance and autocorrelation functions.

2.3.2 Conditions for Stationary Return Process

Mean and Variance Stationary Conditions

Given the assumptions that ε_t is iid distributed variable with $D(0,1)$ and is independent of σ_t, it's easy to see that the return series is mean stationary with $E(r_t) = 0$. To simplify the expression of higher order moments we further assume that $\varepsilon_t \sim iid\ N(0,1)$. Thus the unconditional variance and the fourth moment of return are given by $E(r_t^2) = E(\sigma_t^2)$ and $E(r_t^4) = 3E(\sigma_t^4)$. Obviously to examine the stationarity of the return series we need to check the first and second moments of the conditional variance σ_t^2. The following propositions give the conditions under which the stationary distribution of return, the stationary variance, the finite fourth moment of return process, and stationary covariance exist. Proofs of the propositions are provided in Appendix 1-5.

PROPOSITION 1. *The return series is strict stationary if* $\omega_0 < \infty$, $\omega_1 < \infty$, *and:*

$$(1-\pi)E[ln(\varepsilon_{t-m}^2\alpha_0 + \beta_0)] + \pi E[ln(\varepsilon_{t-m}^2\alpha_1 + \beta_1)] < 0$$

■

Remark. If we assume that $\varepsilon_t \sim iidN(0,1)$, then $\varepsilon_t^2 \backsim \chi^2_{(1)}$. We obtain the following analytical expression for the above strict stationarity condition:

$$(1-\pi)E(ln(\varepsilon_{t-m}^2\alpha_0 + \beta_0)) + \pi E(ln(\varepsilon_{t-m}^2\alpha_1 + \beta_1)) < 0 \Longleftrightarrow$$

$$(1-\pi)F(\alpha_0,\beta_0) + \pi F(\alpha_1,\beta_1) < 0$$

where $F(a,b) =$

$$ln(b) + \frac{b}{2a\sqrt{\pi}}\left(-\frac{2(\gamma + ln\frac{2b}{a})\sqrt{\pi}a}{b} - 2\sqrt{\pi}\,{}_2F_2([1,1];[\frac{3}{2},2];\frac{b}{2a}) + \frac{2\pi^{\frac{3}{2}}\sqrt{\frac{a^2}{b}}erfi(\sqrt{\frac{b}{2a}})}{\sqrt{b}}\right).$$

γ is the Euler's constant = 0.577215665.

${}_2F_2(a,b;c,d;x) = \sum_{n=0}^{\infty}\frac{(a)_n(b)_n}{(c)_n(d)_n}\frac{x^n}{n!}$, $(a)_n = a(a+1)...(a+n-1)$

$erfi(x) = -ierf(ix)$, $erf(x)$ is the error function.

We can now examine the first order stationarity conditions for the conditional variance process in our threshold GARCH model.

PROPOSITION 2. *The return series will be variance stationary if* $\omega_0 < \infty$, $\omega_1 < \infty$, *and:*

$$[(\alpha_0 + \beta_0)(1 - \pi) + (\alpha_1 + \beta_1)\pi] < 1$$

Then the stationary variance is given by:

$$Var(r_t) = E(\sigma_t^2) = \sigma^2 = \frac{\omega_0(1 - \pi) + \omega_1\pi}{1 - [(\alpha_0 + \beta_0)(1 - \pi) + (\alpha_1 + \beta_1)\pi]} .$$

■

Higher Order Moments and Covariance Stationary Conditions

Examining the second moment of σ_t^2, we can obtain the fourth moment of returns.

PROPOSITION 3. *If and only if the following conditions hold:*

$$\omega_0 < \infty, \omega_1 < \infty, [(\alpha_0 + \beta_0)(1 - \pi) + (\alpha_1 + \beta_1)\pi] < 1 \text{ and}$$

$$A = [(2\alpha_0^2 + (\alpha_0 + \beta_0)^2)(1 - \pi) + (2\alpha_1^2 + (\alpha_1 + \beta_1)^2)\pi] < 1$$

The fourth moment of the stationary distribution exists for the return process in our threshold model and is given by

$$\begin{aligned}
E(r_t^4) &= 3E(\sigma_t^4) \\
&= 3c_0^2 \frac{1 + a_0 + b_0}{(1-A)(1-(a_0+b_0))} \\
&\quad +3c_1\pi(1-\pi)\frac{c_1(1-(a_0+b_0)) + 2c_0(a_1+b_1)(1-A)}{(1-A)(1-(a_0+b_0)}
\end{aligned}$$

■

Using the results from the first and second moments of σ_t^2, we can now derive the formulas for the autocovariance and autocorrelation functions of squared returns:

$$\gamma(k) = E(r_t^2 - \sigma^2)(r_{t-k}^2 - \sigma^2) \text{ and } \rho(k) = \frac{\gamma(k)}{\gamma(0)}.$$

PROPOSITION 4. *If the conditions in proposition 3 hold, and let* $\gamma(k) = Cov(r_t^2, r_{t-k}^2)$ *and* $\rho(k) = \dfrac{Cov(r_t^2, r_{t-k}^2)}{Var(r_t^2)}$. *Then, for all* $k \geq 2$*,*

$$\gamma(k) = (a_0 + b_0)\gamma(k-1) \textit{ and for all } k \geq 1 \; \rho(k) = (a_0+b_0)^{k-1}\rho(1)$$

where

$$\begin{aligned}
\rho(1) &= \frac{c_0^2[2a_0 - 2a_0b_0A_0 + A_0(2a_1^2 + (a_1+b_1)^2)\pi(1-\pi)]}{c_0^2(2 + A - 3A_0^2) + 3c_1\pi(1-\pi)[c_1(1-A_0) + 2c_0(a_1+b_1)(1-A)]} \\
&\quad + \frac{(3a_0 + b_0(1+2A_0))\pi(1-\pi)[c_1^2 + 2c_0c_1(a_1+b_1)(1-A)]}{c_0^2(2 + A - 3A_0^2) + 3c_1\pi(1-\pi)[c_1(1-A_0) + 2c_0(a_1+b_1)(1-A)]}
\end{aligned}$$

with

$$A = (2\alpha_0^2 + (\alpha_0 + \beta_0)^2)(1-\pi) + (2\alpha_1^2 + (\alpha_1 + \beta_1)^2)\pi$$
$$A_0 = a_0 + b_0 = (\alpha_0 + \beta_0)(1-\pi) + (\alpha_1 + \beta_1)\pi$$

■

2.3.3 Range of Parameters under Different Stationary Conditions

We recall that π is the probability that the volatility process is in regime 2, α_0, β_0 are parameters in regime 1 and α_1, β_1 are parameters in regime 2. From the stationary conditions derived in the previous section, we note that since the parameter π enters into the conditions, the sum of the parameter values in each regime is no longer required to be less than one. For example to have a strict stationary return process, we allow the sum of the parameters in both regimes to be bigger than one. However, for GARCH type of models we usually require the finite variance of the return process. To obtain a variance stationary process we just need the weighted sum of the sums of parameters in two regimes to be less than one: $[(\alpha_0 + \beta_0)(1-\pi) + (\alpha_1 + \beta_1)\pi] < 1$, therefore we may have a sum of parameters in one regime to be bigger than one. To examine the relationship between π and the range of stationary areas, we graph the stationary areas of parameters in one regime based on different π values for the fixed parameter values in another regime.

We discuss the stationary areas of α_1 and β_1 when π varies from 0.1, 0.5, to 0.9 for four sets of parameter values of α_0 and β_0.[4] According to the stationary

[4] We set $\{\alpha_0, \beta_0\}$ = {0.25, 0.75},{0.25, 0.5}, {0.25, 0.25}, {0.25, 0} respectively.

conditions we derived in last section, the return series will have a variance stationary distribution if $(\alpha_0+\beta_0)(1-\pi)+(\alpha_1+\beta_1)\pi < 1$, and the fourth moment of return series exists if $(2\alpha_0^2 + (\alpha_0 + \beta_0)^2)(1 - \pi) + (2\alpha_1^2 + (\alpha_1 + \beta_1)^2)\pi < 1$. It's easy to verify that if π increases, the weight of $(\alpha_1 + \beta_1)$ increases, therefore the range of $(\alpha_1 + \beta_1)$ will decrease for all cases where $(\alpha_0 + \beta_0) < 1$. In Figure 2.1 we observe the boundaries of the strict stationary area, the variance stationary area, and the fourth moment stationary area move towards the origin when π increases from 0.1 to 0.9. However when $(\alpha_0 + \beta_0) = 1$, there is no clear pattern for the movement of the stationary areas of $(\alpha_1 + \beta_1)$ with different π values.

In Figure 2.1[5] the areas below the solid, dotted, and dashed lines satisfy the three stationary restrictions for π varying from 0.1, 0.5, to 0.9 respectively. For each π value, the areas are shrinking when we impose further restriction on stationarity. For each π value, it was supposed to have three lines corresponding to three stationarity conditions, however it excludes the strict stationary condition for $\pi = 0.1$ since it has a x-axis intercept beyond 90. It is not very surprising, since when $\pi = 0.5$, the strict stationary requirement for α_1 is that its value less than 3.43 when $\beta_1 = 0.1$, if probability π is much smaller, the value of the parameter that satisfies the strict stationary can be very large.

When $(\alpha_0 + \beta_0) = 1$, there is no values of α_1 and β_1 that satisfy the fourth-order stationarity condition when $\pi = 0.1$. Therefore in Figure 2.2, there are only 2 solid lines representing strict stationary and variance stationary areas. Since the sum of parameters in regime one is one, the restriction for variance stationary distribution requires the sum of parameters in regime two to be less than 1 regardless of the value of π, so the boundaries for the variance stationary distribution in the graph are identical for all different π values (represented

[5] In the legend of the graph, SS=strict stationary, 2S=variance stationary, 4S=4th order stationary.

by the solid blue line in Figure 2.2). When $(\alpha_0 + \beta_0) = 1$,there is also no pattern for the stationary areas when π changes, the strict stationary area for $\pi = 0.5$ is larger than that for $\pi = 0.9$, whereas the fourth-order stationary area for $\pi = 0.9$ is larger than that for $\pi = 0.5$. Nevertheless for all three cases in which $(\alpha_0 + \beta_0) < 1$, we observe a clear pattern that when π increases, the stationary areas shrink.

When $\{\alpha_0, \beta_0\}$ = {0.25, 0.25}, {0.25, 0}, the graphs exhibit a similar pattern as the graph in Figure 2.1 for $\{\alpha_0, \beta_0\}$ = {0.25, 0.5}.

2.4 Monte Carlo Simulation

Since the threshold GARCH model is set up as a non-linear model, the volatility process is non-differentiable with respect to the parameters at the point where regime switches from one to another. The derivation of the asymptotic distribution of the maximum likelihood estimator is beyond the scope of this thesis. Instead, we conduct a Monte Carlo simulation study to examine the finite sample properties of the MLE according to different stationarity conditions.

The simulation study is also implemented to investigate the forecasting performance of the threshold GARCH model because the forecasting performance is a key element in selecting a volatility model.

2.4.1 The Simulated Paths of Return Series

In the previous section we derive the stationarity conditions for the return series described by our threshold GARCH model. Now we proceed with a simulation study to examine the estimation performance of this model under different stationary conditions.

For the simulation study, we choose 3 sets of parameters for $\pi = 0.1, 0.5, 0.9$ respectively. The value of ω_0 and ω_1 are set to be 0.02 and 0.01 for all cases. The values of α_0 and β_0 are fixed at 0.25 and 0.5, then β_1 is selected from the different regions in the stationary areas given $\alpha_1 = 0.25$ as shown in Figure 2.1. We choose the parameters in such a way that the regime 1 is always stationary, based on different probabilities with which conditional variance shifts to regime 2, we could have a non-stationary regime 2 but the whole process is still stationary.

Case 1, $\pi = 0.1$:

$$\begin{cases} 1.1\ \textit{Stationary with 4th moment} & \{\alpha_1, \beta_1\} = \{0.25, 1.5\} \\ 1.2\ \textit{Variance Stationary} & \{\alpha_1, \beta_1\} = \{0.25, 2.5\} \\ 1.3\ \textit{Strict Stationary} & \{\alpha_1, \beta_1\} = \{0.25, 3\} \end{cases}$$

Case 2, $\pi = 0.5$:

$$\begin{cases} 2.1\ \textit{Stationary with 4th moment} & \{\alpha_1, \beta_1\} = \{0.25, 0.75\} \\ 2.2\ \textit{Variance Stationary} & \{\alpha_1, \beta_1\} = \{0.25, 0.9\} \\ 2.3\ \textit{Strict Stationary} & \{\alpha_1, \beta_1\} = \{0.25, 1\} \end{cases}$$

Case 3, $\pi = 0.9$:

$$\begin{cases} 3.1\ \textit{Stationary with 4th moment} & \{\alpha_1, \beta_1\} = \{0.25, 0.7\} \\ 3.2\ \textit{Variance Stationary} & \{\alpha_1, \beta_1\} = \{0.25, 0.75\} \\ 3.3\ \textit{Strict Stationary} & \{\alpha_1, \beta_1\} = \{0.25, 0.9\} \end{cases}$$

Using Case 1.1 as an example, the data generating process is described as follows:

$$r_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \begin{cases} 0.02 + 0.25 r_{t-1}^2 + 0.5 \sigma_{t-1}^2 & if \quad y_{t-1} \leq y^* \\ 0.01 + 0.25 r_{t-1}^2 + 1.5 \sigma_{t-1}^2 & if \quad y_{t-1} > y^* \end{cases}$$

ε_t, y_t is drawn independently from standard normal distribution, y^* is chosen in a way such that $p(S_t = 1) = p(y_t > y^*) = 0.1$, and σ_0 is set to 0. We generate 5000 observations using the specified parameters, and to eliminate the possible initial value effect, we drop first 3000 observations.

The paths of return series depend crucially on the parameters in volatility process. Figure 2.3 shows the stationary paths of return series given that parameters are specified as in Case 1.1.[6] In this case we have a stationary return process even when the sum of parameters is substantially bigger than 1 in one regime, but in that regime we have only 10% of the observations. To better describe the volatility process in reality, Figure 2.4 shows the plot of return series given that parameters are specified as in Case 2.1[7] in which case the return series is variance stationary and has equal chance to stay in high or low volatility regime.

2.4.2 The Performance of MLE

In this section we examine the performance of the maximum likelihood estimator. Given that the return series is conditionally normally distributed, the

[6]The parameters used in the simulated path are: $\pi = 0.1$ $\{\omega_0, \alpha_0, \beta_0\} = \{0.02, 0.25, 0.5\}$ $\{\omega_1, \alpha_1, \beta_1\} = \{0.01, 0.25, 1.5\}$

[7]The parameters used in the simulated path are: $\pi = 0.5$ $\{\omega_0, \alpha_0, \beta_0\} = \{0.02, 0.25, 0.5\}$ $\{\omega_1, \alpha_1, \beta_1\} = \{0.01, 0.25, 0.75\}$

log likelihood function for a sample of T observations is:

$$lnL_T(\theta) = -1/2\sum_{t=1}^{T} ln\sigma_t^2 - 1/2\sum_{t=1}^{T}\frac{r_t^2}{\sigma_t^2}$$

where θ is a vector of parameters in the conditional variance process, $\theta = [\omega_0, \omega_1, \alpha_0, \alpha_1, \beta_0, \beta_1]$.

We know that to estimate θ, we need to estimate the threshold value y^* so that the above likelihood function can be formulated. Here we estimate y^* by grid search, the threshold variable y_t is sorted and for each possible threshold value y^* we calculate the corresponding likelihood and the estimated threshold value is the one which maximizes the likelihood:

$$\hat{\theta}(y^*) = \underset{\theta\in\Theta}{argmax} T^{-1} lnL_T(\theta)$$

We conduct a simulation study to analyze the finite sample properties of the maximum likelihood estimator. The maximization problem always converges to a solution in our simulation study.

First, we present the simulation results for 3 sets of parameters in Case 1 when $\pi = 0.1$, considering the sample sizes for 500, 1000, and 2000. The MSE is defined as mean squared errors of estimators from true parameter values $MSE = \frac{1}{T}\sum(\hat{\theta}-\theta)^2$ for $\theta = [\omega_0, \omega_1, \alpha_0, \alpha_1, \beta_0, \beta_1]$. The results are based on 1000 replications. For simplicity we estimate the threshold value by searching over 19 grid points ranging from the $5th$ percentile to the $95th$ percentile point of threshold variable in jumps of 5%.

Table 2.1 presents the simulation results for 3 sets of parameters in Case 1. When $\pi = 0.1$, the probability that the conditional variance changes to regime 2 is small, we just need the sum of parameters to be less than 3.25 to fulfill the requirement for a strict stationary distribution in Case 1.3. Moving down from the upper row section, the sample size increases from 500, 1000, to 2000. While the columns represent different stationary requirements, from left to right, the return series are strict stationary with finite fourth moment (1.1), variance stationary (1.2), and strict stationary (1.3) respectively. The MLE appears to be consistent with the mean value of the estimates approaching the true parameter value when the sample size increases from 500 to 2000. The MSE decrease when sample size increases. We notice that the MSE for α_1 and β_1 are substantially larger than that of α_0 and β_0, this is caused by the nature of non-stationarity in the corresponding regime and small probability to enter that regime. When $\pi = 0.1$, only 10% of the observations belong to the regime associated with α_1 and β_1, it is obvious that we won't be able to obtain a good estimate with only 50 to 200 observations. We also note that the non-existence of moments results in more biased estimators and fatter tails in the distribution of estimators as we move across the Table 2.1 from left to right. The left column has finite first, second, and fourth moments, the middle column has finite first and second moments, while the right column has only finite first moment. Without the existence of finite variance, the MSE can be very large, but here the estimates are reasonably good since only small portion of data is generated by the non-stationary regime.

Figure 2.5-2.7 provide the estimated density of MLE summarized in the above table. The estimated density is computed using kernel smoothing method. The MLE is well behaved even when the variance stationarity condition is violated. As sample size increases from 500 to 2000, the estimates become more accurate with smaller variances and more concentrated around true parameters. We present the density estimates for sample size of 500, 1000, and 2000 in dotted line, dashed line, and dotted and dashed line respectively, while the true parameter values are given by the solid vertical line.

Similar results are obtained for other 2 cases and reported in Table 2.2-2.3. Table 2.2 presents the simulation results for 3 sets of parameters in Case 2. The MLE still has a good limiting behavior as sample size increases. We also observe that the MSE of estimated parameters in each regime are not substantially different as reported in Table 2.1. It may be due to the fact that probability that conditional variance stays in each regime is equal, and we also expect higher MSE for estimated parameters in non-variance stationary regime. Table 2.3 presents simulation results in Case 3. When the probability that conditional variance process in regime 2 equals 0.9, even in the variance stationary case we will no longer have accurate estimates for parameter β_0, so we just report the results for Case 3.1 and 3.2 and skip the strict stationary Case 3.3. With 90% of the observations in a regime without the finite variance we don't expect any reliable estimation result. Since regime 2 is more volatile regime, here the high probability that the conditional variance is in such regime may be the reason that we fail to obtain accurate estimators. We also notice that the estimated parameters in regime 1 turn out to have larger MSE. It confirms our assertion that the small probability in one regime affects the estimation performance in that regime.

Figure 2.8-2.10 provide the estimated density of MLE summarized in Table 2.2. While, Figure 2.11-2.12 provide the estimated density of MLE summarized in Table 2.3.

The threshold value is estimated by the grid search, the limiting behavior is somehow different. We plot the estimated density for y^* in Figure 2.13 and 2.14. In Figure 2.13, we plot the estimated density for the estimated threshold value in the case that the return series has a finite fourth moment when π

changing from 0.1 to 0.9. While in Figure 2.14, we plot the estimated density for y^* in the case that the return series has only finite variance with different probabilities. The estimation for y^* is improved when sample size increases. However when the probability in the volatile regime increases, the estimated thresholds are less accurate and with fatter tails.

2.4.3 Simulation Study for Endogenous Threshold Variable

The well established volume-volatility relationship inspires the use of volume as the threshold variable in our threshold GARCH model, however the high correlation between volume and volatility makes this threshold variable endogenous. The endogeneity of the threshold variable renders the theoretical derivation of the stationarity conditions impossible. Therefore we design a simulation study to examine the effect of endogeneity of the threshold variable on the properties of the return series and the maximum likelihood estimator.

We simply assume that the threshold variable is a linear function of squared returns with some random error.

Under the threshold GARCH model, the demeaned return series and the conditional variance are given by:

$$r_t = \sigma_t \varepsilon_t$$

$$\begin{cases} \sigma_t^2 = \omega_0 + \alpha_0 r_{t-1}^2 + \beta_0 \sigma_{t-1}^2 & if\ y_{t-1} \leq y^* \\ \sigma_t^2 = \omega_1 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2 & if\ y_{t-1} > y^* \end{cases}$$

$$y_t = a r_t^2 + v_t$$

where ε_t follows the independent and identical normal distribution with mean 0 and variance 1: $\varepsilon_t \sim iidN(0,\ 1)$, and v_t is an i.i.d. normal variable with mean 0 and variance σ_v^2. ε_t and v_t are independent.

In this simple data generating process the correlation between the squared return r_t^2 and the threshold variable y_t is governed by the coefficient a when σ_v is small.

$$\begin{aligned}
y_t &= ar_t^2 + v_t \\
E(y_t) &= aE(r_t^2) \\
Var(y_t) &= a^2Var(r_t^2) + \sigma_v^2 \\
Cov(r_t^2,\ y_t) &= E(r_t^2 y_t) - E(r_t^2)E(y_t) \\
&= E(ar_t^4) - E(r_t^2)aE(r_t^2) \\
&= aVar(r_t^2) \\
Corr(r_t^2,\ y_t) &= \frac{Cov(r_t^2,\ y_t)}{\sqrt{Var(r_t^2)Var(y_t)}} \\
&= \frac{aVar(r_t^2)}{\sqrt{Var(r_t^2)(a^2Var(r_t^2) + \sigma_v^2)}} \\
&= (\frac{a^2Var(r_t^2)^2}{a^2Var(r_t^2)^2 + \sigma_v^2 Var(r_t^2)})^{1/2}
\end{aligned}$$

It is obvious that as $a \to 1$ and $\sigma_v \to 0$, $Corr(r_t^2,\ y_t) \to 1$, and as $a \to 0$, $Corr(r_t^2,\ y_t) \to 0$.

We choose the value of $a = (0.1,\ 0.2,\ 0.3,\ 0.4)$ and $\sigma_v = 0.1$, the parameter values in the conditional variance process are set as:

$$\{\omega_0,\ \omega_1,\ \alpha_0,\ \alpha_1,\ \beta_0,\ \beta_1\} = \{0.01,\ 0.02,\ 0.1,\ 0.2,\ 0.55,\ 0.75\}.$$

Corresponding to these parameter values, the average correlation coefficients between the squared return and the threshold variable are:

$$\{0.09, 0.22, 0.36, 0.51\}$$

for the simulated data sets. We generate 1000 bivariate series each with 4000 observations, the first 2000 observations are dropped to eliminate the initial value problem.

We use MSE to evaluate the performance of estimators. The MSE is defined as mean squared errors of estimators from true parameter values $MSE = \frac{1}{T}\sum(\hat{\theta} - \theta)^2$ for $\theta = [\omega_0, \omega_1, \alpha_0, \alpha_1, \beta_0, \beta_1]$.

The results are based on 1000 replications. For simplicity we estimate the threshold value by searching over 19 grid points range from $5th$ percentile to $95th$ percentile points of threshold variable.

Table 2.4 presents the results of estimates when correlation coefficient a ranging from 0.1 to 0.4.

In Table 2.4, we see that the performance of the MLE is improved when sample size increases. The performance of most estimated parameters is also improved when the correlation between the squared returns and the threshold variable increases. When sample size is 1000, the MSE of MLE decreases when a changes from 0.1 to 0.4, except for one parameter α_0. Nevertheless, as the sample size is 2000, the changes in MSE of α_0 become smaller when correlation coefficient a increases.

The results seem inconsistent at the first glance, usually we fail to obtain a consistent estimator when dealing with endogenous variables in solving economic problems. However here the threshold variable is not an explanatory variable in the return or the volatility dynamics, it is an information variable and the higher the endogeneity the more information provided by the threshold variable, therefore the better the performance of the estimator.

2.4.4 Simulation Study for Forecasting Performance

In the previous sections we show that the parameters from the threshold GARCH model can be well estimated. Now we examine the forecasting performance of the threshold model since the predicting power is critical in determining the success of a volatility model. First, assuming that the data is generated by a threshold model we estimate both the threshold model and simple GARCH(1,1) model, then construct the forecasts based on estimated parameters from two models, the results are compared according to 5 common measures. Then, we conduct a model misspecification test. Assuming that data is generated by GARCH model, but we use estimated parameters from threshold model to forecast volatility.

Forecasting Performance of Threshold Model with Threshold DGP

First, we assume that the return data is generated by the threshold GARCH model. In previous sections we show that different from GARCH model, threshold model allows the sum of parameters to be greater than one in one regime but keeps the whole process stationary. Therefore we use two sets of parameters to generate stationary return data, one with stationary parameters in both regimes, another one with non-stationary parameters in one regime but keeping the process stationary.

We generate 1000 return series given the stationary parameters as follows:

$$
\begin{aligned}
r_{i,t} =& \sigma_{i,t}\varepsilon_{i,t} \\
\sigma_{i,t}^2 = 0.02 + 0.15\sigma_{i,t-1}^2\varepsilon_{i,t-1}^2 + 0.55\sigma_{i,t-1}^2 & \qquad \text{if } y_{t-1} \leq y^* \\
\sigma_{i,t}^2 = 0.01 + 0.05\sigma_{i,t-1}^2\varepsilon_{i,t-1}^2 + 0.90\sigma_{i,t-1}^2 & \qquad \text{if } y_{t-1} > y^*
\end{aligned}
$$

where the innovations $\varepsilon_{i,t}$ are independently and identically distributed standard normal random variables, and we use VIX data for the threshold variable y_t with $y^* = mean(y_t) = 19.6696$ for all return series from $i = 1...1000$.

We generate 5000 observations for each return series and drop first 2000 observations to eliminate the initial value effect.

In the sample of 3000 observations, the first 2750 observations are used to estimate the threshold GARCH model, then the estimated parameters and threshold value are used to construct the one-day ahead forecast for the remaining 250 days. In the simulation study, we use the following 5 measures to compare the forecasting performance for ith replication of return series:

$$
\begin{aligned}
ME_i &= \frac{1}{T}\sum(\sigma_{i,t}^2 - \hat{\sigma}_{i,t}^2) \\
MPE_i &= \frac{1}{T}\sum(\sigma_{i,t}^2 - \hat{\sigma}_{i,t}^2)/\hat{\sigma}_{i,t}^2 \\
RMSE_i &= \sqrt{\frac{1}{T}\sum(\sigma_{i,t}^2 - \hat{\sigma}_{i,t}^2)^2} \\
HMSE_i &= \frac{1}{T}\sum(\sigma_{i,t}^2/\hat{\sigma}_{i,t}^2 - 1)^2
\end{aligned}
$$

and R_i^2 obtained from regressing the actual conditional variance σ_i^2 on the forecasts $\hat{\sigma}_i^2$:

$$
\sigma_i^2 = a + b\hat{\sigma}_i^2 + \eta_i
$$

For each measure we compute the average over 1000 replications and compare with the same measure obtained by the standard GARCH(1,1) model estimation. The results are presented in Table 2.5.

The results show that if the data are generated by the threshold GARCH model, then the forecasting performance of the threshold GARCH model is much better than that of the standard GARCH model. Figure 2.15 gives an example of the comparison of the estimated volatility by threshold GARCH model and GARCH model.

The above estimated results of threshold GARCH model is based on a grid search of threshold variable over 19 points ranging from the $5th$ percentile to the $95th$ percentile of the threshold variable in jumps of 5%.

In order to obtain an efficient estimator of the threshold value, we should search over all values in the sample of the threshold variable. Since each grid search involves a minimization problem, it is very computationally costly to search over the entire sample of the threshold variable. Due to the computational burden, we only search the very limited points in the range of the threshold variable. It is expected that as finer grid intervals are used the performance of the estimator will improve. We also examine this effect of number of grid points on the estimation performance through a simulation study. Table 2.6 presents the results of estimated parameters over 3 different jump sizes in the grid points.

It is clear from Table 2.6 that as the finer grid intervals are used, the estimated parameters and threshold value are more closer to that of true parameter values. Nevertheless, the estimation results using a coarser grid interval are not much worse than using a finer grid interval, but the estimation process using the finest grid interval requires much more computing time than that using a coarser interval. Therefore, in the empirical application we search over 37 points ranging from the $5th$ percentile to the $95th$ percentile point of threshold variable in jumps of 2.5%.

Since the threshold model allows the parameters to be non-stationary in one regime while keeping the whole process stationary, we also generate return data according to the following model:

$$
\begin{aligned}
r_{i,t} &= \sigma_{i,t}\varepsilon_{i,t} \\
\sigma^2_{i,t} &= 0.02 + 0.15\sigma^2_{i,t-1}\varepsilon^2_{i,t-1} + 0.55\sigma^2_{i,t-1} \qquad \text{if } y_{t-1} \leq y^* \\
\sigma^2_{i,t} &= 0.01 + 0.25\sigma^2_{i,t-1}\varepsilon^2_{i,t-1} + 0.80\sigma^2_{i,t-1} \qquad \text{if } y_{t-1} > y^*
\end{aligned}
$$

where the innovations $\varepsilon_{i,t}$ are independently and identically distributed standard normal random variables, and we use VIX data for the threshold variable y_t with $y^* = 31$ for all return series from $i = 1...1000$.

Similarly, we generate 5000 observations for each return series and drop first 2000 observations to eliminate the initial value effect. In the sample of 3000 observations, the first 2750 observations are used to estimate the threshold GARCH model, then the estimated parameters and threshold value are used to construct the one-day ahead forecast for the remaining 250 days. Table 2.7 gives the results of forecasting comparison based on 5 measures.

The results show that even if the parameters are non-stationary in one regime, both models can generate reasonable forecasts based on high R^2. Nonetheless, the threshold model still performs significantly better than GARCH model. The return data has only 10% generated by the non-stationary process, which possibly explains why R^2 increases in this case for GARCH model.

Forecasting Performance of Threshold Model with GARCH DGP

We showed that if the data generating process follows the threshold GARCH model, the forecasting performance is quite good using the threshold model. The average R^2 from regressing the true volatility on the estimated volatility over 1000 replication is more than 90%. Now we perform a model misspecification test on the forecasting power of the threshold GARCH model. If the data generating process follows a standard GARCH process, theoretically we will estimate the model well therefore also forecast well. To examine this property, we simulate data according to a GARCH process:

$$r_{i,t} = \sigma_{i,t}\varepsilon_{i,t}$$
$$\sigma^2_{i,t} = 0.02 + 0.05\sigma^2_{i,t-1}\varepsilon^2_{i,t-1} + 0.85\sigma^2_{i,t-1}$$

where the innovations $\varepsilon_{i,t}$ are independently and identically distributed standard normal random variables, and when estimate the threshold model we use an independent standard normal variable as the threshold variable y_t for all return series from $i = 1...1000$.

We generate 5000 observations for each return series and drop first 2000 observations to eliminate the initial value effect. In the remaining sample of 3000 observations, we use 2750 observations for in-sample estimation and the rest 250 for out-sample forecasting. Table 2.8 shows the average of 5 forecasting measures over 1000 replications.

Clearly if the data are generated by a standard GARCH model, the threshold GARCH model is able to provide accurate forecasts very close to the performance of GARCH model.

Figure 2.16 provides an example of the comparison of the estimated volatility by threshold GARCH model and GARCH model given the GARCH data generating process.

2.5 Conclusion

In this chapter, we propose a threshold GARCH model to describe the regime-switching in the volatility process of financial asset returns. The model assumes that there are two regimes in the volatility dynamics, within each regime the volatility follows a GARCH process. To validate the use of a threshold model in finance, we derive various stationarity conditions for the return process under the threshold GARCH model. According to different stationarity conditions, we implement simulation studies to examine the finite sample properties of the maximum likelihood estimator. The results show that the MLE performs well when the stationarity conditions hold. We then conduct simulation experiments to examine the performance of estimators under the assumption that the threshold variable is endogenous. It is suggested that as the correlation between the return and the threshold variable increases we may actually obtain more accurate estimators.

The ability to provide accurate volatility forecasts is a very important criterion to test a volatility model. We further examine the forecasting performance of the threshold GARCH model via simulation studies. The performance of daily out-of-sample volatility forecasts is evaluated using 5 common measures. The threshold GARCH model is able to provide good volatility forecasts regardless the data is generated by threshold model or not. In the next chapter, we apply the threshold GARCH model to empirical data.

2.6 Appendix

Appendix 1: Proof of Proposition 1

We rewrite the conditional variance equation as:

$$\sigma_t^2 = \omega_0(1-S_{t-1})+\omega_1 S_{t-1}+(\alpha_0(1-S_{t-1})+\alpha_1 S_{t-1})r_{t-1}^2+(\beta_0(1-S_{t-1})+\beta_1 S_{t-1})\sigma_{t-1}^2$$

$$= \omega_0 + (\omega_1 - \omega_0)S_{t-1} + [(\alpha_0 + (\alpha_1 - \alpha_0)S_{t-1})\varepsilon_{t-1}^2 + \beta_0 + (\beta_1 - \beta_0)S_{t-1}]\sigma_{t-1}^2 \quad \textbf{(2.1)}$$

Let

$$c_0 = \omega_0(1-\pi) + \omega_1\pi \qquad c_1 = \omega_1 - \omega_0$$
$$a_0 = \alpha_0(1-\pi) + \alpha_1\pi \qquad a_1 = \alpha_1 - \alpha_0$$
$$b_0 = \beta_0(1-\pi) + \beta_1\pi \qquad b_1 = \beta_1 - \beta_0$$
$$B_t = S_t - \pi$$

Then (2.1) can be rewritten as:

$$\sigma_t^2 = c_0 + c_1 B_{t-1} + [\varepsilon_{t-1}^2(a_0 + a_1 B_{t-1}) + b_0 + b_1 B_{t-1}]\sigma_{t-1}^2 \quad \textbf{(2.2)}$$

The transformation makes the random variables in the coefficient have mean zero. Back substitution in (2.2) results in:

$$\begin{aligned}\sigma_t^2 \quad = \quad & c_0 + c_1 B_{t-1} + \sum_{n=1}^{k-1}(c_0 + c_1 B_{t-1-n})\prod_{m=1}^{n}[\varepsilon_{t-m}^2(a_0 + a_1 B_{t-m}) + b_0 + b_1 B_{t-m}] \\ & + \prod_{m=1}^{k}[\varepsilon_{t-m}^2(a_0 + a_1 B_{t-m}) + b_0 + b_1 B_{t-m}]\sigma_{t-k}^2\end{aligned}$$

Following Quinn (1982) and Knight and Satchell (2011) and defining $S_n(t)$ as:

$$S_n(t) = \prod_{m=1}^{n}[\varepsilon_{t-m}^2(a_0 + a_1 B_{t-m}) + b_0 + b_1 B_{t-m}]$$

Then we have

$$ln(S_n(t)) = \sum_{m=1}^{n} ln[\varepsilon_{t-m}^2(a_0 + a_1 B_{t-m}) + b_0 + b_1 B_{t-m}] \tag{2.3}$$

and

$$\frac{1}{n} ln(S_n(t)) \xrightarrow{a.s.} E(ln[\varepsilon_{t-m}^2(a_0 + a_1 B_{t-m}) + b_0 + b_1 B_{t-m}])$$

by the strong law of large numbers.

It's very clear that if

$$E(ln[\varepsilon_{t-m}^2(a_0 + a_1 B_{t-m}) + b_0 + b_1 B_{t-m}]) < 0$$

by the independence assumption between ε and y that is if

$$(1-\pi)E(ln(\varepsilon_{t-m}^2 \alpha_0 + \beta_0)) + \pi E(ln(\varepsilon_{t-m}^2 \alpha_1 + \beta_1)) < 0,$$

then the terms $S_n(t)$ are geometrically bounded as n increases and equation (2.2) has the solution:

$$\sigma_t^2 = c_0 + c_1 B_{t-1} + \sum_{n=1}^{\infty}(c_0 + c_1 B_{t-1-n}) S_n(t) \tag{2.4}$$

Appendix 2: Proof of Proposition 2

Taking expectation on both sides of equation (2.3) we have:

$$E(\sigma_t^2) = E(c_0 + c_1 B_{t-1}) + \sum_{n=1}^{\infty} E(c_0 + c_1 B_{t-1-n}) E(S_n(t)) \tag{2.5}$$

Given that ε_t and S_t are independent, we have:

$$\begin{aligned} E(S_n(t)) &= E[\prod_{m=1}^{n} (\varepsilon_{t-m}^2 (a_0 + a_1 B_{t-m}) + b_0 + b_1 B_{t-m})] \\ &= \prod_{m=1}^{n} [E(\varepsilon_{t-m}^2 (a_0 + a_1 B_{t-m})) + E(b_0 + b_1 B_{t-m})] \\ &= (a_0 + b_0)^n \end{aligned}$$

Provided that $a_0 + b_0 < 1$, that is $(\alpha_0 + \beta_0)(1 - \pi) + (\alpha_1 + \beta_1)\pi < 1$, the equation (2.4) becomes:

$$\begin{aligned} E(\sigma_t^2) &= E(c_0 + c_1 B_{t-1}) + \sum_{n=1}^{\infty} E(c_0 + c_1 B_{t-1-n}) E(S_n(t)) \\ &= c_0 + c_0 \sum_{n=1}^{\infty} (a_0 + b_0)^n \\ &= c_0 \sum_{n=0}^{\infty} (a_0 + b_0)^n \\ &= \frac{c_0}{1 - (a_0 + b_0)} \\ &= \frac{\omega_0 (1 - \pi) + \omega_1 \pi}{1 - [(\alpha_0 + \beta_0)(1 - \pi) + (\alpha_1 + \beta_1)\pi]} \end{aligned}$$

Appendix 3: Proof of Proposition 3

To examine the stationarity conditions for higher order moments of return, we check the second moment of σ_t^2:

$$\begin{aligned}
E(\sigma_t^4) &= E[(c_0 + c_1 B_{t-1}) + \sum_{n=1}^{\infty}(c_0 + c_1 B_{t-1-n})S_n(t)]^2 \\
&= E(c_0 + c_1 B_{t-1})^2 + E[2(c_0 + c_1 B_{t-1})\sum_{n=1}^{\infty}(c_0 + c_1 B_{t-1-n})S_n(t)] \\
&\quad + E[\sum_{n=1}^{\infty}(c_0 + c_1 B_{t-1-n})S_n(t)]^2 \\
&= c_0^2 + c_1^2\pi(1-\pi) + 2c_0^2\sum_{n=1}^{\infty}E(S_n(t)) + 2c_0c_1\sum_{n=1}^{\infty}E(B_{t-1-n}S_n(t)) \\
&\quad + 2c_0c_1\sum_{n=1}^{\infty}E(B_{t-1}S_n(t)) + 2c_1^2E[B_{t-1}\sum_{n=1}^{\infty}B_{t-1-n}S_n(t)] \\
&\quad + E[\sum_{n=1}^{\infty}(c_0 + c_1 B_{t-1-n})S_n(t)]^2
\end{aligned}$$

Note that

$$\begin{aligned}
E(B_{t-1}S_n(t)) &= E\{B_{t-1}[\varepsilon_{t-1}^2(a_0 + a_1 B_{t-1}) + b_0 + b_1 B_{t-1}]S_n(t-1)\} \\
&= (a_1 + b_1)\pi(1-\pi)(a_0 + b_0)^{n-1}
\end{aligned}$$

$$\begin{aligned}
E[\sum_{n=1}^{\infty}(c_0 + c_1 B_{t-1-n})S_n(t)]^2 &= E[c_0\sum_{n=1}^{\infty}S_n(t) + c_1\sum_{n=1}^{\infty}B_{t-1-n}S_n(t)]^2 \\
&= E[c_0^2(\sum_{n=1}^{\infty}S_n(t))^2] + E[2c_0c_1\sum_{n=1}^{\infty}S_n(t)\sum_{n=1}^{\infty}B_{t-1-n}S_n(t)] \\
&\quad + E[c_1\sum_{n=1}^{\infty}B_{t-1-n}S_n(t)]^2
\end{aligned}$$

$$
\begin{aligned}
E[c_0^2(\sum_{n=1}^{\infty}S_n(t))^2] &= c_0^2[\sum_{n=1}^{\infty}E(S_n^2(t)) + 2\sum_{n=1}^{\infty}\sum_{l=n+1}^{\infty}E(S_n(t)S_l(t)] \\
E[2c_0c_1\sum_{n=1}^{\infty}S_n(t)\sum_{n=1}^{\infty}B_{t-1-n}S_n(t)] &= 0 \\
E[c_1\sum_{n=1}^{\infty}B_{t-1-n}S_n(t)]^2 &= c_1^2[\pi(1-\pi)\sum_{n=1}^{\infty}E(S_n^2(t)) \\
&\quad +2\sum_{n=1}^{\infty}\sum_{l=n+1}^{\infty}E(B_{t-1-n}B_{t-1-l}S_n(t)S_l(t)]
\end{aligned}
$$

where

$$
\begin{aligned}
E(S_n^2(t)) &= E\prod_{m=1}^{n}(\varepsilon_{t-m}^2(a_0+a_1B_{t-m})+b_0+b_1B_{t-m})^2 \\
&= E\prod_{m=1}^{n}[\varepsilon_{t-m}^4(a_0+a_1B_{t-m})^2 + 2\varepsilon_{t-m}^2(a_0+a_1B_{t-m})(b_0+b_1B_{t-m}) \\
&\quad +(b_0+b_1B_{t-m})^2] \\
&= [3a_0^2 + a_1^2\pi(1-\pi) + 2a_0b_0 + 2a_1b_1\pi(1-\pi) + b_0^2 + b_1^2\pi(1-\pi)]^n \\
&= [2a_0^2 + (a_0+b_0)^2 + (2a_1^2 + (a_1+b_1)^2)\pi(1-\pi)]^n
\end{aligned}
$$

Let $A = 2a_0^2 + (a_0+b_0)^2 + (2a_1^2 + (a_1+b_1)^2)\pi(1-\pi)$, then

$$
\begin{aligned}
E(S_n(t)S_l(t) &= E[\prod_{m=1}^{n}(\varepsilon_{t-m}^2(a_0+a_1B_{t-m})+b_0+b_1B_{t-m})^2 \\
&\quad \prod_{m=n+1}^{\infty}(\varepsilon_{t-m}^2(a_0+a_1B_{t-m})+b_0+b_1B_{t-m})] \\
&= A^n(a_0+b_0)^{l-n}
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
E[c_0^2(\sum_{n=1}^{\infty} S_n(t))^2] &= c_0^2(\sum_{n=1}^{\infty} A^n + 2\sum_{n=1}^{\infty}\sum_{l=n+1}^{\infty} A^n(a_0+b_0)^{l-n}) \\
E[c_1\sum_{n=1}^{\infty} B_{t-1-n}S_n(t)]^2 &= c_1^2\pi(1-\pi)\sum_{n=1}^{\infty} A^n
\end{aligned}
$$

Provided that

$$
\begin{aligned}
A &= 2a_0^2 + (a_0+b_0)^2 + (2a_1^2 + (a_1+b_1)^2)\pi(1-\pi) \\
&= (2\alpha_0^2 + (\alpha_0+\beta_0)^2)(1-\pi) + (2\alpha_1^2 + (\alpha_1+\beta_1)^2)\pi < 1
\end{aligned}
$$

The above equations can be simplified as:

$$
\begin{aligned}
E[c_0^2(\sum_{n=1}^{\infty} S_n(t))^2] &= c_0^2(\sum_{n=1}^{\infty} A^n(1 + 2\sum_{j=1}^{\infty}(a_0+b_0)^j)) \\
&= c_0^2\frac{A(1+(a_0+b_0))}{(1-A)(1-(a_0+b_0))} \\
E[c_1\sum_{n=1}^{\infty} B_{t-1-n}S_n(t)]^2 &= c_1^2\pi(1-\pi)\sum_{n=1}^{\infty} A^n \\
&= c_1^2\pi(1-\pi)\frac{A}{(1-A)}
\end{aligned}
$$

Substitute back to the expression of $E(\sigma_t^4)$, we have:

$$
\begin{aligned}
E(\sigma_t^4) &= c_0^2 + c_1^2\pi(1-\pi) + 2c_0^2\sum_{n=1}^{\infty}(a_0+b_0)^n + c_1^2\pi(1-\pi)\frac{A}{(1-A)} \\
&\quad +2c_0c_1\pi(1-\pi)(a_1+b_1)\sum_{n=1}^{\infty}(a_0+b_0)^{n-1} + c_0^2\frac{A(1+a_0+b_0)}{(1-A)(1-(a_0+b_0))} \\
&= c_0^2 + c_1^2\pi(1-\pi) + \frac{2c_0^2(a_0+b_0)}{1-(a_0+b_0)} + 2c_0c_1\pi(1-\pi)\frac{a_1+b_1}{1-(a_0+b_0)} \\
&\quad +c_0^2\frac{A(1+a_0+b_0)}{(1-A)(1-(a_0+b_0))} + c_1^2\pi(1-\pi)\frac{A}{(1-A)} \\
&= c_0^2\frac{1+a_0+b_0}{(1-A)(1-(a_0+b_0))} \\
&\quad +c_1\pi(1-\pi)\frac{c_1(1-(a_0+b_0)) + 2c_0(a_1+b_1)(1-A)}{(1-A)(1-(a_0+b_0)}
\end{aligned}
$$

Appendix 4: Proof of Proposition 4

We have $\sigma^2 = E(r_t^2) = \dfrac{c_0}{1-(a_0+b_0)}$, subtract mean from equation (2.2), we get:

$$\begin{aligned}
\sigma_t^2 - \sigma^2 &= c_1 B_{t-1} + [\varepsilon_{t-1}^2(a_0 + a_1 B_{t-1}) + b_0 + b_1 B_{t-1}]\sigma_{t-1}^2 - \sigma^2(a_0 + b_0) \\
&= c_1 B_{t-1} + (a_0 + a_1 B_{t-1}) r_{t-1}^2 + (b_0 + b_1 B_{t-1})\sigma_{t-1}^2 - \sigma^2(a_0 + b_0) \\
r_t^2 - \sigma^2 &= c_1 B_{t-1} + (a_0 + a_1 B_{t-1}) r_{t-1}^2 + (b_0 + b_1 B_{t-1})\sigma_{t-1}^2 - \sigma^2(a_0 + b_0) - \sigma_t^2 + r_t^2 \\
&= [a_0 + b_0 + (a_1 + b_1)B_{t-1}](r_{t-1}^2 - \sigma^2) + (\sigma^2(a_1 + b_1) + c_1)B_{t-1} \\
&\quad -(b_0 + b_1 B_{t-1})(r_{t-1}^2 - \sigma_{t-1}^2) + r_t^2 - \sigma_t^2 \\
&= [a_0 + b_0 + (a_1 + b_1)B_{t-1}](r_{t-1}^2 - \sigma^2) + (\sigma^2(a_1 + b_1) + c_1)B_{t-1} \\
&\quad +\sigma_t^2(\varepsilon_t^2 - 1) - (b_0 + b_1 B_{t-1})\sigma_{t-1}^2(\varepsilon_{t-1}^2 - 1)
\end{aligned}$$

The above expression of $r_t^2 - \sigma^2$ is very similar to the equation (3.6) in Ding and Granger (1996). Interestingly, the above expression represents $r_t^2 - \sigma^2$ as an random coefficient ARMA(1,1) process if we write the compound error as an MA(1) process; letting $\sigma_t^2(\varepsilon_t^2 - 1) = v_t$:

$$\sigma_t^2(\varepsilon_t^2 - 1) - (b_0 + b_1 B_{t-1})\sigma_{t-1}^2(\varepsilon_{t-1}^2 - 1) = v_t - (b_0 + b_1 B_{t-1}) v_{t-1}$$

with $E(v_t) = 0$ and $E(v_t v_s) = 0$ for all $t \neq s$. Therefore, without any formal proof, the information already gives us some idea of the behaviors of the auto covariances.

Multiply both sides by $(r_{t-1}^2 - \sigma^2)$:

$$
\begin{aligned}
(r_t^2-\sigma^2)(r_{t-1}^2-\sigma^2) &= [a_0+b_0+(a_1+b_1)B_{t-1}](r_{t-1}^2-\sigma^2)^2 \\
&\quad +(\sigma^2(a_1+b_1)+c_1)B_{t-1}(r_{t-1}^2-\sigma^2)+(r_t^2-\sigma_t^2)(r_{t-1}^2-\sigma^2) \\
&\quad -(b_0+b_1B_{t-1})(r_{t-1}^2-\sigma_{t-1}^2)(r_{t-1}^2-\sigma^2)
\end{aligned}
$$

Taking expectation on both sides

$$
\begin{aligned}
E(r_t^2-\sigma^2)(r_{t-1}^2-\sigma^2) &= (a_0+b_0)E(r_{t-1}^2-\sigma^2)^2-b_0E(r_{t-1}^2-\sigma_{t-1}^2)(r_{t-1}^2-\sigma^2) \\
&\quad +E(r_t^2-\sigma_t^2)(r_{t-1}^2-\sigma^2) \\
\gamma(1) &= (a_0+b_0)\gamma(0)-2b_0E(\sigma_{t-1}^4)
\end{aligned}
$$

where $\gamma(1)=E(r_t^2-\sigma^2)(r_{t-1}^2-\sigma^2)$ is the covariance between r_t^2 and r_{t-1}^2, and $\gamma(0)=E(r_{t-1}^2-\sigma^2)^2$ is the variance of r_{t-1}^2. If we assume that the second moment of conditional variance or the fourth moment of residual exists, that is when the conditions in proposition 3 satisfied, then it can be shown that:

$$
\begin{aligned}
\rho(1) &= \frac{c_0^2[2a_0-2a_0b_0A_0+A_0(2a_1^2+(a_1+b_1)^2)\pi(1-\pi)]}{c_0^2(2+A-3A_0^2)+3c_1\pi(1-\pi)[c_1(1-A_0)+2c_0(a_1+b_1)(1-A)]} \\
&\quad +\frac{(3a_0+b_0(1+2A_0))\pi(1-\pi)[c_1^2+2c_0c_1(a_1+b_1)(1-A)]}{c_0^2(2+A-3A_0^2)+3c_1\pi(1-\pi)[c_1(1-A_0)+2c_0(a_1+b_1)(1-A)]}
\end{aligned}
$$

where $A_0=a_0+b_0$

It is easy to show that:

$\gamma(k)=(a_0+b_0)\gamma(k-1)$ for $k\geq 2$, and $\rho(k)=(a_0+b_0)^{k-1}\rho(1)$

Appendix 5:

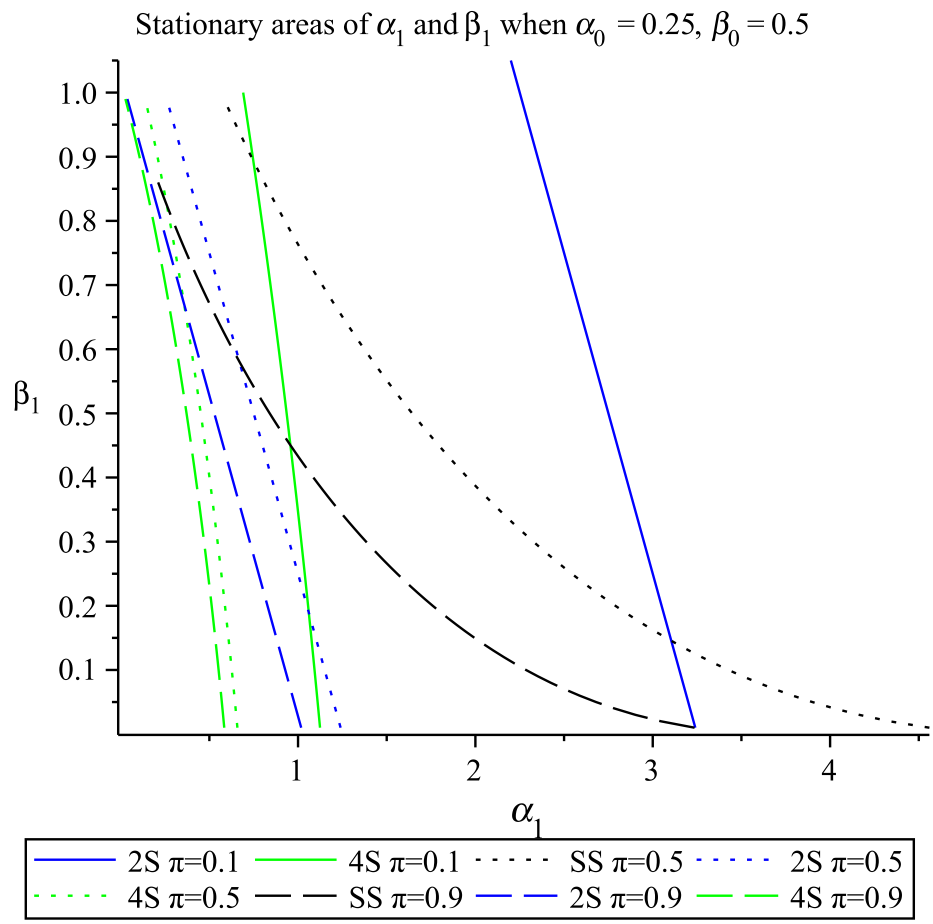

Figure 2.1: Stationary Areas of α_1 and β_1 given (α_0, β_0)=(0.25, 0.5)

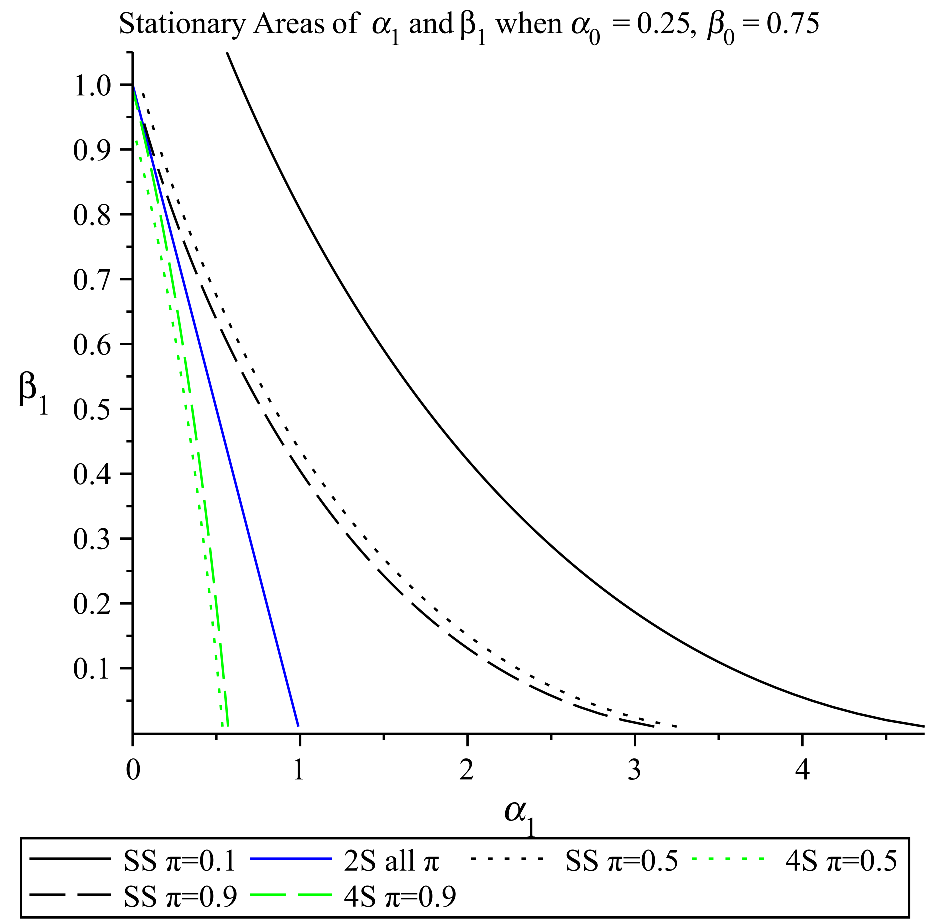

Figure 2.2: Stationary Areas of α_1 and β_1 given (α_0, β_0)=(0.25, 0.75)

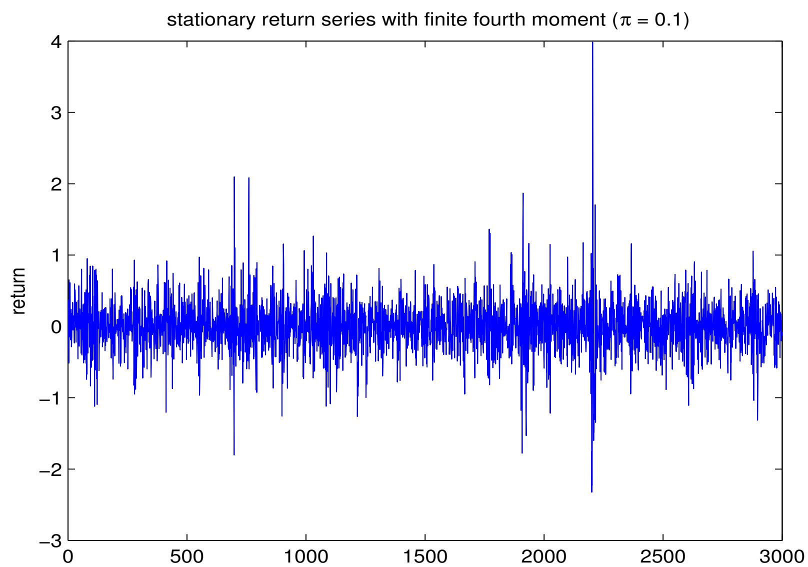

Figure 2.3: Simulated Paths of the Return Series in Case 1.1

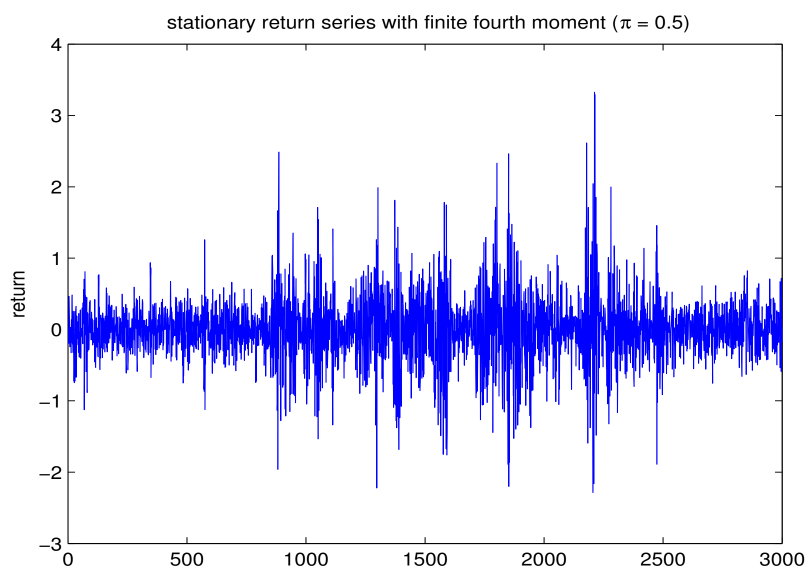

Figure 2.4: Simulated Paths of the Return Series in Case 2.1

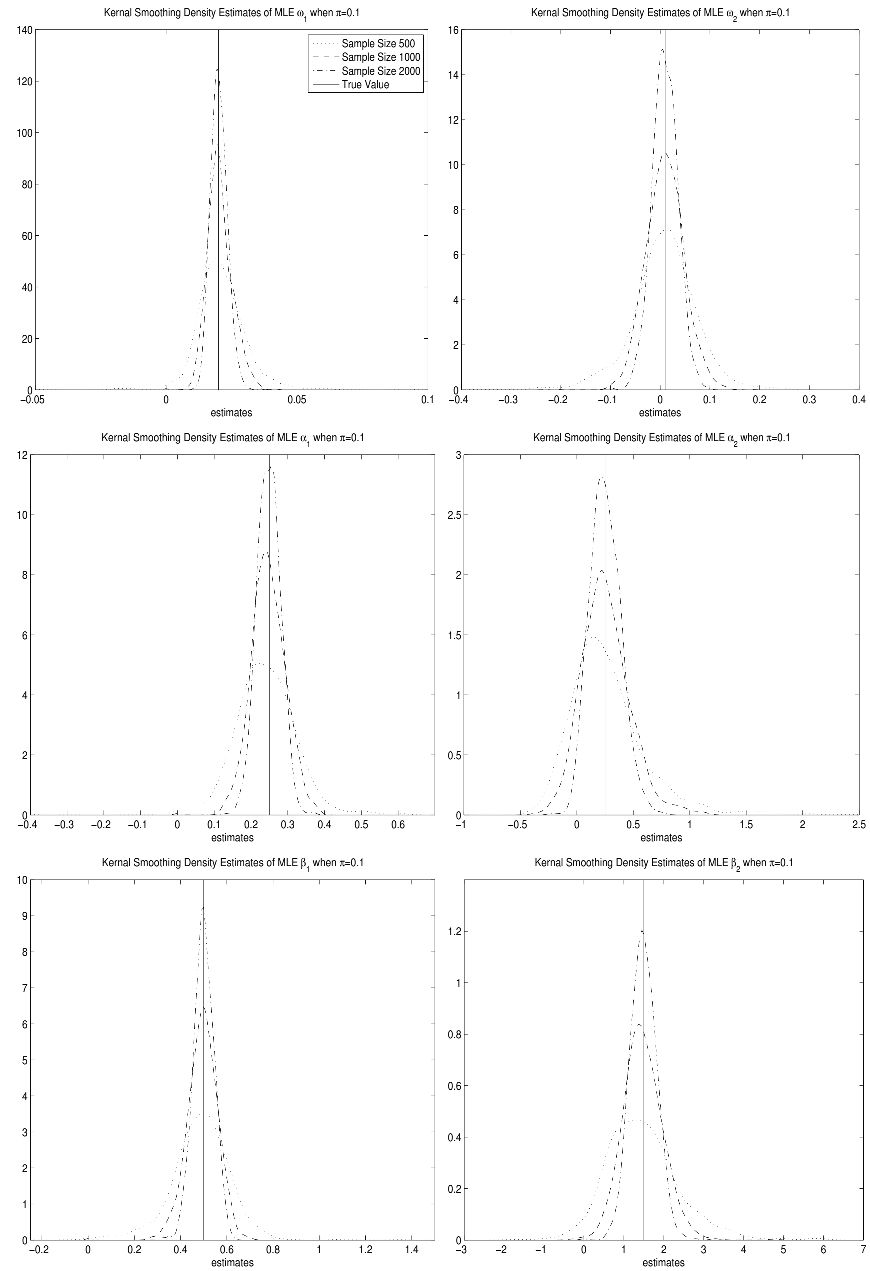

Figure 2.5: Kernel Smoothing Density Estimates of MLE (Case 1.1)

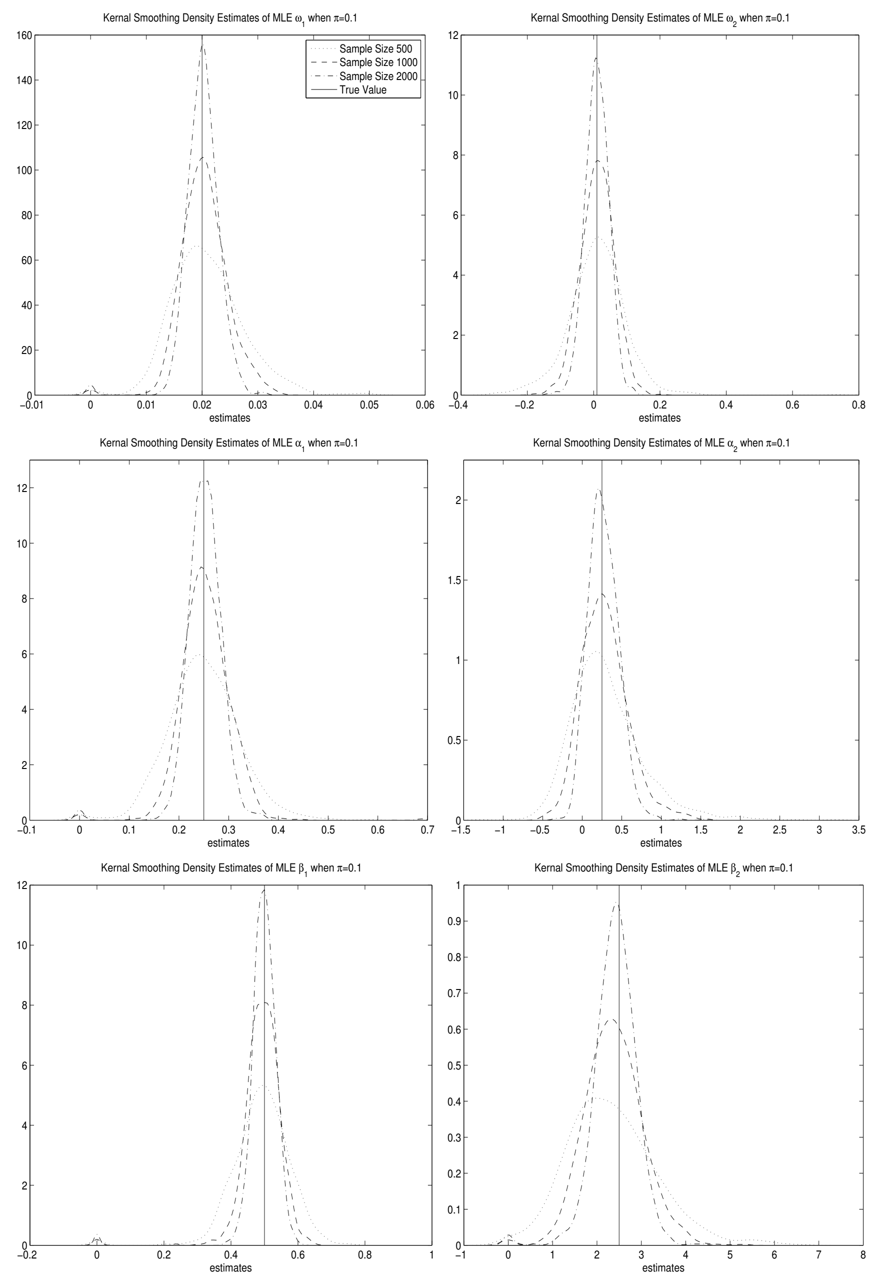

Figure 2.6: Kernel Smoothing Density Estimates of MLE (Case 1.2)

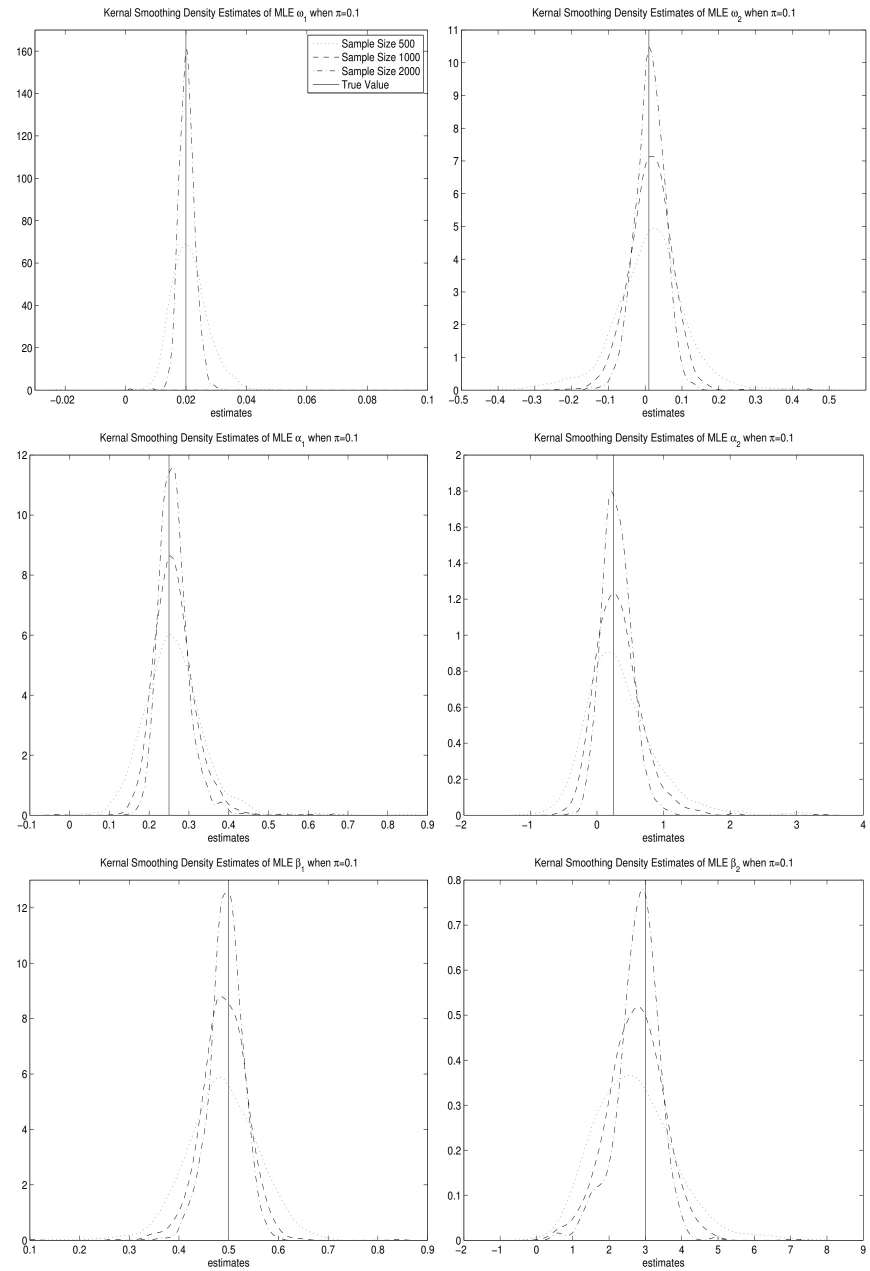

Figure 2.7: Kernel Smoothing Density Estimates of MLE (Case 1.3)

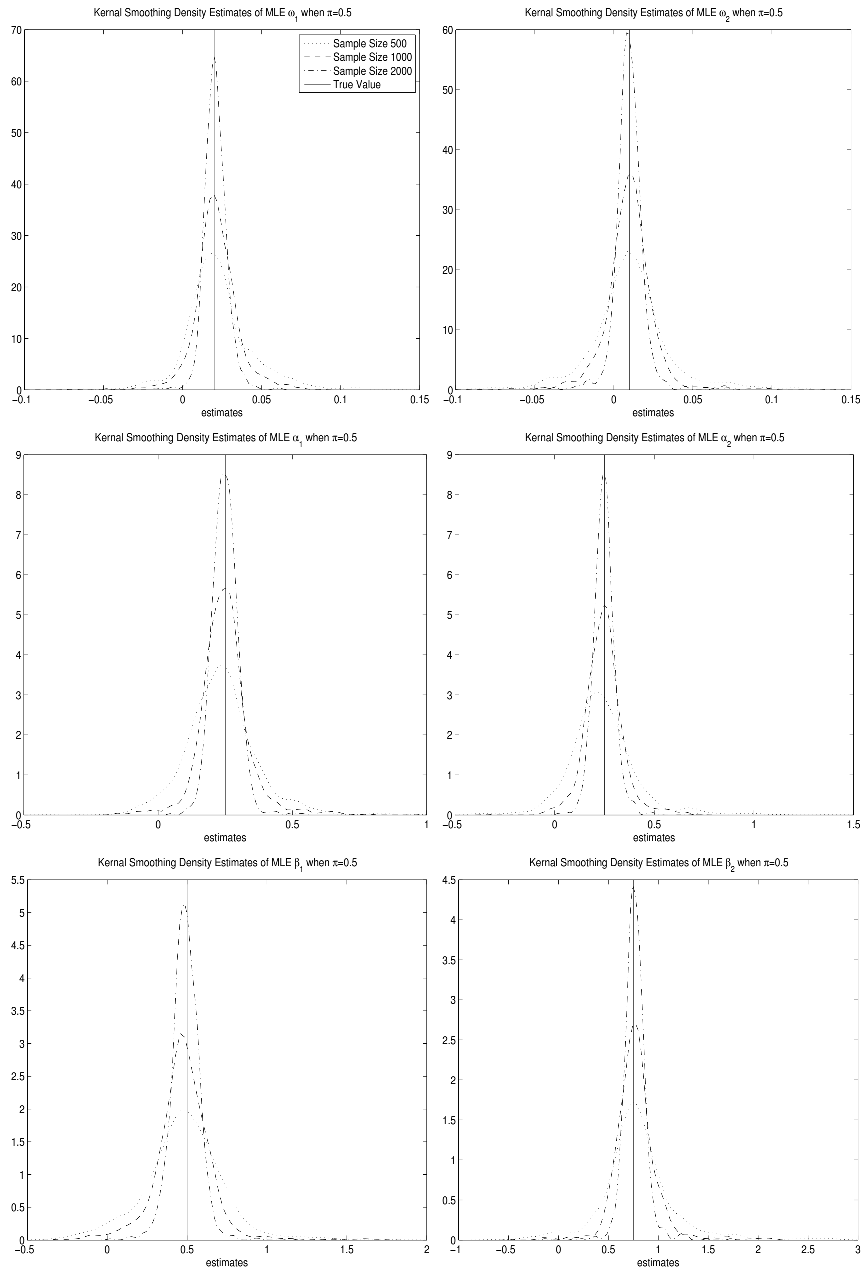

Figure 2.8: Kernel Smoothing Density Estimates of MLE (Case 2.1)

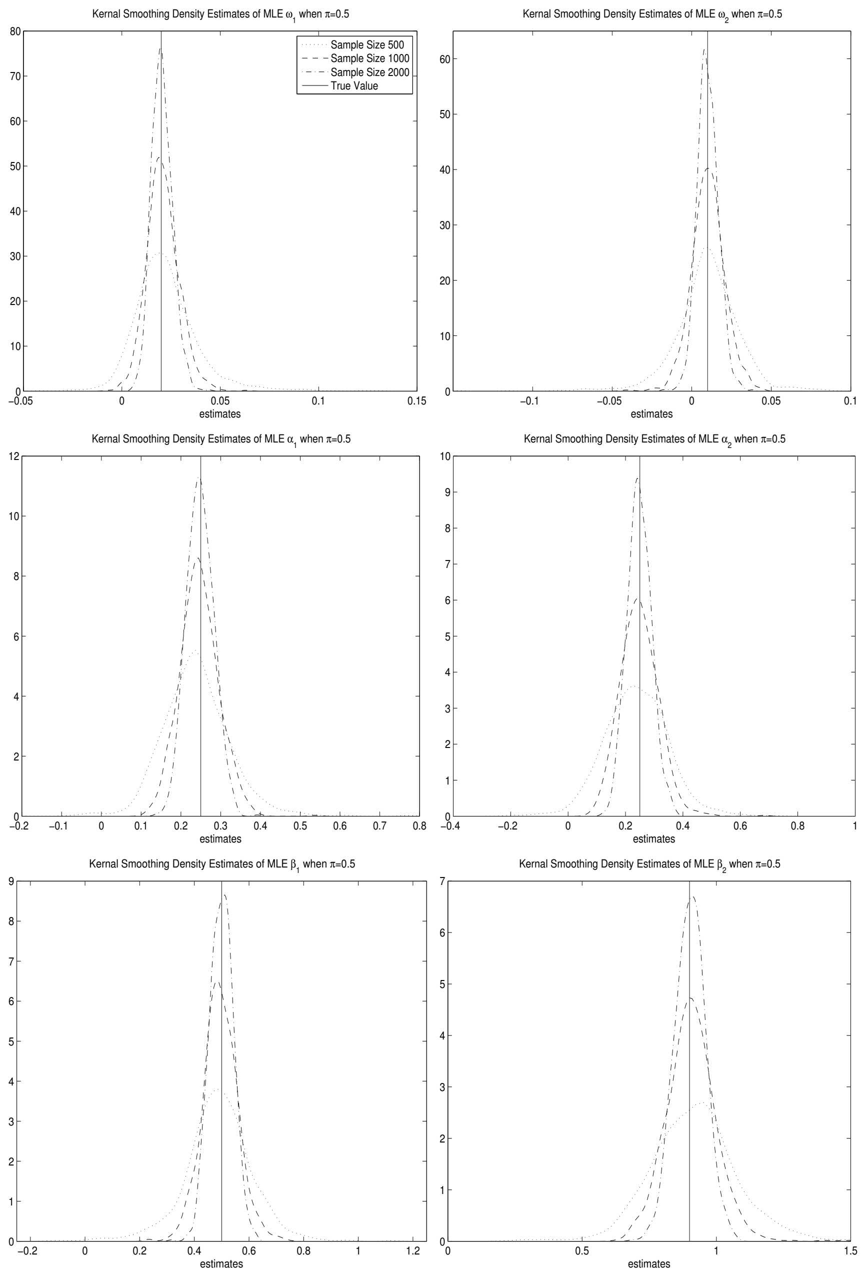

Figure 2.9: Kernel Smoothing Density Estimates of MLE (Case 2.2)

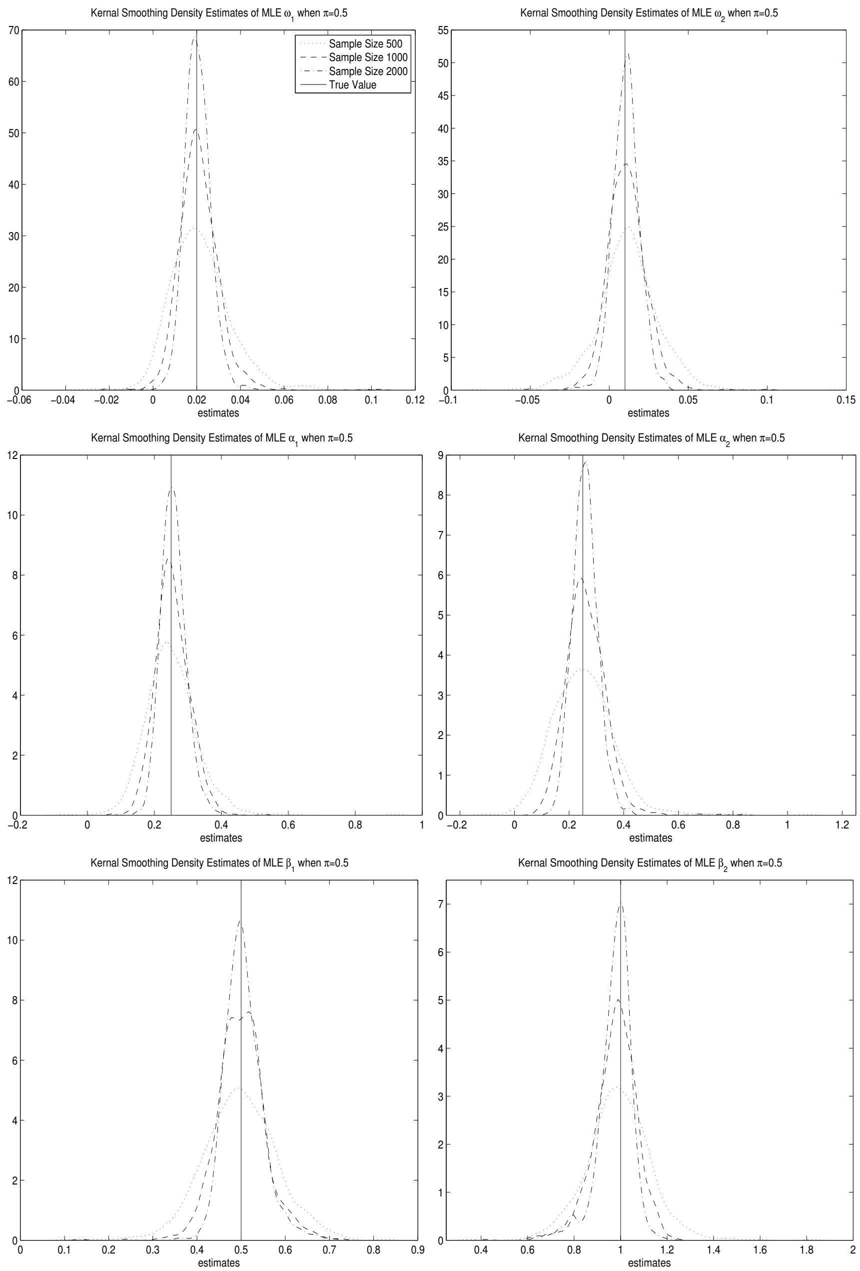

Figure 2.10: Kernel Smoothing Density Estimates of MLE (Case 2.3)

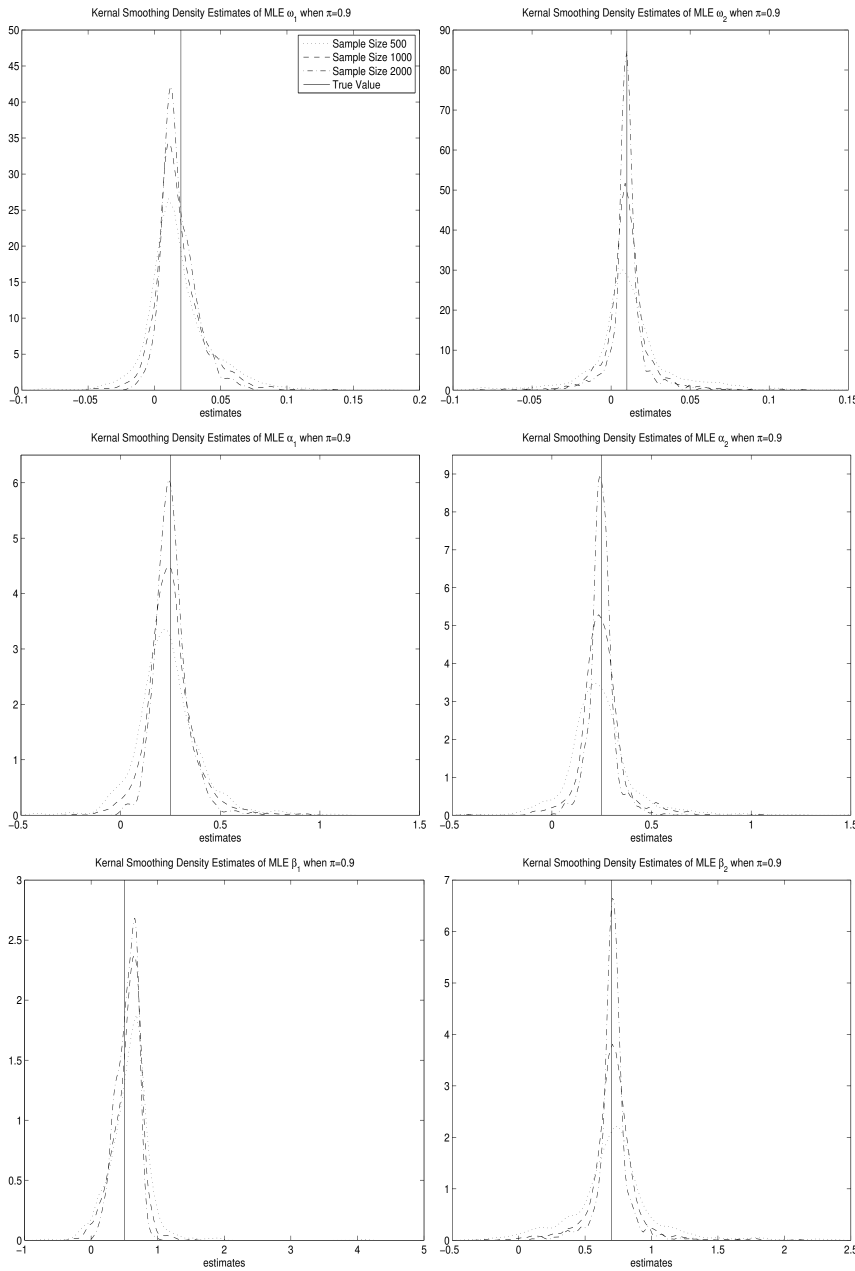

Figure 2.11: Kernel Smoothing Density Estimates of MLE (Case 3.1)

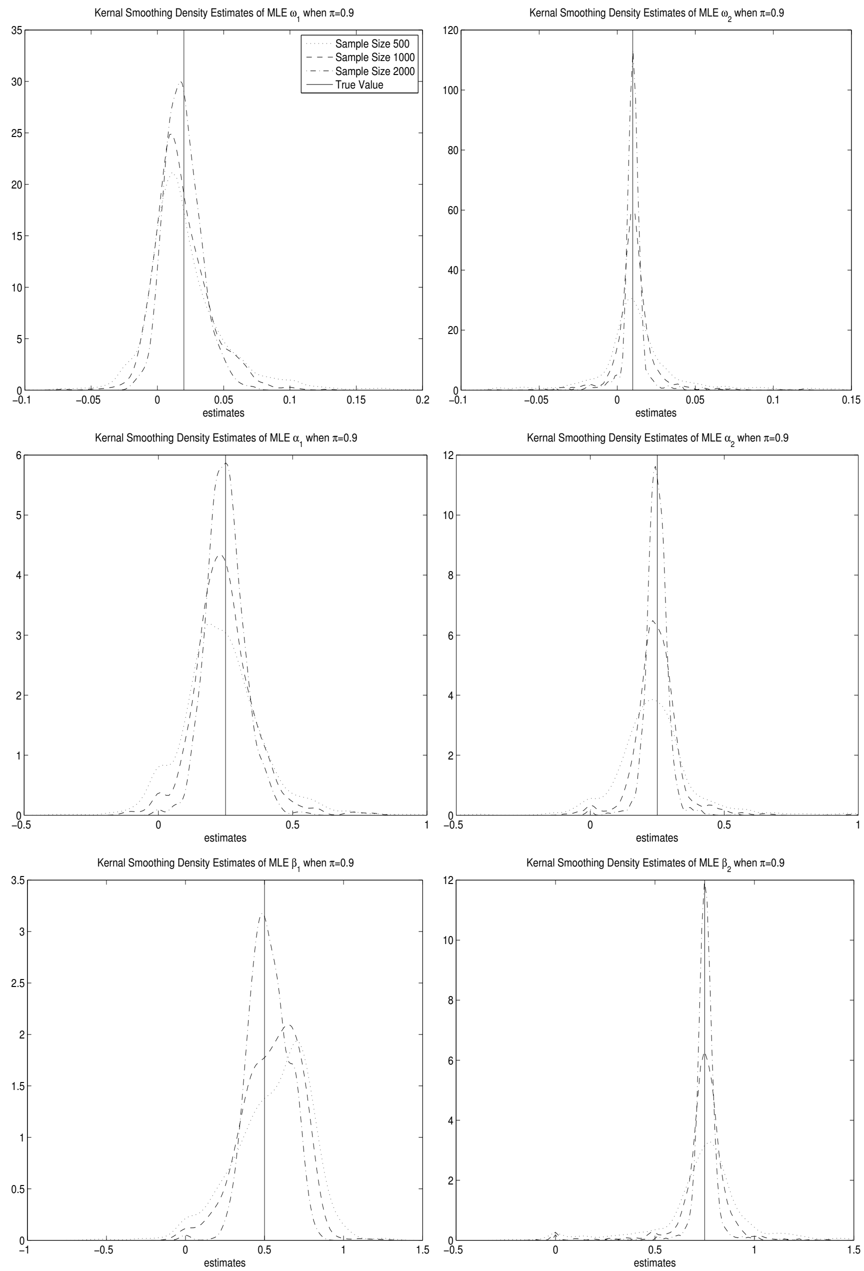

Figure 2.12: Kernel Smoothing Density Estimates of MLE (Case 3.2)

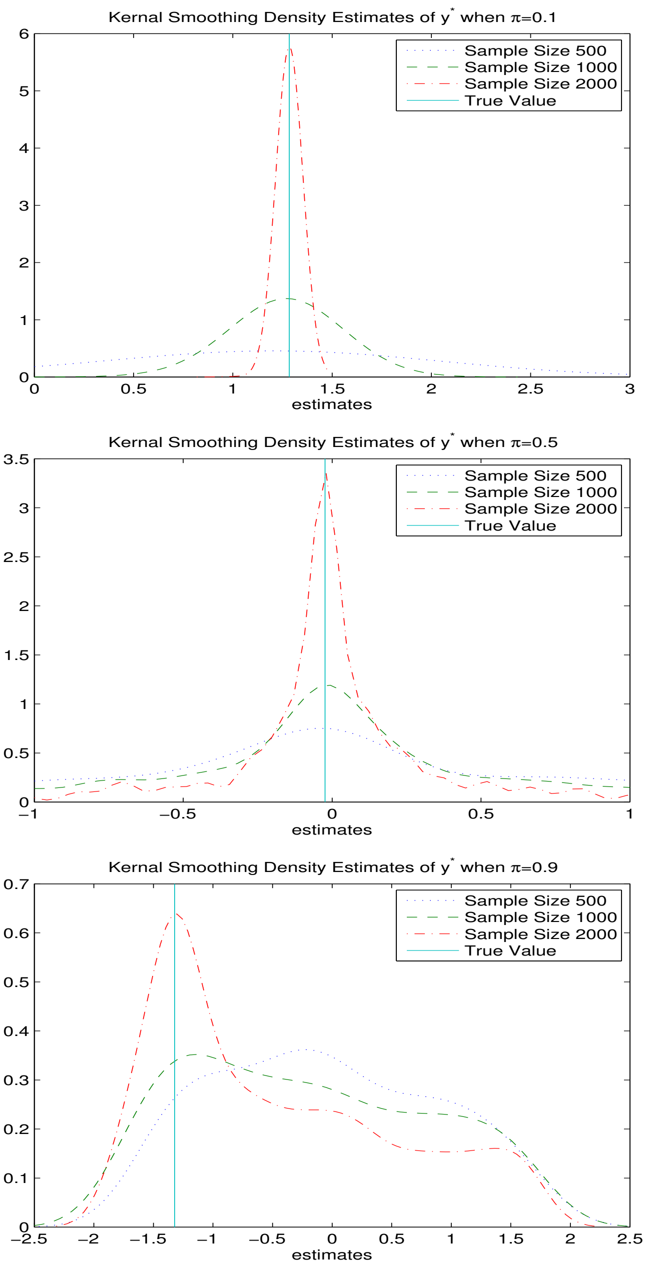

Figure 2.13: Kernel Smoothing Density Estimates of y^* When Finite Fourth Moment Exists

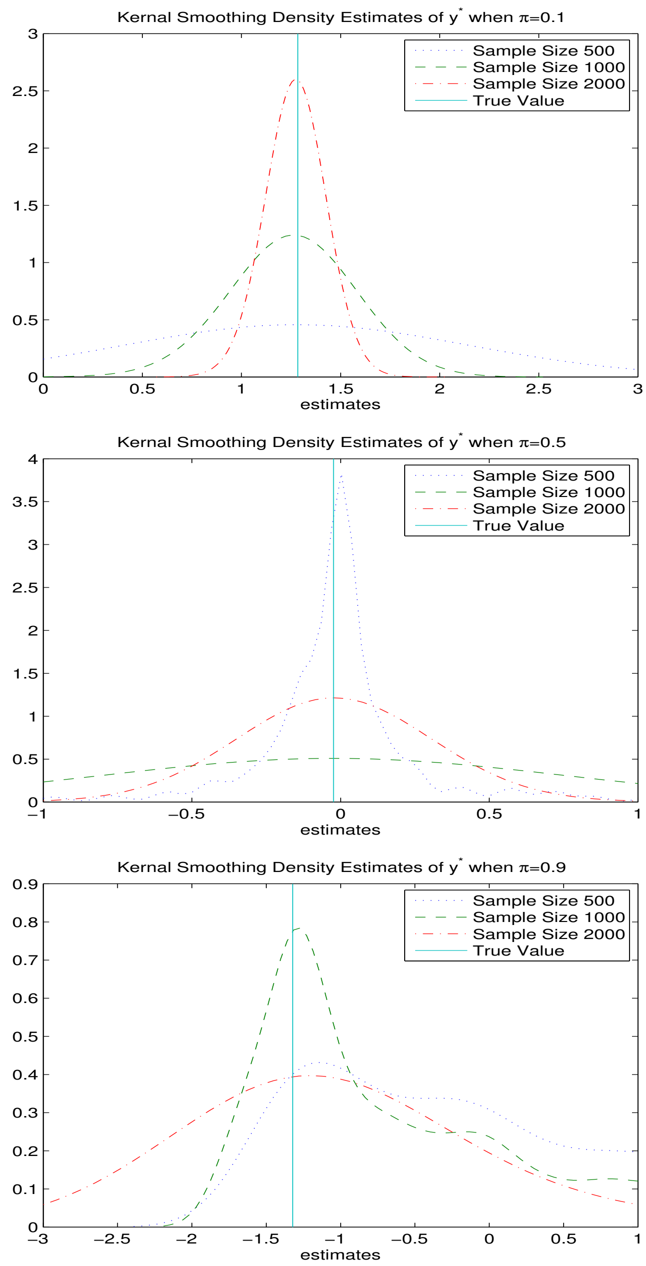

Figure 2.14: Kernel Smoothing Density Estimates of y^* When Finite Variance Exists

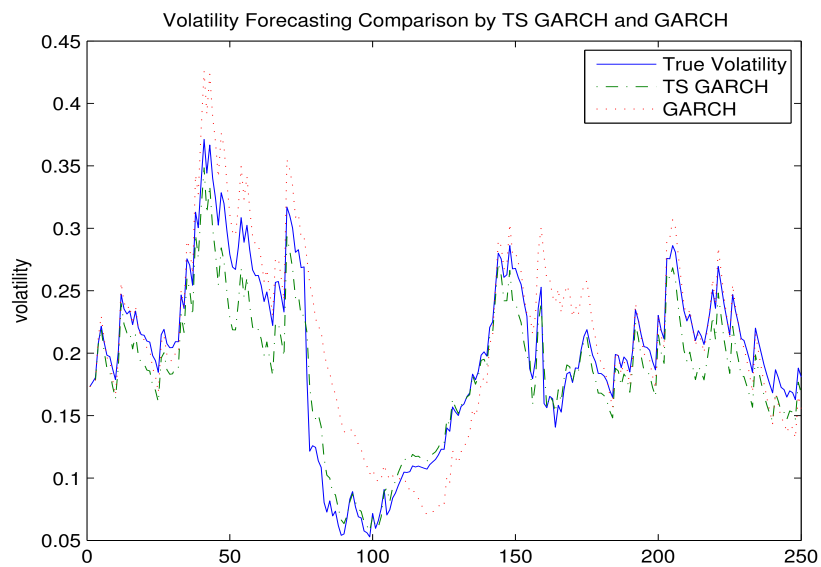

Figure 2.15: Forecasting Performance based on TS GARCH DGP

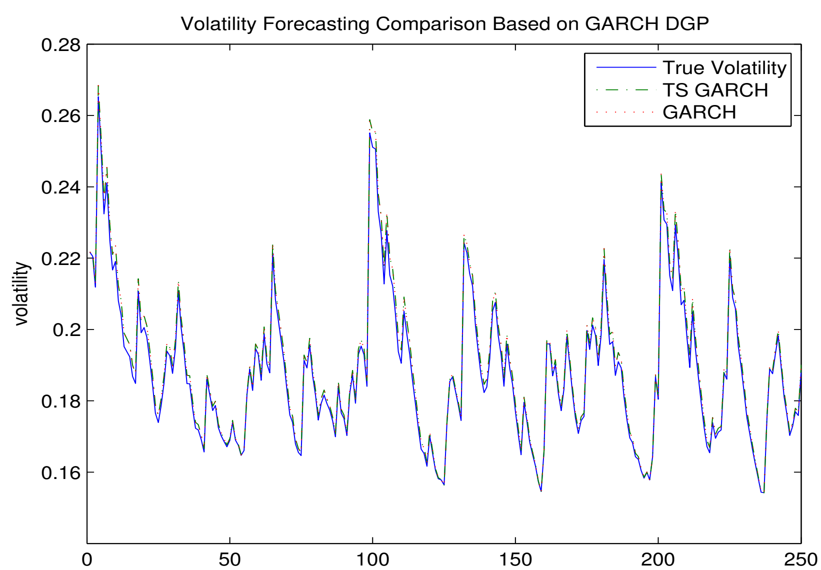

Figure 2.16: Forecasting Performance based on GARCH DGP

Table 2.1: The MLE for parameters in Case 1

T	Par	True Value (1.1)	$\bar{\hat{\theta}}$	MSE	True Value (1.2)	$\bar{\hat{\theta}}$	MSE	True Value (1.3)	$\bar{\hat{\theta}}$	MSE
	ω_0	0.02	.0210	.0001	0.02	.0209	.0000	0.02	.0205	.0001
	ω_1	0.01	.0072	.0045	0.01	.0096	.0077	0.01	.0114	.0094
500	α_0	0.25	.2389	.0076	0.25	.2487	.0051	0.25	.2508	.0077
	α_1	0.25	.2711	.1194	0.25	.3077	.2158	0.25	.3356	.2801
	β_0	0.5	.4905	.0232	0.5	.4908	.0077	0.5	.4725	.0136
	β_1	1.5	1.4712	.8277	2.5	2.3337	1.0359	3	2.6001	1.5691
	ω_0	0.02	.0206	.0000	0.02	.0204	.0000	0.02	.0238	.0131
	ω_1	0.01	.0097	.0015	0.01	.0107	.0025	0.01	.0136	.0032
1000	α_0	0.25	.2474	.0021	0.25	.2527	.0023	0.25	.2555	.0045
	α_1	0.25	.2569	.0464	0.25	.2843	.0880	0.25	.3160	.1246
	β_0	0.5	.4949	.0042	0.5	.4929	.0037	0.5	.4776	.0093
	β_1	1.5	1.4762	.2526	2.5	2.3752	.4346	3	2.6380	.9153
	ω_0	0.02	.0202	.0000	0.02	.0200	.0000	0.02	.0196	.0000
	ω_1	0.01	.0094	.0007	0.01	.0100	.0012	0.01	.0127	.0018
2000	α_0	0.25	.2479	.0011	0.25	.2513	.0016	0.25	.2525	.0038
	α_1	0.25	.2519	.0185	0.25	.2700	.0376	0.25	.2910	.0568
	β_0	0.5	.4993	.0022	0.5	.4952	.0031	0.5	.4780	.0096
	β_1	1.5	1.4917	.1096	2.5	2.4091	.2396	3	2.6786	.6989

Table 2.2: The MLE for parameters in Case 2

T	Par	True Value (2.1)	$\bar{\hat{\theta}}$	MSE	True Value (2.2)	$\bar{\hat{\theta}}$	MSE	True Value (2.3)	$\bar{\hat{\theta}}$	MSE
	ω_0	0.02	.0229	.0005	0.02	.0226	.0003	0.02	.0201	.0002
	ω_1	0.01	.0090	.0010	0.01	.0096	.0005	0.01	.0107	.0004
500	α_0	0.25	.2342	.0192	0.25	.2383	.0081	0.25	.2243	.0111
	α_1	0.25	.2415	.0313	0.25	.2370	.0123	0.25	.2265	.0183
	β_0	0.5	.4768	.0649	0.5	.4880	.0159	0.5	.4546	.0291
	β_1	0.75	.7973	.1577	0.9	.9314	.0439	1	.9020	.1091
	ω_0	0.02	.0212	.0002	0.02	.0209	.0001	0.02	.0194	.0001
	ω_1	0.01	.0093	.0003	0.01	.0103	.0001	0.01	.0103	.0001
1000	α_0	0.25	.2482	.0092	0.25	.2463	.0024	0.25	.2307	.0074
	α_1	0.25	.2477	.0096	0.25	.2483	.0044	0.25	.2364	.0096
	β_0	0.5	.4811	.0375	0.5	.4949	.0042	0.5	.4611	.0235
	β_1	0.75	.7740	.0509	0.9	.9063	.0085	1	.8962	.0914
	ω_0	0.02	.0207	.0001	0.02	.0204	.0000	0.02	.0203	.0000
	ω_1	0.01	.0096	.0001	0.01	.0101	.0000	0.01	.0107	.0001
2000	α_0	0.25	.2487	.0023	0.25	.2482	.0011	0.25	.2567	.0015
	α_1	0.25	.2473	.0030	0.25	.2477	.0018	0.25	.2599	.0027
	β_0	0.5	.4897	.0090	0.5	.4987	.0018	0.5	.5050	.0021
	β_1	0.75	.7651	.0152	0.9	.9012	.0033	1	.9754	.0061

Table 2.3: The MLE for parameters in Case 3

T	Par	True Value (3.1)	$\bar{\hat{\theta}}$	MSE	True Value (3.2)	$\bar{\hat{\theta}}$	MSE
	ω_0	0.02	.0175	.0007	0.02	.0210	.0009
	ω_1	0.01	.0118	.0007	0.01	.0123	.0010
500	α_0	0.25	.2361	.0257	0.25	.2353	.0215
	α_1	0.25	.2437	.0304	0.25	.2412	.0236
	β_0	0.5	.5780	.0973	0.5	.5562	.0528
	β_1	0.7	.7088	.0988	0.75	.7505	.0658
	ω_0	0.02	.0179	.0003	0.02	.0185	.0004
	ω_1	0.01	.0104	.0002	0.01	.0112	.0002
1000	α_0	0.25	.2482	.0143	0.25	.2444	.0124
	α_1	0.25	.2501	.0119	0.25	.2491	.0080
	β_0	0.5	.5543	.0472	0.5	.5452	.0349
	β_1	0.7	.7075	.0352	0.75	.7425	.0200
	ω_0	0.02	.0180	.0002	0.02	.019	.0002
	ω_1	0.01	.0099	.0001	0.01	.0103	.0001
2000	α_0	0.25	.2510	.0069	0.25	.2488	.0051
	α_1	0.25	.2489	.0040	0.25	.2475	.0019
	β_0	0.5	.5434	.0264	0.5	.5251	.0156
	β_1	0.7	.7086	.0151	0.75	.7503	.0060

Table 2.4: The Performance of MLE with Endogenous Threshold

			$a = 0.1$		$a = 0.2$		$a = 0.3$		$a = 0.4$	
T	Par	True Value	$\hat{\theta}$	MSE	$\hat{\theta}$	MSE	$\hat{\theta}$	MSE	$\hat{\theta}$	MSE
	ω_0	.01	.0107	.0001	.0101	.0003	.0102	.0001	.0101	.0001
	ω_1	.02	.0202	.0007	.0207	.0004	.0224	.0003	.0226	.0003
500	α_0	.1	.0894	.0072	.0932	.0087	.0942	.0092	.0951	.0146
	α_1	.2	.1804	.0330	.1754	.0168	.1788	.0134	.1800	.0098
	β_0	.55	.5493	.0468	.5566	.0403	.5510	.0316	.5545	.0276
	β_1	.75	.7695	.3637	.7634	.2016	.7438	.1315	.7311	.0898
	ω_0	.01	.0101	.0000	.0101	.0000	.0100	.0000	.0100	.0000
	ω_1	.02	.0208	.0001	.0211	.0001	.0213	.0001	.0214	.0001
1000	α_0	.1	.0983	.0016	.0984	.0017	.0993	.0021	.0976	.0028
	α_1	.2	.1888	.0070	.1903	.0054	.1923	.0046	.1941	.0037
	β_0	.55	.5489	.0114	.5487	.0096	.5514	.0077	.5530	.0070
	β_1	.75	.7406	.0551	.7357	.0356	.7336	.0284	.7309	.0188
	ω_0	.01	.0100	.0000	.0101	.0000	.0101	.0000	.0101	.0000
	ω_1	.02	.0205	.0000	.0207	.0000	.0205	.0000	.0206	.0000
2000	α_0	.1	.1010	.0007	.1008	.0008	.1020	.0010	.1016	.0012
	α_1	.2	.1961	.0029	.1968	.0023	.1985	.0019	.1987	.0017
	β_0	.55	.5506	.0044	.5498	.0039	.5490	.0034	.5502	.0030
	β_1	.75	.7358	.0184	.7370	.0134	.7380	.0094	.7392	.0072

Table 2.5: Forecasting Performance Based on Threshold GARCH DGP

	ME	MPE	$RMSE$	$HMSE$	R^2
TS GARCH	-.0004	.0049	.0881	.3949	.9256
GARCH	.0075	.0991	.1712	3.9838	.6905

Table 2.6: Estimation Performance with Different Jump Size of Grid Points in Threshold Variable

	True	5%			2.5%			1%		
	Value	$\hat{\theta}$	Std	MSE	$\hat{\theta}$	Std	MSE	$\hat{\theta}$	Std	MSE
ω_0	0.02	.0184	.0061	.0000	.0193	.0061	.0000	.0196	.0059	.0000
ω_1	0.01	.0127	.0052	.0000	.0125	.0053	.0000	.0126	.0056	.0000
α_0	0.15	.1343	.0348	.0015	.1397	.0346	.0013	.1394	.0346	.0013
α_1	0.05	.0530	.0171	.0003	.0524	.0171	.0003	.0517	.0173	.0003
β_0	0.55	.5895	.1083	.0133	.5711	.1073	.0119	.5654	.1035	.0110
β_1	0.9	.8829	.0370	.0017	.8840	.0382	.0017	.8845	.0394	.0018
y_0	19.67	19.87	.4235	.2211	19.81	.4110	.1888	19.81	.3778	.1638

Table 2.7: Forecasting Performance Based on Threshold GARCH DGP with Non-Stationary Parameters in One Regime

	ME	MPE	$RMSE$	$HMSE$	R^2
TS GARCH	.0259	.0363	2.0194	4.0616	.9708
GARCH	.2990	.2504	4.7271	38.8685	.8287

Table 2.8: Forecasting Performance Based on GARCH DGP

	ME	MPE	$RMSE$	$HMSE$	R^2
TS GARCH	-.0001	.0010	.0724	.2079	.9554
GARCH	-.0001	.0009	.0719	.2043	.9607

2.7 Bibliography

AKGIRAY, V. (1989): "Conditional heteroscedasticity in time series of stock returns: evidence and forecast," *Journal of Business*, 62, 55–80.

AMEMIYA, T. (1977): "A note on a heteroscedasticity model," *Journal of Econometrics*, 6(3), 365–70.

ANDERSEN, T., AND T. BOLLERSLEV (1997): "Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns," *Journal of Finance*, 52, 975–1005.

——— (1998): "DM-Dollar volatility: intraday activity patterns, macroeconomic announcements, and long run dependencies," *Journal of Finance*, 53, 219–265.

ANDERSEN, T., T. BOLLERSLEV, F. DIEBOLD, AND P. LABYS (2003): "Modeling and forecasting realized volatility," *Econometrica*, 71(2), 579–625.

ANDERSEN, T. G. (1994): "Stochastic autoregressive volatility: A framework for volatility modeling," *Mathematical Finance*, 4, 75–102.

——— (1996): "Return volatility and trading volume: An information flow interpretation of stochastic volatility," *Journal of Finance*, 51, 169–204.

BAILLIE, R. T., AND T. BOLLERSLEV (1992): "Prediction in dynamic models with time dependent conditional variances," *Journal of Econometrics*, 52, 91–113.

BAILLIE, R. T., T. BOLLERSLEV, AND H. MIKKELSEN (1996): "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 74, 3–30.

BAKSHI, G., C. CAO, AND Z. CHEN (1997): "Empirical performance of alternative option pricing models," *Journal of Finance*, 52, 2003–49.

BAUWENS, L., AND S. LAURENT (2005): "A new class of multivariate skew densities with application to GARCH models," *Journal of Business and Economic Statistics*, 23(3), 346–54.

BEKAERT, G., AND G. WU (2000): "Asymmetric volatility and risk in equity markets," *Review of Financial Studies*, 13, 1–42.

BLACK, F. (1976): "Studies of stock price volatility changes," *Proceedings of the American Statistical Association,Business and Economics Statistics Section*, pp. 177–81.

BOLLEN, B., AND B. INDER (2002): "Estimating daily volatility in financial markets utilizing intraday data," *Journal of Empirical Finance*, 9, 551–62.

BOLLERSLEV, T. (1986): "Generalized autoregressive conditional heteroscedasticity," *Journal of Econometrics*, 51, 307–327.

BOLLERSLEV, T., R. Y. CHOU, AND K. F. KRONER (1992): "ARCH modeling in finance: A review of the theory and empirical evidence," *Journal of Econometrics*, 52, 5–59.

BOLLERSLEV, T., AND R. ENGLE (1993): "Common persistence in conditional variances," *Econometrica*, 61, 167–86.

BOLLERSLEV, T., AND P. D. JUBINSKI (1999): "Equity trading volume and volatility: latent information arrivals and common long-run dependencies," *Journal of Business and Economic Statistics*, 17, 9–21.

BOLLERSLEV, T., AND H. ZHOU (2006): "Volatility puzzles: A simple framework for gauging return-volatility regressions," *Journal of Econometrics*, 131(1-2), 123–50.

BROOKS, C. (1998): "Predicting stock index volatility: can market volume help," *Journal of Forecasting*, 17, 59–80.

CAMPBELL, J., A. LO, AND A. C. MACKINLAY (1997): *The Econometrics of Financial Markets*. Princeton University Press.

CAMPBELL, J. Y., AND L. HENTSCHEL (1992): "No news is good news: An asymmetric model of changing volatility in stock returns," *Journal of Financial Economics*, 31, 281–318.

CAMPBELL, J. Y., AND A. S. KYLE (1993): "Smart money, noise trading, and stock price behavior," *The Review of Economic Study*, 60(1), 1–34.

CHEN, N.-F., R. ROLL, AND S. A. ROSS (1986): "Economic forces and the stock market: testing the APT and alternative asset pricing theories," *Journal of Business*, 59(3), 383–403.

CHERNOV, M., A. GALLANT, E. GHYSELS, AND G. TAUCHEN (2003): "Alternative models for stock price dynamics," *Journal of Econometrics*, 116, 225–57.

CHRISTIE, A. (1982): "The stochastic behavior of common stock variances: value, leverage and interest rate effects," *Journal of Financial Economics*, 10, 407–32.

CLARK, P. K. (1973): "A Subordinated stochastic process model with fixed variance for speculative prices," *Econometrica*, 41, 135–156.

DIEBOLD, F. X., AND R. S. MARIANO (1995): "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13(3), 253–65.

DING, Z., AND C. W. J. GRANGER (1996): "Modeling volatility persistence of speculative returns: A new approach," *Journal of Econometrics*, 73, 185–215.

DING, Z., C. W. J. GRANGER, AND R. ENGLE (1993): "A long memory property of stock market returns and a new model," *Journal of Empirical Finance*, 1, 83–106.

DROST, F. C., AND T. E. NIJMAN (1993): "Temporal aggregation of GARCH processes," *Econometrica*, 61, 909–27.

ENGLE, R., AND A. J. PATTON (2001): "What good is a volatility model," *Quantitative Finance*, 1(2), 237–245.

ENGLE, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation," *Econometrica*, 50, 987–1008.

——— (2001): "GARCH 101: the use of ARCH/GARCH models in applied econometrics," *Journal of Economic Perspectives*, 15, 157–68.

ENGLE, R. F., AND T. BOLLERSLEV (1986): "Modeling the persistence of conditional variances," *Econometric Reviews*, 5, 1–50.

ENGLE, R. F., AND V. K. NG (1993): "Measuring and testing the impact of news on volatility," *Journal of Finance*, 48, 1749–78.

EPPS, T., AND M. EPPS (1976): "The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distribution hypothesis," Manuscript.

FAMA, E. F. (1963): "Mandelbrot and the stable paretian hypothesis," *The Journal of Business*, 36(4), 420–429.

FLEMING, J., AND C. KIRBY (2003): "A closer look at the relation between GARCH and stochastic autoregressive volatility," *Journal of Financial Econometrics*, 1, 365–419.

FLEMING, J., C. KIRBY, AND B. OSTDIEK (2001): "The economic value of volatility timing," *Journal of Finance*, 56, 329–52.

——— (2003): "The economic value of volatility timing using realized volatility," *Journal of Financial Economics*, 67, 473–509.

FRENCH, K. R., G. W. SCHWERT, AND R. F. STAMBAUGH (1987): "Expected stock returns and volatility," *Journal of Financial Economics*, 15, 265–302.

GALLANT, A. R., P. E. ROSSI, AND G. TAUCHEN (1992): "Stock price and volume," *The Review of Financial Studies*, 5(2), 199–242.

GIRAITIS, L., P. KOKOSZKA, AND R. LEIPUS (2000): "Stationary ARCH models: dependence structure and central limit theorem," *Econometric Theory*, 16, 3–22.

GLOSTEN, L. R., R. JAGANNATHAN, AND D. E. RUNKLE (1993): "On the relation between the expected value and the volatility of the nominal excess return on stocks," *Journal of Finance*, 48, 1779–1801.

GRAY, S. F. (1996): "Modeling the conditional distribution of interest rates as a regime-switching process," *Journal of Financial Economics*, 42, 27–62.

HAMILTON, J. D., AND R. SUSMEL (1994): "Autoregressive conditional heteroskedasticity and changes in regimes," *Journal of Econometrics*, 64, 307–333.

HARVEY, A. C., AND E. SHEPHARD (1996): "Estimation of an asymmetric model of asset prices," *Journal of Business and Economic Statistics*, 14, 429–34.

HENTSCHEL, L. (1995): "All in the family: nesting symmetric and asymmetric GARCH models," *Journal of Financial Economics*, 39, 71–104.

HSIEH, D. A. (1989): "Modeling heteroskedasticity in foreign exchange rates," *Journal of Business and Economic Statistics*, 7, 307–17.

KARPOFF, J. M. (1987): "The relationship between price changes and trading volume: A survey," *Journal of Financial and Quantitative Analysis*, 22, 109–26.

KLIVECKA, A. (2004): "Random coefficient GARCH(1,1) model with IID coefficients," *Lithuanian Mathematical Journal*, 44, 374–85.

KNIGHT, J., AND S. E. SATCHELL (2011): "Some new results for threshold AR(1) models," *Journal of Time Series Econometrics*, 3(2), Article1.

KOURTELLOS, A., T. STENGOS, AND C. M. TAN (2007): "THRET: Threshold regression with endogenous threshold variables," Working paper.

LAMOUREUX, C. G., AND W. D. LASTRAPES (1990): "Persistence in variance, structural change and the GARCH model," *Journal of Business and Economic Statistics*, 8(2), 225–34.

——— (1990a): "Heteroskedasticity in stock return data: Volume versus GARCH effects," *Journal of Finance*, 45, 221–29.

LEE, S. W., AND B. E. HANSEN (1994): "Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator," *Econometric Theory*, 10, 29–52.

LIESENFELD, R. (1998): "Dynamic bivariate mixture models: modeling the behavior of prices and trading volume," *Journal of Business and Economic Statistics*, 16, 101–9.

——— (2001): "A generalized bivariate mixture model for stock price volatility and trading volume," *Journal of Econometrics*, 104, 141–78.

LOPEZ, J. (2001): "Evaluating the predictive accuracy of volatility models," *Journal of Forecasting*, 20(2), 87–109.

LUMSDAINE, R. L. (1996): "Consistency and asymptotic normality of the quasi-maximum likelihood estimator in GARCH(1,1) and covariance stationary GARCH(1,1) models," *Econometrica*, 64, 575–96.

MANDELBROT, B. (1963): "The variation of certain speculative prices," *Journal of Business*, 36(4), 394.

MCNEIL, A. J., R. FREY, AND P. EMBRECHTS (2005): "Quantitative risk management: concepts, techniques and tools," Princeton Series in Finance.

NELSON, D. B. (1988): "The time series behavior of stock market volatility and retruns," Unpublished ph.d. dissertation, MIT, Cambridge, MA.

——— (1990): "Stationarity and persistence in the GARCH(1,1) model," Discussion paper.

——— (1991): "Conditional heteroskedasticity in asset pricing: A new approach," *Econometrica*, 59, 347–70.

PAGAN, A. (1988): "The econometrics of financial markets," Discussion paper.

QUINN, B. G. (1982): "A note on the existence of strictly stationary solutions to bilinear equations," *Journal of Time Series Analysi*, 3, 249–52.

RABEMANANJARA, R., AND J. M. ZAKOIAN (1993): "Threshold ARCH models and asymmetries in volatility," *Journal of Applied Econometrics*, 8, 31–49.

TAUCHEN, G. E., AND M. PITTS (1983): "The price variability-volume relationship on speculative markets," *Econometrica*, 51, 485–505.

TAYLOR, S. J. (1986): "Modeling Financial Time Series," John wiley, chichester.

WAGNER, N. F., AND T. A. MARSH (2005): "Surprise volume and heteroskedasticity in equity market returns," *Quantitative Finance*, 5(2), 153–68.

WEISS, A. A. (1986): "Asymptotic theory for ARCH models: estimation and testing," *Econometric Theory*, 2, 107–31.

YU, J. (2002): "Forecasting volatility in the New Zealand stock market," *Applied Financial Economics*, 12, 193–202.

ZAKOIAN, J. M. (1994): "Threshold heteroskedastic models," *Journal of Economic Dynamics and Control*, 18(5), 931–55.

Chapter 3

Threshold GARCH Model: Empirical Application

3.1 Introduction

Volatility modeling and forecasting are very important in financial markets, since volatility is a key component in pricing derivative securities, risk management, and making monetary policy.

In the previous chapter we propose a threshold GARCH model to describe the regime shifting in the volatility process of financial asset returns. We derive theoretical conditions for the existence of various moments of return series and examine the properties of MLE via simulation studies. Now we want to investigate the performance of the threshold GARCH model via an empirical application. The simulation results reveal that the parameters are very well estimated when the return series has a finite variance. In this chapter we show that estimated parameters in all empirical applications satisfy the variance stationary requirement without imposing restrictions in the estimation process. The ARCH and GARCH parameters are significantly different from 0 with exogenous or endogenous threshold variables.

The success of a volatility model is determined crucially by its out-of-sample predicting power. Therefore, extensive research has been devoted to this subject. In the 2003 survey on the volatility forecasting literature, Poon and Granger (2003) reviewed 93 published and working papers that study the forecasting performance of various volatility models. The comparisons among different forecasting models show a mixed picture. Poon and Granger conclude that the overall ranking favors ISD (option implied standard deviation) model, while HISVOL (historical volatility models) and GARCH models are roughly equal. However, they also mention that the success of the implied volatility models is benefited from using a larger information set, but they are less practical due to the availability of options. GARCH models perform well in forecasting volatility as described in Hansen and Lunde (2005), they show that simple GARCH model performs well in forecasting currency exchange rates, while in analysis of IBM returns, the models that can accommodate a leverage effect perform better. Taking into account the parameter size in the threshold model, we compare the forecasting performance of the threshold model with GARCH(3,2) and GJR(2,1,2) models.

In volatility forecasting literature there is also a big concern on how should the true volatility be measured. In fact the accuracy of measures of actual volatility has significant effect on the outcomes in comparing volatility models. Most of the early works use the daily squared return to proxy actual daily volatility. However, as shown in Lopez (2001), while squared return is an unbiased estimator of daily variance, it is a very noisy measure of true variance. Taking this into account, besides using daily squared return, we also compare our volatility forecast with realized volatility constructed from intra-day high frequency data.

In this chapter we apply the threshold model to empirical data and find good fit of threshold model in terms of in-sample estimation as well as out-of-sample forecasting.

The rest of this chapter is organized as follows. Section 3.2 briefly describes empirical data and different variables used as the threshold to identify regime-switching in the volatility process. In Section 3.3, we estimate the threshold GARCH model using stock and currency exchange data. We also employ the volatility index as an exogenous threshold variable, and the volume turnover and number of trades as endogenous threshold variables. In Section 3.4 we examine the forecasting performance of the threshold GARCH model. The forecasting performance is evaluated using 5 common measures. A brief conclusion is contained in the last section.

3.2 Data

The empirical data consists of stocks from MMI and NASDAQ and three currency exchange rates.

The first data set consists of 20 stocks in the major market index (MMI)[1]. We obtain the data of most stocks for the period from Jan. 2, 1970 to Dec. 31, 2008, except for AXP and T. The data for AXP and T start from May 18, 1977 and Jan 2, 1984 respectively. We choose the stocks from MMI because

[1]The firms in the MMI are American Express (AXP), AT&T (T), Chevron (CHV), Coca-Cola (KO), Disney (DIS), Dow Chemical (DOW), Du Pont (DD), Eastman Kodak (EK), Exxon (XOM), General Electric (GE), General Motors (GM), International Business Machines (IBM), International Paper (IP), Johnson & Johnson (JNJ), McDonald's (MCD), Merck (MRK), 3M (MMM), Philip Morris (MO), Procter and Gamble (PG), and Sears (S).

they are well known and highly capitalized stocks representing a broad range of industries and they generally exhibit a high level of trading activity. Return data are obtained from daily stock file of the Center for Research in Security Prices (CRSP) and accessed from Wharton Research Data Services (WRDS).

The exogenous threshold variable we used in this empirical study is the Volatility Index (VIX). The Chicago Board Options Exchange (CBOE) Volatility Index is a key measure of market expectations of near-term (30-day) volatility conveyed by S&P 500 stock index option prices. It is a weighted blend of prices for a range of options on the S&P 500 index. The volatility index is calculated and disseminated in real-time by CBOE. We obtain the data from CBOE website from Jan. 2 1990 to Dec. 31 2008. Since the volatility index measures the market expectations for the future volatility, it is reasonable to assume the independence between VIX and the current volatility. It is shown in the data that the sample correlation coefficient between squared return and VIX for 20 stocks in MMI ranges from 0.03 to 0.09, while average correlation coefficient between squared return and volume is around 0.5. We can thus treat VIX as a weakly exogenous variable.

The summary statistics for the returns in MMI are presented in Table 3.1 in Appendix. The columns report the sample minimum, maximum, mean, standard deviation, coefficient of skewness, and coefficient of kurtosis. We notice that all the return series have large kurtosis comparing to a normal distribution, and most of the returns are negatively skewed.

Since the simulation study of the endogenous threshold model suggests that the MLE perform reasonably well for large endogeneity coefficient, we also apply our model to the volume data. In addition, the use of the threshold model to describe the conditional variance dynamics is motivated by the volume-volatility correlation, we want to examine whether the endogenous threshold variable volume provides more information on the regime shifts in the conditional variance process. Nonetheless the volume variable reveals the trading activities for the individual stocks, while the VIX variable just gives the information for the market as a whole. The volume data are also obtained from WRDS. To remove the trend in the volume series, we define the adjusted volume series by taking the log of the trading volume and then removing the 100-day moving average from the log volume series. The resulting series have an average correlation coefficient around 0.5. Since the volume series, even with detrending adjustment, is still very noisy, we also search for other endogenous threshold variables. Since the adjusted volume does not provide much information in regime switching, we didn't report the estimation results.

To further explore the usefulness of the endogenous threshold, we also obtain the second data set for 4 most active stocks in NASDAQ[2], since the number of trades data is only available for NASDAQ stocks. The number of trades data is available for most of the stocks from Jan. 03, 1995 to Dec. 31, 2010 except YHOO and GOOG, they have the number of trades available from Apr. 15, 1996 and Aug. 20, 2004 respectively. There is always concern about the noise in trading activity variables, volume is one very noisy trading variable. We tried to adjust the volume series by taking log and removing time trend for MMI stocks. Here we use another volume variable for NASDAQ stocks. The volume variable is defined as the ratio of trading volume over the total shares outstanding (Volume/SHOUT) for the stock. This is actually the turnover of the daily stock and it is stationary. For the number of trades, if there is a clear

[2]4 most active stocks in NASDAQ: Yahoo! Inc. (YHOO), Apple Inc. (AAPL), Google Inc. (GOOG), QUALCOMM Incorporated (QCOM).

time trend, we detrend the series by removing the best straight-line fit from the series. The data descriptions for return, volume/SHOUT and number of trades are available in Table 3.2 and Table 3.3 in Appendix.

Since we only obtain the daily price data for MMI and NASDAQ stocks, we use the squared daily return as a proxy for the actual volatility when evaluating the forecasting performance. In the volatility forecasting literature how should the true volatility be measured is a big concern. However, it is shown that the daily squared return is a very noisy measure to approximate the actual daily volatility, even though the squared return is an unbiased estimator of daily variance. Taking this into account, besides using daily squared return, we also compare our volatility forecast with realized volatility constructed from intra-day high frequency data. Thanks to Dinghai Xu, we are able to obtain intra-day high frequency data for IBM and GE stocks from Mar. 03, 2005 to Sep. 24, 2008 as well as three currency exchange rates, namely CAD/USD, USD/JPY, and GBP/USD. The high-frequency intra-day transaction prices for currencies are available from Apr. 13, 1998 to July 28, 2006. We summarize the sample statistics for three currency exchange returns and the realized volatility constructed from HF data in Table 3.4 to Table 3.6.

3.3 Estimation of the Threshold GARCH Model

3.3.1 Exogenous Threshold Variable and MMI Stocks

In this section we first apply the threshold GARCH model to the data set that contains 20 stocks from MMI. We assume that the return series follows the threshold GARCH model where the trigger variable is either exogenous (VIX) or endogenous (volume).

Since the return series in our threshold GARCH model is assumed to be a zero mean process, we first remove the mean from the returns. In addition to a constant we also filter the AR effect to order 5:

$$r_t = R_t - \mu - \sum_{j=1}^{5} \delta_j R_{t-j}$$

where R_t is the observed return, μ is the mean, and δ_j is the coefficient of AR variables.

The threshold variable (VIX or volume) is given in this threshold GARCH model, so we also need to determine the threshold value for this variable. Now the model that needs to be estimated has a set of parameters as a function of the threshold value.

$$r_t = \sigma_t \varepsilon_t$$

$$\begin{cases} \sigma_t^2 = \omega_0 + \alpha_0 r_{t-1}^2 + \beta_0 \sigma_{t-1}^2 & if\ y_{t-1} \leq y^* \\ \sigma_t^2 = \omega_1 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2 & if\ y_{t-1} > y^* \end{cases}$$

To estimate the threshold value we divide the sample of threshold variable into 40 intervals and the 39 grid points correspond to $2.5th$ percentile point to $97.5th$ percentile point. We use only the first lag of VIX and volume as the threshold variable since we believe the most recent observation of them provides the most up-to-date information on the condition of market and individual stocks. The robust standard errors we compute for the volatility parameters are Bollerslev-Wooldridge standard errors. Since we use the grid search to obtain the threshold value y^*, we are not able to compute its standard error.

There are other estimation methods available for the threshold model that provide such statistics. We may explore alternative methods in the future studies.

The estimation results in Table 3.7 are based on 4787 observations from Jan. 02, 1990 when VIX is available. The estimated threshold value and the probability that volatility is in regime 2 is given by y^*, for example the estimated threshold is $y_{92.5}$ for IBM, it means that the threshold value is the 92.5 percentile point of VIX price, so the volatility is in a volatile regime with probability of 7.5%. The parameters in the threshold GARCH model are significant for most of the stocks. For some stocks the sum of estimated parameters is greater than 1 in one regime, but consider the probability π, we will still have the stationary process. The probability is given by the location of the threshold value in the sample space of the threshold variable, it is very clear in our estimation results since we use 2.5 percentile as the increment. For example the estimated parameters of PG in regime 2 have a sum of 1.0026, but the threshold value estimated is y_{80}, that means there is only 20% chance that the conditional variance shifts to the regime 2, therefore with a stationary regime 1 the stationarity condition for return process still holds. We observe that for some stocks the estimated parameters in 2 regimes are very similar, it is not surprising because we use VIX as a threshold variable for all 20 stocks in our sample, some stocks may follow closely with the market, while others may be less affected by the market conditions. Nonetheless when we use the market condition as a threshold, it indeed separates the returns in low volatility regime from that in the high volatility regime.

We plot the MMM return series in different regimes in Figure 3.1 and 3.2. Figure 3.1 contains a graph of return series that is divided into 2 regimes, while Figure 3.2 provides the threshold value of the VIX price to separate two regimes.

From Figure 3.1 we see that given the threshold value, the high volatility regime identifies the 3 periods of clustering of extreme returns. The volatile periods are separated from the less volatile periods, and the clustering of volatile periods confirm the presence of the disruptive events during the periods, such as the clustering of volatile periods at the starting point of the graph corresponds to the 1990-1991 Persian Gulf Crisis, the clustering of volatile periods at the middle of the graph corresponds to the periods from 1997 Asian crisis to 2000 dot.com bubble, and the clustering at the end of the graph corresponds to the 2008 subprime mortgage crisis. The use of the VIX as the threshold variable enable us to find the periods in which the VIX market and the stock are volatile since the markets tend to move together, meanwhile we will miss the stock specific information so that some volatile events are ignored simply because of the involatile VIX at that point in time. Nevertheless those stock specific events tend to be non-persistent, we observe that some large negative returns are not identified as in the high volatility regime simply because it is a rare event. Figure 3.2 shows the VIX price and regimes divided by the VIX price. The estimated threshold value for VIX price is 20.17, when the price of VIX is above the threshold value, people view the market as unstable, therefore we observe large price movement in the VIX prices. The empirical results reveal the potential of the threshold variable to identify the regimes in the volatility process hence provide better forecast.

3.3.2 The Estimation with Endogenous Threshold Variables

When using volume as the threshold variable, we capture almost all the large positive or negative returns in the volatile periods, however we fail to identify the clustering of the volatile periods. Using MMM as an example, we plot the regimes in return as well as the threshold value in the adjusted volume series. In comparison to the previous graph using VIX as the threshold, the only volatile period confirmed by volume is the 1990-1991 Gulf Crisis. The reason that volume does not provide better identification of regime switching may be due to the transformation we made on the volume variable. When we transform the volume series to a stationary process the important information may also be removed. We do observe that the volume series is very noisy.

The estimation results question the use of volume as an effective threshold variable. We now consider other 2 candidates for the endogenous threshold variables: the number of trades and the volume turnover. Table 3.8 presents the estimation results for 4 NASDAQ stocks using the two endogenous threshold variables.

The estimated parameters using volume turnovers behave much better than using adjusted volume as the threshold. Figure 3.5 to 3.8 show how regimes in QCOM returns are separated by using volume turnover and number of trades as threshold variable.

It is obvious that for QCOM the volume turnover is more capable of identifying the volatile periods in the sample. The ratio of volume over shares outstanding is quite low after 2000 dot.com bubble, therefore labels the period between 2002 and 2008 as the tranquil period. It is consistent with the smaller positive and negative outliers in return series during the period. The number of trades during the same period is also lower than the estimated threshold value which indicate that the period is less volatile, however the time period before 2008 crisis is identified as a volatile regime since the number of trades is substantially large during the period. But, the period is characterized with smaller variation in the return process. It just reminds us there may be noisy trading activities even when the market is calm.

The volume turnover or number of trades convey information that can be used to identify the volatility regimes. But the performance of threshold models depends on how relevant and useful the threshold variable is, we see that VIX as an indicator of market conditions certainly help us to separate the volatile regime from a less volatile regime. We also observe the similar effect when using variables that reveal trading activities of individual stocks. We may find more variables or the combination of variables that provide even more precise information on regime shifting. We also present the estimation results of NASDAQ stocks using VIX as a threshold variable. The estimation results are shown in Table 3.9. Since the constants are very close to zero for all stocks, we just exclude the estimates from the estimation results.

The estimated parameters for both sets of stock data appear to be stationary and have a finite variance without imposing any restriction in estimation process. We also estimate the model for three currency exchange rates, for the currency exchange data, we don't know which variable will provide helpful information in identifying regime shifts. Currency exchange rates may not follow the move of the stock market closely, so the volatility index may not provide useful information. Therefore we simply employ the spot price of three exchange rates and the volatility index as possible threshold variables. We would like to investigate which variable contains more information. The results are presented in Table 3.10, we actually find that the volatility index works well for GBP/USD, well the spot prices are better indicators for CAD/USD and USD/JPY series. Of course there are many other good candidates for threshold variables, such as interest rate, or oil price which are more likely to affect the currency exchange rate. But clearly the available variables are quite informative for separating regimes in currency exchange rates. We plot the regimes in the CAD/USD series in Figure 3.9.

3.4 Forecasting Performance of the Threshold Model

We have shown the good in-sample fit of the threshold model. In this section we test the out-of-sample predicting power of the threshold model. To construct the volatility forecasts, we estimate the threshold parameters using in-sample data for MMI and NASDAQ stocks and three currency exchange rates.

Since the detrended log volume does not perform well in estimation, we only consider using VIX as an exogenous trigger, while volume turnover and number of trades as endogenous triggers. We construct ten-day-ahead daily volatility forecasts for 250 days for stocks and 100 days for currency exchange rates.

3.4.1 Forecasting Evaluation

There are many criteria that have been used in the volatility forecasting literature. However, there is no consensus on which measure is more appropriate. The ranking of predicting power of models may be different if different measures are used in evaluating these models. Therefore, in our empirical study we choose 5 popular measures instead of only one measure.

The commonly used simple measures in forecasting evaluation include: ME, MSE, MPE, and RMSE. The mean error (ME) and mean percentage error (MPE) are signed measures of error which indicate whether the forecasts are biased. Or, in other words, are they disproportionately positive or negative. Both ME and MPE allow the positive and negative errors to cancel each other, while MPE is a relative measure and a scaled measure. The most commonly used measure is the mean squared error (MSE) measure. MSE contains both bias and variance of errors:

$$E(MSE) = E[(y - \hat{y})^2] = Var(y - \hat{y}) + (E[y] - E[\hat{y}])^2$$

Usually the root mean squared error (RMSE) is reported rather than MSE, because the RMSE is measured in the same unit as the data, and is representative of the size of a typical error.

Bollerslev and Ghysels (1996) further suggest a heteroskadasticity-adjusted version of MSE called HMSE, where

$$HMSE = \frac{1}{T}\sum(\sigma^2_{t+k}/\hat{\sigma}^2_{t+k} - 1)^2$$

There is also a popular regression-based method for forecasting evaluation. It measures the explanatory power of the regression of actual series on the forecasts.

Following most studies, we employ 5 measures in evaluating forecasting accuracy of the threshold GARCH model. For any given stock i the 5 measures are:

$$ME_i = \frac{1}{T}\sum(\sigma^2_{i,t} - \hat{\sigma}^2_{i,t})$$
$$MPE_i = \frac{1}{T}\sum(\sigma^2_{i,t} - \hat{\sigma}^2_{i,t})/\hat{\sigma}^2_{i,t}$$
$$RMSE_i = \sqrt{\frac{1}{T}\sum(\sigma^2_{i,t} - \hat{\sigma}^2_{i,t})^2}$$

$$HMSE_i = \frac{1}{T}\sum(\sigma^2_{i,t}/\hat{\sigma}^2_{i,t} - 1)^2$$

and R^2_i obtained from regressing the actual conditional variance σ^2_i on the forecasts $\hat{\sigma}^2_i$:

$$\sigma^2_i = a + b\hat{\sigma}^2_i + \eta_i$$

3.4.2 Forecasting Results of MMI Stocks Return Volatility

We report the forecasting evaluations in Table 3.11. For 18 out of 20 stocks, the 2008 financial crisis is included in the out-of-sample forecasting period. We found in Table 3.11 that the threshold GARCH model outperforms GARCH model in all 5 measures for 12 out of 18 stocks and provides a better forecast for the volatility surge in the period of financial crisis. For the remaining 8 stocks the evidence is mixed, 3 stocks with only one measure favoring GARCH model, and only 1 stock with four measures supporting GARCH model.

Figure 3.10 and 3.11 provide illustrations of the forecasting comparison between threshold GARCH and GARCH models. It is clear in Figure 3.10 that the forecasts from threshold GARCH model pick up more variation in squared returns than GARCH model. Also there is a quicker response to volatility spike by the threshold model forecasts. Even for the stocks with more forecasting measures against threshold model, for example in Figure 3.11, we have forecasts plot for GM, there are still much more changes in squared returns explained by threshold volatility forecasts than GARCH forecasts. It suggests that the measures are not always accurate in providing reference on model comparison. We conclude that the better performance by threshold model is benefited from a large ARCH effect in the volatile periods. When regime changes from low volatility to high volatility in the out-of-sample period, the volatility forecast is computed using a large ARCH coefficient in the volatile period, therefore the effect of the increase in the previous day squared return is magnified in the current day volatility forecast.

We also need to mention that in our experiment only one threshold variable is used to identify the regime shifting for all 20 stocks, it is possible that some stocks follow the market more closely than the others, in that case we will get much better forecasts by using threshold model. On the other hand some stocks are more likely to be affected by factors other than market risk, in that situation using VIX as a threshold may not help much in volatility forecasting. But it is not evidence against the use of the threshold model, we can always search for variables that provide more information related to the trading activities of individual stocks.

3.4.3 Realized Volatility as Proxy of Actual Volatility

The empirical results show a better fit of threshold volatility forecast with data variation. However there is still a large proportion of variation in data not explained by the threshold model (See Figure 3.10 and 3.11 and low R^2 in Table 3.11). With the wider availability of high frequency data, researchers propose that the poor performance of volatility forecasting models may be caused by the use of a proxy of actual volatility. In the empirical application we compare the volatility forecast with squared daily return which is a proxy of actual daily volatility. Though the squared return is a unbiased estimator of true volatility, it is a very noisy measure due to the nature of its distribution. Shown by Lopez (2001), almost 75% of the time the squared de-meaned returns are either 50% larger or smaller than true variance:

Let $r_t = \mu + \varepsilon_t$, $\quad \varepsilon_t = \sigma_t z_t$,

given $z_t \sim N(0,1)$. Then

$$E(\varepsilon_t^2|\phi_{t-1}) = \sigma_t^2 E(z_t^2|\phi_{t-1}) = \sigma_t^2$$

since $z_t^2 \sim \chi^2_{(1)}$, we actually can show that:

$$Prob(\varepsilon_t^2 \in [\tfrac{1}{2}\sigma_t^2,\ \tfrac{3}{2}\sigma_t^2]) = Prob(z_t^2 \in [\tfrac{1}{2},\ \tfrac{3}{2}]) = 0.2588$$

Based on the above discussion we would like to ask if the poor performance of the volatility forecasts is actually a result of using bad proxy of true volatility. We use intra-day high frequency data of IBM and GE to construct the daily realized volatility as a measure of true volatility. The data contains 938 daily observations from March 03, 2005 to September 24, 2008. Also even though the GARCH(1,1) model is proved to be a robust volatility model for forecasting, taking into account that our threshold GARCH model contains more parameters than that of GARCH(1,1), we now use GARCH(3,2):

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{3}\beta_i\sigma_{t-i}^2 + \sum_{j=1}^{2}\alpha_j r_{t-j}^2$$

and GJR(2,1,2):

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{2}\beta_i\sigma_{t-i}^2 + \delta I(r_{t-1} < 0)r_{t-1}^2 + \sum_{j=1}^{2}\alpha_j r_{t-j}^2$$

models as competing models.

Realized Volatility Estimator and Its Distribution

The realized volatility is defined as the sum of high-frequency intra-day squared returns in a trading day to approximate the daily quadratic variation of the log price process.

Consider a discrete-time model in which the daily asset return is expressed as:

$$r_t = \sigma_t v_t$$

where $v_t \sim iidN(0,1)$.

For a given trading day t, the prices, $P_{t,d}, d = 1, \ldots, D$, are observed tick-by-tick. D refers to the total number of observations in day t. The dth intra-period return at day t can be calculated by taking the difference between logarithmic price at d and that at $d-1$:

$$r_{t,d} = log(P_{t,d}) - log(P_{t,d-1})$$

Assume $r_{t,d} = \sigma_{t,d} v_{t,d}$ where $v_{t,d} \sim iidN(0, \frac{1}{D})$

Then, the daily return is the sum of all intra-period returns:

$$r_t = log(P_{t,D}) - log(P_{t,0}) = \sum_{d=0}^{D} r_{t,d}$$

and

$$\sigma_t = \frac{1}{D} \sum_{d=1}^{D} \sigma_{t,d}$$

The squared daily return is:

$$r_t^2 = \sum_{d=0}^{D} r_{t,d}^2 + \sum_{i \neq j}^{D} r_{t,j} r_{t,j}$$

The daily realized volatility (or variance) is computed as:

$$RV_t = \sum_{d=0}^{D} r_{t,d}^2$$

If the intra-day returns are uncorrelated, then

$$Var(r_t) = E(r_t^2) = E(\sum_{d=0}^{D} r_{t,d}^2) + 0 = E(RV_t) = \sigma_t^2$$

Realized volatility is an unbiased estimator for the variance of daily asset returns. Theoretically, the measure provides a consistent estimator of the latent volatility and becomes more efficient if the most frequent interval is used to compute the measure. Empirically, due to the inefficiency of the financial market, the noisy trading activities may bias the realized volatility and the problem worsens when higher and higher frequencies are used. Here we follow Andersen and Bollerslev (1998) and employ the empirically optimal sampling frequency at 5-minute interval. The construction of realized volatility is explained as follows, given the closing price $P_{h,t}$ at any 5-minute interval at day t, the 5-minute return $r_{h,t}$ for $h = 1, 2..., H; \quad t = 1, 2, ..., T$ is defined as:

$$r_{h,t} = 100(logP_{h,t} - logP_{h-1,t})$$

Then, the realized volatility is simply the sum of squared 5-minute returns in a day:

$$RV_t = \sum_{h=1}^{H} r_{h,t}^2$$

To reduce the effect of noise in the high frequency trading, we further apply the scaling correction to the above realized volatility measure as proposed by Martens (2002):

$$RV_t = w \sum_{h=1}^{H} r_{h,t}^2$$

where $w = 1 + \frac{v_1}{v_2}$, and

$$v_1 = \frac{10000}{T}\sum_{t=1}^{T}(logP_{H,t} - logP_{0,t})^2 \quad v_2 = \frac{10000}{T}\sum_{t=1}^{T}(logP_{0,t} - logP_{H,t-1})^2$$

The sample statistics for IBM and GE returns and realized volatility are reported in Table 3.12. The plots of IBM and GE return series and realized volatility are provided in Figure 3.12-3.15. It is not surprising that the realized volatility has smaller variance than that of squared returns.

Since the sample size is much smaller than that in the earlier empirical study, we only forecast 100 daily volatilities using threshold GARCH, GARCH(3,2) and GJR(2,1,2) models. The estimating and forecasting process is same as in the previous section. Table 3.13 and Figure 3.16-3.19 present the forecasting performance of three models.

The use of realized volatility to approximate the latent volatility definitely improves the forecasting performance of threshold GARCH and other GARCH models. All 5 measures in Table 3.13 support the threshold GARCH model and in Figure 3.16-3.19 we are glad to see a much better fit to the data variation when realized volatility is used. We also confirm that the threshold GARCH model produces better forecasts when there is a sudden change in market conditions. Our findings are similar to that of realized volatility literature in forecasting, the R^2 computed from threshold GARCH model is 0.5532, which is more than 2 times of that obtained by using daily squared returns as true volatility, 0.2026. The improvement of R^2 is more significant for other GARCH models, with a more than 3 times increase from 0.1268 in GARCH(1,1) model to 0.5145 in GJR model. The results are consistent with the findings in Blair, Poon, and Taylor (2001), they report a three to four times increase of R^2 when intra-day 5-minute squared returns are used to approximate the true volatility.

Forecasting Volatility for Currency Exchange Rates

Since the currency exchange rates experience different patterns in volatility, comparing with that of stocks the volatility is much volatile during the sample for three currencies. We plot the returns as well as the realized volatility constructed from high frequency intra-day data in Figure 3.20-3.25. While the forecasting performance is shown in Table 3.14 and Figure 3.26-3.31. Though the performance of the threshold GARCH model is very close to other GARCH models in some measures, it outperforms GARCH(3,2) and GJR model across all 5 measures.

3.5 Conclusion and Future Research

In Chapter 2, we propose a threshold GARCH model to explain the time dependencies in volatility dynamics of financial asset returns. We theoretically investigate the properties of the model as a valid volatility model. In this chapter we examine the estimation and forecasting performance of the threshold GARCH model empirically. We employ three sets of data, including MMI stocks, NASDAQ stocks, and currency exchange rates. The estimated parameters are all significant and satisfy the variance stationary requirement. We use the volatility index as an exogenous threshold variable, it is proved to be an useful information variable in identifying regime shifts in the volatility process. The endogenous threshold variable volume turnover and number of trades may also help in separating regimes for stocks not following closely to the market. We further explore the forecasting performance of the threshold GARCH model. We use 5 measures to evaluate the models. The forecasting performance of the threshold GARCH model is better than competing models when the financial crisis is included in the forecasting period. The threshold GARCH forecasts are able to pick up the volatility spike faster. When a more accurate volatility proxy - realized volatility is used in forecasting evaluation, the forecasting performance of all models improved, while the threshold GARCH model still performs better across all measures.

The exogenous threshold variable VIX helps in improving the forecasting performance for many stocks as well as an exchange rate series. There may exist more variables that can be used in the threshold GARCH model to identify the regime switching. We assume 2 regimes just for simplicity, it is possible to extend the model to accommodate more regimes and more threshold variables.

It is showed in our empirical applications that many stocks are moving together and affected by a single factor that representing the market risk. The multivariate modeling of parallel financial time series becomes more and more popular in financial studies. In the next chapter we introduce a new method to extracting underlying factors from multivariate signals. The method is the independent component analysis and it may help in financial analysis of multivariate time series.

3.6 Appendix

:

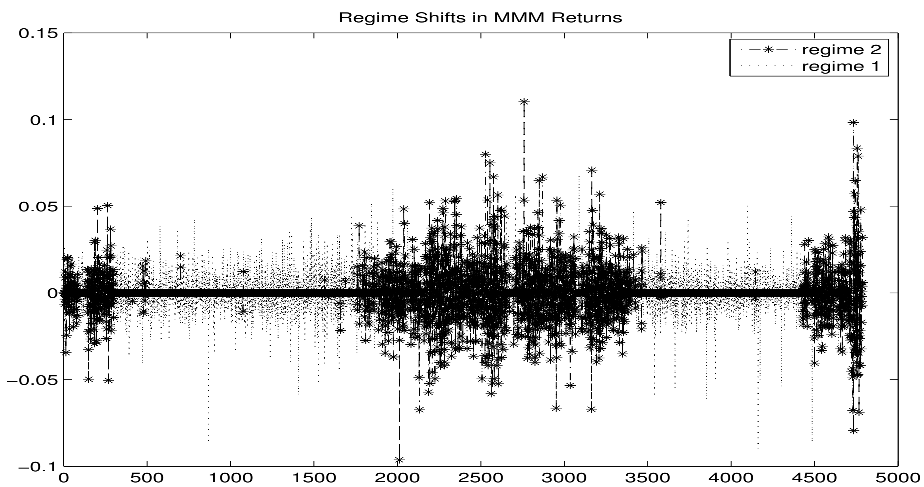

Figure 3.1: Regime Shifts in MMM Returns

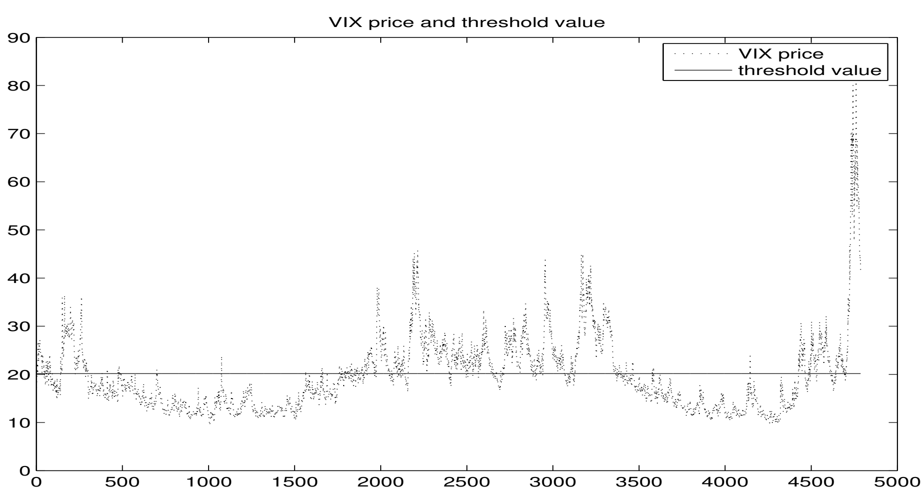

Figure 3.2: MMM Threshold Value in VIX Price

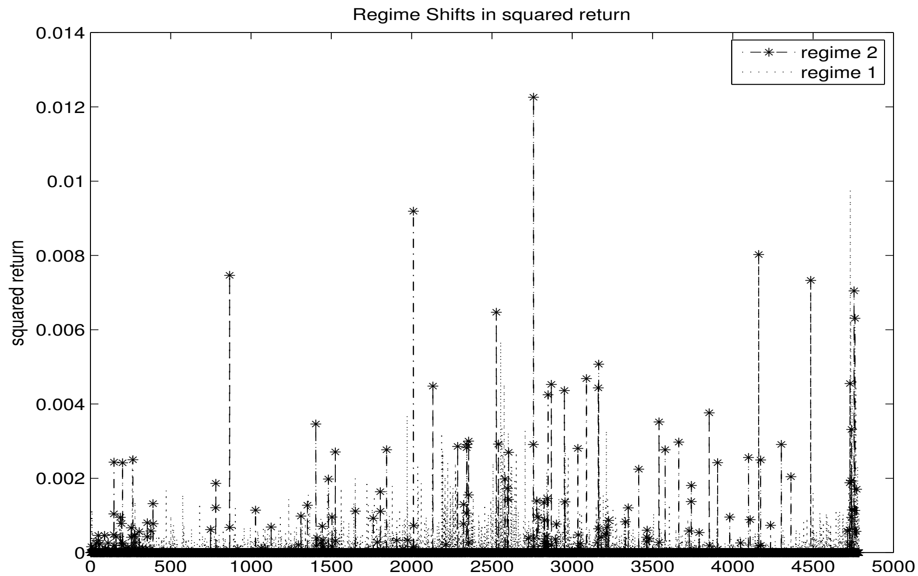

Figure 3.3: Regime Shifts in MMM Return

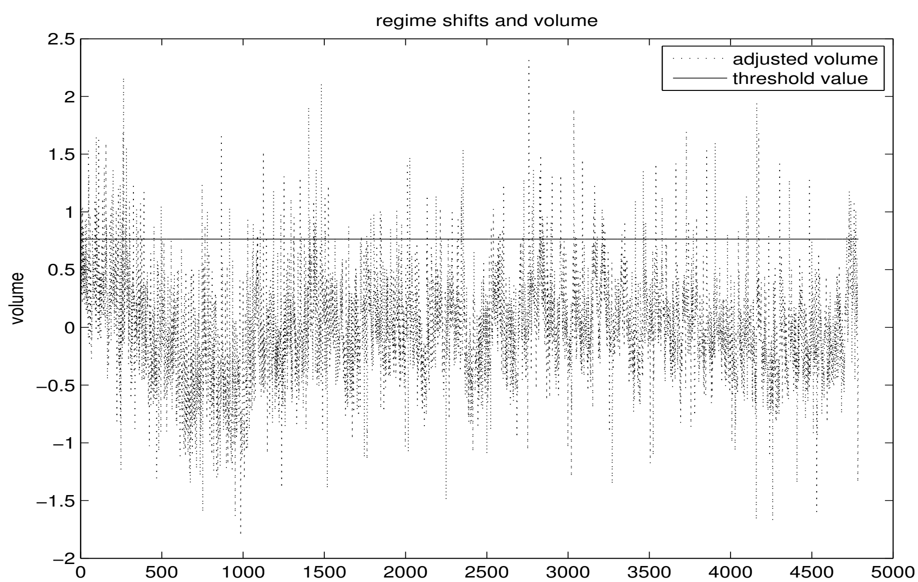

Figure 3.4: MMM Threshold Value in Volume

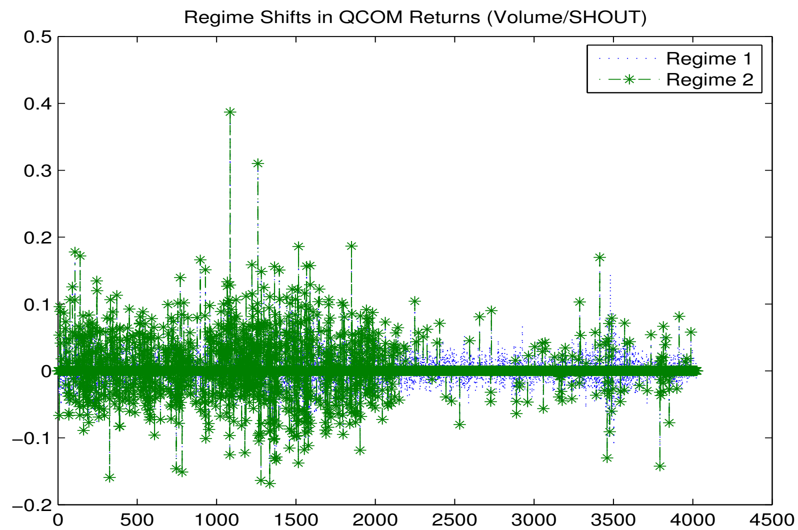

Figure 3.5: Regime Shifts in QCOM Return

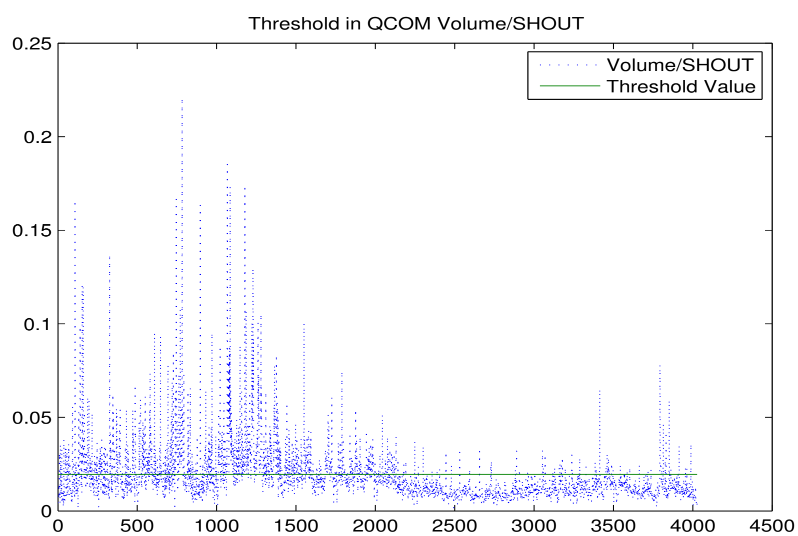

Figure 3.6: QCOM Threshold Value in VOL/SHOUT

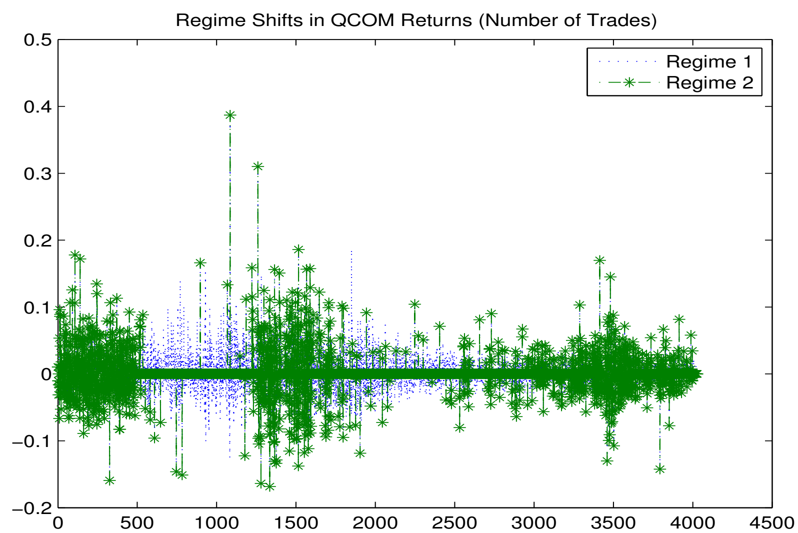

Figure 3.7: Regime Shifts in QCOM Returns

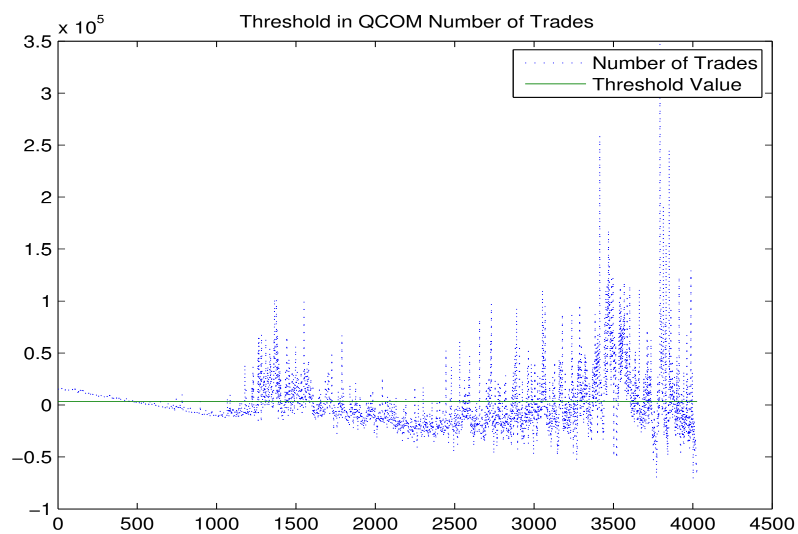

Figure 3.8: QCOM Threshold Value in Number of Trades

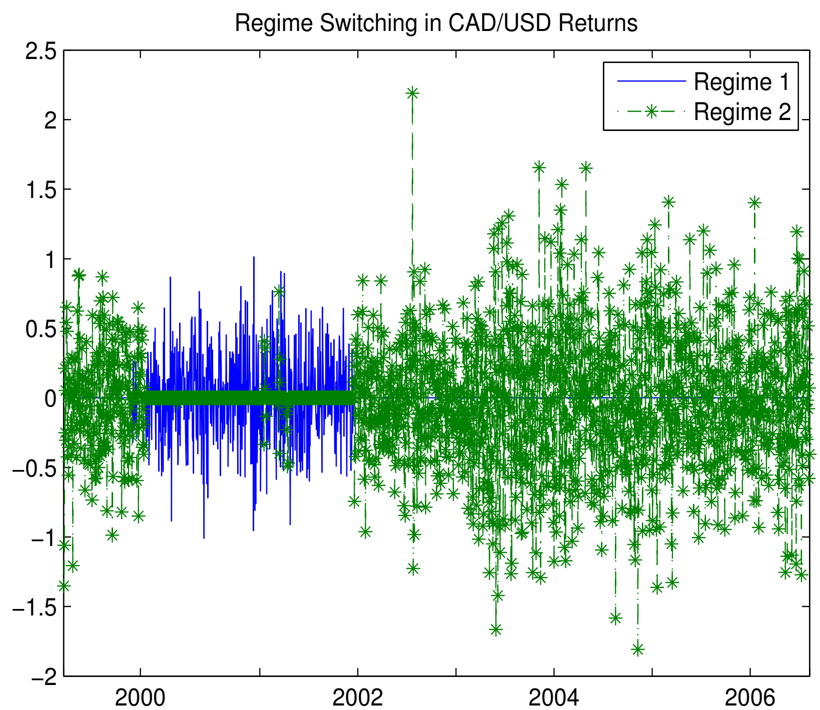

Figure 3.9: Regime Shifts in CAD/USD Returns

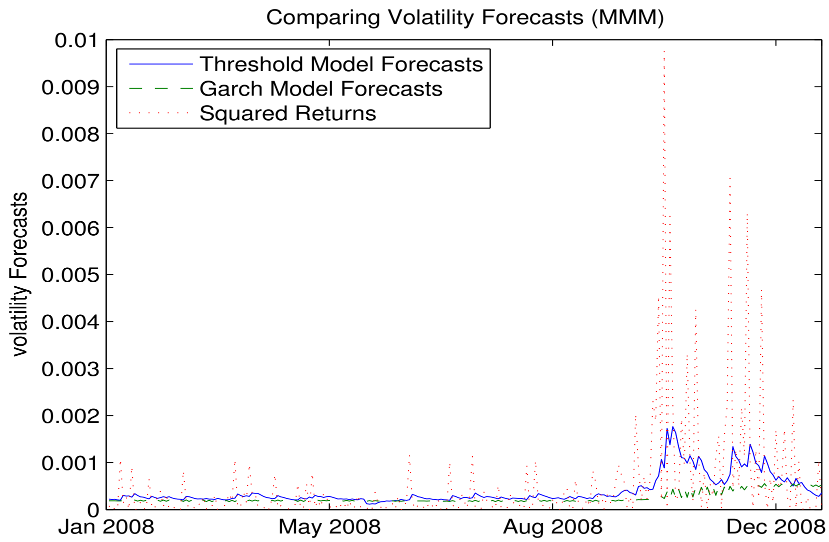

Figure 3.10: Volatility Forecast MMM

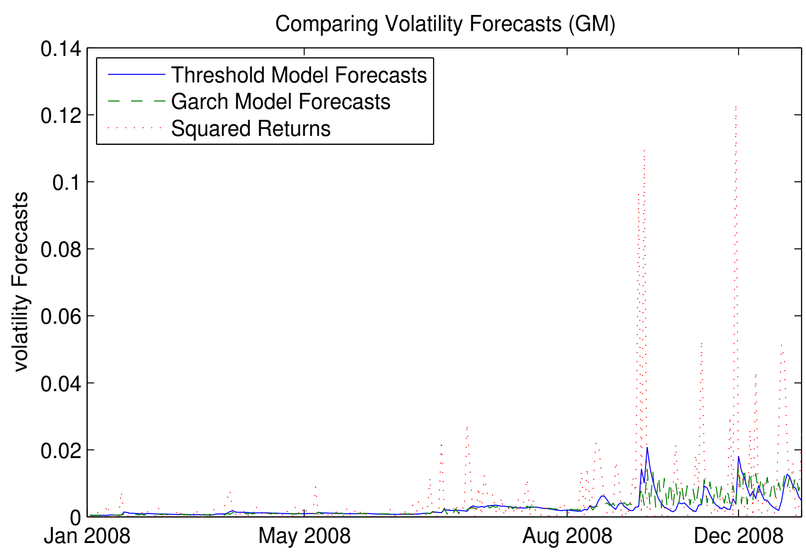

Figure 3.11: Volatility Forecast GM

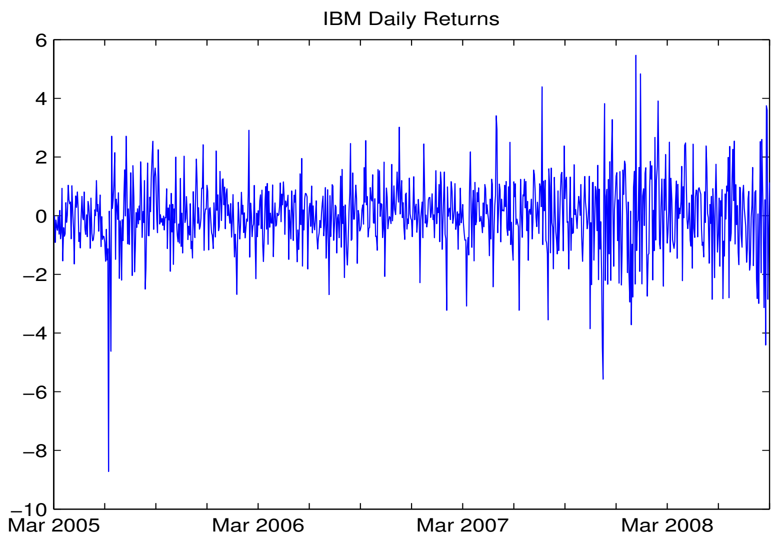

Figure 3.12: IBM Daily Returns

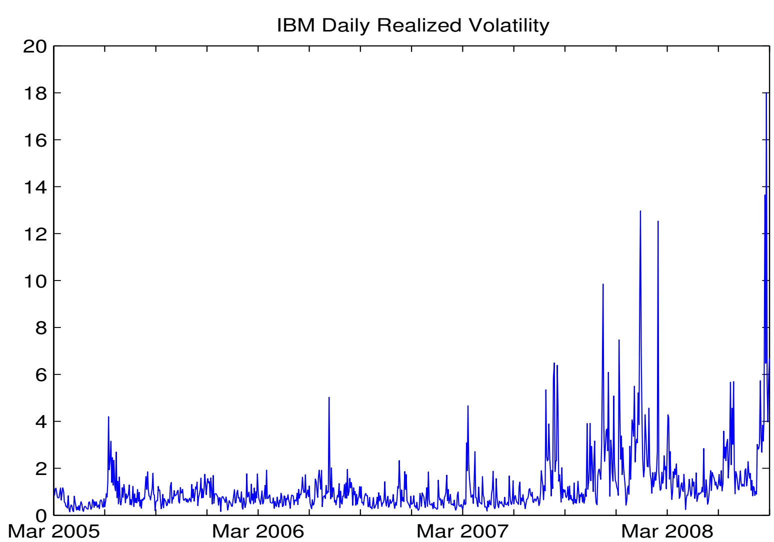

Figure 3.13: IBM Daily Realized Volatility

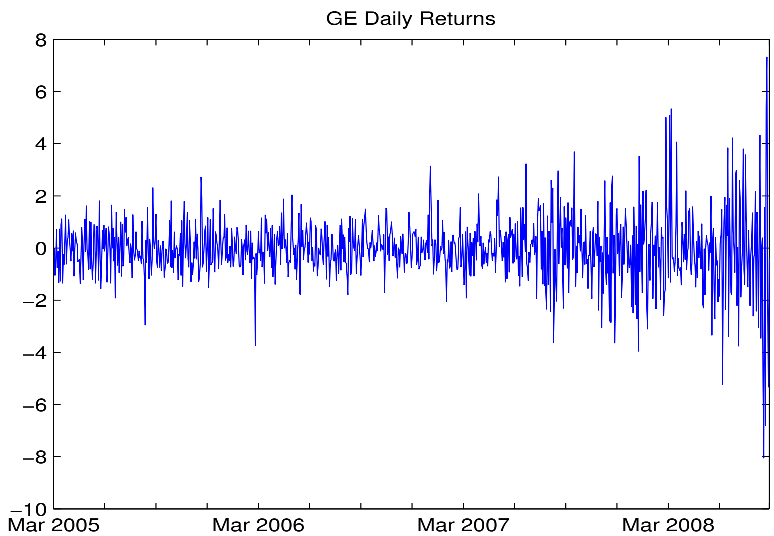

Figure 3.14: GE Daily Returns

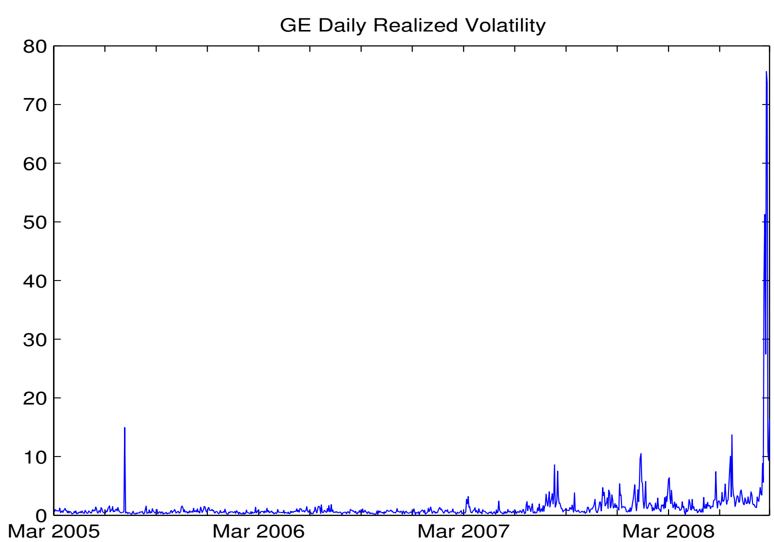

Figure 3.15: GE Daily Realized Volatility

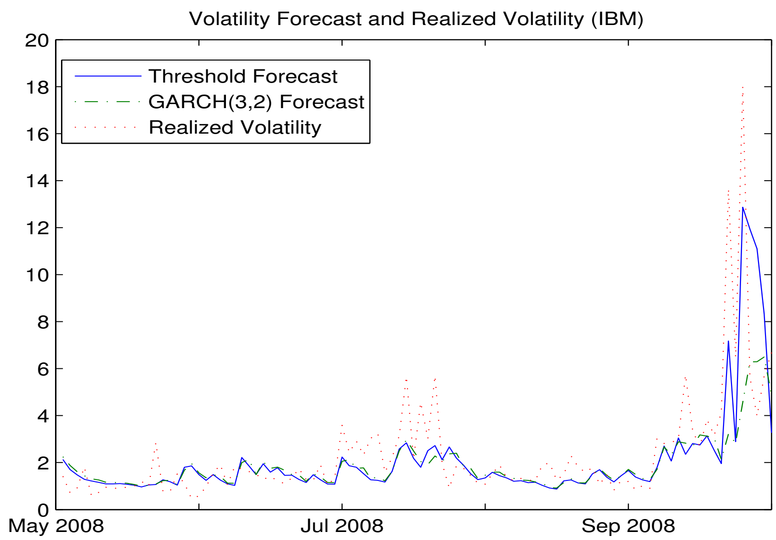

Figure 3.16: IBM Volatility Forecasts Comparison 1

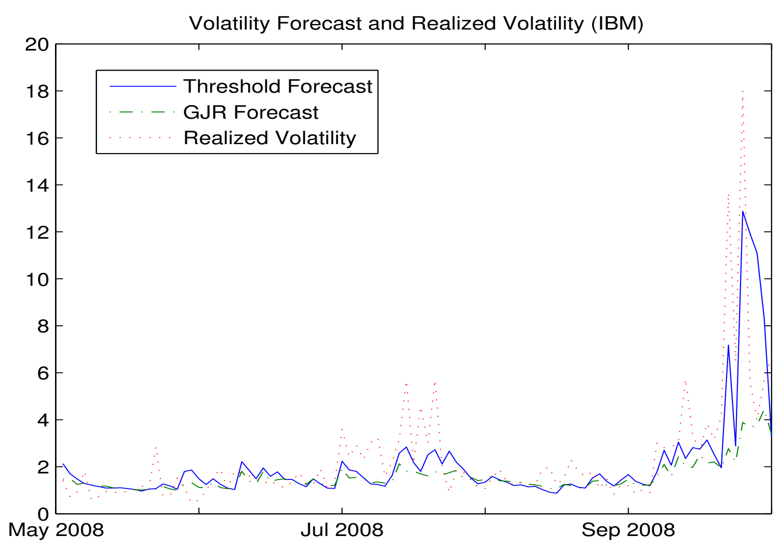

Figure 3.17: IBM Volatility Forecasts Comparison 2

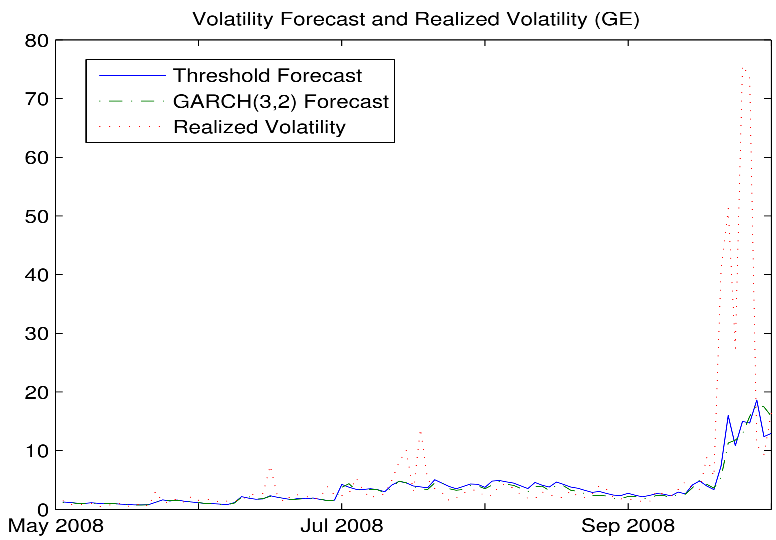

Figure 3.18: GE Volatility Forecasts Comparison 1

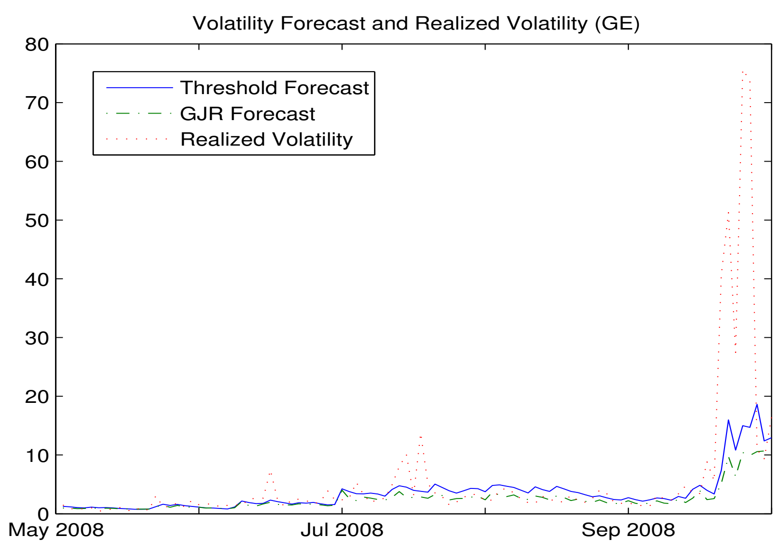

Figure 3.19: GE Volatility Forecasts Comparison 2

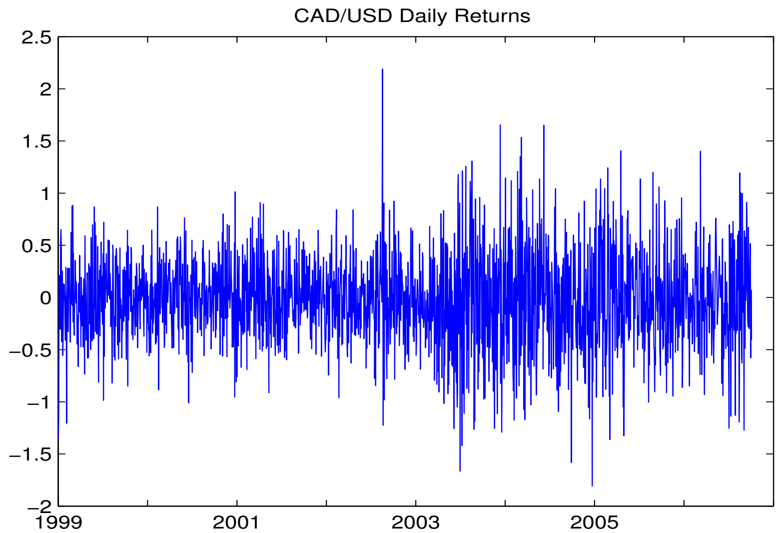

Figure 3.20: CAD/USD Daily Returns

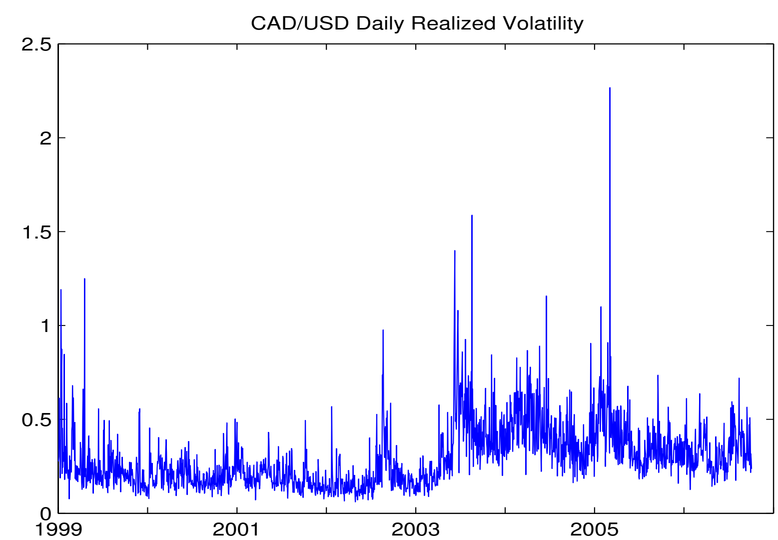

Figure 3.21: CAD/USD Daily Realized Volatility

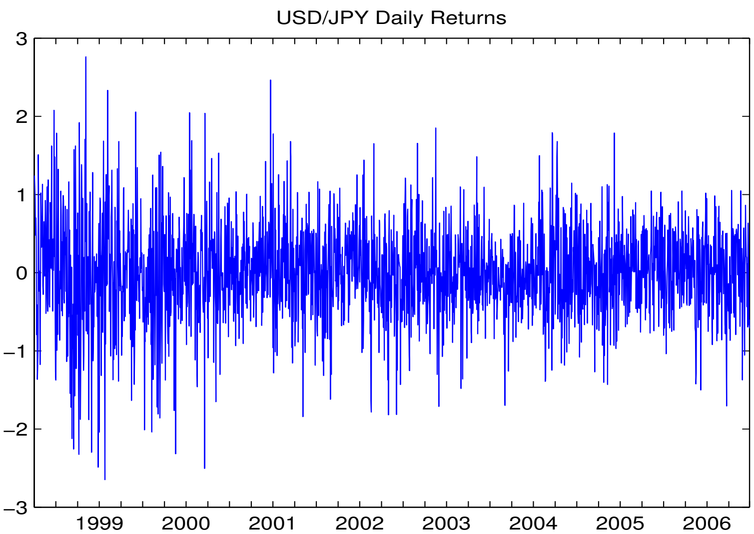

Figure 3.22: USD/JPY Daily Returns

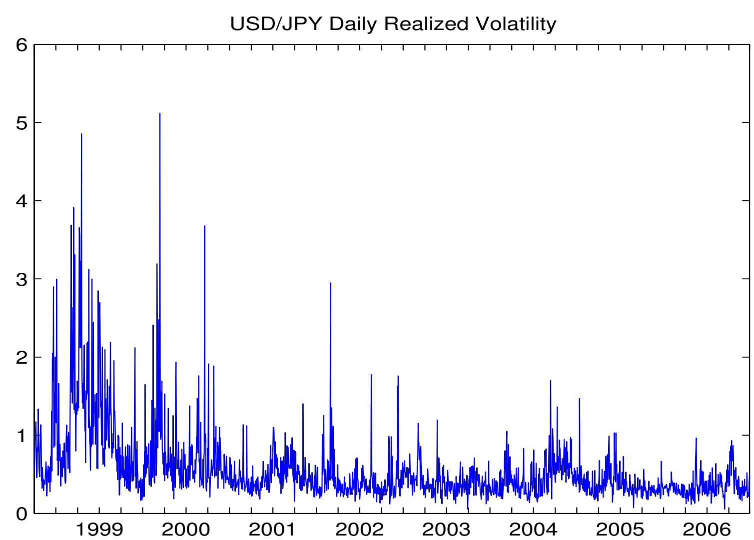

Figure 3.23: USD/JPY Daily Realized Volatility

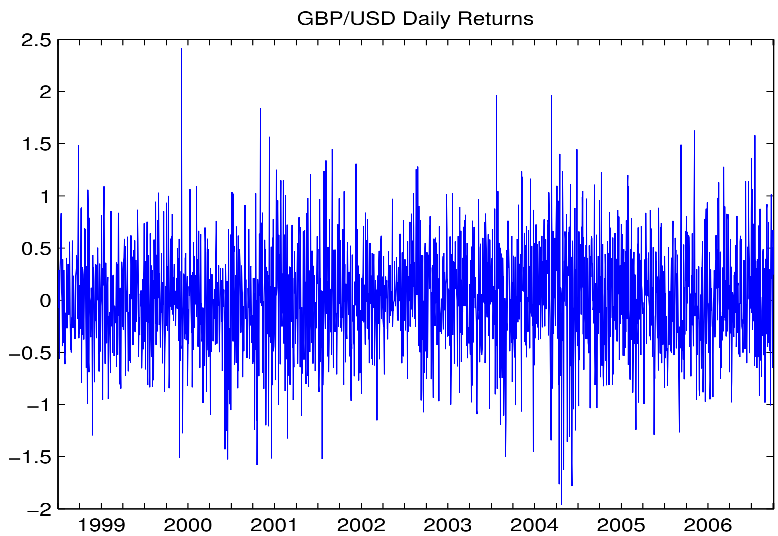

Figure 3.24: GBP/USD Daily Returns

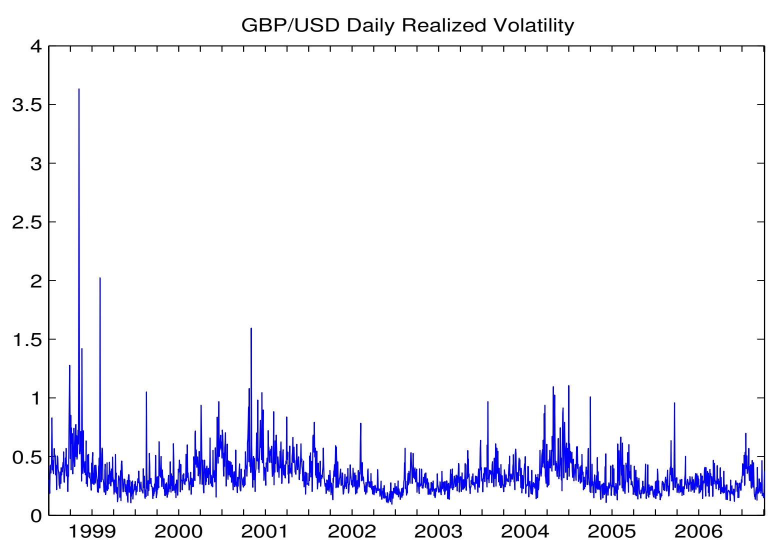

Figure 3.25: GBP/USD Daily Realized Volatility

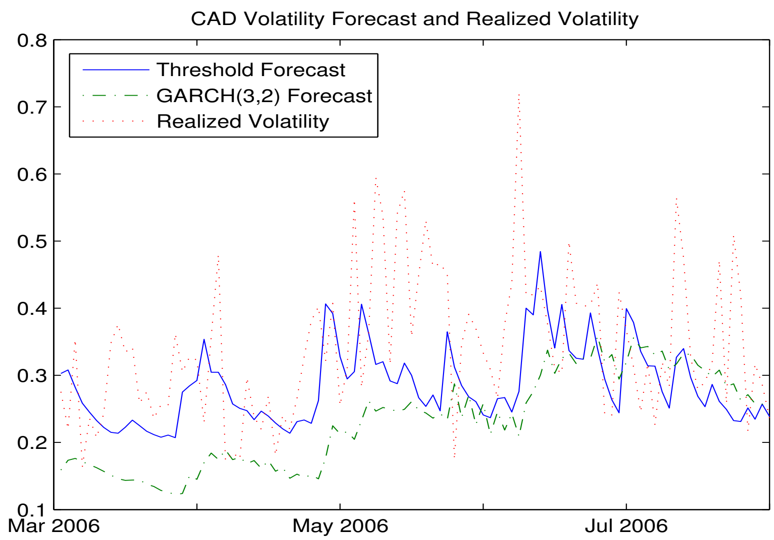

Figure 3.26: CAD/USD Volatility Forecasts Comparison 1

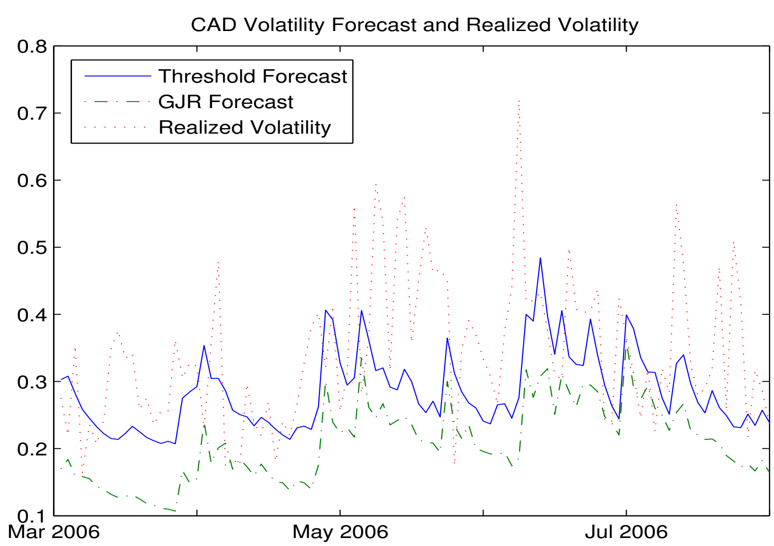

Figure 3.27: CAD/USD Volatility Forecasts Comparison 2

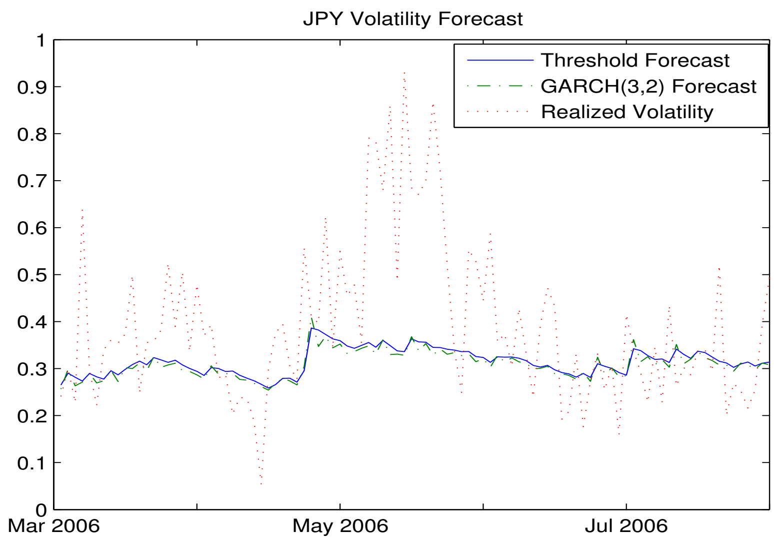

Figure 3.28: USD/JPY Volatility Forecasts Comparison 1

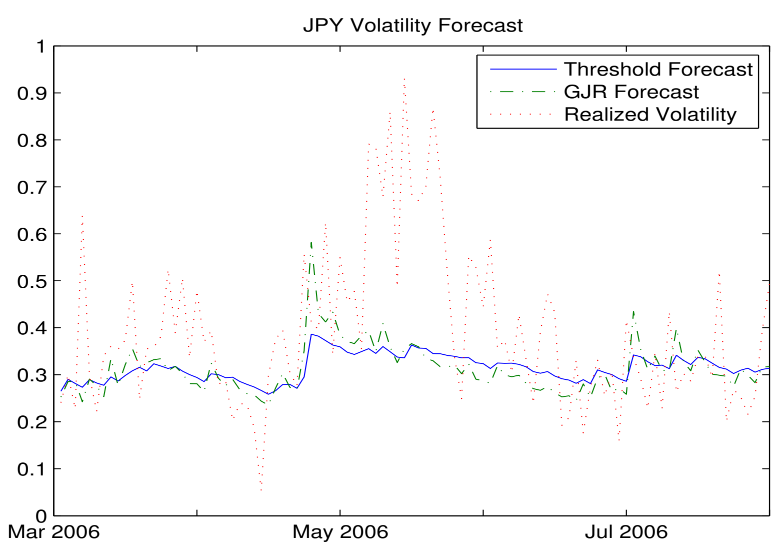

Figure 3.29: USD/JPY Volatility Forecasts Comparison 2

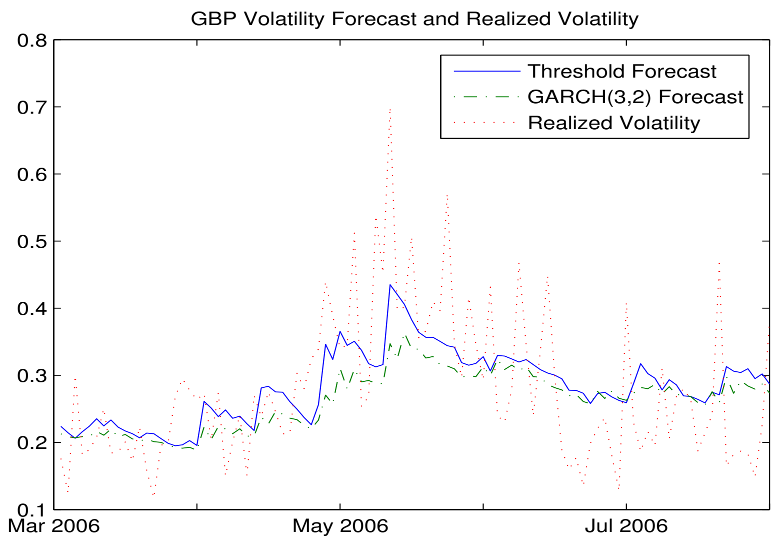

Figure 3.30: GBP/USD Volatility Forecasts Comparison 1

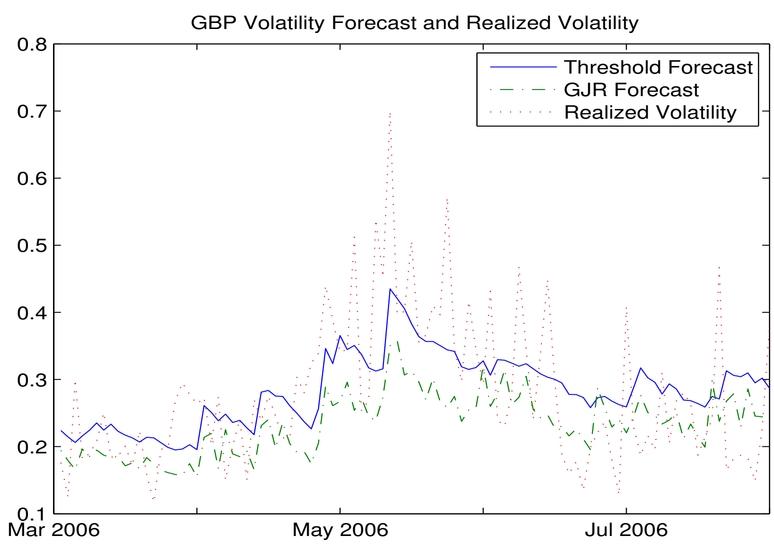

Figure 3.31: GBP/USD Volatility Forecasts Comparison 2

Table 3.1: Descriptive Statistics for MMI Returns

Stocks	# of obs	min	max	mean	std	skewness	kurtosis
AXP	7890	-.2623	.1856	5.91e-04	.0215	-.1186	11.0912
T	6274	-.2194	.2025	6.04e-04	.0173	.1174	14.7662
CHV	9845	-.1667	.2085	6.44e-04	.0166	.2082	10.5984
KO	9845	-.2469	.1957	5.70e-04	.0160	-.0172	15.5269
DIS	9845	-.2909	.1910	6.24e-04	.0209	-.1985	13.1012
DOW	9845	-.1932	.1524	5.03e-04	.0181	-.1165	10.0208
DD	9845	-.1827	.1147	4.45e-04	.0166	-.0177	8.1179
EK	9844	-.3024	.2406	1.84e-04	.0191	-.2118	22.0107
XOM	9845	-.2343	.1790	6.77e-04	.0146	.0071	19.2241
GE	9845	-.1749	.1361	5.63e-04	.0164	.0250	9.8170
GM	9844	-.3111	.3511	2.25e-04	.0251	.4792	31.6393
IBM	9842	-.2296	.1316	4.02e-04	.0169	.0614	12.9150
IP	9845	-.2695	.1912	3.37e-04	.0186	-.1214	13.2700
JNJ	9845	-.1835	.1223	5.87e-04	.0155	-.0489	8.9948
MCD	9844	-.1942	.1622	7.35e-04	.0184	0174	9.7235
MRK	9845	-.2678	.1303	5.49e-04	.0166	-.4530	14.5407
MMM	9845	-.2598	.1154	4.49e-04	.0148	-.3594	16.3442
MO	9843	-.2300	.1638	8.36e-04	.0174	-.1798	12.6358
PG	9845	-.3138	.2220	5.85e-04	.0148	-1.1529	42.4104
S	9845	-.2713	.2866	3.47e-04	.0221	-.1551	27.8540

Table 3.2: Descriptive Statistics for NASDAQ Returns

Stocks	# of obs	minima	maxima	mean	std	skewness	kurtosis
AAPL	4027	-.5187	.3323	.0014	.0324	-.4584	25.2202
GOOG	1604	-.1161	.1999	.0014	.0232	.7889	11.0641
QCOM	4027	-.1685	.3871	.0015	.0351	.8636	10.8992
YHOO	3705	-.2184	.4797	.0015	.0413	.7366	11.4068

Table 3.3: Summary Statistics of VOL/SHOUT and Number of Trades of NASDAQ Stocks

	Volume/SHOUT			Number of Trades		
Stocks	min	max	mean	min	max	mean
AAPL	.0020	.3974	.0264	477	646258	58606
GOOG	.0039	.6385	.0409	6600	227399	38013
QCOM	.0011	.2210	.0181	79	437540	40373
YHOO	2.39e-04	.3279	.0242	27	876185	40028

Table 3.4: Summary Statistics of CAD/USD Return and Realized Volatility

CAD/USD	r_t	r_t^2	RV
Mean	-0.0164	0.1966	0.2955
Var	0.1964	0.1129	0.0286
Skewness	0.0124	4.3994	2.3822
Kurtosis	3.9254	35.3875	17.0873
ACF1	-0.0589	0.0573	0.5927
ACF2	-0.0089	0.0988	0.5596
ACF3	0.0139	0.0740	0.5159
ACF4	-0.0299	0.0746	0.5081
ACF5	-0.0307	0.1247	0.5225

Table 3.5: Summary Statistics of GBP/USD Return and Realized Volatility

GBP/USD	r_t	r_t^2	RV
Mean	0.0099	0.2556	0.3375
Var	0.2550	0.1821	0.0295
Skewness	0.0263	4.2372	5.2897
Kurtosis	3.7833	32.2126	77.1422
ACF1	0.0284	0.0431	0.5127
ACF2	-0.0396	0.028	0.3934
ACF3	-0.0164	0.033	0.3594
ACF4	-0.005	0.0113	0.3531
ACF5	0.0541	0.0737	0.3730

Table 3.6: Summary Statistics of USD/JPY Return and Realized Volatility

USD/JPY	r_t	r_t^2	RV
Mean	-0.0006	0.514	0.556
Var	0.3962	1.9856	0.2071
Skewness	-0.119	9.0588	3.7217
Kurtosis	4.3379	118.7738	23.2468
ACF1	0.0222	0.4375	0.6445
ACF2	0.0254	0.3831	0.584
ACF3	-0.0103	0.4142	0.5827
ACF4	-0.001	0.342	0.5397
ACF5	0.0515	0.3529	0.5343

Table 3.7: Estimation Results of MMI Returns Using VIX as Threshold

	y^*	α_0	α_1	β_0	β_1
AXP	$y_{72.5}$	.0418 (.0067)	.0997 (.0189)	.9497 (.0081)	.8585 (.0281)
T	y_{65}	.0366 (.0102)	.1269 (.0311)	.9448 (.0137)	.7941 (.0379)
CHV	y_{60}	.033 (.0098)	.0863 (.0151)	.9276 (.0217)	.89 (.0205)
KO	$y_{92.5}$	.0279 (0048)	.1306 (0352)	.9704 (0051)	.8613 (.0309)
DIS	$y_{82.5}$	.0634 (.0180)	.1023 (.0278)	.7402 (.0539)	.8515 (.0307)
DOW	$y_{67.5}$	.0523 (.0112)	.0620 (.0120)	.9260 (.0132)	.9375 (.0115)
DD	$y_{77.5}$	.0194 (.004)	.1044 (.0247)	.9757 (.0054)	.8416 (.0381)
EK	$y_{62.5}$	.1290 (.0552)	.2969 (.0964)	.2063 (.1375)	.4656 (.0897)
XOM	y_{95}	.028 (.007)	.0959 (.0152)	.9588 (.01)	.8778 (.0182)
GE	$y_{77.5}$	.0213 (.0048)	.1207 (.0227)	.9721 (.0059)	.8397 (.0307)
GM	$y_{67.5}$	.0352 (.0067)	.0916 (.0160)	.9454 (.0120)	.9076 (.0155)
IBM	$y_{92.5}$	.0285 (.0056)	.1918 (.0495)	.9694 (.0062)	.7479 (0579)
IP	$y_{57.5}$	.0305 (.0109)	.1276 (.0247)	.8713 (.0321)	.8539 (.0235)
JNJ	y_{85}	.0469 (.0067)	.1629 (.0755)	.9513 (.0070)	.7723 (.0674)
MCD	y_{60}	.0335 (.0096)	.0408 (.0095)	.9400 (.0140)	.9516 (.0097)
MRK	$y_{67.5}$	.0106 (.0129)	.0726 (.0234)	.9197 (.0327)	.8947 (.0338)
MMM	y_{60}	.0321 (.0164)	.1048 (.0281)	.7524 (.0630)	.8191 (.1291)
MO	y_{15}	.0167 (.0418)	.0514 (.0115)	.2022 (.0199)	.9396 (.0164)
PG	y_{80}	.0167 (.0042)	.1488 (.0515)	.9809 (.0043)	.8538 (.0385)
S	y_{50}	.1037 (.0223)	.1119 (.0369)	.7863 (.0432)	.8471 (.0313)

(The small numbers under the estimates are the estimated standard errors)

Table 3.8: Estimation Results of NASDAQ Returns Using VOL/SHOUT and Number of Trades as Threshold

		Volume/SHOUT			
Stocks	y^*	α_0	α_1	β_0	β_1
AAPL	y_{95}	.0827 0250	.0051 0056	.9319 0191	.8824 0631
GOOG	$y_{52.5}$	.0284 0106	.1550 0717	.9363 1174	.8832 0157
QCOM	y_{70}	.1261 024	.0163 0103	.8858 0205	.9699 0167
YHOO	$y_{72.5}$	.0108 0093	.2705 0521	.9123 0330	.8014 0319

		Number of Trades			
Stocks	y^*	α_0	α_1	β_0	β_1
AAPL	y_{60}	.1437 058	.0441 0113	.7981 0502	.944 0132
GOOG	y_{80}	1668 0444	.0866 07	.6073 0702	.0501 2497
QCOM	y_{35}	.1515 0508	.0558 0109	.7508 0562	.9336 0132
YHOO	$y_{47.5}$	.1488 0340	.1238 0564	.8300 0358	.8176 0539

(The small numbers under the estimates are the estimated standard errors)

Table 3.9: Estimation Results of NASDAQ Returns Using VIX as Threshold

Stocks	y^*	α_0	α_1	β_0	β_1
AAPL	$y_{22.5}$	.1616 079	.0791 025	.172 0202	.9228 1813
GOOG	y_{50}	.1059 044	.0464 0152	.815 0637	.9564 0161
QCOM	$y_{32.5}$	.0112 005	.0791 0154	.9878 0072	.9147 0152
YHOO	y_{85}	.0051 0026	.1008 0501	.9904 0038	.9030 0362

(The small numbers under the estimates are the estimated standard errors)

Table 3.10: Estimation Results of Exchange Returns

	Threshold	ω_0	ω_1	α_0	α_1	β_0	β_1
CAD/USD	Spot Rate	.0243 0225	.0028 0018	.0577 0207	.0362 0098	.7172 2389	.952 0151
GBP/USD	VIX	.0105 0226	.0042 003	.0466 0247	.0201 0092	.9174 0616	.9626 0186
USD/JPY	Spot Rate	.0728 0394	.0014 0015	.0998 0439	.0195 0064	.6869 1393	.977 0084

(The small numbers under the estimates are the estimated standard errors)

Table 3.11: Forecasting Results for 20 MMI Stocks

		ME^1	MPE	$RMSE$	$HMSE$	R^2
AXP	TS GARCH	.0569	.3970*	.0090	6.0986*	.0722*
	GARCH	.0404*	.4136	.0064*	6.7201	.0413
T	TS GARCH	-.004*	-.3086*	.0006*	3.3388	.0111*
	GARCH	-.0056	-.3472	.0009	2.5066*	.0051
CHV	TS GARCH	.0361*	.3499*	.0057*	6.2621*	.1499*
	GARCH	.0449	.7507	.0071	21.1730	.0890
KO	TS GARCH	.0198	.3719*	.0031	6.4860*	.1346*
	GARCH	.0167*	.5707	.0026*	17.6212	.0591
DIS	TS GARCH	.0250*	.2001*	.0039*	5.8526*	.0708
	GARCH	.0349	.5260	.0055	12.2574	.1220*
DOW	TS GARCH	.0287*	.2673*	.0046*	10.2805	.0781
	GARCH	.0322	.3842	.0051	8.7226*	.0841*
DD	TS GARCH	.0196*	.1871*	.0031*	4.1163*	.1390*
	GARCH	.0413	.7065	.0065	12.1604	.1179
EK	TS GARCH	.0715*	1.0969*	.0113*	18.287*	.0072*
	GARCH	.0768	1.4259	.0121	32.8882	.0059
XOM	TS GARCH	.0217*	.2818*	.0034*	5.8539*	.1906*
	GARCH	.0335	.6573	.0053	17.8408	.1026
GE	TS GARCH	.035*	.3401*	.0055*	11.405*	.1509*
	GARCH	.0443	.6309	.0070	15.5842	.1280
GM	TS GARCH	.27	1.0918	.0431	17.3807*	.0551*
	GARCH	.24*	1.0234*	.0376*	19.4617	.0509
IBM	TS GARCH	.0592*	.0578*	.0001*	2.0658*	.1735*
	GARCH	.0916	.1689	.0014	3.2271	.1392
IP	TS GARCH	.0235*	.2692*	.0037*	8.0684*	.1122*
	GARCH	.0562	.6689	.0089	23.5101	.0961
JNJ	TS GARCH	.0023*	.0034*	.0004*	3.4455*	.1165*
	GARCH	.0075	.4593	.0012	20.2049	.0705
MCD	TS GARCH	.0066*	.1593*	.0011*	5.5311*	.1082*
	GARCH	.0104	.3084	.0016	7.1577	.0655
MRK	TS GARCH	.0263*	.5576*	.0042*	27.5117*	.0278
	GARCH	.0488	1.0612	.0077	37.7295	.0340*
MMM	TS GARCH	.0109*	.088*	.0017*	2.6553*	.1735*
	GARCH	.0261	.8644	.0041	16.0783	.0916
MO	TS GARCH	.005*	.1373*	.0001*	9.0653*	.0361*
	GARCH	.0152	.5348	.0024	24.8221	.0137
PG	TS GARCH	-.0014*	.108*	.0002*	3.5269*	.01466*
	GARCH	.0123	.4415	.0019	8.1946	.1006
S	TS GARCH	.0202	.5497	.0032	54.4468	.0068*
	GARCH	.0095*	.3065*	.0015*	27.8286*	.0016

(1: The value of ME measure should be multiplied by 0.01.)
*: The model performed better based on a forecasting measure.

Table 3.12: Sample Statistics of HF IBM and GE Returns

IBM	r_t	r_t^2	RV
Mean	0.0190	1.498	1.2499
Var	1.5001	14.1777	2.0106
Skewness	-0.4659	10.4747	5.0680
Kurtosis	7.3347	176.5405	42.2725
ACF1	-0.0059	0.1132	0.6033
ACF2	0.0243	0.1072	0.5326
ACF3	0.0424	0.1615	0.4434
ACF4	-0.0139	0.093	0.4132
ACF5	-0.0396	0.325	0.3771

GE	r_t	r_t^2	RV
Mean	-0.0302	1.5805	1.4525
Var	1.5813	19.8735	18.8073
Skewness	-0.085	7.7857	14.6402
Kurtosis	8.9542	83.9779	289.6924
ACF1	-0.0773	0.2977	0.7208
ACF2	-0.0638	0.4186	0.5376
ACF3	-0.0331	0.2578	0.5611
ACF4	0.0144	0.3978	0.3495
ACF5	-0.0004	0.2047	0.1707

Table 3.13: Forecasting Performance of IBM and GE

		ME	MPE	RMSE	HMSE	R^2
	TS	.2534	.1635	2.5337	.3113	.5532
IBM	GARCH(3,2)	.4289	.1742	4.289	.4494	.3851
	GJR(2,1,2)	.7464	.3505	7.4636	.7098	.5145
	TS	2.0059	.2213	20.059	.9214	.556
GE	GARCH(3,2)	2.163	.3069	21.63	.1.2894	.4471
	GJR(2,1,2)	2.9807	.6021	29.807	2.1961	.5399

Table 3.14: Forecasting Performance of Three Currencies

		ME	MPE	RMSE	HMSE	R^2
	TS	.0567	.2225	.5672	.2021	.1102
CAD/USD	GARCH(3,2)	.1096	.5688	1.0963	.648	.0729
	GJR(2,1,2)	.1338	.7255	1.3378	.8828	.10
	TS	-.0043	-.0204	.0436	.0931	.3556
GBP/USD	GARCH(3,2)	.0155	.0536	.1547	.1173	.277
	GJR(2,1,2)	.0447	.1968	.447	.1965	.2315
	TS	.0773	.2313	.7731	.2767	.2721
USD/JPY	GARCH(3,2)	.0829	.2549	.8289	.3016	.2324
	GJR(2,1,2)	.0791	.248	.7908	.2912	.1506

3.7 Bibliography

AKGIRAY, V. (1989): "Conditional heteroscedasticity in time series of stock returns: evidence and forecast," *Journal of Business*, 62, 55–80.

ANDERSEN, T., AND T. BOLLERSLEV (1997): "Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns," *Journal of Finance*, 52, 975–1005.

——— (1998a): "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts," *International Economic Review*, 39(4), 885–905.

——— (1998b): "DM-Dollar volatility: intraday activity patterns, macroeconomic announcements, and long run dependencies," *Journal of Finance*, 53, 219–265.

ANDERSEN, T., T. BOLLERSLEV, P. CHRISTOFFERSEN, AND F. DIEBOLD (2005): "Volatility forecasting," .

ANDERSEN, T., T. BOLLERSLEV, F. DIEBOLD, AND P. LABYS (2003): "Modeling and forecasting realized volatility," *Econometrica*, 71(2), 579–625.

ANDERSEN, T. G., T. BOLLERSLEV, AND N. MEDDAHI (2005): "Correcting the errors: volatility forecast evaluation using high-frequency data and realized volatilities," *Econometrica*, 73(1), 279–96.

ARMSTRONG, J. S., AND R. FILDERS (1995): "On the selection of error measures for comparisons among forecasting wethods," *Journal of Forecasting*, 14, 67–71.

BAILLIE, R. T., AND T. BOLLERSLEV (1992): "Prediction in dynamic models with time dependent conditional variances," *Journal of Econometrics*, 52, 91–113.

BAKSHI, G., C. CAO, AND Z. CHEN (1997): "Empirical performance of alternative option pricing models," *Journal of Finance*, 52, 2003–49.

BARRUCI, E., AND R. RENO (2002): "On measuring volatility and the GARCH forecasting performance," *Journal of International Financial Markets, Institutions and Money*, 12, 182–200.

BLAIR, B. J., S.-H. POON, AND S. J. TAYLOR (2001): "Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns," *Journal of Econometrics*, 105, 5–26.

BRAILSFORD, T. J., AND R. W. FAFF (1996): "An evaluation of volatility forecasting technique," *Journal of Banking and Finance*, 20, 419–38.

BROOKS, C. (1998): "Predicting stock index volatility: can market volume help," *Journal of Forecasting*, 17, 59–80.

CHRISTOFFERSEN, P., AND F. DIEBOLD (2000): "How relevant is volatility forecasting for financial risk management," *The Review of Economics and Statistics*, 82(1), 12–22.

CORNISH, R. (2007): "A comparison of the properties of realized variance for the FTSE 100 and FTSE 250 equity indices," *Forecasting Volatility in the Financial Markets*, 3ed, Satchell, S. and Knight. J., Elsevier Butterworth-Heinemann.

DE BRUIN, P., AND P. H. FRANSES (1999): "Forecasting power-transformed time series data," *Journal of Applied Statistics*, 26(7), 807–15.

DIEBOLD, F. X., AND J. A. LOPEZ (1996): "Forecast evaluation and combination," Working paper.

DIEBOLD, F. X., AND R. S. MARIANO (1995): "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13(3), 253–65.

ENGLE, R., AND A. J. PATTON (2001): "What good is a volatility model," *Quantitative Finance*, 1(2), 237–245.

GRANGER, C. W. J. (1996): "Can we improve the perceived quality of economic forecasts?," *Journal of Applied Econometrics*, 11(5), 455–473.

GRANGER, C. W. J., AND P. NEWBOLD (1976): "Forecasting transformed series," *Journal of the Royal Statistical Society*, 38(2), 189–203.

GRANGER, C. W. J., AND M. H. PESARAN (1999): "A decision theoretic approach to forecast evaluation," Working paper.

HANSEN, P. R., AND A. LUNDE (2005): "A forecast comparison of volatility models: Does anything beat a GARCH (1,1)?," *Journal of Applied Econometrics*, 20, 873–89.

LOPEZ, J. (2001): "Evaluating the predictive accuracy of volatility models," *Journal of Forecasting*, 20(2), 87–109.

MARTENS, M. (2002): "Measuring and forecasting S&P 500 index futures volatility using high-frequency data," *Journal of Futures Markets*, 22(6), 497–518.

MCNEIL, A. J., R. FREY, AND P. EMBRECHTS (2005): "Quantitative risk management: concepts, techniques and tools," Princeton Series in Finance.

PATTON, A. (2010): "Volatility forecast comparison using imperfect volatility proxies," *Journal of Econometrics*, forthcoming.

PATTON, A., AND K. SHEPPARD (2007): "Evaluating volatility forecasts," *Handbook of Financial Time Series*, Andersen, T.G., R.A.Davis, J.P. Kreiss, and T. Mikosch, eds. Springer-Verlag.

POON, S., AND C. W. J. GRANGER (2003a): "Forecasting volatility in financial markets: A review," *Journal of Economic Literature*, 41, 478–539.

POON, S.-H., AND C. GRANGER (2003b): "Practical issues in forecasting volatility," *Financial Analysts Journal*, 61(1), 45–56.

YU, J. (2002): "Forecasting volatility in the New Zealand stock market," *Applied Financial Economics*, 12, 193–202.

Chapter 4

Applications of Independent Component Analysis in Finance: Does Time Structure Matter?

4.1 Introduction

Independent component analysis (ICA) belongs to a class of blind source separation (BSS) methods. ICA is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals.

From its introduction in early 1980s, ICA has been widely applied in the fields of signal processing, artificial networks, statistics, etc. In the late 1990s, researchers start employing this new technique in the field of finance. The independent component analysis is closely related to the well known statistical technique - principal component analysis (PCA), which has many applications in economics and finance. Both PCA and ICA assume that the observed signals are linear transformation of components. Given a set of multivariate measurement x, the PCA extracts a smaller set of uncorrelated components that would represent the original measurements as well as possible. ICA, however, reveals components with strong property - statistical independence. To obtain uncorrelated components, PCA requires only the information on second order statistics, while in ICA higher order statistics are used to achieve independence. For Gaussian variables, if they are uncorrelated, then they are also independent. Therefore if the components are Gaussian variables, then PCA is capable of finding them by extracting uncorrelated Gaussian components. Because of this equivalence between uncorrelatedness and independence for Gaussian variables, it is usually assumed that the components are non-Gaussian in ICA.

It is now well acknowledged that financial series move together more or less closely over time and cross markets. Therefore, multivariate modeling framework has been the subject of extensive academic research. However, substantial computational difficulties prevent success of such modeling in practice. The ability of finding the independent components (ICs) from correlated signals may help us reveal some driving mechanisms that otherwise remain hidden. Also we know that most of the financial time series are not normally distributed, thus the underlying sources that generate these time series need not to be normally distributed. If the sources are not normally distributed, then we must use ICA instead of PCA to identify them.

The ICA model can be estimated by many approaches such as maximum likelihood, tensorial methods, and nonlinear decorrelaion. A simple and intuitive estimation principle - maximization of non-Gaussianity is described later. Non-Gaussianity can be achieved by maximizing kurtosis, entropy, or minimizing higher order cross-cumulants. Most applications of ICA in economics assume the components are randomly distributed such that the order of elements in each component does not matter, therefore use higher order statistics in estimation. However, the signals we deal with in economics are actually time series that contain more structures than just simple random variables, for example, the autocovariances of the independent components are well-defined statistics and can be used in estimating ICA model. Prior to applying ICA in financial modeling, we first examine the effect of time structure on estimating ICA model, and hope to obtain some guidance in choosing appropriate ICA estimation method for financial application.

The rest of this paper is organized as follows. In Section 4.2 we discuss the ICA model and its estimation. Section 4.3 presents the empirical results of comparing the ICA models using stock price data. In Section 4.4 we discuss the procedure of data generation and present the Monte Carlo experiment results. A brief conclusion is contained in the last section.

4.2 ICA Model

4.2.1 Introduction

ICA can be viewed as an extension to PCA and factor analysis, but is more powerful in terms of its capability of finding the independent underlying factors or sources. In connection with ICA, PCA is a useful preprocessing step.

The basic ICA model is defined as follows. We observe n random variables $x_1, x_2, ..x_n$, which are modeled as linear combinations of n random source variables $s_1, s_2, ..., s_n$:

$$x_i = a_{i,1}s_1 + a_{i,2}s_2 + ... + a_{i,n}s_n \text{ or } x = As$$

where $a_{i,j}$, $i, j = 1, ..., n$ and s_i are some real coefficients and the independent components respectively. Here we focus on linear ICA models only, but there are many non-linear ICA models which may be of interest in future research. The basic ICA model is identifiable under two restrictions: the independent components are statistically independent and they must have non-Gaussian distribution. Usually the mixing matrix A is assumed to be square for simplicity.

A simple application of ICA is the "cocktail party problem", where the underlying speech signals are separated from a sample data consisting of people talking simultaneously in a room. Each underlying speech signal can be seen as independent of other speech signals since each person has a distinctive voice. Therefore, if we record N observations of mixed sound waves from N microphones in the room, where at most N people are talking in the same time. Usually the problem is simplified by assuming no time delays or echoes. Then ICA can extract all distinctive voice signals from the observed mixtures.

In finance we usually observe multivariate time series with correlation, e.g. the price of stocks in financial market. The price of each stock follows a distinctive price path, at the same time it is correlated with other stocks or market indexes in some degree. There is a large number of stocks in any financial market, however we believe the price movement of any stock is mainly affected by a small number of factors that represent the underlying market conditions. If the underlying sources are independent, ICA is certainly a good tool that can help us to find them.

4.2.2 Estimation of ICA Model

The method of estimating ICA model described below is based on maximization of non-Gaussianity. This principle is very intuitive in ICA estimation. The fundamental idea of the method is motivated by the central limit theorem. It states that under certain conditions the distribution of a sum of independent random variables converges to a Gaussian distribution. Therefore if the central limit theorem applies, even a sum of two independent random variables has a distribution that is closer to Gaussian than any of of the two original random variables.

Assume the observed signals x are distributed according to ICA model:$x = As$, the invertibility of A ensures:

$$x = As \Longrightarrow s = A^{-1}x$$

Thus to estimate one of the ICs, consider a linear combination of the x_is, denoted by $y = b^T x = \sum_i b_i x_i$, where b is a vector to be estimated. Now we have y as a certain linear combination of the s_i's:

$$y = b^T x = b^T As = q^T s = \sum_i q_i s_i$$

Therefore we vary the coefficients in q and see how the distribution of $y = q^T s$ changes. If the central limit theorem applies, a sum of even two independent random variables is more Gaussian than any of the original ones, thus $y = q^T s$ is usually more Gaussian than any of the s_is and becomes least Gaussian when it in fact equals one of the s_i. Because $q^T s = b^T x$, we can obtain one estimated IC by maximizing the non-Gaussianity of $b^T x$. The problem turns out to be a maximization problem.

Even though the above ICA model can be estimated easily, there always exist two ambiguities or indeterminacies. First, we cannot determine the variances of the independent components because the observed signals remain unchanged if we multiply any source s_i by a scalar and divide the corresponding column a_i of A by the same scalar, say α_i: $x = \sum_i (a_i, 1/\alpha_i)(s_i\alpha_i)$. We may restrict each independent component has unit variance, however it still leaves the ambiguity of sign. Second, we cannot determine the order of the independent components since change of the order in the sum does not affect the output.

The basic ICA model that is described above assumes a linear mixture of independent random variables. However in economics or finance, we often deal with a mixture of time series. If the independent components are time series, they contain more structures than simple random variables. For instance, the autocovariances (covariances over different time lags) of the ICs are then well defined statistics, and therefore can be used to estimate ICA model.

4.2.3 Application of ICA

Back and Weigend(1997) performed the earliest financial application of ICA in finance. They use the joint approximate diagonalization of eigenmatrices (JADE) algorithm which assumes the components are random variables and uses the fourth order cumulants to estimate ICA model. Since then, more attentions are drawn to this new technique. Kiviluoto and Oja (1998) apply ICA to the cashflow in 40 stores that belong to the same retail chain and identify the fundamental factors that affect the cashflow. Chin, Weigend, and Zimmermann (2000) identify the distribution of portfolio returns by approximating the distributions of independent components, then derive analytic solutions to three risk measures: VaR,SVaR,and LPM. Subsequent work of ICA application has extended to various topics like factor analysis, forecasting, and risk management. However, there are still many potential areas that ICA can play a role and we hope to explore the possibility in improving economic modeling by this powerful new tool.

4.3 Empirical Study of the Effect of Time Structure on Estimation of ICA model

Most of the applications of ICA in finance used the basic ICA algorithm, which assumes the ICs are independent random variables even when the time series data are analyzed. The ignorance of time structure may affect the results of estimation of ICA model, however we don't know the direction and magnitude of the effect. Also since the underlying time-dependence structure of the sources is unknown, and different assumptions on time-dependence structure result in the use of different ICA methods, there exists a question as to which method should be used in what situation. Here we apply two ICA algorithms to a sample of Japanese stock returns, one of them assumes the basic ICA model, while another considers the ICs as time signals. We want to know how sensitive is the estimation in response to the consideration of time structure.

4.3.1 Data

To answer the question how does time structure affect estimating ICA model, we conduct an empirical study on time series data. The ICA model is applied to a portfolio of stocks from the Tokyo Stock Exchange (TSE). The portfolio consists of 27 largest firms in the TSE from August 1986 to October 1989. We choose this data set to compare with the study conducted in Back and Weigend (1997). The continuous compounded returns of stocks are defined by the difference of log prices:

$$x(t) = log(p_t) - log(p_{t-1}).$$

Table 4.1 in Appendix contains the basic descriptions of return data. Figure 4.1 shows the stock prices of the largest company in the TSE, the Bank of Tokyo-Mitsubishi. Figure 4.2-4.6 shows the prices of the largest 6 stocks in TSE.

4.3.2 Algorithm

JadeR Algorithm

The JadeR algorithm developed by Cardoso (1993) is designed for the basic ICA model, which assumes that the ICs are independent random variables with non-Gaussian distribution. Compared to the other basic ICA algorithms, JadeR is an efficient version of the standard two-stage procedure approach. The first stage is performed by computing the sample covariance matrix, giving the second order statistics of the observed data. From this, a matrix is computed by eigen-decomposition which whitens the data. The second stage consists of finding a rotation matrix which jointly diagonalizes eigen-matrices formed from the fourth order cumulants of the whitened data. It scales the ICs to unit variance and extracts the ICs in an order that “the most energetically significant” components appear first.

Statistical independence is achieved and the transformation matrix is found by minimizing the sum of squared fourth order cross-cumulants of ICs. For the zero-mean variables x_i, x_j, x_k, x_l the fourth order cross-cumulant is given by:

$$E(x_i x_j x_k x_l) - E(x_i x_j)E(x_k x_l) - E(x_i x_k)E(x_j x_l) - E(x_i x_l)E(x_k x_j) \quad \forall i, j, k, l \neq i, i, i, i$$

The statistical independence is achieved to the degree of the minimization of the fourth order cross-cumulants (reducing them to the values close to zero).

ThinICA Algorithm

The ThinICA algorithm developed by Cruces and Cichocki (2003) is made for a time series ICA model, it is considered an extension of the AMUSE (Tong, 1991) and SOBI (Belouchrani et al. 1997) algorithms. The algorithm is based on the criteria that jointly perform maximization of several cumulants of the output and/or the second order time delay covariance matrices. The employed contrast function combines the robustness of the joint approximate diagonalization techniques with the flexibility of the methods for blind signal extraction.

Instead of the higher order information, for time series data the information in a time-lagged covariance matrix can be used in estimation of ICA model. For any independent time series random variable y_i, y_j, in addition to the zero instantaneous covariances, the lagged covariances are all zero as well:

$$E[y_i(t)y_j(t-\tau)] = 0, \quad \forall i, j, \tau$$

4.3.3 Estimation Comparison of ICA models

Structure of ICs

Assume the returns are generated by ICA model:$x(t) = As(t)$, we then estimate ICA model according to different assumptions on ICs: random variables or having time structure, and obtain the estimated ICs using demixing matrix W: $y(t) = Wx(t)$.

Figure 4.7-4.8 plots the return series of the largest 6 stocks in TSE. While Figure 4.9 and Figure 4.10 plot the first 6 independent components extracted by the basic ICA model and the time series ICA model respectively.

The dominant ICs are defined as follows to demonstrate the contributions of the ICs to any given stocks. Any give stock return x_i is just the weighted sum of the independent components, where the weight vector is the transpose of the corresponding row of the mixing matrix. The weighted ICs are thus obtained by multiplying the corresponding row of the mixing matrix with the ICs. Those ICs with the largest maximum signal amplitudes are dominant ICs. Figure 4.11 and Figure 4.12 provide the weighted ICs obtained from the basic ICA model and the time series ICA model for the Bank of Tokyo-Mitsubishi.

To compare estimation results from two ICA models, we reconstruct the returns and therefore the prices by estimated ICs (where mixing matrix A is obtained by the inverse of demixing matrix W). For simplicity we assume there exists the same number of sources as that of observations, so the mixing matrix is always square and invertible.

$$\widehat{x_i}(t) = Ay(t) = \sum_{k=1}^{n} a_{i,k} y_k(t)$$

Then we define the weighted ICs for ith stock as:

$$\overline{y_k}(t) = a_{i,k} y_k(t) \; k = 1, ..., n$$

Since $\widehat{x_i}(t) = \sum_{k=1}^{n} \overline{y_k}(t)$, we sort the weighted ICs by their $L\infty$ norms to demonstrate their contributions to the maximum level change in a given stock. It is shown from Figure 4.13 and Figure 4.14 that most stocks are well constructed by only few dominant weighted ICs.

Estimation Results

We compare two models using the root mean square errors between reconstructed prices and the original prices:

$$RMSE = \sqrt{1/NT \sum_{i=1}^{N} \sum_{t=1}^{T} (p_{i,t} - \widehat{p_{i,t}})^2}$$

We compute RMSE by using 4 and 8 dominant ICs for both basic and time ICA models in Table 4.2. The results suggest some improvement by using time structure assumption on ICA estimation.

4.4 Simulation Study on the Effect of Time Structure on Estimation of ICA model

In order to capture the effect of time structure on estimating ICA model, we simulate return data according to ICA model where the underlying sources are generated to contain time structure:

$$x(t) = As(t)$$

The mixing matrix A that we used here is the estimated mixing matrix (the inverse of the demixing matrix) from previous empirical work. While the sources $s(t)$ are generated to contain some time structures, like AR process. Then ICA models are estimated according to different assumptions on the sources: random variables or time signals, and the estimation results are compared. There are 3 different structures of sources considered in the simulation study: AR(1), MA(1), and GARCH(1,1). All the results are based on 1000 replications.

4.4.1 The Sources Are Generated by AR(1) Process

In the first experiment , we consider the case where the sources follow AR(1) process with normally distributed errors:

$$s_{i,t} = \delta_i s_{i,t-1} + \epsilon_{i,t}, \text{ for } i = 1, .., N \text{ and } t = 1, .., T$$

We choose $N = 20$ and $T = 1000$, ϵ_i is generated independently from standard normal distribution. The initial value s_0 is also normally distributed with mean 0, the variance σ_0 is calculated from the relationship $\sigma_{0,i} = \sigma / \sqrt{(1 - \delta_i{}^2)}$ to satisfy the stationary condition, while δ_i is chosen randomly from a uniform distribution on interval [0.1 0.9]. Use ϵ and s_0, the rest sources are generated by the recursive formula: $s_{i,t} = \delta_i s_{i,t-1} + \epsilon_{i,t}$. For the given underlying sources, we obtain the return series by ICA model: $x(t) = As(t)$. Then the simulated data are estimated by ICA model based on the different assumptions of the sources, and the reconstructed returns using all the estimated ICs and 4 dominant ICs are compared with original returns. The criterion used in comparison is the root mean square error:

$$RMSE = \sqrt{1/NT \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{i,t} - \widehat{x_{i,t}})^2}$$

4.4.2 The Sources Are Generated by MA(1) Process

In the second experiment , we consider the case where the sources follow MA(1) process with student t-distributed errors:

$$s_{i,t} = \epsilon_{i,t} + \delta_i \epsilon_{i,t-1}, \text{ for } i = 1, .., N, \text{ and } t = 1, .., T$$

We choose $N = 20$ and $T = 1000$, ϵ_i is generated independently from student t-distribution. As before δ_i is chosen randomly from a uniform distribution on interval [0.1 0.9]. Use ϵ, the rest sources are generated by the recursive formula: $s_{i,t} = \epsilon_{i,t} + \delta_i \epsilon_{i,t-1}$. Then we generate returns by ICA model and estimation results are presented for different degrees of freedom of t-distribution.

4.4.3 The Sources Are Generated by GARCH(1,1) Process

In the third experiment , we consider the case where the sources follow GARCH(1,1) process:

$$s_{i,t} = \sigma_{i,t} z_{i,t}$$

$$\sigma_{i,t}^2 = \kappa_i + \alpha_i \sigma_{i,t-1}^2 + \beta_i {s_{i,t-1}}^2 \text{ for } i = 1, .., N \text{ and } t = 1, .., T$$

We choose $N = 20$ and $T = 1000$, z_i is generated independently from standard normal distribution. The GARCH parameters κ,α, and β are estimated from ICs obtained from previous empirical work.

4.4.4 Simulation Results

The simulation results somehow support our conjecture that time structure of independent components will affect the estimation of ICA models. In the first case reported in Table 4.3 when the sources are generated by AR Processes, the estimation improved when the autocorrelation is considered, however the improvement is not substantial. The basic model with random ICs performs reasonably well.

While for the MA sources, except the case where the errors in the MA process are t-distributed with degree of freedom 5, in the rest of cases the autocorrelation of sources improve the estimation of ICA model. The results are presented in Table 4.4.

In the GARCH case, showed in Table 4.5, however, if the sources are generated by a GARCH process, then we simply assume that the sources themselves are not autocorrelated, but squared sources experience high autocorrelations. Therefore if we use the autocorrelation in the sources to estimated ICA model, we won't be able to improve the estimation.

The results show that if the sources contain time dependency structure, then we must take into account of it when estimating ICA model. The time structure we assumed on sources is the autocorrelation. If in fact the time structure of sources are not from sources themselves, but the higher orders of sources such as described by a GARCH model, then autocorrelation assumption will not help much in estimating ICA model. There are reasons to search for estimation methods that deal with more complicated structure than just simple autocorrelation, maybe the method copes with non-stationary variance could probably capture more characteristics of financial time series in real life.

4.5 Conclusion and Extension

The independent component analysis is a statistical method that reveals the underlying hidden components from a set of multivariate signals. It is widely used in many field such as signal processing, artificial neural networks, statistics, etc. From the late 1990s, ICA has been applied to problems in economics and finance. Since there are many situations in which parallel economic and financial time series are being analyzed, ICA might help us to reveal some driving mechanisms that otherwise remain hidden. In most of the ICA applications in finance, the independent components are assumed to be random variables and the time structure is usually ignored. In this paper we examine the effect of time structure on the estimation performance of ICA models.

We employ 27 stocks in Tokyo Stock Exchange and estimate the basic ICA model and time series ICA model. The empirical study shows that very few of the latent components can be used to reconstruct the original price movement well. The results also support the use of time series ICA model. However, a firmer conclusion should be examined through more complicated data generating process.

In finance we often deal with correlated multivariate time series and we are very keen to know the underlying factors that generate these data. ICA, as a statistical method to extracting the underlying latent factors, has a great potential in financial time series analysis.

4.6 Appendix

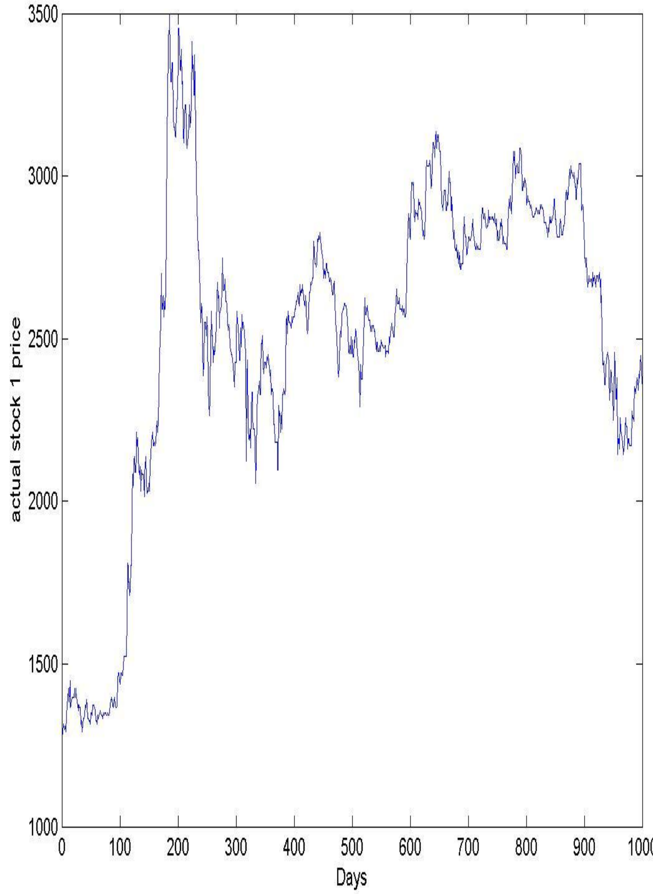

Figure 4.1: The price series of the Bank of Tokyo-Mitsubishi

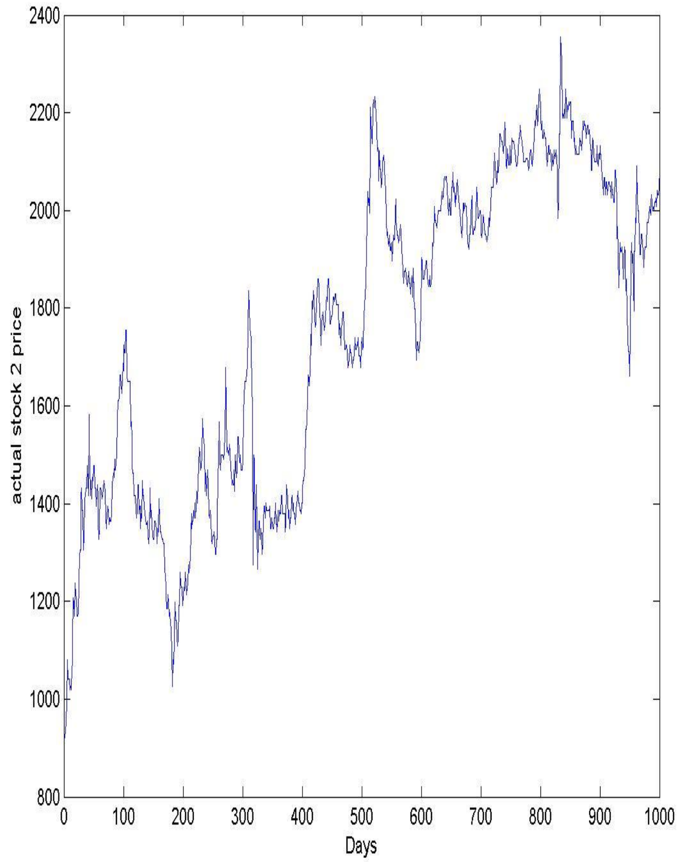

Figure 4.2: The price series of the Bank of Toyota Motor

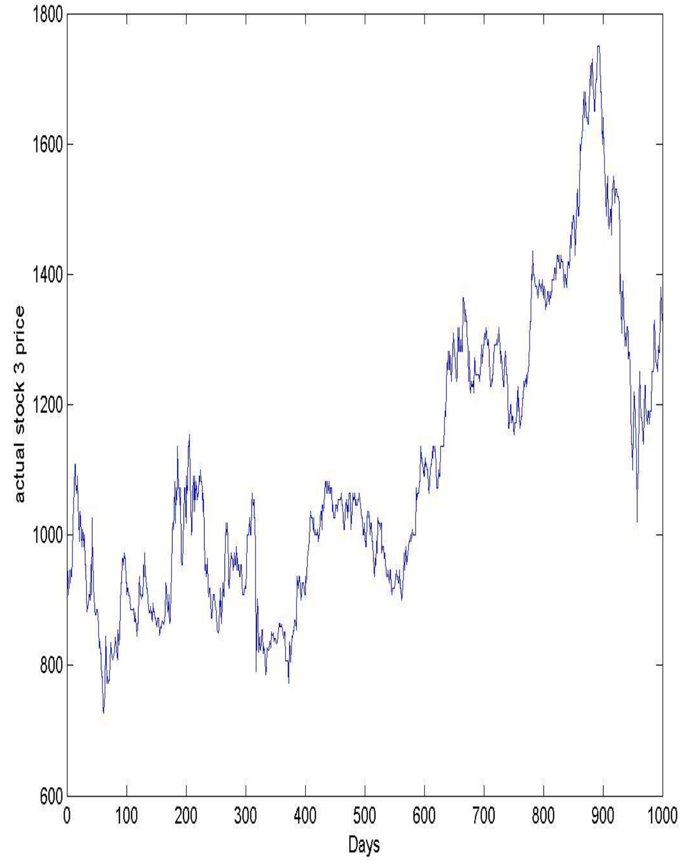

Figure 4.3: The price series of the Bank of Sumitomo Bank

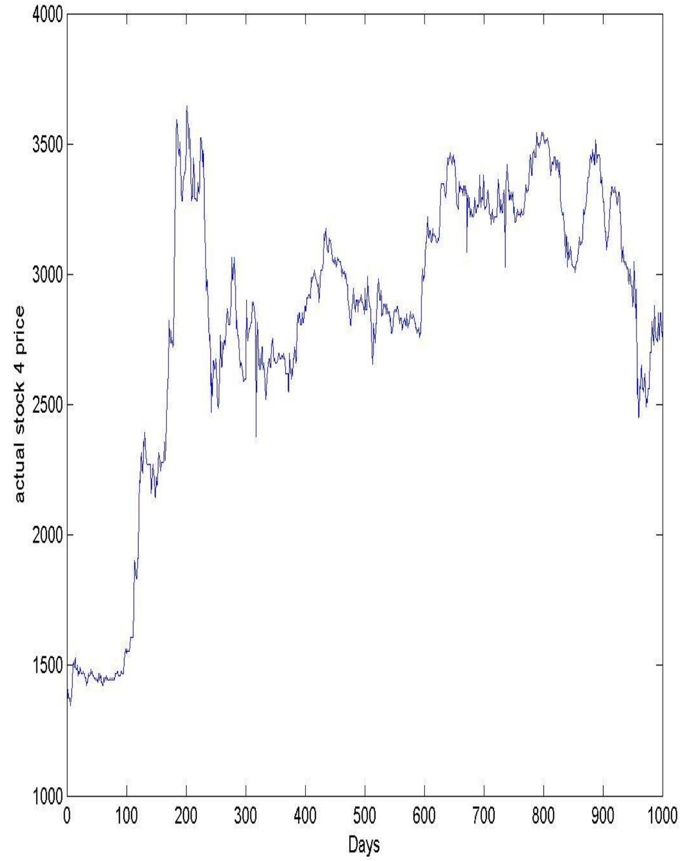

Figure 4.4: The price series of the Bank of Fuji Bank

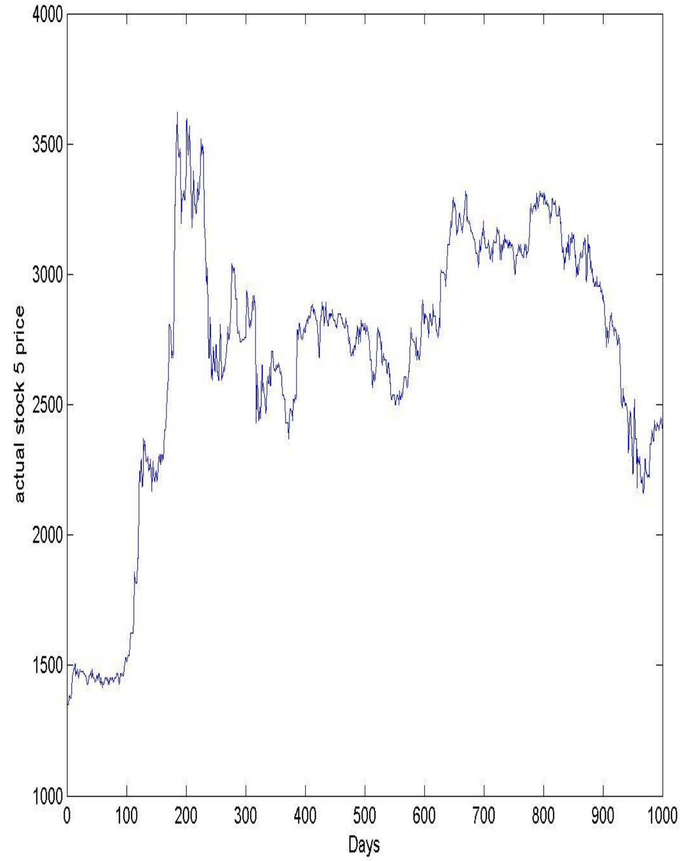

Figure 4.5: The price series of the Bank of Dai-Ichi Kangyo

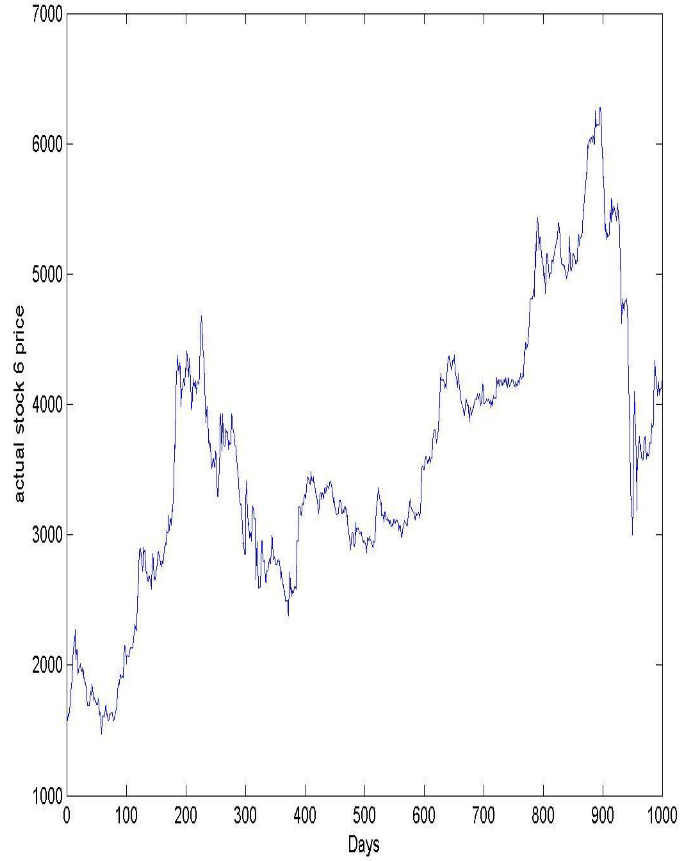

Figure 4.6: The price series of the Industrial Bank of Japan

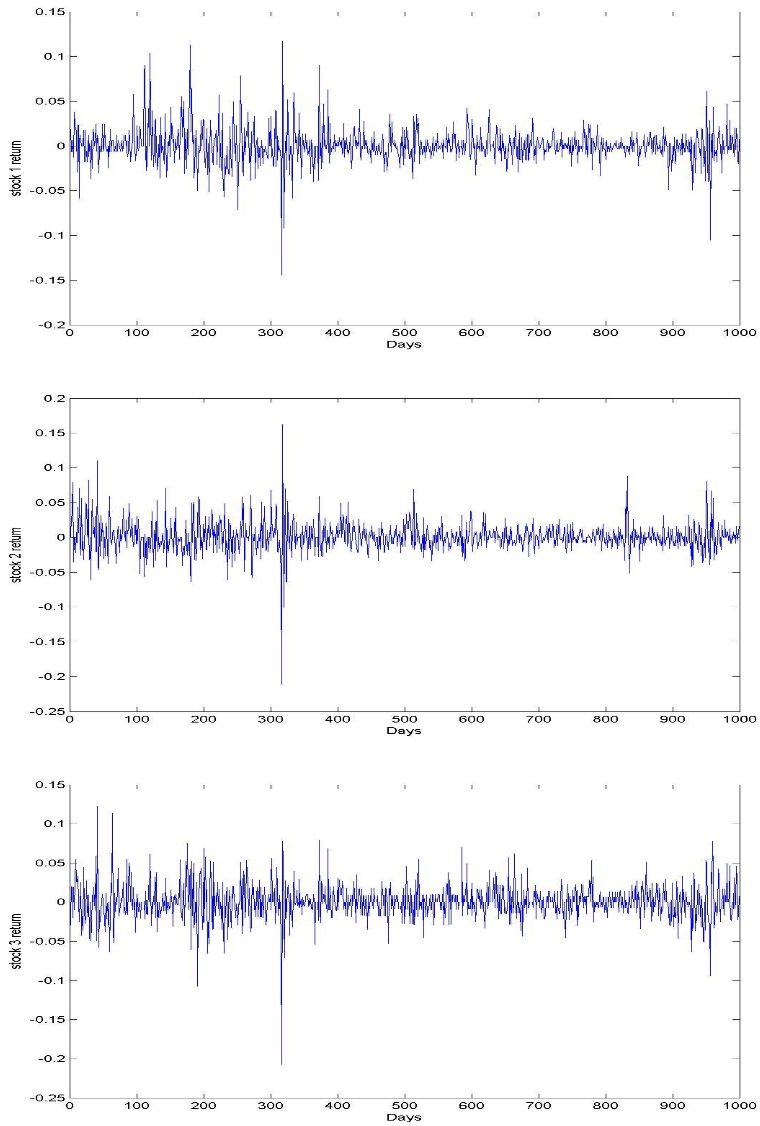

Figure 4.7: The return series of largest 3 stocks in the TSE

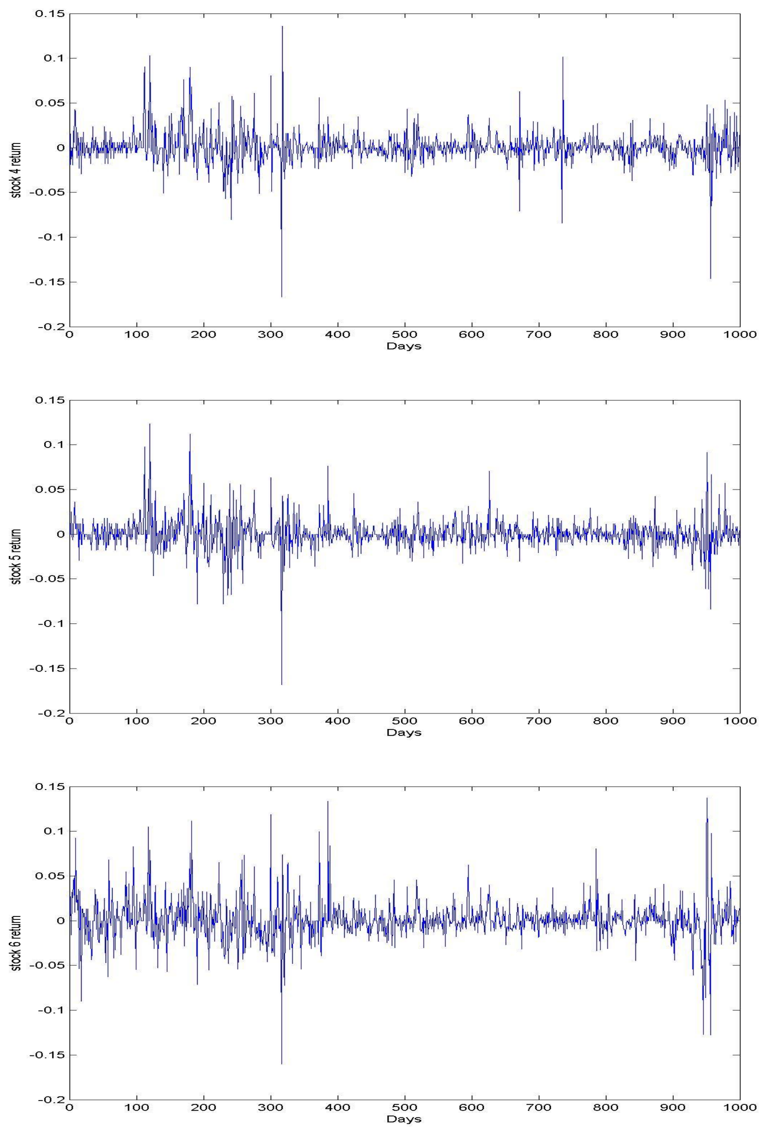

Figure 4.8: The return series of the 4^{th}, 5^{th}, and 6^{th} largest stocks in the TSE

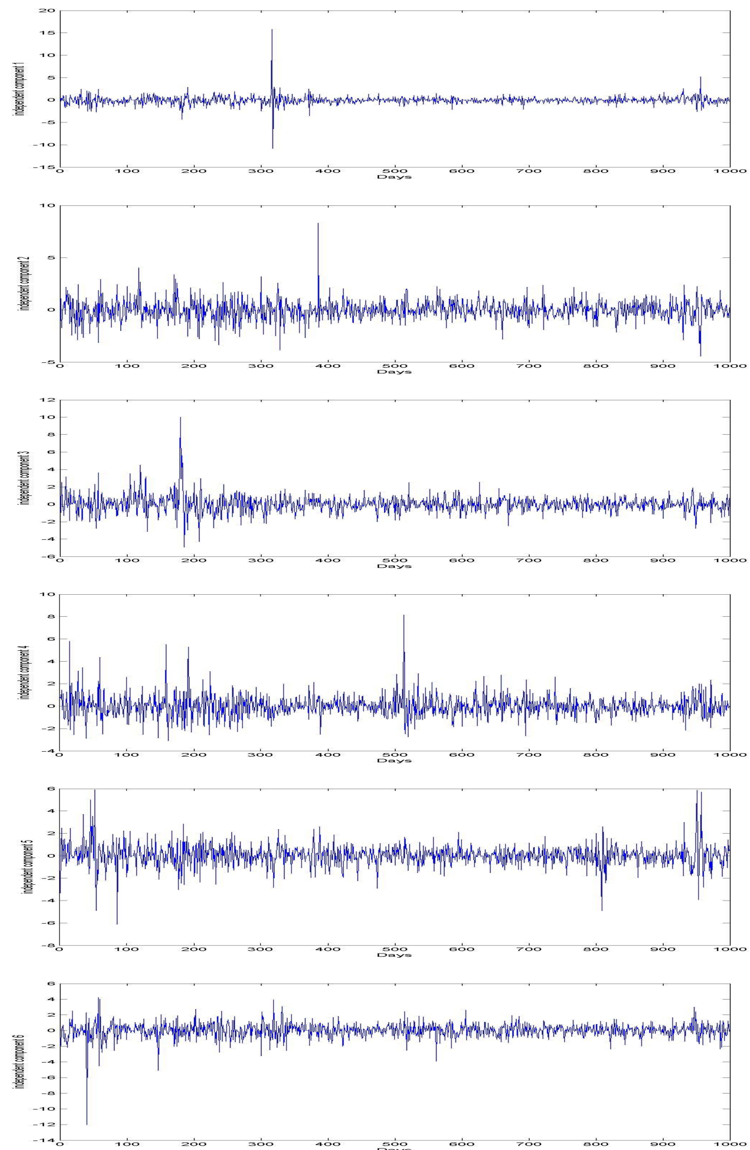

Figure 4.9: The first 6 independent components estimated by the basic ICA model

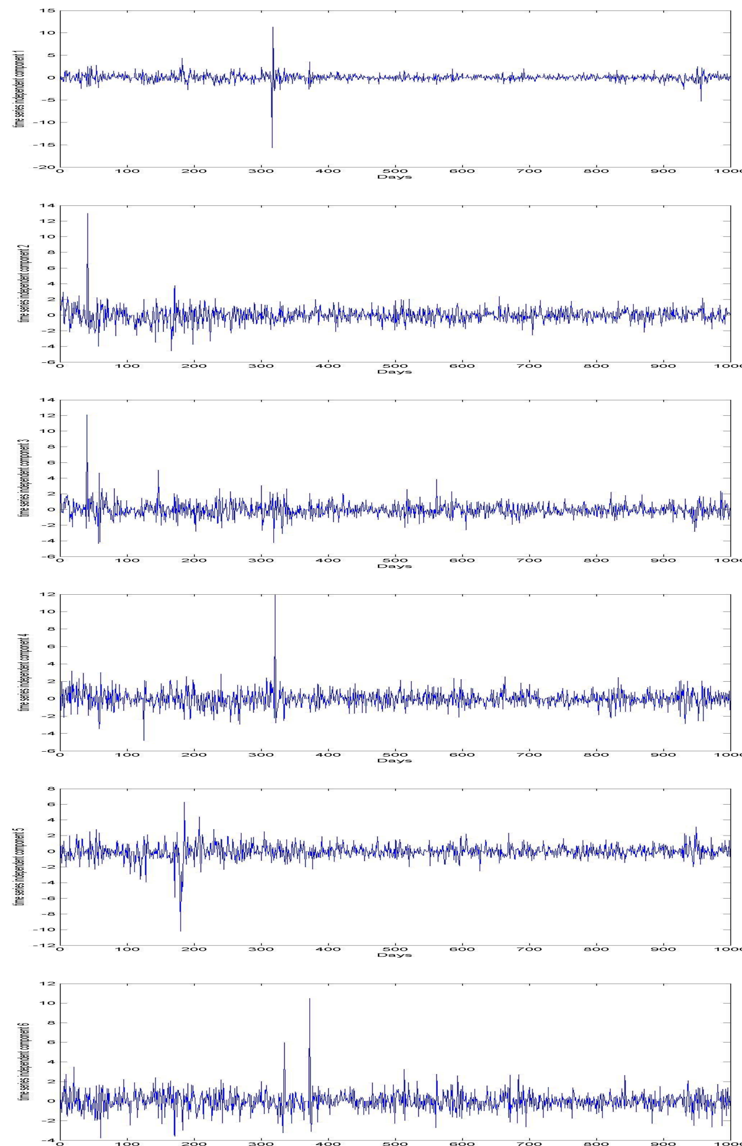

Figure 4.10: The first 6 independent components estimated by the time series ICA model

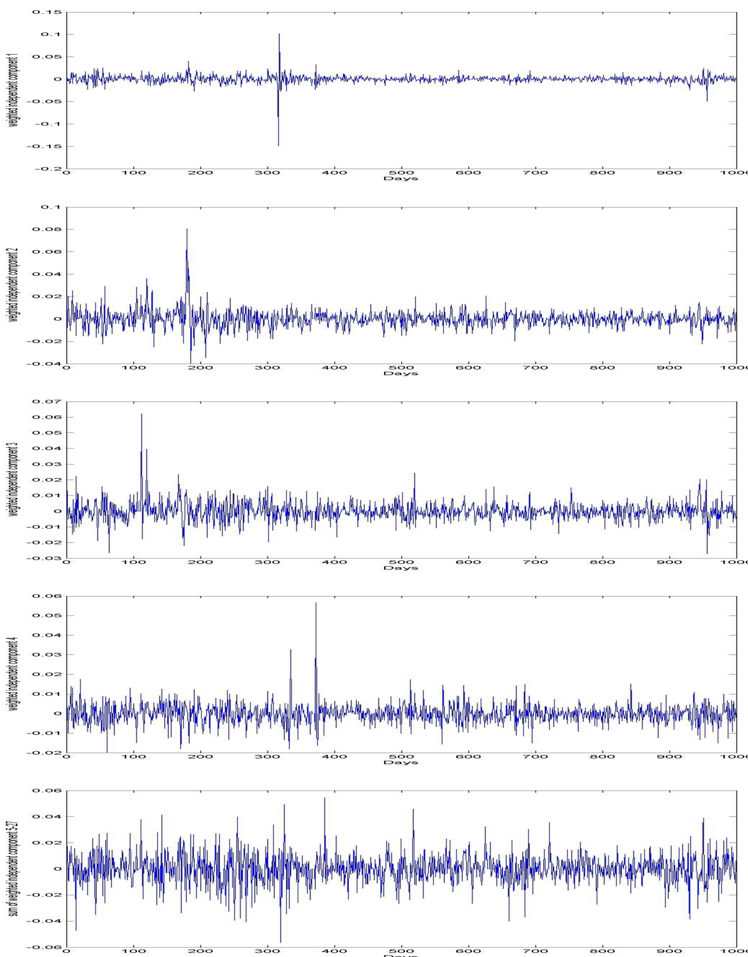

Figure 4.11: The 4 most dominant ICs and the sum of the remaining 23 least dominant ICs for the returns of the Bank of Tokyo-Mitsubishi estimated by the basic ICA model

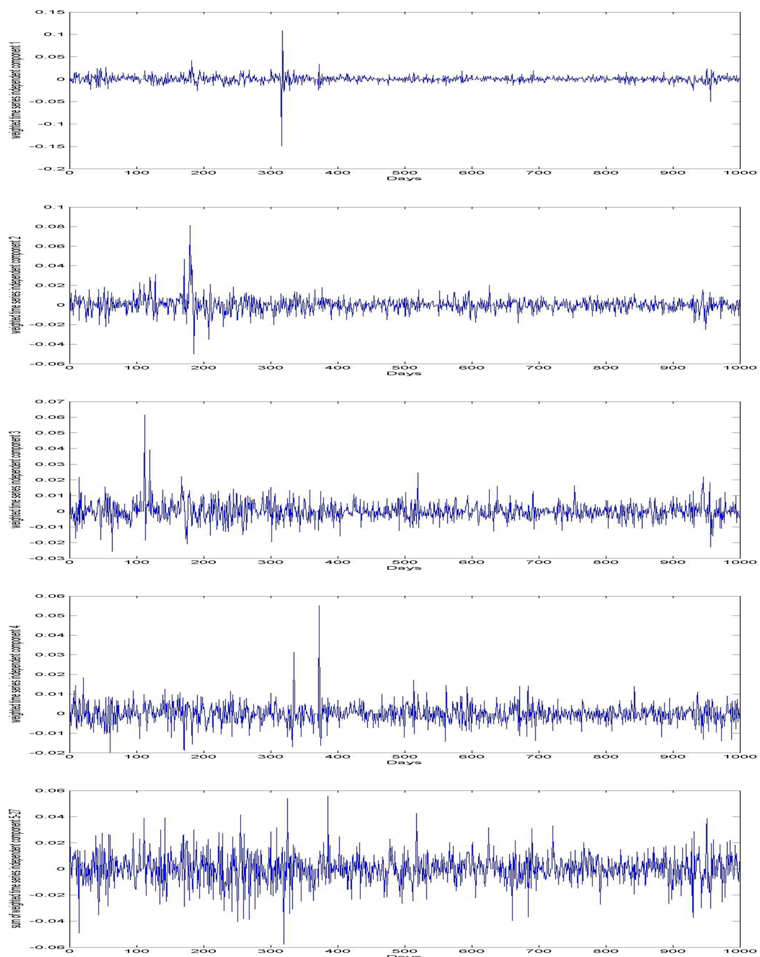

Figure 4.12: The 4 most dominant ICs and the sum of the remaining 23 least dominant ICs for the returns of the Bank of Tokyo-Mitsubishi estimated by the time series ICA model

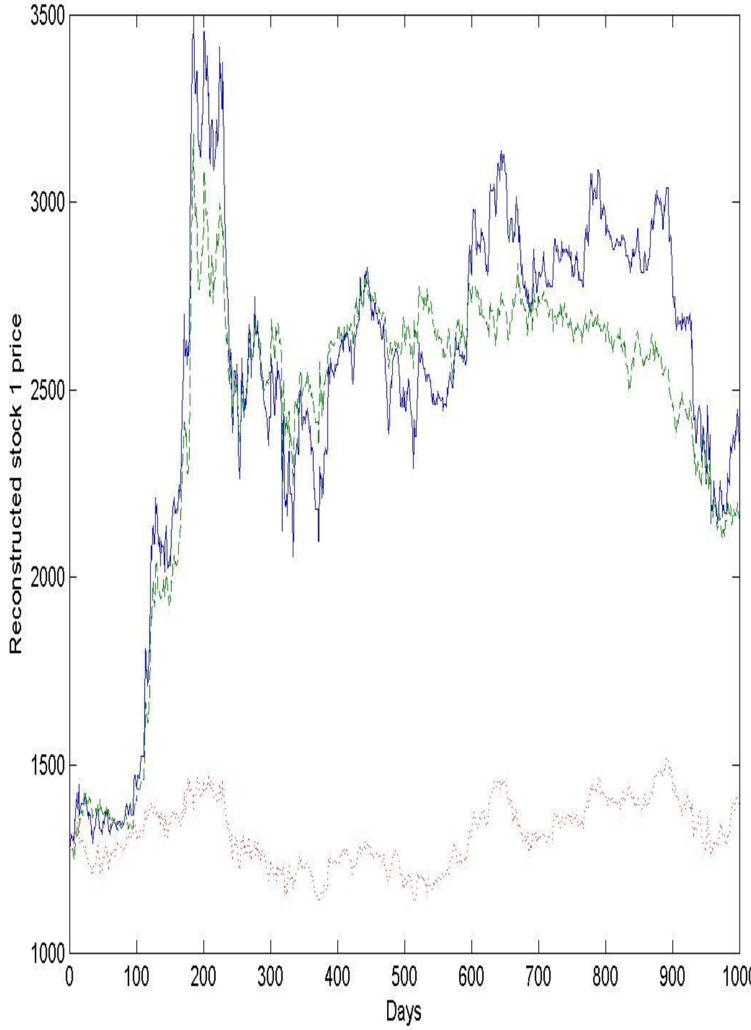

Figure 4.13: The reconstructed price series of the Bank of Tokyo-Mitsubishi using the basic ICA model. The solid line on the top is the original stock price. The dashed line in the middle is the reconstructed price using the four most dominant weighted ICs. The dotted line on the bottom is the reconstructed residual stock price using the sum of the remaining 23 weighted ICs.

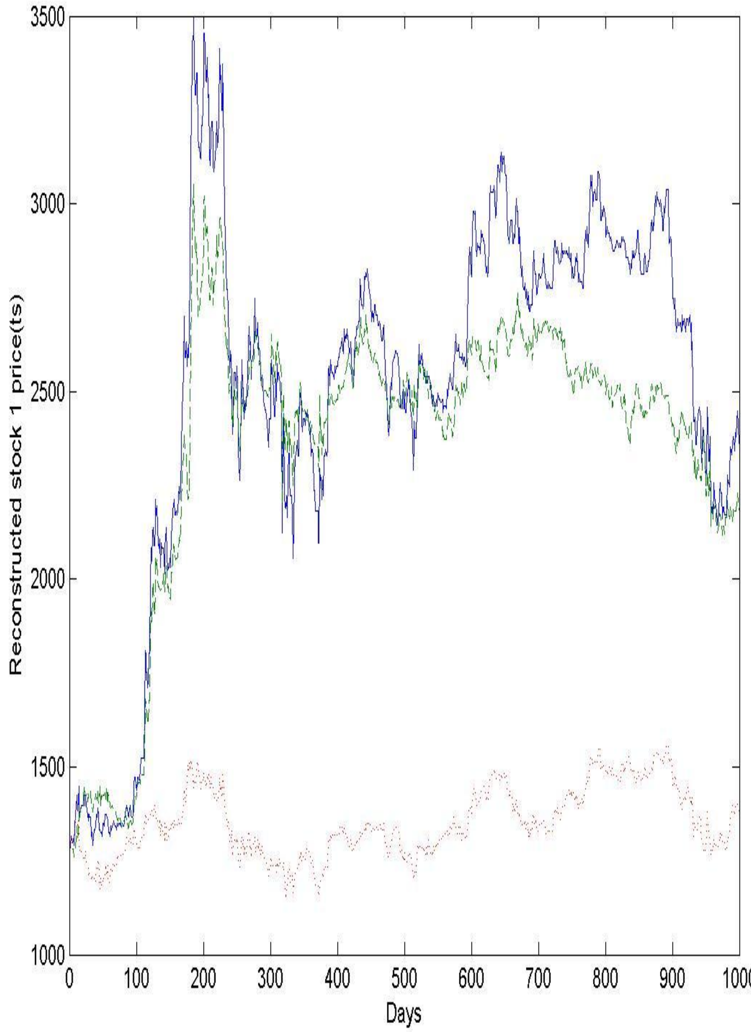

Figure 4.14: The reconstructed price series of the Bank of Tokyo-Mitsubishi using the time series ICA model. The solid line on the top is the original stock price. The dashed line in the middle is the reconstructed price using the four most dominant weighted ICs. The dotted line on the bottom is the reconstructed residual stock price using the sum of the remaining 23 weighted ICs.

Table 4.1: The Sample Statistics of TSE Return Data

TSE stocks	Mean	Variance	Skewness	Kurtosis
Bank of Tokyo-Mitsubishi	.0006	.0004	.31	12.22
Toyota Motor	.0008	.0005	-.03	16.32
Sumitomo Bank	.0003	.0005	-.37	11.81
Fuji Bank	.0007	.0004	-.04	18.22
Dai-Ichi Kangyo	.0006	.0003	.06	16.7
Industrial Bank of Japan	.001	.0006	.39	10.78
Sanwa Bank	.0007	.0004	-.04	12
Matsushita Electric	.0004	.0006	.37	10.6
Sakura Bank	.0008	.0005	.37	16
Nomura Securities	0	.0005	.57	8.24
Tokyo Electric	-.0001	.0005	.29	11.71
Hitachi	.0007	.0006	.13	6.49
Mitsubishi Industries	.0008	.0006	.31	7.98
Asahi Bank	.0007	.0005	-.04	12.62
Tokai Bank	.0007	.0004	-.24	19.22
Honda Motor	.0006	.0006	.17	13.22
Sony	.0011	.0004	.73	6.94
Seibu Railway	-.0001	.0008	.41	9.14
Toshiba	.0008	.0006	.21	6.13
Ito-Yokado	.0001	.0003	-.43	10.83
Kansai Electric	.0001	.0006	.33	8.97
Nippon	.0011	.0006	.22	6.14
Mitsubishi Trust	.0002	.0006	.33	8.9
Nissan Motor	.0008	.0005	.57	8.28
Denso	.0006	.0005	-.28	12.27
Mitsubishi	.0006	.0006	-.27	8.26
Tokyo Marine	.0002	.0006	-.31	19.73

Table 4.2: Empirical results for TSE stocks

Models	RMSE(4 ICs)	RMSE(8 ICs)
Basic ICA Model	552	442
Time ICA model	437	330

Table 4.3: Results for AR sources

Model	RMSE(All ICs)	RMSE (4 ICs)
Basic ICA model	2.47e-17	0.024
Time ICA model	2.27e-17	0.020

Table 4.4: Results for MA sources

df	RMSE(All ICs)	RMSE(TS)(All ICs)	RMSE(4 ICs)	RMSE(TS)(4 ICs)
5	9.96e-18	1.0e-17	0.0102	0.0106
6	9.50e-18	9.44e-18	0.01	0.0099
7	9.26e-18	9.1e-18	0.0099	0.0096
8	9.12e-18	8.9e-18	0.0097	0.0093
9	8.97e-18	8.72e-18	0.0096	0.0091
10	8.93e-18	8.66e-18	0.0095	0.009

Table 4.5: Results for GARCH sources

Model	RMSE(All ICs)	RMSE (4 ICs)
Basic ICA model	1.40e-17	0.014
Time ICA model	1.42e-17	0.015

4.7 Bibliography

BACK, A. D., AND A. S. WEIGEND (1997): "A first application of independent component analysis to extracting structure from stock returns," *International Journal of Neural Systems*, 8, 473–84.

BELOUCHRANI, A., K. ABED-MARAIM, J.-F. CARDOSO, AND E. MOULINES (1997): "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process*, 45, 434–44.

CARDOSO, J.-F. (1998): "Blind signal separation: statistical principles," *Proc. IEEE*, 86, 2009–25.

CHA, S.-M., AND L.-W. CHAN (2000): "Applying independent component analysis to factor model in finance," *Lecture Notes in Computer Science*, 1983.

CHIN, E., A. WEIGEND, AND H. ZIMMERMANN (1999): "Computing portfolio risk using gaussian mixtures and independent component analysis," *Computational Intelligence for Financial Engineering*, New York.

CICHOCKI, A., S. R. STANSELL, Z. LEONOWICZ, AND J. BUCK (2005): "Independent variable selection: application of independent component analysis to forecasting a stock index," *Journal of Asset Management*, 2.

CRUCES, S. A., AND A. CICHOCKI (2003): "Combining blind source extraction with joint approximate diagonalization: thin algorithm for ICA," *Proc. of the Fourth Sysposium on Independent Component Analysis and Blind Source Separation*, pp. 463–69.

HYVARINEN, A., AND E. OJA (1998): "Independent component analysis for parallel financial time series," *Proc. ICONIP'98*, pp. 895–9.

LIZIERI, C., S. SATCHELL, AND Q. ZHANG (2007): “The underlying return generating factors for REIT returns: An application of independent component analysis,” *Real Estate Economics*, 35, 569–98.

LU, C.-J., T.-S. LEE, AND C.-C. CHIU (2009): “Financial time series forecasting using independent component analysis and support vector regression,” *Decision Support Systems*, 47(2), 115–25.

MOODY, J., AND H. YANG (1999): “Term structure of interactions of foreigh exchange rates,” *Computational Finance-Proceedings of the Sixth Int'l Conference, New York*.

Chapter 5

Conclusion

My thesis focuses on the financial asset return and volatility modeling.

In order to explain the regime switching in the volatility dynamics of financial asset returns, we propose a threshold GARCH model to utilize the information provided by an observable threshold variable. In the first essay we theoretically investigate the usefulness of the threshold model. We derive theoretical conditions, which ensure that the return process in the threshold model is strictly stationary, as well as conditions for the existence of various moments. A simulation study is further conducted to examine the finite sample properties of the maximum likelihood estimator. Simulation results reveal that the maximum likelihood estimator is well behaved for modest sample sizes when the stationarity conditions hold and the variance of the return series exists.

The second essay investigates the estimation and forecasting performance of the threshold GARCH model. We find that the volatility index is useful in identifying the regime shifting in the volatility process for stock returns as well as currency exchange rates. The variable is especially helpful when there is a sudden spike in volatility during the forecasting period. We also examine the forecasting performance for different volatility proxies. The forecasting performance of our threshold model is improved significantly when the realized volatility is used as a proxy for the actual volatility. The forecasting performance of three currency exchange rates also support the use of the threshold model in volatility forecasting.

The third essay examine the effect of time structure on the estimation performance of ICA models and provide guidance in applying the ICA model to time series data. We compare the performance of the basic ICA model to the time series ICA model in which the cross-autocovariances are used as a measure to achieve independence. We conduct a simulation study to evaluate the time series ICA model under different time structure assumptions about the underlying components that generate financial time series. Moreover, the empirical study supports the use of the time series ICA model.

CURRICULUM VITAE

Name:	Jing Wu
Place of Birth:	Urumuqi, China
Education and Degrees:	The University of Western Ontario London, Canada 2004–2005 M.A.
	The University of Western Ontario London, Canada 2005–2012 Ph.D.
Honors and Awards:	Ontario Graduate Scholarship The University of Western Ontario 2006
	Western Graduate Research Scholarship The University of Western Ontario 2005, 2007, 2008
Related Work Experience:	Instructor The University of Western Ontario 2008-2010
	Instructor University of Waterloo 2010
	Research and Teaching Assistant The University of Western Ontario 2005-2009