November 2012

Volatility, Duration, and Value-at-Risk

Pujun Liu
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

VOLATILITY, DURATION, AND VALUE-AT-RISK

(Spine title: Volatility, Duration, and Value-at-Risk)
(Thesis format: Integrated Article)

by

Pujun Liu

Graduate Program in 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
OCTOBER 2012

THE UNIVERSITY OF WESTERN ONTARIO
SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES

CERTIFICATE OF EXAMINATION

Supervisor	Examiners
Dr. John Knight	Dr. Dinghai Xu
Supervisory Committee	Dr. Hao Yu
Dr. Youngki Shin	Dr. Tim Conley
Dr. Martijn Van Hasselt	Dr. Youngki Shin

The thesis by

Pujun Liu

entitled

Volatility, Duration, and Value-at-Risk

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

Date ________________________

Chair of the Thesis Examining Board

Abstract

The thesis consists of three essays dealing with the modeling of volatility in financial markets, trade durations, and Value-at-Risk (VaR). The first essay models nonlinearities in the return series to estimate time-varying volatility by incorporating both regime changes and jumps. Two types of regime-switching GARCH-jump models with autoregressive jump intensity are presented. The first model follows the traditional Markov regime-switching model proposed in Hamilton (1989). As the unknown regimes in the Markov model lead to difficulty in forecasting, a threshold GARCH-jump model, in which regimes are known after observing the threshold variable in the previous period, is also proposed. The second essay models the intraday durations between two adjacent trade transactions by considering the impact of unaccounted structural changes on parameter estimates. Monte Carlo simulations show that the observed high persistence in trade durations can be spurious and caused by unaccounted structural changes in the data generating process. The third essay investigates the use of realized moments in VaR forecasting, which is an important issue in risk management. Many VaR models rely only on the mean and volatility and ignore higher moments of returns, which leads to underestimation of VaR due to the unaccounted fat-tail property of the return series. Applying the Cornish-Fisher expansion to incorporate realized higher moments constructed from high frequency data, the proposed realized moment models outperform the realized volatility model and the traditional RiskMetrics model, especially during the financial crisis period (2008-09).

Acknowledgements

It is an absolute pleasure to thank all the people who made this thesis possible.

First and foremost, I wish to express my deepest gratitude to my supervisor, Professor John Knight, for his continuous support of my Ph.D study and research, for his patience, encouragement, enthusiasm, as well as his immense knowledge and research experience. His dedication to research and students has taught me a lot about both work and life.

I would like to thank Youngki Shin and Martijn van Hasselt for being members of my committee and giving valuable suggestions to my papers and research. I would also like to thank Dinghai Xu for providing data used in the thesis, as well as his useful comments. My sincere thanks also goes to Lance Lochner, Maria Ponomareva, Yvonne Adams, Audra Bowlus, Chris Robinson, Hiroyoki Kasahara, Tim Conley, Peter Streufert, Yufeng Wang, Jin Zhou, Chris Mitchell and Yao Li.

Finally, I would like to thank my parents for their deepest love and continuous support throughout my life.

London, Canada Pujun Liu

August 12th, 2012

To my parents

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

Introduction

This dissertation consists of three essays dealing with the modeling of volatility in financial markets, financial asset trade durations, and Value-at-Risk. The thesis is related to both regularly spaced and irregularly spaced financial data. Chapter 2 takes into consideration non-linearity issues in the return series to estimate volatility by incorporating both regime changes and jumps. Chapter 3 deals with irregularly spaced financial data by considering the impact of unaccounted structural changes on parameter estimates of intraday trade duration process. An important issue in risk management is the forecasting of market Value-at-Risk (VaR). In Chapter 4, two new VaR forecasting models are proposed. Realized higher moments are constructed to provide better VaR forecasts, taking advantage of the information conveyed in high frequency data.

In Chapter 2, *Regime-Switching GARCH-Jump Models with Autoregressive*

Jump Intensity, two types of regime-switching GARCH-jump models with autoregressive jump intensity are proposed to model the non-linearity in return series and the associated volatility. Chan and Maheu (2002) present an autoregressive jump intensity model to explain the jump clustering phenomenon. However, the forecasts of their model are inaccurate when the out-of-sample period differs from the in-sample period in the frequency of jumps. To solve this problem, regime shifts are incorporated in both the smoothly changing GARCH term and the infrequent jump term. The first model is a Markov regime-switching model which generalizes the GARCH model by distinguishing two regimes with different GARCH volatility and jump intensity levels. As the regimes are unknown to the econometrician in the Markov regime-switching model, which leads to difficulty in forecasting, a threshold GARCH-jump model with an exogenous threshold variable is also proposed. The stationarity conditions and moments of returns are derived for the threshold GARCH-jump model. Using Japanese YEN-US Dollar exchange rates, it is shown that both types of regime-switching models have better performance than the traditional GARCH model for the in-sample period. Constructing realized volatility from 5-minute intraday data for evaluation, the threshold GARCH-jump model outperforms the single regime autoregressive jump intensity model to provide volatility forecasts.

The rapid development in computer technology has led to the availability of ultra high frequency data, which arrive in irregular time intervals, making traditional econometric techniques inapplicable. To solve this problem, Engle and Russell (1998) build a linear autoregressive conditional duration (ACD) model to account for stochastic clustering of durations between two adjacent trades. In Chapter 3, *Autoregressive Conditional Duration Models with Structural Changes*, we find that high persistence of trade durations noted in the literature, i.e., the sum of estimated autoregressive coefficients on lagged durations and conditional expected durations are close to one, may come from unaccounted structural shifts in the data generating process. Monte Carlo experiments are conducted to show that even a temporary change in one parameter of the ACD model for a relatively short time period can lead to a big bias in the estimates of the autoregressive parameters, which converge to one as jump size increases. The sample mean of the conditional expected duration is derived for ACD model with structural changes. Finally, we estimate Boeing transaction duration data using a threshold ACD model and find that it fits the data better than the single-regime ACD model.

Under the Basel II and Basel III Accords, banks are required to maintain regulatory capital for market risk according to their assets' riskiness, which is defined as the $\alpha\%$ VaR, such that the loss of a specific asset within a future time period will only be surpassed for $(1-\alpha)\%$ of the time. Many VaR models rely only on the mean and volatility of the return series and ignore higher moments, which often leads to underestimation of VaR due to the unaccounted fat-tail property of the return series. Aiming to solve this issue, Chapter 4, *Value-at-Risk Estimation via Realized Higher Moments using High Frequency Data*, investigates the impact of realized higher moments constructed from high frequency data on VaR forecasts. Recently, Amaya et. al (2011) proved that, under realistic assumptions of an affine jump-diffusion process with stochastic volatility, the realized moments converge in mean square to the integrated moments up to the fourth moment. The well-known realized variance is a special example of realized moments, i.e., it is the realized second moment. As realized moments are *ex post* measures, two new models are proposed to provide one-step-ahead forecasts for realized moments, after exploring the characteristics of realized moments. We find that the logarithmic realized fourth moment is significantly autocorrelated and often displays long memory properties. The first model extends the exponentially weighted moving average (EWMA) procedure to realized volatility and the logarithmic realized fourth moment. The second model applies an autoregressive fractionally integrated moving average (ARFIMA) model to both the logarithmic realized volatility and logarithmic realized fourth moment according to their autoregressive and long-memory characteristics. After calculating skewness using forecasts of realized moments, we apply the Cornish Fisher approximation to incorporate the time-varying volatility and kurtosis in the VaR forecasting. In an empirical study, we compare the performance of realized moments models with other VaR models such as the Riskmetrics model widely used in the financial industry, concluding that the realized moments models provide accurate forecasts and outperform the Riskmetrics model and the realized volatility model, especially during the financial crisis period around 2008.

1.1 Bibliography

AMAYA, D., P. CHRISTOFFERSEN, K. JACOBS, AND A. VASQUEZ (2011): "Does Realized Skewness and Kurtosis Predict the Cross-section of Equity Returns?," *working paper*.

CHAN, W., AND J. M. MAHEU (2002): "Conditional jump dynamics in stock market returns," *Journal of Business & Economic Statistics*, 20(3), 377–389.

ENGLE, R., AND J. RUSSELL (1998): "Autoregressive conditional duration: a new model for irregularly spaced transaction data," *Econometrica*, 66, 1127–1162.

Chapter 2

Regime-Switching GARCH-Jump Models with Autoregressive Jump Intensity

2.1 Introduction

Estimating and forecasting volatility is an important task in financial markets. Volatility, interpreted as uncertainty of return and calculated as the standard deviation or variance of the return series, is a key variable in derivative pricing, portfolio rebalancing and risk management. As it is widely perceived that volatility of asset returns is changing over time, it is important to investigate the characteristics of the volatility process.

Jump diffusion models are a class of volatility models which have received wide-spread acceptance for their ability to model both continuous small changes and infrequent large movements in financial return series since the seminal work by Press (1967) and Merton (1976). In the discrete version of the jump diffusion models, GARCH models and stochastic volatility models are used to account for the diffusion part in return, or the smoothly changing movements that might be caused by normal news events as well as liquidity trading. Jumps refer to the infrequent large movements in return that are caused by the unusual and important news events, such as earning surprises. For the jump part, jump intensity, which refers to the arrival rate of jumps, is usually assumed to be independent, partly because of the difficulty of estimation of stochastic jump intensity models without a closed-form likelihood function. Recently, Chan and Maheu (2002) and Maheu and McCurdy (2004) model the return series as a combination of jumps and smoothly changing components, in which the conditional jump intensity is autoregressive. They find that the jump intensity is strongly rejected to be constant. The autoregressive parameter in the jump intensity is positively significant and high for individual stock returns, which supports the phenomenon of jump clustering.

While jump diffusion models present a parametric way to model abnormal or jump innovations as well as normal innovations, the harnessing of high frequency data in the last decade has also led to separate analysis of diffusive and jump components using a non-parametric approach. Daily realized volatility, constructed by Andersen and Bollerslev (1998) and Bandorff-Nielsen and Shephard (2001) using the summation of squared intraday returns, is a consistent estimator of the quadratic variation in a continuous jump-diffusion setting with a bounded jump intensity. This provides a good proxy for daily volatility after dealing with intraday pattern and microstructure noise. Moreover, Bandorff-Nielsen and Shephard (2004) present realized bipower variation constructed from high frequency data, which is consistent for the integrated variance in the same jump-diffusion setting. As the difference of quadratic variation and the integrated variance is the cumulative squared jumps, the result renders feasible statistical tests for the presence and impact of jumps. Huang and Tauchen (2005) show that a test for jumps has good power and detection capacities using Monte Carlo analysis, and indicate strong empirical evidence for jumps to account for stock market price variance.

Another line of literature deals with non-linearity using regime switching models. Lamoureux and Lastrapes (1990) show that the high persistence of the conditional variance using GARCH model may be overstated due to the failure of recognizing structural changes in the model. Gray (1996) develops a generalized regime switching GARCH model and finds that it outperforms single-regime models out-of-sample using short-term interest rate. More recently, Hillebrand (2005) show that the convergence of the sum of estimated autoregressive parameters holds for all common estimators of GARCH. Thus, in the presence of neglected parameter changes, GARCH is no longer a suitable model to measure persistence.

In this chapter we model structural breaks and jumps together by building regime switching GARCH-jump models based on Chan and Maheu (2002)'s autoregressive jump intensity (ARJI) model. The motivation is that although the ARJI model provides good in-sample estimation, the out-of-sample forecasting ability is not as good especially when the jump frequency in the out-of-sample period differs from the in-sample period. For example, when the out-of-sample period is a relatively tranquil period which contains less jumps, using parameters estimated from the relatively volatile in-sample period will overestimate the jump frequency and lead to inaccurate forecasting. In this chapter we show that the out-of-sample forecasting performance is not as good as GARCH model for Japanese Yen-US Dollar exchange rate. Thus, it is necessary to distinguish between volatile period and tranquil period for jumps. In addition, Maheu and McCurdy (2004) plot the time-series of conditional variance components of IBM estimated by their generalized autoregressive jump intensity (GARJI) model, in which both GARCH component and jump component of the conditional variance are higher than normal in some periods, while in other periods both of them are less volatile. The phenomenon suggests that both smooth changes and jumps may be governed by regime changes. Furthermore, the high persistence in conditional variance may be spurious due to latent structural changes in the data generating process. Thus, we model the GARCH volatility and jump intensity process in different regimes in order to improve the out-of-sample forecasting performance.

Two types of regime-switching GARCH-jump model are developed. The first one follows the traditional Markov regime-switching model proposed in Hamilton (1989), which has good stationarity conditions but latent regimes. The difficulty to introduce regimes into conditional jump intensity in this type of model is that the jump intensity will depend on the entire regime path from the beginning of the period to the current period as it is autoregressive, which leads to computational complexity. To circumvent this problem, the jump intensity is assumed to depend only on its current regime state. However, since the regimes are unknown this results in poor forecasting in the Markov regime-switching models. Consequently, we also consider a threshold GARCH-jump model, in which regimes are known after the observation of the threshold variable at the previous period. Recently, Knight and Satchell (2011) derive sufficient and necessary conditions for the existence of a stationary distribution for a threshold AR (1) model with exogenous threshold variable. We extend their research and find stationarity conditions for the threshold GARCH-jump model with regimes in both GARCH type conditional variance and jump intensity.

The chapter is organized as follows. Section 2.2 presents the Markov regime-switching GARCH-jump model and proposes an estimation mechanism after a brief review of the ARJI model by Chan and Maheu (2002). In section 2.3, the threshold GARCH-jump model with an exogenous trigger is developed and stationarity conditions are derived. Empirical analysis is conducted in sections 2.4 and 2.5. The Japanese Yen-US Dollar spot exchange rate series are used for the estimation, in which realized volatility constructed from 5-minute intraday data are used as proxy of volatility for evaluation of forecasts of different models. Section 2.6 contains a brief conclusion.

2.2 Markov Regime-Switching GARCH-Jump (RS-GARJI) Model

Regimes are incorporated into both GARCH variance and jump intensity. For a Markov regime-switching GARCH (1,1) with jump intensity as an AR (1) process, which is denoted by RSGARJI model, the model is given by

$$R_t = \mu + \epsilon_{1,t} + \sum_{k=1}^{N(t)} Y_{t,k}$$

$$\epsilon_{1,t} = z_t \sigma_t$$

$$Y_{t,k} \sim i.i.d \quad N(\theta, \delta^2)$$

$$P(N(t) = j|\Phi_{t-1}) = exp(-\lambda_t)\lambda_t^j / j! \qquad \text{for j= 0, 1, 2...}$$

In regime s_t , for $s_t = 1, 2$,

$$\sigma_t^2 = \omega_{s_t} + a_{s_t}\epsilon_{t-1}^2 + b_{s_t}E[\sigma_{t-1}^2|\Phi_{t-1}, s_t] \qquad (2.1)$$

$$\lambda_t = \alpha_{s_t} + \rho_{s_t} E[\lambda_{t-1}|\Phi_{t-1}, s_t] + \gamma_{s_t} E[\xi_{t-1}|\Phi_{t-1}, s_t] \tag{2.2}$$

$$\xi_{t-1} = E[N(t-1)|\Phi_{t-1}, S_{t-1}] - \lambda_{t-1}$$

R_t denotes the return at time t, which is the first difference of logarithmic price. $\epsilon_{1,t}$ is a GARCH component with an autoregressive conditional variance σ_t^2, and $\sum_{k=1}^{N(t)} Y_{t,k}$ is the jump innovation which is a compound Poisson process. As in the GARCH model, z_t follows the normal distribution with mean 0 and variance 1. The GARCH component can explain the continuous small changes in the return series. $N(t)$ is the number of jumps happening at time t, which follows a Poisson process with autoregressive jump intensity λ_t. Jumps happen occasionally. If $N(t) = 0$, there is no jumps at time t. So, the jump innovation, can explain for the infrequent large movements in the return series. s_t denotes the regime at time t, which can take value of 1 or 2 referring to two different regimes. S_t is the entire regime path $\{s_t, s_{t-1}, ...\}$. Φ_{t-1} denotes the information set until time $t-1$. Smoothly changing components are represented by $\epsilon_{1,t}$, which follows a GARCH process with different parameters in different regimes, corresponding to $(\omega_{s_t}, a_{s_t}, b_{s_t})$ in regime s_t. The jump intensity λ_t follows an approximate AR(1) process in each regime, with parameters $(\alpha_{s_t}, \rho_{s_t}, \gamma_{s_t})$ in regime s_t. We use the approximate AR(1) process introduced by Chan and Maheu (2002) to model the jump intensity, as it can circumvent the problem that the likelihood function has no closed form when the jump intensity follows an ARMA process. ξ_{t-1} can be viewed as an approximate error term. It will be discussed later.

When regimes are assumed to be exogenous, i.e, explanatory variables in the conditional intensity process σ_t contain no information about s_t beyond that contained in Φ_{t-1}, s_t follows a first-order Markov process as in Hamilton (1989).

$$P(s_t = j|S_{t-1}) = P(s_t = j|s_{t-1} = i) = p_{ij} \tag{2.3}$$

The transition matrix is

$$P = \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{bmatrix}$$

p_{11} and p_{22} are respectively the persistence of regime 1 and regime 2.

This is the model setting when the regimes follows a Markov process. In the following subsections, we discuss why the model is built this way. First we present the autoregressive jump intensity (ARJI) model of Chan and Maheu (2002) and discuss its properties. Then we elaborate on the way of constructing the regime-switching model in both conditional variance and jump intensity.

2.2.1 ARJI model by Chan and Maheu (2002)

Previous literature suggests that jump intensity is time-varying and may depend on its lagged values. For example, Knight and Satchell (1998) model a self-exciting jump intensity process which depends on past volatility and a stochastic deviation from fundamentals. By substitution the jump intensity can be expressed as an autoregressive form together with non-negative error term. Chan and Maheu (2002) and Maheu and McCurdy (2004) generate an autoregressive conditional jump intensity (ARJI) model and derive conditional moments of the returns. In applications to several individual firms, the persistence parameter for the arrival of jump events is quite high, up to 0.924 for Texaco. Their model is a single-regime discrete-time GARCH-jump model with time dependent jump intensity with the following specification.

$$\lambda_t = \alpha + \rho\lambda_{t-1} + \gamma\xi_{t-1}$$

$$\xi_{t-1} = E[N(t-1)|\Phi_{t-1}] - \lambda_{t-1}$$

$E[N(t-1)|\Phi_{t-1}]$ is the *ex post* assessment of the expected number of jumps given information set Φ_{t-1}, while λ_{t-1} is the *ex ante* assessment. So ξ_t can be viewed as the change in the conditional forecast of the jumps after the infomation set is updated. It is a martingale difference sequence with respect to $\{\Phi_{t-1}\}$, as $E[\xi_{t-1}|\Phi_{t-2}] = E[E[N(t-1)|\Phi_{t-1}] - E[N(t-1)|\Phi_{t-2}]|\Phi_{t-2}] = E[N(t-1)|\Phi_{t-2}] - E[N(t-1)|\Phi_{t-2}] = 0$, which implies that there is no autocorrelation in the intensity residual and the unconditional expectation is zero. Thus it can be viewed as an error term in the jump intensity process. If $|\rho| < 1$, then the jump intensity is covariance stationary. The conditional mean of return is $E(R_t|\Phi_{t-1}) = \mu + \theta\lambda_t$, and the conditional variance is $Var(R_t|\Phi_{t-1}) =$

$\sigma_t^2 + (\theta^2 + \delta^2)\lambda_t$, which is a combination of the GARCH conditional variance component and the jump component.

Maheu and McCurdy (2004) describe some empirical results found by the model. They find strong evidence of time dependence in jump intensities for both stock indices and several individual stocks. The average proportion of conditional variance explained by jumps varies from 20% to 40%, at times as much as 90%. It is much higher than Huang and Tauchen (2005)'s finding that jumps account for 7 percent of stock market price variance using $S\&P$ index.

The ARJI model provides better out-of-sample forecasts following large negative moves in the market. However, the forecasts may be worse than the GARCH (1,1) model when the out-of-sample period differs with the in-sample period in frequency of jumps. As jumps are infrequent and hard to predict, there is no reason to assume that the out-of-sample period has similar frequency of jumps as the in-sample period. When the in-sample period has a relatively lower jump intensity than the out-of-sample period, the forecasts of volatility based on the in-sample parameters will be underestimated. Thus, it's important to take into account regime changes in the data generating process.

2.2.2 Construction of Markov Regime-Switching GARCH-Jump Model (RSGARJI Model)

The reported high persistence of conditional variance by GARCH models, together with high level of jump clustering revealed in Maheu and McCurdy (2004), may be spurious and due to structural shifts in the data generating process, such as deterministic changes in the intercept parameter of autoregressive jump intensity process. What's more, there are periods during which few jumps happen and other periods when jumps cluster. In an attempt to solve the problem, we now introduce regime shifts into the conditional jump intensity. To incorporate structural changes in the data generating process, a popular approach is the Markov regime-switching model applied to dependent processes by Hamilton (1989). State 1 and state 2 refer to low jump intensity regime and high jump intensity regime respectively, and a Markov process is used to govern the switches between regimes. The jump intensity depends on its own lagged value within each regime.

The regime-switching model is based on ARJI model for two reasons. Firstly, the autoregressive jump intensity setting can account for clustering of jumps and also incorporate shocks. By introducing regimes into GARCH-type conditional variance and jump intensity, we can explore if persistence of conditional variance and jumps vary for different regimes and whether the high persistence is spurious due to structural changes. Secondly, jump diffusion models with stochastic jump intensity are hard to estimate as the likelihood function has no closed form. The ARJI model avoids this problem by assuming approximate autoregressive jump intensity structure with a filter to infer the ex post distribution of jumps.

Let $\{s_t\}$ be a Markov Chain with 2-dimensional state space. In state 1, jumps are not so frequent, while jumps are more likely to happen in state 2. S_t denotes the regime path at time t, $\{s_t, s_{t-1}, s_{t-2}, ...\}$. Φ_{t-1} refers to the information set at time $t-1$. To clarify, ϵ_t refers to $R_t - \mu$, which is the summation of $\epsilon_{1,t}$ and the jump part.

The first specification for the conditional jump intensity λ_t is

$$P(N(t) = j|\Phi_{t-1}, S_t, x_t) = \exp(-\lambda_t)\lambda_t^j/j! \qquad \text{for } j = 0, 1, 2... \tag{2.4}$$

$$\lambda_t = \alpha_1 + \rho_1\lambda_{t-1} + \gamma_1\xi_{t-1} \qquad \text{if } s_t = 1 \tag{2.5}$$

$$\lambda_t = \alpha_2 + \rho_2\lambda_{t-1} + \gamma_2\xi_{t-1} \qquad \text{if } s_t = 2 \tag{2.6}$$

The conditional jump intensity, $\lambda_t = E(N(t)|\Phi_{t-1}, S_t)$, has an autoregressive form and depends on contemporaneous conditional variance and trading volume. Hereby $\xi_{t-1} = E[N(t-1)|\Phi_{t-1}, S_{t-1}] - \lambda_{t-1}$, which also depends on the state space of regimes. It is easy to show that ξ_t is still a martingale difference sequence with respect to $\{\Phi_{t-1}, S_{t-1}\}$, where Φ_{t-1} is the information set up to $t-1$,which is composed of past values of R_t. Therefore it is a well defined residual.

However, this specification for jump intensity has computational difficulties. Although the current regime only determines the parameters α,ρ and γ, the dependence of λ_t on both the current regime and its own lagged value λ_{t-1} makes λ_t depend on the entire regime path $\{s_t, s_{t-1}, s_{t-2}, ...\}$, by iterative substitution. As the regimes are latent, the inability to observe them leads to the need of integrating out all possible paths when calculating the sample likelihood. It makes the estimation practically intractable. This problem is similar to that which arises in regime-switching GARCH models, as noted by Klaaseen (2002), which models the GARCH type volatility as a regime-switching process.

To circumvent the problem of path dependence, we model the conditional jump intensity as

$$\lambda_t = \alpha_{s_t} + \rho_{s_t} E[\lambda_{t-1}|\Phi_{t-1}, s_t] + \gamma_{s_t} E[\xi_{t-1}|\Phi_{t-1}, s_t] \tag{2.7}$$

where $\lambda_t = E[N(t)|\Phi_{t-1}, s_t]$. Note that λ_t depends only on s_t instead of S_t now. The idea is inspired by Klaaseen (2002) to integrate out the regime path S_{t-1} out of the right hand side of the equation. After S_{t-1} is integrated out for conditional intensity of the last period, the right hand side only depends on the current regime s_t. In addition, this is equivalent to integrating out s_{t-1},the regime at time $t-1$, as the lag of equation (2.7) implies that λ_{t-1} only depends on s_{t-1} and is independent of S_{t-2}. s_t is included in the conditioning variables because it may contain some information about the last period regime s_{t-1}.

In order for λ_t to be positive for all t, a sufficient condition is that $\alpha_{s_t} > 0$, $\rho_{s_t} - \gamma_{s_t} \geq 0$ and $\gamma_{s_t} \geq 0$. This specification of the conditional jump intensity removes the problem of regime path dependence, and allows the jump intensity to be autocorrelated, which can help explain the phenomenon of jump clustering around significant news events.

2.2.3 Properties of the Model with Exogenous Regime and an Estimation Mechanism

The steady state probabilities of the regimes 1 and 2 at time $t-1$,$P(s_{t-1}=1)$ and $P(s_{t-1}=2)$, are derived in Hamilton (1989),

$$P(s_{t-1}=1)=\frac{1-p_{22}}{2-p_{11}-p_{22}} \tag{2.8}$$

$$P(s_{t-1}=2)=\frac{1-p_{11}}{2-p_{11}-p_{22}} \tag{2.9}$$

Proposition 2.1. *If the unconditional jump intensity exists for both regime 1 and 2, denoted by λ_1 and λ_2 respectively, then*

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = B^{-1} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \tag{2.10}$$

with

$$B=\begin{bmatrix} 1-\rho_1 p_{11} & -\rho_1(1-p_{11}) \\ -\rho_2(1-p_{22}) & 1-\rho_2 p_{22} \end{bmatrix} \tag{2.11}$$

The proof of Proposition 2.1 is in Appendix A. From Proposition 2.1, for the existence of the unconditional jump intensity, the inverse of B , the unconditional mean of GARCH-type conditional variance, needs to exist. In order for the unconditional jump intensity to be strictly positive for all t, the four elements of the inverse of B should be positive.

$$B^{-1} = \frac{1}{1-\rho_1 p_{11} - \rho_2 p_{22} - \rho_1\rho_2(1-p_{11}-p_{22})}\begin{bmatrix} 1-\rho_2 p_{22} & \rho_1(1-p_{11}) \\ \rho_2(1-p_{22}) & 1-\rho_1 p_{11} \end{bmatrix} \quad (2.12)$$

For the unconditional jump intensity to be strictly positive, $1-\rho_1 p_{11} - \rho_2 p_{22} - \rho_1\rho_2(1-p_{11}-p_{22}) > 0$, $1-\rho_1 p_{11} > 0$, and $1-\rho_2 p_{22} > 0$.

Then the conditional variance of R_t is

$$\begin{aligned} Var(R_t|\Phi_{t-1}) &= \sum_{s_t=1,2} P(s_t|\Phi_{t-1}) Var(R_t|s_t,\Phi_{t-1}) \\ &= \sum_{s_t=1,2} P(s_t|\Phi_{t-1})(\sigma_t^2 + (\theta^2+\delta^2)\lambda_t) \end{aligned} \quad (2.13)$$

The estimation of the model can be conducted by maximum likelihood based on the estimation mechanism of the ARJI model using an iterative algorithm. As there are two latent variables, the regime variable s_t and number of jumps N_t, the conditional density function is computed by integrating out the regime variable and number of jumps step by step.

$$\begin{aligned} f(R_t|\Phi_{t-1}) &= f(R_t|\Phi_{t-1}, s_t=1)P(s_t=1|\Phi_{t-1}) + f(R_t|\Phi_{t-1}, s_t=0)P(s_t=0|\Phi_{t-1}) \\ &= [\sum_{j=0}^{\infty} f(R_t|N(t)=j,\Phi_{t-1},s_t=1)P(N(t)=j|\Phi_{t-1},s_t)]P(s_t=1|\Phi_{t-1}) \\ &+ [\sum_{j=0}^{\infty} f(R_t|N(t)=j,\Phi_{t-1},s_t=0)P(N(t)=j|\Phi_{t-1},s_t)]P(s_t=0|\Phi_{t-1}) \end{aligned} \quad (2.14)$$

When there is an infinite summation in the likelihood function, I truncate it at 20. For the model estimates, it's found that there is zero probability in the tail of the conditional Poisson distribution for jump numbers larger than 15, which is in accordance with Maheu and McCurdy (2004)'s finding. The first part of the right hand side of the conditional sample likelihood function, $f(R_t|N(t) = j, \Phi_{t-1}, s_t)$,can be derived as follows. Note that σ_t^2 depends on s_t.

$$f(R_t|N(t) = j, \Phi_{t-1}, s_t) = \frac{1}{\sqrt{2\pi(j\delta^2 + \sigma_t^2)}} \exp(-\frac{(R_t - \mu - j\theta)^2}{2(j\delta^2 + \sigma_t^2)}) \tag{2.15}$$

Then the expression of $P(N(t) = j|\Phi_{t-1}, s_t)$ and $P(s_t|\Phi_{t-1})$ is needed.

From the model specification, it's known that

$$P(N(t) = j|\Phi_{t-1}, s_t) = \exp(-\lambda_t)\lambda_t^j/j! \qquad \text{for } j = 0, 1, 2... \tag{2.16}$$

where λ_t is a function of s_t, which is not straightforward to compute because of integrating out of regime path S_{t-1} in $E[\lambda_{t-1}|\Phi_{t-1}, s_t]$.

$$\lambda_t = \alpha_{s_t} + \rho_{s_t} E[\lambda_{t-1}|\Phi_{t-1}, s_t] + \gamma_{s_t} E[\xi_{t-1}|\Phi_{t-1}, s_t] \tag{2.17}$$

As λ_{t-1} is a function of s_{t-1},

$$E[\lambda_{t-1}|\Phi_{t-1}, s_t] = \sum_{s_{t-1}=1,2} \lambda_{t-1}(s_{t-1}) P(s_{t-1}|\Phi_{t-1}, s_t) \tag{2.18}$$

$\xi_t = E[N(t)|\Phi_t, s_t] - \lambda_t$ is also a function of s_t. The density of the expectation part of ξ_t is

$$P(N(t) = j|\Phi_t, s_t) = f(R_t|N(t) = j, \Phi_{t-1}, s_t) P(N(t) = j|\Phi_{t-1}, s_t)/f(R_t|\Phi_{t-1}, s_t) \tag{2.19}$$

Then

$$\xi_{t-1} = \sum_{j=0}^{\infty} jP(N(t-1) = j|\Phi_{t-1}, s_{t-1}) - \lambda_{t-1} \tag{2.20}$$

$$E[\xi_{t-1}|\Phi_{t-1}, s_t] = \sum_{s_{t-1}=1,2} \xi_{t-1}(s_{t-1})P(s_{t-1}|\Phi_{t-1}, s_t) \tag{2.21}$$

After getting $P(s_t|\Phi_{t-1})$ and $P(s_{t-1}|\Phi_{t-1}, s_t)$, the sample likelihood can be resolved. According to Bayes' rule,

$$P(s_{t-1}|\Phi_{t-1}, s_t) = \frac{P(s_{t-1}|\Phi_{t-1})P(s_t|s_{t-1}, \Phi_{t-1})}{P(s_t|\Phi_{t-1})} \tag{2.22}$$

$$P(s_t|\Phi_{t-1}) = \sum_{s_{t-1}=1,2} P(s_t|s_{t-1}, \Phi_{t-1})P(s_{t-1}|\Phi_{t-1}) \tag{2.23}$$

where $P(s_t|s_{t-1}, \Phi_{t-1}) = P(s_t|s_{t-1})$ is the transition probability. The computation of ex post regime probability $P(s_{t-1}|\Phi_{t-1})$ and ex ante regime probability $P(s_t|\Phi_{t-1})$ is discussed in Hamilton (1994) and works by applying a first-order recursive mechanism. That is,

$$P(s_{t-1}|\Phi_{t-1}) = \frac{f(R_{t-1}|s_{t-1}, \Phi_{t-2}) \sum_{s_{t-2}=1,2}(P(s_{t-2}|\Phi_{t-2})P(s_{t-1}|s_{t-2}, \Phi_{t-2}))}{f(R_{t-1}|\Phi_{t-2})} \tag{2.24}$$

and

$$f(R_{t-1}|s_{t-1}, \Phi_{t-2}) = \sum_{j=0}^{\infty} f(R_{t-1}|N(t-1) = j, \Phi_{t-2}, s_{t-1})P(N(t-1) = j|\Phi_{t-2}, s_{t-1}) \tag{2.25}$$

Thus, the iterative procedure is as follows:

Step 1. Set the initial values of $P(s_0|\Phi_0)$, σ_0^2 and λ_0 as their unconditional means respectively. ξ_0 is set to be 0.

Step 2. Given $P(s_0|\Phi_0)$, $P(s_1|\Phi_0)$ is computed using equation (2.23), which is then used to calculate $P(s_0|\Phi_0, s_1)$ via equation (2.22).

Step 3. With the above regime probabilities available, σ_1 can be computed and used to calculate $f(R_1|N(1) = j, \Phi_0, s_1)$ using equation (2.15). Together with λ_1 derived from equation (2.17), $f(R_1|\Phi_0)$ can be computed using equation (2.14).

Step 4. $f(R_1|\Phi_0, s_1)$ is derived via equation (2.25). Consequently, $P(N(1) = j|\Phi_1, s_1)$ is available with equation (2.19), and $P(s_1|\Phi_1)$ is derived from equation (2.24). Together with the value of λ_1, $E[\xi_1|\Phi_0, s_1]$ and $E[\lambda_1|\Phi_0, s_1]$ can be computed. Then λ_2 is available to compute the density $f(R_2|\Phi_1)$. The sample likelihood can be computed using this iterative procedure.

As the log likelihood function has a closed form expression, maximum likelihood method can be applied to estimate the model when the regime variable is exogenous.

2.3 Threshold GARCH-Jump model with Exogenous Trigger

2.3.1 Model

One main shortcoming of the hidden Markov regime switching model, discussed in the last section, is that the regimes of each period are not known to the econometrician, and this leads to difficulty with forecasting, especially when used to forecast volatility a few days later. Threshold models with observable triggers can solve this problem. In threshold models, when the trigger is below the threshold value, the economy is in regime 1, while it is in regime 2 otherwise. Thus, in each period, the regime can be observed, which is an exogenous variable. While estimation of threshold models poses no difficulty, there are limited theoretical results available concerning the stationarity conditions, and existence of moments.

Recently, Knight and Satchell (2011) derived necessary and sufficient conditions for the existence of a stationary distribution of threshold-AR (1) model with an exogenous trigger variable. As our GARCH-jump model is a GARCH model with AR (1) jump intensity, these conditions for the threshold-AR (1) can be applied to the threshold GARCH-jump model with an exogenous trigger. The threshold GARCH-jump model is

$$R_t = \mu + \epsilon_{1,t} + \sum_{k=1}^{N(t)} Y_{t,k} \tag{2.26}$$

$$\epsilon_{1,t} = z_t \sigma_t \tag{2.27}$$

$$Y_{t,k} \sim N.I.D(\theta, \delta^2) \tag{2.28}$$

$$P(N(t) = j|\Phi_{t-1}) = \exp(-\lambda_t)\lambda_t^j / j! \qquad \text{for } j = 0, 1, 2... \tag{2.29}$$

$$\sigma_t^2 = \omega_1 + a_1 \epsilon_{t-1}^2 + b_1 \sigma_{t-1}^2 \qquad \text{if } \nu_{t-1} \leq \nu_0 \tag{2.30}$$

$$\lambda_t = \alpha_1 + \rho_1 \lambda_{t-1} + \gamma_1 \xi_{t-1} \qquad \text{if } \nu_{t-1} \leq \nu_0 \tag{2.31}$$

$$\xi_{t-1} = E[N(t-1)|\Phi_{t-1}] - \lambda_{t-1} \tag{2.32}$$

$$\sigma_t^2 = \omega_2 + a_2 \epsilon_{t-1}^2 + b_2 \sigma_{t-1}^2 \qquad \text{if } \nu_{t-1} > \nu_0 \tag{2.33}$$

$$\lambda_t = \alpha_2 + \rho_2 \lambda_{t-1} + \gamma_2 \xi_{t-1} \qquad \text{if } \nu_{t-1} > \nu_0 \tag{2.34}$$

Φ_{t-1} denotes the information up to time $t-1$, which includes the time series of R_t and threshold variable ν_t up to time $t-1$. In this model setting, there are two regimes as well as in the previous hidden Markov regime-switching model. The parameters in the GARCH conditional variance and jump intensity depends on the threshold variable ν_t. In hidden Markov model regimes in each period are unknown to the econometricians both before and after the estimation, however, in the threshold model the regimes are known to the econometrician, which is very helpful in the estimation as well as in forecasting. For example, there is no need to integrate out the previous regimes in the threshold model, which simplifies the estimation algorithm. Let $s_t = 0$ if $\nu_t \leq \nu_0$, and $s_t = 1$ if $\nu_t > \nu_0$, where we assume that s_t follows a i.i.d Bernoulli distribution with $P(s_t = 1) = \pi$.

2.3.2 Stationarity Conditions and Moments of Returns

Proposition 2.2. *λ_t is strictly stationary if* $\ln|\rho_1|(1-\pi)+\ln|\rho_2|\pi < 0$. *The return series is covariance-stationary if* $|\rho_1|(1-\pi)+|\rho_2|\pi < 1$ *and* $|(a_1+b_1)|(1-\pi)+|(a_2+b_2)|\pi < 1$. *The mean of return is given by*

$$E(R_t) = \mu + \theta E(\lambda_t) = \mu + \theta\left(\frac{\alpha_1(1-\pi)+\alpha_2\pi}{1-\rho_1(1-\pi)-\rho_2\pi}\right)$$

The variance of return is given by

$$\begin{aligned} Var(R_t) &= \frac{\omega_1(1-\pi)+\omega_2\pi}{1-(a_1+b_1)(1-\pi)-(a_2+b_2)\pi} \\ &+ \frac{(\theta^2+\delta^2)(\alpha_1(1-\pi)+\alpha_2\pi)(1-b_1(1-\pi)-b_2\pi)}{(1-\rho_1(1-\pi)-\rho_2\pi)(1-(a_1+b_1)(1-\pi)-(a_2+b_2)\pi)} \end{aligned}$$

The proof of Proposition 2.2 is in Appendix B. It shows that both the mean and the variance of return depends on the parameters and probability of each regime. The conditional skewness and kurtosis are given by

$$Skewness(R_t|\Phi_{t-1}) = \frac{\lambda_t(\theta^3+3\theta\delta^2)}{(\sigma_t^2+\lambda_t\delta_t^2+\lambda_t\theta^2)^{3/2}} \tag{2.35}$$

$$Kurtosis(R_t|\Phi_{t-1}) = 3 + \frac{\lambda_t(\theta^4+6\theta^2\delta^2+3\delta^4)}{(\sigma_t^2+\lambda_t\delta^2+\lambda_t\theta^2)^2} \tag{2.36}$$

The derivation of the above conditional moments is given in Das and Sundaram (1997). The skewness is positive if $\theta > 0$. Both skewness and kurtosis depend on the conditional jump intensity λ_t, the jump size's mean θ and variance δ^2. The conditional kurtosis is larger than 3 in the presence of jumps, as the existence of outliers leads to fatter tails in the return distribution.

The estimation of the threshold GARCH-jump model in-sample is conducted by MLE. After the threshold value ν_0 is estimated, the regime of each observation is known by comparison of the threshold variable and ν_0. Given ν_0, the MLE estimator is obtained by maximizing the log likelihood function. Thus, both the MLE estimator and the log likelihood value are functions of ν_0.

Construction of the likelihood function is similar to that of GARCH-jump model. The conditional density of returns is normal given j jumps occurring,

$$f(R_t|N(t)=j,\Phi_{t-1}) = \frac{1}{\sqrt{2\pi(j\delta^2+\sigma_t^2)}}\exp(-\frac{(R_t-\mu-j\theta)^2}{2(j\delta^2+\sigma_t^2)}) \quad \textbf{(2.37)}$$

Then the conditional density of returns can be found by integrating out the number of jumps occurring,

$$f(R_t|\Phi_{t-1}) = \sum_{j=0}^{\infty} f(R_t|N(t)=j,\Phi_{t-1})P(N(t)=j|\Phi_{t-1}) \quad \textbf{(2.38)}$$

When constructing the jump intensity, the ex post filter can be built via Bayes' rule as,

$$P(N(t)=j|\Phi_t) = f(R(t)|N(t)=j,\Phi(t-1))P(N(t)=j|\Phi(t-1))/f(R(t)|\Phi(t-1)) \quad \textbf{(2.39)}$$

for $j=0,1,2,....$ Then the jump intensity residual is available and the autoregressive jump intensity can be constructed. In order to find ν_0 which maximizes the log likelihood value, the sample of the threshold variable is divided into 20 intervals with 19 grid points from 5 percentile point to 95 percentile point.

Table 2.1: Descriptive Statistics of Daily Returns of Japanese Yen

Statistics	1990-2005	1990-2004	2004-2005
Obs	3751	3500	251
Mean	-0.009	-0.009	-0.012
Std. Deviation	0.703	0.707	0.651
Skewness	-0.508	-0.552	0.277
Kurtosis	7.027	7.176	3.903
Min	-5.630	-5.630	-1.550
Max	3.240	3.240	2.452
$\|R\| > 2$	59	58	1
$\|R\| > 3$	13	13	0

2.4 Data and Estimation

2.4.1 Data

The data used are the Japanese Yen- US Dollar spot exchange rate series. The in-sample period contains 3500 daily observations,which starts from January 2nd,1990 to January 9th, 2004. The data is accessed from Wharton Research Data Service (WRDS), and is obtained from Bank of Japan. The return R_t is calculated to be 100 times the log difference of exchange rate R_t. 251 observations from January 12th, 2004 to January 11th, 2005 are used as the out-of-sample period for forecasting purposes after eliminating weekends and holidays. Table 2.1 provides summary statistics for returns for daily Japanese Yen exchange rate according to different sample periods.

The threshold variable used in the chapter is Chicago Board Option Exchange (CBOE) S&P 500 Volatility Index (VIX), which is a key measure of market expectations of near-term 30-day implied volatility built by S&P 500 stock index option prices. It is reasonable to assume that when VIX is high, the market has an expectation of high volatility of stocks. When VIX is higher than some particular value, we assume that the market enters into a regime that is more volatile. Furthermore,as VIX is a 30-day expectation built on a market index, it is also reasonable to assume that it is exogenous in relation to the current volatility of the return of a specific stock or exchange rate. The VIX series is also accessed from Wharton Research Data Service (WRDS) from January 2nd,1990 to January 11th, 2005.

2.4.2 Estimation

Table 2.2 reports the MLE estimates with standard errors in brackets, and corresponding log likelihood values using threshold GARCH (1,1)-jump AR (1) model (TS-GARJI), threshold GARCH (1,1) model (TS-GARCH),and regime switching GARCH (1,1)-jump AR (1) model (RS-GARJI), together with the results using GARCH (1,1) model. Akaike's information criterion and Bayesian information criterion are also included to compare goodness of fit of models.

Table 2.2 shows that the parameters in the threshold GARJI model are significant except the intercepts in the GARCH variance term and the jump intensity AR (1) term. For the regime switching GARJI model, the parameters of jump intensity in the first regime are insignificant. As the first regime

Table 2.2: Estimates and log likelihood values using different models

Parameter	TS-GARJI	TS-GARCH	RS-GARJI	GARCH(1,1)	ARJI
μ	0.035	-0.006	0.038	-0.007	0.034
	(0.012)	(0.011)	(0.012)	(0.011)	(0.012)
ω_1	0.003	0.011	0.014	0.009	0.003
	(0.002)	(0.002)	(0.004)	(0.001)	(0.001)
a_1	0.011	0.033	0.019	0.041	0.014
	(0.005)	(0.005)	(0.006)	(0.004)	(0.004)
b_1	0.966	0.938	0.961	0.941	0.968
	(0.013)	(0.008)	(0.013)	(0.006)	(0.008)
ω_2	0.009	0.006	0.001		
	(0.006)	(0.004)	(0.002)		
a_2	0.011	0.062	0.008		
	(0.006)	(0.008)	(0.004)		
b_2	0.956	0.920	0.984		
	(0.024)	(0.012)	(0.031)		
α_1	0.006		0.063		0.017
	(0.005)		(0.104)		(0.008)
ρ_1	0.970		0.640		0.901
	(0.024)		(0.582)		(0.048)
γ_1	0.081		-0.077		0.184
	(0.048)		(0.111)		(0.066)
α_2	0.004		0.026		
	(0.003)		(0.014)		
ρ_2	0.987		0.935		
	(0.011)		(0.098)		
γ_2	0.146		0.964		
	(0.058)		(0.423)		
θ	-0.219		-0.265		-0.292
	(0.066)		(0.071)		(0.082)
δ	0.912		0.917		0.984
	(0.075)		(0.083)		(0.088)
ν_0	20.235	22.765			
p_{11}			0.983		
			(0.017)		
p_{22}			0.953		
			(0.053)		
logLLF	-3426.2	-3563.9	-3426.3	-3571.4	-3437.2
AIC	6882.4	7141.8	6886.6	7150.8	6892.4
BIC	6974.8	7184.9	6991.3	7175.4	6947.8

is the less volatile regime with a much smaller unconditional variance, it implies that there are no jumps in the less volatile period. Therefore the regime switching GARJI model is re-estimated with only jumps in the more volatile period. Both AIC and BIC suggest that the threshold GARJI model has the best fit of data among all the models, while GARCH (1,1) model has the worst. The parameters in threshold GARJI model satisfy the stationarity condition, and in each regime the summation of a and b are less than one. The threshold GARCH (1,1) model has a threshold value at 70%, which implies that there is 30% chance that the conditional variance shifts to regime 2. For the threshold GARJI model, the threshold value is at 55%, implying that there is chance of 45% for the conditional variance and jump intensity to shift to regime 2. Including jump term better fits the data, as the threshold GARJI model outperforms the threshold GARCH (1,1) model. Figure 2.1 and Figure 2.2 plot the time series of the threshold variable, VIX, and the estimated threshold value for the threshold GARCH model and the threshold GARJI model respectively, showing that they have similar estimates in magnitude for the threshold value.

The mean reported in Table 2.2 for the threshold GARJI and the regime-switching GARJI model are not the mean of the return, which is the reason that it is different from the mean of return in the GARCH (1,1) model. The mean of return is $E(R_t) = \mu + \theta E(\lambda_t)$ instead of μ for the jump models. The persistence parameters in the jump intensity process are high for both regimes in the threshold GARJI model. For the threshold GARCH (1,1) model, we find that the persistence parameter in the diffusive conditional variance process is lower in each regime than in the GARCH (1,1) model. However, after jumps are incorporated, the persistence parameter in the conditional variance process is higher in each regime for both the threshold GARJI model and the regime switching GARJI model than in the GARCH (1,1) model. It implies that separating the effects of jumps and diffusive volatility makes the volatility more persistent, while the high persistence of diffusive volatility of the GARCH (1,1) model may come from the reason that different regimes are not identified.

2.5 Forecasting

When forecasting volatility it is difficult to know what to compare the forecasts with, since volatility is unobserved. Consequently, we use realized volatility as the proxy ex post daily volatility to measure the forecasting performance of regime switching GARCH-Jump models. The data set for constructing the realized volatility contains the five-minute transaction price for the Japanese Yen-US dollar spot exchange rate from January 12th, 2004 to January 11th, 2005. I use five minute data as it is considered the highest frequency at which prices are less distorted by the market microstructure noise. Following Andersen and Bollerslev (1998), the trading day t starts from 21:00 GMT on day $t-1$ to 21:00 GMT on day t, which ensures that all transactions on day t of

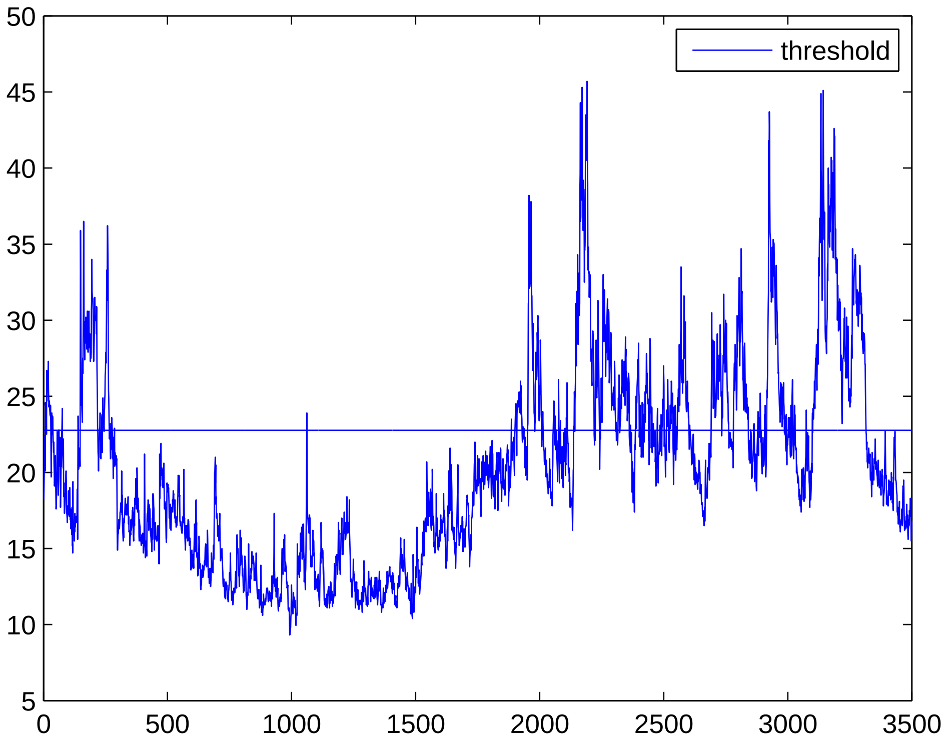

Figure 2.1: VIX with estimated threshold value for the threshold GARCH model

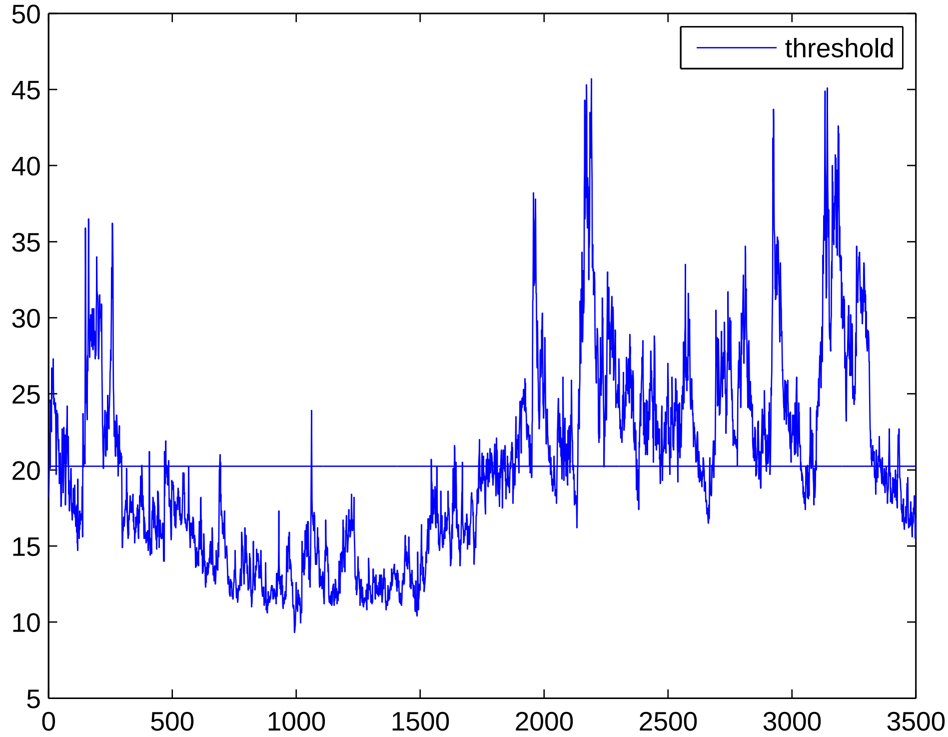

Figure 2.2: VIX with estimated threshold value for the threshold GARCH-jump model

local time take place during this period. Weekend days and major holidays are deducted for the reason of too many missing values or slower trading pattern. 251 days are left after the deduction.

The m*th* five-minute exchange rate on day t is denoted by $P_{m,t}$, for $m = 1, 2, ..., M$. For five-minute data, $M = 288$. The five minute return $r_{m,t}$ is constructed as $r_{m,t} = 100(\ln P_{m,t} - \ln P_{m-1,t})$, for $m = 1, 2, ..., M$, and $t = 1, 2, ..., 251$. The realized volatility is obtained by summing up the squared intra-day 5-minute returns as $RV_t = \sum_{m=1}^{M} r_{m,t}^2$. Andersen and Bollerslev (1998) note that, under a jump-diffusion semi-martingale setting for the price process with bounded jump intensity λ_t, the realized volatility is consistent for quadratic variation of the logarithmic price process. As the quadratic variation consists of both the diffusive volatility and the cumulative squared jumps, the realized volatility is a good proxy for volatility when jumps are taken into consideration as in our models. Figure 2.3 plots the evolution of the realized volatility constructed using 5-minute intraday returns in the out-of-sample period.

The respective parameters estimated from in-sample period are used to conduct conditional variance forecasts for all the models. The threshold variable is still chosen as VIX and the threshold value is taken as the in-sample estimate. A rolling scheme is used, that is, the in-sample period contains 3500 observations and moves forward every 50 observations. Figure 2.4, Figure 2.5 and Figure 2.6 depict the out-of-sample one-step-ahead forecasts of conditional variance of GARCH(1,1) model, TSGARJI (1,1) model, and RSGARJI (1,1) model respectively. From the figures we note that conditional variances conducted from the Markov regime-switching GARCH-jump model has a bigger variation than those from the GARCH(1,1) model and threshold GARCH-jump model. Although the Markov regime-switching GARCH-jump model fits the data better sometimes when there is a peak in the realized volatility, it also makes some worse forecasts. When realized volatility is quite high, conditional variance from the threshold GARCH-jump model cannot catch up with it. One possible reason is that jump size could be an increasing function of volatility. As $Var(R_t|\Phi_{t-1}) = \sigma_t^2 + (\theta^2 + \delta^2)\lambda_t$, θ and δ can also play a role in determine the conditional variance. They can be functions of σ_t or λ_t.

For evaluating the forecasts, we run a linear regression of realized volatility on its forecast. Then the coefficient of determination, R^2, provides a guide to the accuracy of volatility forecasts. The one-day-ahead out-of-sample volatility forecasts are evaluated using the following regression,

$$RV_t = c + dVar(R(t)|\Phi_{t-1}) + error_t \qquad (2.40)$$

where $Var_{t-1}(R(t))$ is the out-of-sample conditional variance forecast for day t of the corresponding model. R^2 can be used to evaluate forecasting models as it shows how much of the variation in realized volatility can be explained by the variation of conditional variance forecasts. Table 2.3 reports the R^2 for different models.

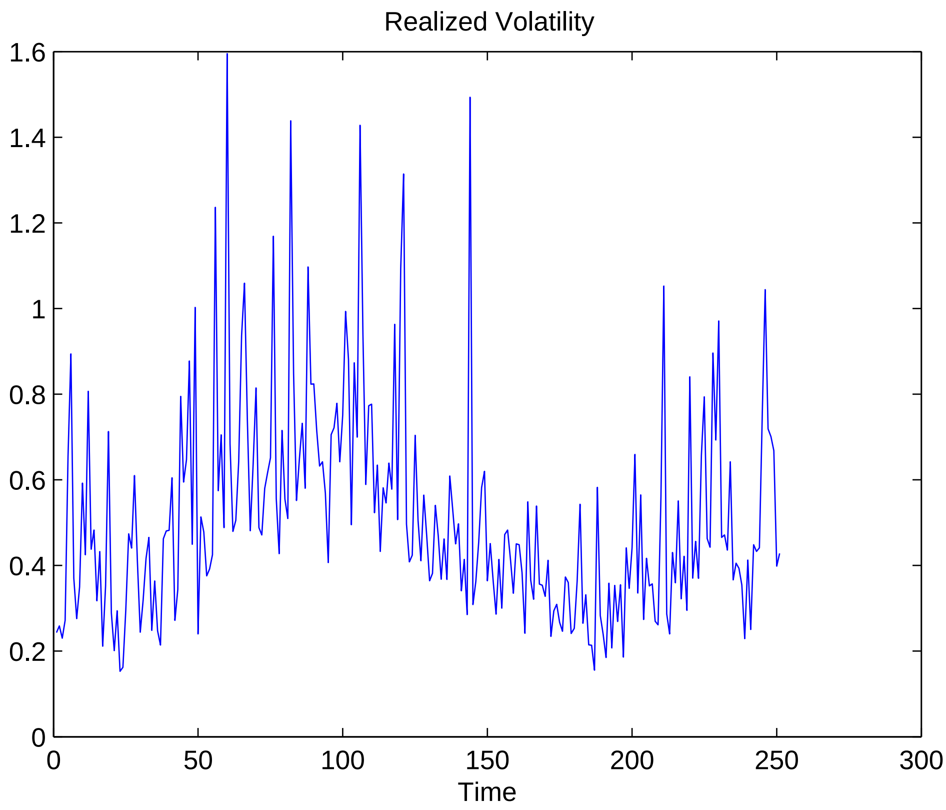

Figure 2.3: Out-of-sample realized volatility from Jan 2004 to Jan 2005

Table 2.3: Out-of-sample evaluation statistics of conditional variance forecasts

	TS-GARJI	TS-GARCH	RS-GARJI	ARJI	GARCH (1,1)
R^2	0.2077	0.1903	0.1030	0.1499	0.1895

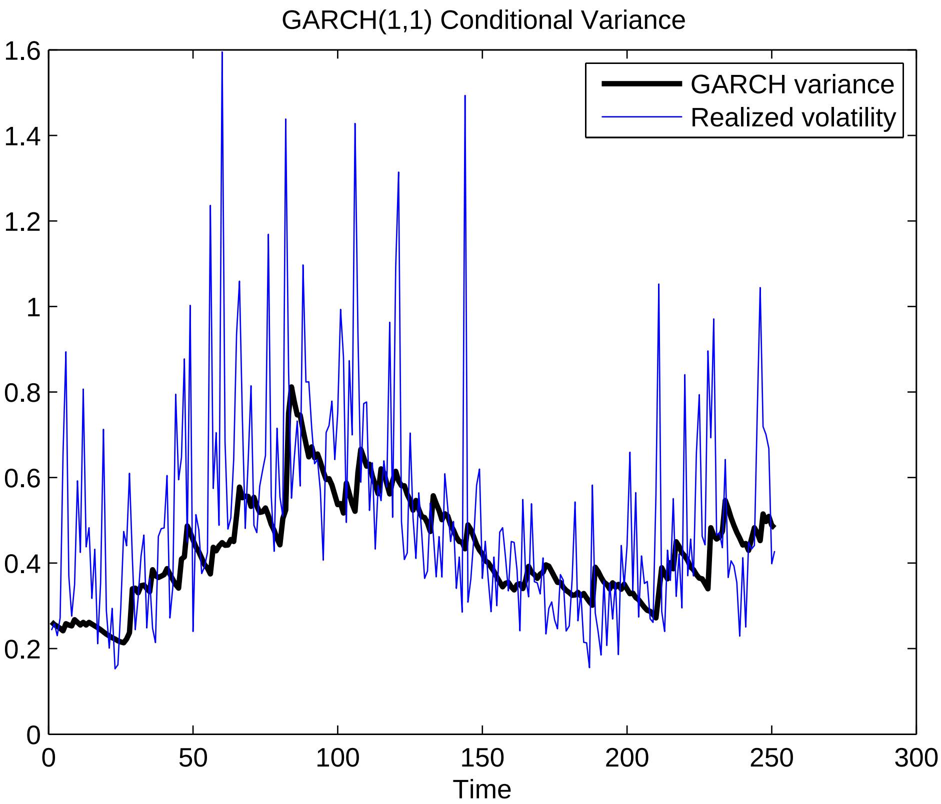

Figure 2.4: Out-of-sample conditional variance forecast using GARCH (1,1) model

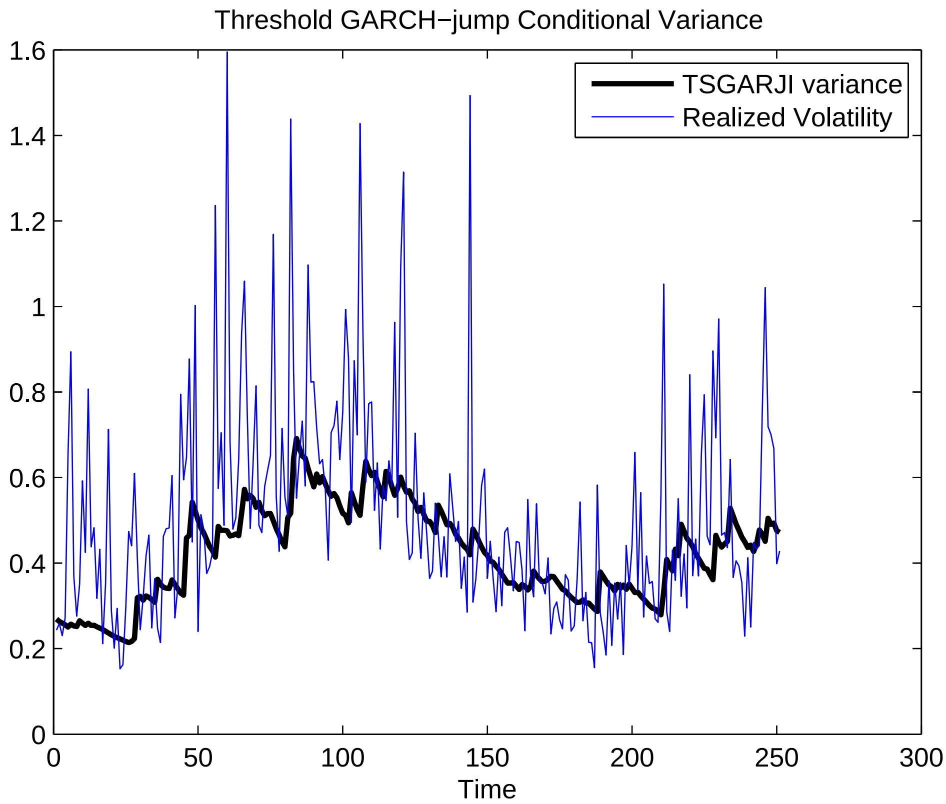

Figure 2.5: Out-of-sample conditional variance forecast using threshold-GARJI model

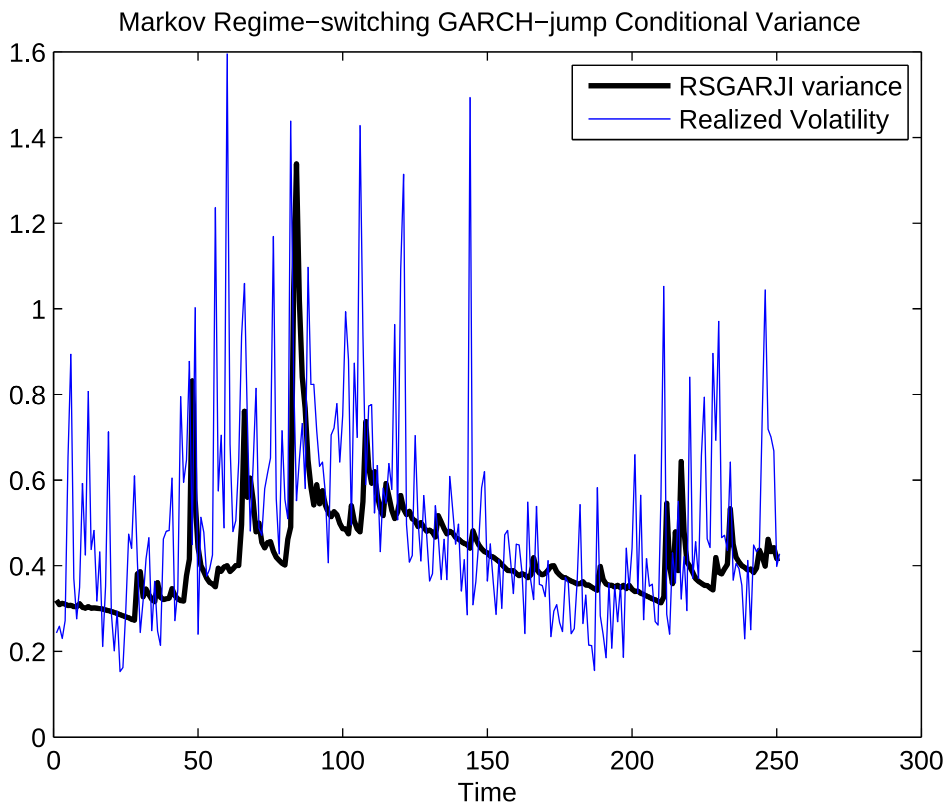

Figure 2.6: Out-of-sample conditional variance forecast using Regime-switching GARJI model

From the table we find that threshold GARJI model performs better to explain the variation of the data out-of-sample in terms of R^2 than the single-regime ARJI model. The reason is that the out-of-sample period is a tranquil period, in which only one out of 251 observations has an absolute value that is larger than 2; however, in the in-sample period, there are 58 out of 3500 observations whose absolute values are larger than 2. Thus the ARJI model overestimates the jump part and leads to inaccurate forecasts. The threshold GARJI model has the advantage of distinguishing the tranquil period from the volatile period, leading to more accurate forecasting performance. Although the single-regime ARJI model does not perform as well as the GARCH (1,1) model, it does not imply that there is no need to incorporate jumps, as the threshold GARJI model has a higher R^2 than the GARCH (1,1) model. The performance of the Markov regime-switching GARJI model is not good, which is generally found in the exchange rate literature.

2.6 Conclusion

In this chapter we have developed two models to model jumps and regime switching at the same time. The data generating process is assumed to be a combination of a GARCH process capturing small and smooth changes and a compound poisson process with autoregressive jump intensity modeling large and abrupt changes in return. Therefore, the volatility process is affected by both components. Meanwhile, we present switching regimes in the models to account for the phenomenon that there are tranquil periods when jumps rarely happen and volatile periods when jumps are more likely to happen. Regimes are incorporated either using a hidden first-order Markov process or through an exogenous threshold variable. For the Markov regime-switching GARJI model, the jump intensity is assumed to depend on the expected lagged intensity conditional on the current state, in order to solve the problem of regime path dependence. In each regime, both the GARCH variance and the jump intensity process have different parameter settings.

In the Markov regime-switching GARJI model, regimes are unknown to the econometrician. As the one-day-ahead state of regime is unknown, it leads to difficulty in forecasting. Therefore, a threshold GARCH-jump model is also proposed. The one-day-ahead state of regime is known to the econometrician by comparing the threshold variable and the threshold value. The stationarity conditions and moments of returns are derived for the threshold GARCH-jump model with an exogenous threshold variable. Both models are estimated using maximum likelihood method, and the regime-switching GARCH-jump model are more computationally intensive than the threshold GARCH-jump model.

We use the Japanese Yen-US Dollar spot exchange rate for estimation, and realized volatility constructed from 5-minute intraday data as proxy of volatility for evaluation of forecasts of different models. The empirical results indicate that jump intensity has a significant level of persistence, and the regime switching GARCH-jump models outperform GARCH model for the in-sample period. We also find that the persistence of diffusive volatility is lower for a threshold GARCH (1,1) than a single regime GARCH model, which is in accordance with previous literature. Out-of-sample forecasts suggest that threshold GARCH-jump model has a good ability to forecast volatility of Japanese Yen-US Dollar exchange rate and it outperforms the single regime ARJI model, indicating that it's necessary to incorporate regimes into jump models.

2.7 Bibliography

ANDERSEN, T., AND T. BOLLERSLEV (1998): "Answering skeptics:Yes,standard volatility models do provide accurate forecasts," *International Economic Reviews*, 39, 115–158.

BARNDOFF-NIELSEN, O., AND N. SHEPHARD (2001): "Non-Gaussian Ornstein-Unlenbeck models and some of their uses in financial economics," *The Royal Statistical Society*, B63, 167–241.

BARNDOFF-NIELSEN, O., AND N. SHEPHARD (2004): "Power and bipower variation with stochastic volatility and jumps," *Journal of Financial Econometrics*, 2, 1–37.

CHAN, W., AND J. M. MAHEU (2002): "Conditional jump dynamics in stock market returns," *Journal of Business & Economic Statistics*, 20(3), 377–389.

DAS, S., AND R. K. SUNDARAM (1997): "Taming the skew: Higher-order moments in modelling asset price processes in finance, NBER Working Paper 5976," .

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. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

HAMILTON, J. D. (1994): "Chapter 22: Modelling time series with changes in regimes," *Time Series Analysis*, pp. 677–703.

HILLEBRAND, E. (2005): "Neglecting parameter changes in GARCH models," *Journal of Econometrics*, 129, 121–138.

HUANG, X., AND G. TAUCHEN (2005): "The relative contribution of jumps to total price variation," *Journal of Financial Econometrics*, 3, 456–499.

KLAASEEN, F. (2002): "Improving GARCH volatility forecasts with regime switching GARCH," *Empirical Economics*, 27, 363–394.

KNIGHT, J. L., AND S. SATCHELL (1998): "GARCH processes, some exact results, some difficulties and a suggested remedy," *Forecasting Volatility in Financial Markets, J. Knight and S. Satchell(editors),Oxford:Butterworth-Heineman*, pp. 321–346.

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

LAMOUREUX, C., AND W. LASTRAPES (1990): "Persistence in variance, structural change, and the GARCH model," *Journal of Business and Economics Statistics*, 8, 225–234.

MAHEU, J. M., AND T. M. MCCURDY (2004): “News arrival,Jump dynamics, and volatility components for individual stock returns,” *Journal of Finance*, 59(2), 755–793.

MERTON, R. (1976): “Option pricing when underlying stock returns are discontinuous,” *Journal of Financial Econometrics*, 3, 125–144.

PRESS, S. (1967): “A compound events model for security prices,” *Journal of Business*, 40, 317–335.

Chapter 3

Autoregressive Conditional Duration Models with Structural Changes

3.1 Introduction

Over the last two decades, the rapid development in computer technology has led to the availability of ultra high frequency data or tick-by-tick data in financial markets, as every transaction time together with other information such as volume, bid-ask spread, and price can be recorded for many types of data. While most financial data are regularly spaced, the ultra high frequency data arrive in irregular time intervals, which makes the traditional econometric techniques no longer directly applicable. One way to solve this problem is to model the process of duration between adjacent trade transactions. In addition, recent literature on market microstructure suggests that the frequency of transactions carries important information and thus should be modeled.

To model the duration process, Engle and Russell (1998) build a linear autoregressive conditional duration (ACD) model which accounts for stochastic clustering in durations, i.e., long durations are usually followed by long durations and short durations are usually followed by short durations. Following their seminal work, there is a growing literature in modeling durations, such as the fractionally integrated ACD model of Jasiak (1998) which account for long memory, the stochastic conditional duration (SCD) model introduced by Bauwens and Veredas (2004) and so on. A survey provided by Pacurar (2008) discussses both theoretical developments of ACD models and relevant empirical results using financial transaction data.

Analogous to the GARCH model of Bollerslev (1986), the ACD model assumes that the conditional expected duration follows an autoregressive process, depending on past durations and conditional expected durations. However, test statistics show that there are still excess dispersions and nonlinearities in the standardized residuals that cannot be fully captured by the autoregressive process. Aiming to explain the excess nonlinearities, Zhang, Russell and Tsay (2001) build a K-regime threshold ACD model with different parameter values and distributions of error term in each regime. Existence of moments are established and multiple structural breaks are recognized in the data which match with some economic events.

Meanwhile, many studies based on the linear ACD model and its extensions reveal high persistence of trade durations, as the sum of estimated autoregressive coefficients on lagged durations and conditional expected durations are close to one. As we know, the high persistence of volatility in the GARCH model may be spurious and due to structural changes in the data generating process, according to Lamoureux and Lastrapes (1990) and Hillebrand (2005) among others. Whether the high persistence often found in estimated GARCH and ACD models is an artifact of the data or caused from structural shifts remains an interesting question.

In this chapter, Monte-Carlo simulations are conducted to investigate whether the unaccounted shifts in the parameters of the ACD model can lead to spurious high persistence of trade durations. The contribution of the chapter is three-fold. First, Monte Carlo simulations show that both permanent changes and temporary changes in any of the three parameters in the conditional duration process that are not accounted for in global estimations can lead to strong bias in the parameter estimates. The high persistence of durations found in the literature may be due to unaccounted for structural changes in the conditional duration process. For example, if there are two regimes with low but slightly different autoregressive parameters in the data generating process while the changing point is the midpoint of the sample, using a single-regime ACD model for estimation will lead to a very high autoregressive parameter estimate. Therefore, the single-regime ACD model is not a suitable model to measure duration persistence in this case. Second, the unconditional mean of the conditional expected duration of the ACD model with unaccounted structural changes is derived. Third, a threshold ACD model is applied to Boeing trade duration data when the innovation follows the exponential distribution or the Weibull distribution.

The chapter is organized as follows. Section 3.2 provides a brief review of the existing literature on ACD models. Section 3.3 describes the ACD framework and the estimation methodology. Section 3.4 conducts the Monte Carlo simulations for both temporary and permanent shifts in the parameters of the conditional duration process. Section 3.5 presents the unconditional mean of the ACD model with unaccounted structural changes in the data generating process. Section 3.6 derives the stationarity condition of a threshold ACD model with an exogenous threshold variable and provides empirical estimation. Section 3.7 concludes.

3.2 Literature Review

Traditional econometric models in time series deal with regularly-spaced data, such as daily price data using the first or last observation of a trading day, or 5-minute data usually used in realized volatility models. However, the rapid development in increasing computer power leads to the recording of every trading transaction together with the transaction's characteristics such as volume, price and so on. The distinctive feature of the transaction data or tick-by-tick data is that the transactions are irregularly spaced, consequently, the use of traditional econometric models is not available. Since the introduction of the autoregressive conditional duration (ACD) model by Engle and Russell (1998), it has been widely used, along with its extensions, in modeling durations between trade transactions. The standardized durations are assumed to be independent and identically distributed with a unit mean. Engle and Russell (1998) use the standard exponential distribution for the distribution of the standardized durations, as it provides quasi maximum likelihood (QML) estimators for the parameters even if the distribution is not exponential. The QML estimates are consistent and asymptotically normal when the distribution is within the standard Gamma family, as shown in Drost and Werker (2004). For greater flexibility of a changing hazard function, the standardized Weibull Distribution with scale parameter equal to one is also used in Engle and Russell (1998). Grammig and Maurer (2000) apply the Burr distribution which reduces to the exponential distribution and Weibull distribution with special parameter values, relaxing the monotonicity of the hazard function.

High persistence in trade durations are often revealed in empirical analysis, as the estimated coefficients of the linear ACD models on the lagged variables sum to one, under different distribution assumptions of the standardized durations. In addition, a hyperbolic decay of autocorrelations is shown in many financial durations series. To account for the long memory property, Jasiak

(1998) propose the Fractionally Integrated ACD (FIACD) model, which is analogous to the FIGARCH model of Baillie, Bollerslev, and Mikkelsen (1996). However, the long memory phenomenon can also occur as a result of the presence of unaccounted regime-switching or structural changes in the time series. Zhang, Russell, and Tsay (2001) present a threshold ACD model to distinguish between heavier and thinner trading periods. Moreover, Meitz and Teräsvirta (2006) introduce the time-varying ACD model and the smooth transition ACD (STACD) model and discuss their properties. In the STACD model, the transition between states is driven by a suitably chosen non-negative and bounded transition function, such as the logistic transition function. Using trade durations between transactions of IBM shares in 2002, the STACD (1,1) model of orders 1 and 2 outperforms the ACD (1,1) model.

3.3 The ACD Framework

Let $t_0, t_1, ..., t_n, ...$ be a sequence of transaction times with $0 = t_0 \leq t_1 \leq ... \leq t_n \leq ....$ The number of transactions that happen before time t is denoted by $N(t)$. Let $x_i = t_i - t_{i-1}$ be the duration between trades. The linear $ACD(p,q)$ model of Engle and Russell (1998) deals with the durations after removing the daily seasonality of trade durations.

$$x_i/\psi_i = \epsilon_i \tag{3.1}$$

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} \tag{3.2}$$

where ψ_i is the conditional expectation of the ith duration. $\psi_i = E(x_i|x_{i-1},...,x_1)$. ϵ_i follows an i.i.d distribution with density $p(\epsilon;\phi)$. Then the error term ϵ_i needs an expectation of 1. As the error term is i.i.d, all past information influence the current duration through the conditional durations. The stochastic clustering phenomenon of trade durations can be explained by the autoregressive structure of the conditional durations. The density of the error term can take a lot of forms, such as the exponential distribution and Weibull distribution used in Engle and Russell (1998) and Burr distribution used in Grammig and Maurer (2000). The sufficient conditions for $x(i)$ to be covariance stationary is $\alpha_1 + \beta_1 < 1$ for the ACD (1,1) model. Assume that the process is covariance-stationary, the unconditional expected value of ψ is

$$E(\psi) = \frac{\omega}{1-\alpha_1-\beta_1} \tag{3.3}$$

Engle and Russell (1998) report that the autoregressive coefficients sum close to 1 in empirical analysis, i.e. ,$\alpha_1 + \beta_1$ is close to one in ACD(1,1) case, which indicates a high persistence of trade durations.

The vector of parameters is denoted by $\theta = (\omega, \{\alpha_i\}, \{\beta_i\}, \phi)$. When the distribution of the error term ϵ is exponential, the log likelihood function is given by

$$L(\theta) = -\sum_{i=1}^{N(T)} (log\psi_i + x_i/\psi_i) \tag{3.4}$$

Engle and Russell (1998) establish the QMLE properties of exponential ACD models. Under some regularity conditions, even if the distribution of ϵ

is not exponential, consistent and asymptotically normal estimates of θ can be obtained by maximizing the function (3.4). Due to the similarity between the ACD model and the GARCH model, the standard errors need to be adjusted following Bollerslev and Wooldridge (1992).

However, the assumption of exponential distribution for ϵ_i implies that the model has a constant hazard function. The hazard function of a random variable X is defined as

$$h(x) = \frac{f(x)}{S(x)} \tag{3.5}$$

where $f(x)$ is the probability density function of X and $S(x) = 1 - P(X \leq x)$ is the survival function of X. The hazard function equals one if the innovation follows a standard exponential distribution. As trading duration in the financial market is inversely related to trading intensity, and the trading intensity is dependent on the arrival of new information, it is not reasonable to assume the hazard function of trading duration to be constant over time.

To overcome this problem, alternative distributions for the innovation are proposed in the literature, as mentioned earlier. The Weibull distribution is widely used for its simplicity and its power to generate a dynamic hazard function. If the innovation follows a standardized Weibull distribution with scale parameter equal to one and shape parameter equal to γ, the log likelihood function is

$$\sum_{i=1}^{N(T)} log(\frac{\gamma}{x_i}) + \gamma log(\frac{\Gamma(1+\frac{1}{\gamma})x_i}{\psi_i}) - (\frac{\Gamma(1+\frac{1}{\gamma})x_i}{\psi_i})^{\gamma}$$

where $\Gamma()$ is the Gamma function. The Weibull distribution reduces to the exponential distribution if γ equals to 1. If $\gamma > 1$, the hazard function is a monotonically increasing function. If $0 < \gamma < 1$, the hazard function is a monotonically decreasing function.

3.4 Monte Carlo Simulation

In order to check if the high persistence of trade durations found in the linear ACD literature is due to structural shifts in the parameters, we conduct Monte Carlo simulations for separate parameter changes when the error term follows the exponential distribution.

3.4.1 Individual Shift in the Intercept Parameter ω

Firstly, we simulate a ACD(1,1) model with relatively low $\alpha + \beta$ value, and a small number of deterministic shifts in the intercept parameter ω. The data generating process is

$$\psi_i = \omega_j + \alpha x_{i-1} + \beta \psi_{i-1} \tag{3.6}$$

where ω_j is the intercept parameter in regime j. Let $\alpha = 0.05$ and $\beta = 0.55$ with $\alpha + \beta = 0.6$. we consider two kinds of shifts in ω. In the first experiment, there are two regimes in the series. The first regime is from the beginning to the midpoint of the time series with $\omega_1 = 0.5$. The intercept parameter decreases at the midpoint, with $\omega_2 = 0.3$ in the second regime. Other parameters remain constant. Note that we choose values of ω to ensure that the deseasonalized durations after adjusting for daily periodicity have a mean of approximately unity following Engle and Russell (1998). For actively traded stocks, hundreds or even thousands of durations exist for each calendar day. Thus, we choose sample sizes of 3000, 6000, 10000 and 30000. For all experiments, we use 2000 replications, as we find that 2000 replications, 5000 replications and 10000 replications lead to very similar mean and standard deviations of estimates.

Table 3.1 reports the single-regime exponential ACD (EACD) estimates when there is a unaccounted structural change for ω at the midpoint of the sample. In order to check the reliability of the estimates, the results without any parameter change are also generated for different sample sizes, with $\omega_1 = \omega_2 = 0.4$. The first 4 rows show that there is a small sample bias in the estimates. When the sample size is small, β is underestimated and ω is overestimated. As the sample size increases, the sample bias decreases. For sufficiently large sample size, i.e., when the sample size is 30000, the parameter estimates become very accurate. The standard deviation is high when the sample size is small, and becomes smaller gradually when sample size increases, especially when the sample size equals 30000.

Row 5 to row 8 report the estimates when $\omega_1 = 0.5$ in the first regime and decreases to 0.3 in the second regime. We find that even if the sample size is quite small, i.e., $N = 3000$, which is less than a week for actively traded stocks, there is a big difference between the estimates and the true values of the parameters. The sum of autoregressive parameter estimates converges to unity,

Table 3.1: Effects of one structural change in ω with $\alpha + \beta = 0.6$

ω_1	ω_2	sample size	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha} + \hat{\beta}$
0.4	0.4	3000	0.4360(0.217)	0.0514(0.018)	0.5129(0.224)	0.5643
0.4	0.4	6000	0.4266(0.162)	0.0508(0.013)	0.5228(0.168)	0.5736
0.4	0.4	10000	0.4173(0.121)	0.0506(0.010)	0.5320(0.126)	0.5826
0.4	0.4	30000	0.4047(0.067)	0.0502(0.006)	0.5452(0.070)	0.5954
0.5	0.3	3000	0.0008(0.0067)	0.0123(0.0051)	0.9869(0.0108)	0.9992
0.5	0.3	6000	0.0002(0.0003)	0.0082(0.0019)	0.9916(0.0021)	0.9998
0.5	0.3	10000	0.0001(0.0001)	0.0063(0.0010)	0.9936(0.0011)	0.9999
0.5	0.3	30000	0.0000(0.0000)	0.0036(0.0005)	0.9964(0.0005)	1.0000

Note: The change point is at the midpoint of the series. Standard deviations are reported in parentheses.

while the intercept estimate is close to zero. It implies that a unaccounted structural change in the intercept parameter leads to spuriously persistent autoregressive parameter.

While experiment 1 deals with the case that there is a permanent parameter change in the data generating process, the question of what effects a temporary parameter change may bring on the estimates is also interesting. Market microstructure literature has suggested the economics of trade duration clustering as follows. There are two types of traders: informed traders who can observe private information and only enter the market when receiving a private signal, and liquidity traders who arrive randomly. When some private information is released, informed traders will enter the market and trade until the information loses its value. The specialists learn of the information by observing order flow and adjust the price slowly. Therefore, it's reasonable to assume that after some private information releases, the frequency of trading will increase and the trade durations will decrease due to the actions of informed traders. When the information becomes public, even liquidity traders will increase the intensity of trading and thus the expectation of durations will decrease. After the information loses its value, the expectation of duration will return to its normal value.

The unconditional expectation of duration is given by $E(x_i) = \frac{\omega}{1-\alpha-\beta}$, which is an increasing function of the intercept parameter ω. So we model the change of expectation of durations by a parameter change in ω. We simulate the effects of a piece of private information release by assuming there are three regimes in the sample. In regime 1, which is the first half of the sample, no significant news releases and the intercept parameter is ω_1. In the middle of the sample, a private signal appears and the informed traders increase the intensity of trading. The decrease in the durations is modeled by the intercept parameter ω_2, which is less than ω_1. Regime 2, which is the fast transaction regime, lasts for a period that is not long, which is $1/30$ of the sample size. If the sample size is 30000, then $1/30$ of the sample size is 1000 durations, which takes 1 or 2 days to happen for actively traded stocks. After the information loses its value, the intercept parameter ω_3 returns to its previous value in regime 3 for the remaining $7/15$ of the sample.

Table 3.2 reports the effects of a temporary structural change in ω on the estimates. Similar to experiment 1, we set the true values as $\alpha = 0.05$ and $\beta = 0.55$. The first two rows show the estimates when $\omega_1 = 2\omega_2 = \omega_3$ and the

Table 3.2: Simulation results of a temporary change in ω with $\alpha + \beta = 0.6$

ω_1	$\omega_2(N/30)$	ω_3	sample size	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$
0.4068	0.2034	0.4068	12000	0.2043(0.136)	0.0451(0.019)	0.7507(0.153)
0.4068	0.2034	0.4068	30000	0.1857(0.115)	0.0426(0.019)	0.7718(0.133)
0.4091	0.1364	0.4091	12000	0.0155(0.059)	0.0193(0.012)	0.9652(0.072)
0.4091	0.1364	0.4091	30000	0.0010(0.009)	0.0106(0.003)	0.9885(0.012)

Note: Standard deviations are reported in parentheses.

mean of durations is 1. The estimate for β is highly positively biased. Row 3 and 4 shows that if $\omega_1 = 3\omega_2 = \omega_3$, then the sum of autoregressive parameter estimates converges to unity in large sample. The standard deviation decreases gradually when sample size increases. When the sample size equals 30000, the standard deviation to the autoregressive parameter is very small, if $\omega_1 = 3\omega_2 = \omega_3$. Although the impact of the news does not last for a long time, it has a big effect on the estimates of autoregressive parameters especially β. Thus, neglecting structural changes in the intercept parameter will lead to spurious high persistence in durations.

We conduct experiments for parameter changes in α and β as follows.

3.4.2 Individual Shift in α

In order to assess a single change at the midpoint of the sample in parameter α, we consider the ACD model with the following parameter values

$$\psi_i = 0.1 + \alpha_j x_{i-1} + 0.6\psi_{i-1} \tag{3.7}$$

In the first regime, $\alpha_1 = 0.1$, and it changes to other values which vary from 0.1 to 0.39 in the second regime. Table 3.3 reports the simulation results for

Table 3.3: Simulation results of a permanent change in α ($\omega = 0.1$ and $\beta = 0.6$)

α_1	α_2	$\overline{\alpha+\beta}$	sample size	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}+\hat{\beta}$
0.1	0.1	0.7	6000	0.102(0.019)	0.100(0.013)	0.595(0.066)	0.696
0.1	0.1	0.7	12000	0.102(0.013)	0.100(0.009)	0.595(0.046)	0.696
0.1	0.1	0.7	30000	0.100(0.008)	0.100(0.006)	0.599(0.029)	0.699
0.1	0.2	0.75	6000	0.071(0.011)	0.161(0.015)	0.667(0.037)	0.829
0.1	0.2	0.75	12000	0.071(0.008)	0.161(0.011)	0.668(0.027)	0.829
0.1	0.2	0.75	30000	0.071(0.005)	0.161(0.007)	0.669(0.017)	0.830
0.1	0.3	0.8	6000	0.042(0.006)	0.224(0.017)	0.707(0.024)	0.931
0.1	0.3	0.8	12000	0.042(0.004)	0.224(0.011)	0.708(0.016)	0.932
0.1	0.3	0.8	30000	0.042(0.003)	0.224(0.008)	0.708(0.010)	0.932
0.1	0.35	0.83	6000	0.035(0.004)	0.257(0.017)	0.705(0.019)	0.961
0.1	0.35	0.83	12000	0.034(0.003)	0.257(0.012)	0.705(0.014)	0.962
0.1	0.35	0.83	30000	0.034(0.002)	0.257(0.008)	0.706(0.009)	0.963

Note: Standard deviations are reported in parentheses.

different sample size. As jump size in α increases, the estimated sum of autoregressive parameters moves towards unity, although the convergence speed is much slower than changes in the constant ω. As $\psi_i = 0.1 + \alpha_j \epsilon_{i-1}\psi_{i-1} + 0.6\psi_{i-1}$, the effect of the α coefficient is multiplied by ϵ_{i-1}, and thus is smaller than the effect of the constant term and the β coefficient.

Table 3.4 reports the simulation results of a temporary change in α at the midpoint of the sample. In row 1 to 3, α jumps from 0.1 to 0.3 for a period of $1/30$ length of the sample size. The sample mean of $\alpha + \beta$ is 0.707, but the estimated sum of autoregressive parameters without considering the regime change is around 0.82. As jump size increases to 0.39, the estimated sum of autoregressive parameters increases to 0.902 in row 6, when the sample size is 30,000. Similar to previous experiments, the standard deviation decreases gradually when sample size increases. When the sample size equals 30000, the standard deviation to the autoregressive parameter is very small, especially

Table 3.4: Simulation results of a temporary change in α ($\omega = 0.1$ and $\beta = 0.6$)

α_1	$\alpha_2(N/30)$	α_3	$\overline{\alpha+\beta}$	sample size	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}+\hat{\beta}$
0.1	0.3	0.1	0.707	6000	0.063(0.016)	0.115(0.016)	0.704(0.058)	0.819
0.1	0.3	0.1	0.707	12000	0.061(0.011)	0.115(0.011)	0.712(0.041)	0.827
0.1	0.3	0.1	0.707	30000	0.059(0.008)	0.114(0.007)	0.716(0.027)	0.830
0.1	0.39	0.1	0.710	6000	0.042(0.011)	0.125(0.019)	0.760(0.045)	0.885
0.1	0.39	0.1	0.710	12000	0.038(0.008)	0.120(0.014)	0.775(0.032)	0.895
0.1	0.39	0.1	0.710	30000	0.036(0.005)	0.118(0.009)	0.784(0.021)	0.902

Note: Standard deviations are reported in parentheses.

when there is a large change in parameters.

3.4.3 Individual Shift in β

To investigate changes in α, we consider the ACD (1,1) model with conditional expected duration process given by

$$\psi_i = 0.1 + 0.1x_{i-1} + \beta_j \psi_{i-1} \tag{3.8}$$

For a permanent change in β, there are two segments and the change point is at the midpoint of sample. Table 3.5 shows the simulation results. The estimated sum $\hat{\alpha} + \hat{\beta}$ moves towards unity with growing jump size. In row 6 to 9, when the sample mean of $\alpha + \beta$ is 0.8, the estimated sum of autoregressive parameters $\hat{\alpha} + \hat{\beta}$ is larger than 0.99. The increase in sample size does not have much influence on the estimates.

For a temporary change in β, there are three segments and the first change point is at one third of the sample. The second segment is relatively short as its length is $\frac{1}{30}$ of the sample, and then for the third segment the value of β returns to previous value 0.6. Table 3.6 shows the simulation results. Similar to

Table 3.5: Simulation results of a permanent change in β ($\omega = 0.1$ and $\alpha = 0.1$)

β_1	β_2	$\overline{\alpha+\beta}$	sample size	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}+\hat{\beta}$
0.6	0.6	0.7	6000	0.102(0.018)	0.100(0.013)	0.595(0.063)	0.696
0.6	0.6	0.7	12000	0.102(0.013)	0.100(0.010)	0.595(0.045)	0.696
0.6	0.6	0.7	30000	0.100(0.008)	0.100(0.006)	0.599(0.028)	0.699
0.6	0.7	0.75	6000	0.037(0.016)	0.089(0.023)	0.821(0.060)	0.910
0.6	0.7	0.75	12000	0.037(0.011)	0.090(0.017)	0.822(0.044)	0.912
0.6	0.7	0.75	30000	0.036(0.007)	0.091(0.009)	0.822(0.026)	0.913
0.6	0.8	0.8	6000	0.004(0.001)	0.059(0.015)	0.935(0.017)	0.994
0.6	0.8	0.8	12000	0.003(0.001)	0.057(0.011)	0.938(0.013)	0.995
0.6	0.8	0.8	30000	0.003(0.001)	0.055(0.008)	0.940(0.009)	0.995

Note: Standard deviations are reported in parentheses.

Table 3.6: Effects of a temporary change in β ($\omega = 0.1$ and $\alpha = 0.1$)

β_1	$\beta_2(N/30)$	β_3	$\overline{\alpha+\beta}$	sample size	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}+\hat{\beta}$
0.6	0.8	0.6	0.707	6000	0.041(0.013)	0.094(0.017)	0.789(0.055)	0.883
0.6	0.8	0.6	0.707	12000	0.038(0.009)	0.091(0.013)	0.800(0.040)	0.891
0.6	0.8	0.6	0.707	30000	0.036(0.006)	0.089(0.009)	0.809(0.026)	0.898
0.6	0.89	0.6	0.710	6000	0.022(0.004)	0.100(0.015)	0.842(0.026)	0.942
0.6	0.89	0.6	0.710	12000	0.017(0.003)	0.086(0.012)	0.870(0.020)	0.956
0.6	0.89	0.6	0.710	30000	0.012(0.002)	0.070(0.008)	0.899(0.015)	0.969

Note: The change point is at the midpoint of the sample. Standard deviations are reported in parentheses.

the permanent change, the estimated sum of autoregressive parameters moves towards unity as jump size increases.

3.5 Theoretical Results

The previous section shows that both unaccounted permanent and temporary structural changes in the data generating process can lead to spuriously high autocorrelation parameter. In this section, we derive the unconditional mean of the ACD (1,1) model when there are unaccounted structural changes in the data generating process. The results are inspired by Hillebrand (2005), in which the GARCH model with neglected parameter changes are considered. Suppose that there are $K-1$ unaccounted switches in the parameters of the data generating process. In the k*th* segment, the conditional duration process is

$$\psi_i = \omega_k + \alpha_k x_{i-1} + \beta_k \psi_{i-1} \qquad i = I_{k-1}, ..., I_k \tag{3.9}$$

where $k = 1, 2, ..., K$, $I_0 = 0$ and $I_K = I$. The first observation in the k*th* segment is I_{k-1}, and the last observation is I_k. The whole sample is estimated ignoring the structural changes as

$$\psi_i = \omega + \alpha x_{i-1} + \beta \psi_{i-1} \tag{3.10}$$

Then we have:

Proposition 3.1. *Assume that for $k = 1, 2, ..., K$, the segment length $I_k - I_{k-1} \to \infty$ as the sample size $I \to \infty$. Let $E(\psi_i(k))$ denote the unconditional expected value of the conditional duration within segment k. If ψ_i is generated by equation (3.9) and estimated by (3.10), then the sample mean of conditional duration is*

$$\overline{\psi} = \frac{1}{I}\sum_{k=1}^{K}(I_k - I_{k-1})E(\psi_i(k)) + o(1)_I \tag{3.11}$$

where $o(1)_I \to 0$ as $I \to \infty$.

The proof of Proposition 3.1 is in Appendix C. It implies that the unconditional mean of the conditional duration process is a weighted average of the unconditional mean of each segment, if the ACD model has different parameter values in each segment of the duration process. Although the proposition applies when the segment length converges to infinity, we find that the unconditional expectations follow the results of the proposition in all the simulation experiments of the previous section. For example, when there is a temporary change in the autoregressive parameter β in the previous section, we find that the unconditional mean of the duration process is the weighted average of the unconditional mean in the first segment, the second segment and the third segment of data, taking the segment size as weight.

3.6 Threshold ACD model and Empirical Estimation

Having shown the effects of ignoring possible regime shifts, we now consider modeling them via a threshold ACD model. Further, in this section we will use Boeing duration data to estimate the standard ACD model along with the threshold ACD model and compare the results.

In the threshold ACD (1,1) model with two regimes, the conditional expected duration process is given by

$$\psi_i = \omega_1 + \alpha_1 x_{i-1} + \beta_1 \psi_{i-1} \qquad \text{for } 0 < y_{i-1} \leq y_0 \tag{3.12}$$

$$\psi_i = \omega_2 + \alpha_2 x_{i-1} + \beta_2 \psi_{i-1} \qquad \text{for } y_{i-1} > y_0 \tag{3.13}$$

The regime-switching ACD parameters satisfy that $\omega(j) > 0$, $\alpha_j \geq 0$ and $\beta_j \geq 0$, for $j = 1, 2$. The regime or state of the process is determined by the threshold variable, which can be either endogenous or exogenous. y_{i-1} is the threshold variable, which can be observed one period earlier, and y_0 is the threshold value. The threshold variable y_{i-1} could be the lagged duration, the associated volume of the trade, or some economic variables. The probability of regime 1 is given by $P(y_{i-1} \leq y_0) = 1 - \pi$, and the probability of regime 2 is given by $P(y_{i-1} > y_0) = \pi$. For different regimes, the conditional mean and duration persistence are different.

Conditions for geometric ergodicity and existence of moments, when the last period duration is used as the threshold variable, can be derived for TACD (1,1) models, as shown in Zhang, Russell, and Tsay (2001). However, it's difficult to generate the existence of moments conditions for higher order TACD models. The covariance stationarity condition of the single-regime ACD model may not apply here, as it's possible for the duration process to be stationary in one regime, and not stationary in the other regime, but still stationary for the threshold duration process. Following Knight and Satchell (2009)'s work about the stationarity condition of a threshold autoregressive process, we can derive the unconditional mean of the duration and the stationarity condition for the TACD (1,1) model given that the threshold variable is exogenous.

Proposition 3.2. *If the threshold variable is exogenous,* $\omega_1 < \infty$*,* $\omega_2 < \infty$ *and* $(\alpha_1 + \beta_1)(1 - \pi) + (\alpha_2 + \beta_2)\pi < 1$*, then the unconditional mean of the duration exists and is given by*

$$E(x_i) = \frac{\omega_1(1 - \pi) + \omega_2\pi}{1 - (\alpha_1 + \beta_1)(1 - \pi) - (\alpha_2 + \beta_2)\pi} \tag{3.14}$$

The conditional duration process is stationary if $|(\alpha_1+\beta_1)|(1-\pi)+|(\alpha_2+\beta_2)|\pi < 1$.

The proof of the proposition is very similar to the proof of Proposition 2.2 in Chapter 2. It shows that for the TACD (1,1) model, the unconditional mean of the duration depends on the parameters and probabilities of both regimes. It is possible for the TACD (1,1) model to be stationary if $\alpha_1 + \beta_1 > 1$ in regime 1 and $\alpha_2 + \beta_2 < 1$ in regime 2, and vice versa.

The Boeing transaction duration data is from September 1, 2000 to October 31, 2000, which contains 90136 observations after the seasonal adjustments to remove the diurnal pattern of daily trading activities and the day-of-week effects. [1]To eliminate the diurnal pattern, 13 knots are chosen over each trading day with 30 minutes apart. The first knot is at 10:00 AM to remove the effect of opening auction, and the last one is at 4:00 PM. The duration at each knot is computed by averaging the durations around the knot within 15-minutes either side. Then the daily seasonal factor is computed as the average of the duration at each knot in the two-month sample period. The adjusted duration is the duration divided by the daily seasonal factor. Figure 3.1 plots the adjusted duration for the whole sample. As Figure 3.1 contains too many observations and thus is hard to observe the pattern of durations, Figure 3.2 plots the first 5000 adjusted durations. The mean of the adjusted duration is 1. Figure 3.3 shows the autocorrelation function of the duration series. It clearly shows that there exists temporal dependence in the duration series.

[1]I would like to thank Dinghai Xu and John Knight for providing the cleaned data set. For detailed description of the adjustment, see Knight and Ning (2008).

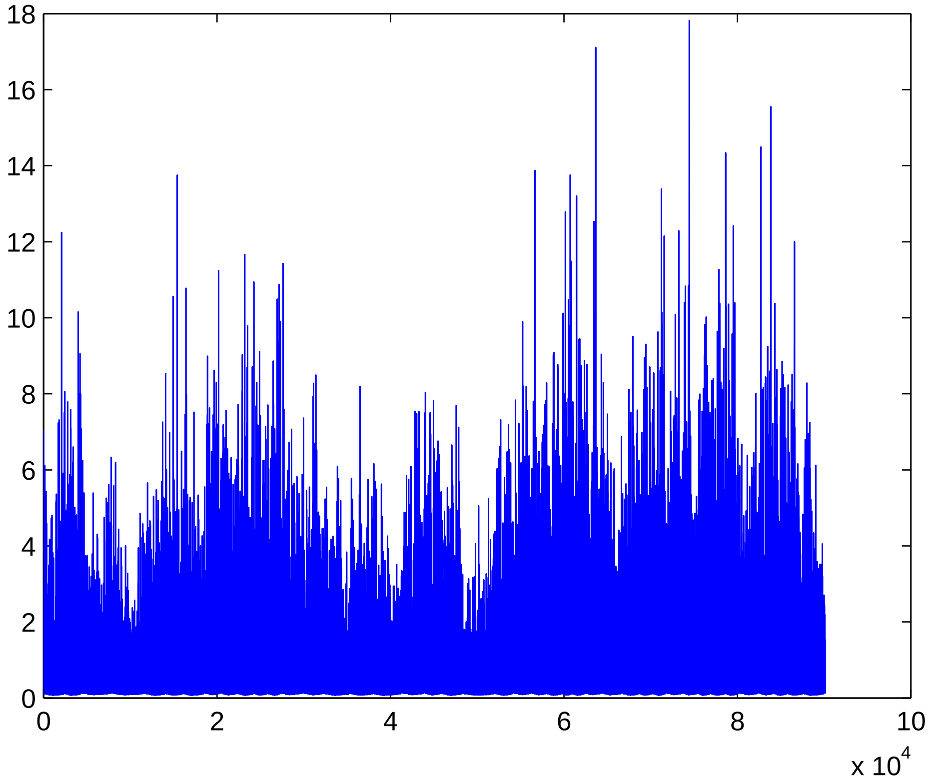

Figure 3.1: Adjusted duration from September 1, 2000 to October 31, 2000

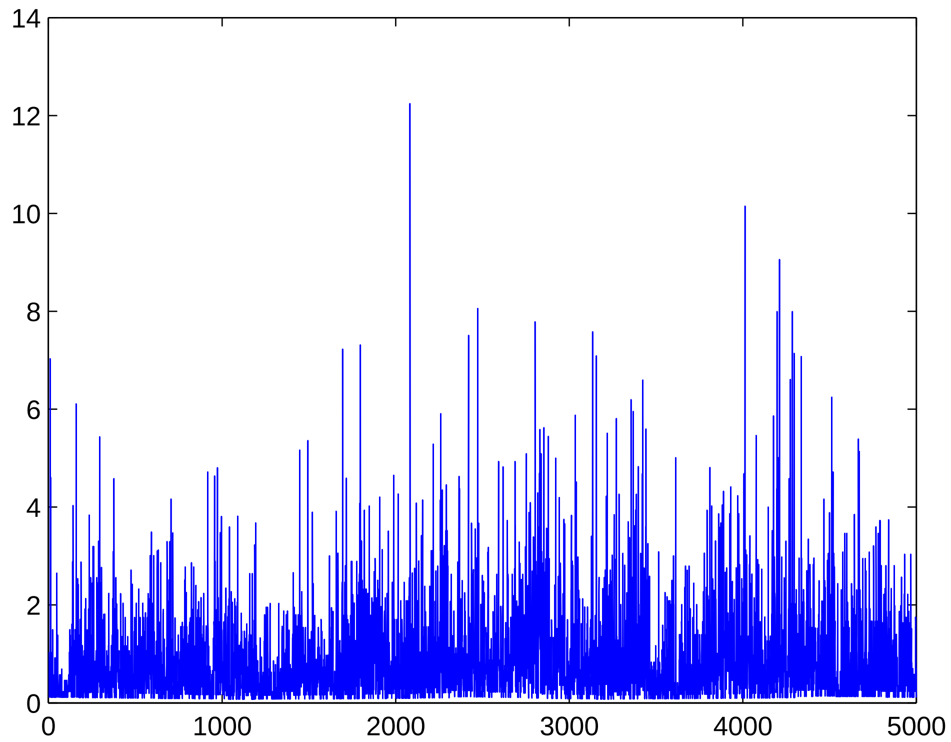

Figure 3.2: Adjusted duration with sample size of 5000

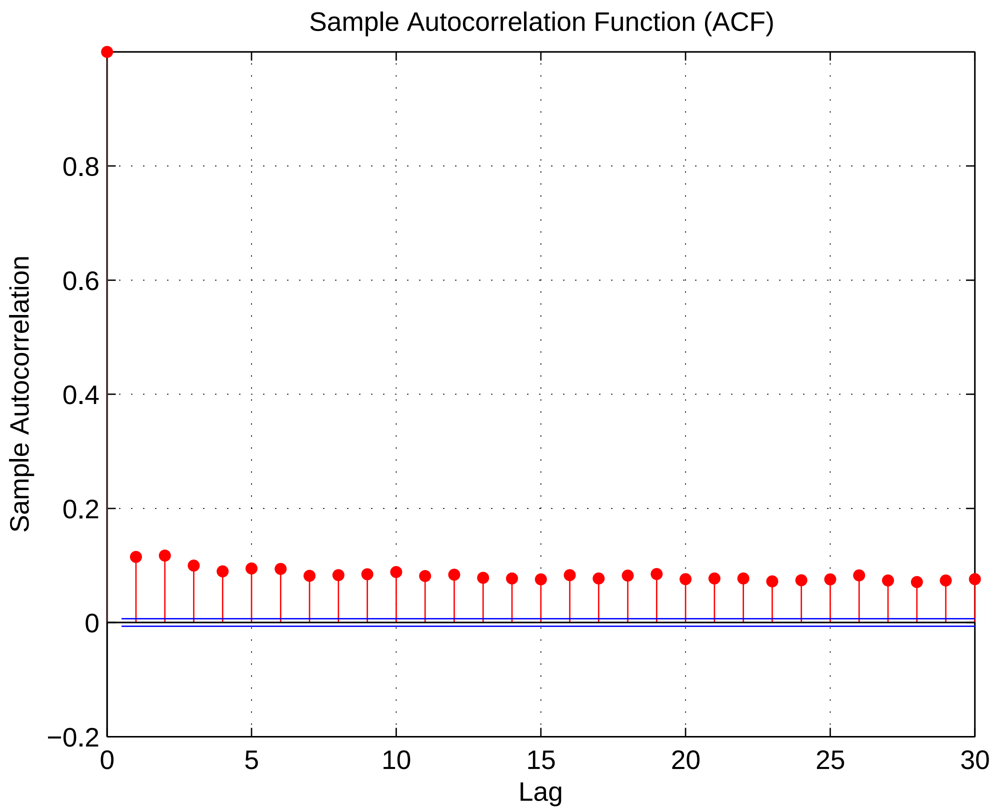

Figure 3.3: Sample autocorrelation function of Boeing transaction durations from September 1, 2000 to October 31, 2000

For Boeing duration data, we use the lagged duration x_{i-1} as the threshold variable, such that the threshold variable is observed in the last period. Maximum likelihood estimation is adopted for the ACD model and the threshold ACD model. The threshold value is estimated by grid search, as the sample of the threshold variable is divided into 40 intervals with 39 grid points from 2.5 percentile point to 97.5 percentile point. At each grid point, the likelihood value is calculated and compared with each other to find the maximum likelihood value. Then the corresponding grid point is the estimate of the threshold value. Table 3.7 reports the estimates, corresponding Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) of the TACD (1,1) model and the single-regime ACD (1,1) model when the innovations follow the exponential distribution. All the parameters are significant except the intercept parameter in regime 1 of the TACD model. For the ACD model, $\alpha + \beta = 0.997$, which is very close to 1. We find that the TACD model has higher value of $\alpha + \beta$ in regime 1 than the single-regime ACD model, and lower value of $\alpha + \beta$ in regime 2 than the single-regime ACD model. The TACD model has lower values of both AIC and BIC than the single-regime ACD model, implying that the exponential TACD model fits the data better than the exponential ACD model.

Table 3.8 reports the estimates of the TACD (1,1) model and the single-regime ACD (1,1) model when the innovations follow the Weibull distribution. In the Weibull TACD model, the shape parameter can be different in regime 1

Table 3.7: Estimation results of exponential ACD model and exponential TACD model

parameters	ACD$(1,1)$	TACD$(1,1)$
ω_1	0.003(0.001)	0.000(0.001)
α_1	0.028(0.001)	0.038(0.001)
β_1	0.969(0.001)	0.962(0.002)
ω_2		0.011(0.001)
α_2		0.024(0.001)
β_2		0.971(0.001)
y_0		0.336
π		0.7
Log-likelihood	-85992	-85933
AIC	171990	171880
BIC	172020	171950

Note: Standard errors are reported in the parentheses.

and regime 2. That is, the hazard function of trading duration can be different in different regimes, which is more favorable in practice. We find that the estimates of autoregressive parameters are very similar in the Weibull ACD model and the exponential ACD model; however, the scale parameter in the Weibull distribution does not equal to one. Similar to the case of the exponential TACD model, the Weibull TACD model has higher value of $\alpha + \beta$ in regime 1 than the single-regime Weibull ACD model, and lower value of $\alpha + \beta$ in regime 2 than the single-regime Weibull ACD model.

3.7 Conclusion

If an ACD model is estimated on a series of durations which contains parameter changes in the conditional expected duration process and these parameter changes are not accounted for, the observed high persistence is spurious. The

Table 3.8: Estimation results of Weibull ACD model and Weibull TACD model

parameters	ACD$(1,1)$	TACD$(1,1)$
ω_1	0.003(0.001)	0.001(0.001)
α_1	0.028(0.001)	0.036(0.001)
β_1	0.970(0.001)	0.963(0.002)
γ_1	1.081(0.003)	1.071(0.005)
ω_2		0.010(0.001)
α_2		0.023(0.001)
β_2		0.971(0.001)
γ_2		1.086(0.003)
y_0		0.336
π		0.7
Log-likelihood	-85519	-85453

Note: Standard errors are reported in the parentheses.

summation of the estimates of the autoregressive parameters of the conditional expected duration process is overestimated and converges to one. In Monte Carlo simulations of the ACD model, we show that the effect occurs for realistic sample sizes and jump sizes in financial market duration series. Monte Carlo simulation experiments are conducted for both permanent changes and temporary changes in the intercept and slope parameters, α, β and ω. Even a temporary change in the parameters for a relatively short time period can lead to a big bias in the estimates of the autoregressive parameters. The sample mean of the conditional expected duration is derived for the ACD model with unaccounted structural changes.

Finally, we estimate Boeing transaction duration data using both the ACD model and a threshold ACD model, when the innovation follows the exponential distribution or the Weibull distribution. We find that the threshold model fits the data better than the single-regime ACD model.

3.8 Bibliography

BAILLIE, R., T. BOLLERSLEV, AND M. H.O. (1996): "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," 52, 91–113.

BAUWENS, L., AND V. D (2004): "The stochastic conditional duration model: a latent factor model for the analysis of financial durations," *Journal of Econometrics*, 119, 381–412.

BOLLERSLEV, T., AND W. J. (1992): "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances," 11, 143–172.

DROST, F., AND B. WERKER (2004): "Semiparametric Duration Models," 22, 40–50.

EASLEY, D., AND M. O'HARA (1992): "Time and the process of security price adjustment," *The Journal of Finance*, 19, 69–90.

ENGLE, R., AND J. RUSSELL (1998): "Autoregressive conditional duration: a new model for irregularly spaced transaction data," *Econometrica*, 66, 1127–1162.

GRAMMIG, J., AND M. K. (2000): "Non-Monotonic Hazard Functions and the Autoregressive Conditional Duration Model," 3, 16–38.

GRAMMIG, J., AND K. MAURER (2000): "Non-monotonic hazard functions and the autoregressive conditional duration models," *Econometrics Journal*, 3, 16–38.

HILLEBRAND, E. (2005): "Neglecting parameter changes in GARCH models," *Journal of Econometrics*, 129, 121–138.

JASIAK, J. (1998): "Persistence in intertrade durations," *Finance*, 19, 166–195.

KNIGHT, J., D. XU, AND T. WIRJANTO (2011): "Asymmetric Stochastic Conditional Duration Model- A Mixture-of-Normal Approach," *Journal of Financial Econometrics*, 9(3), 469–488.

KNIGHT, J. L., AND C. NING (2008): "Estimation of the Stochastic Conditional Duration Model via Alternative Methods," 11, 573–592.

LAMOUREUX, C., AND W. LASTRAPES (1990): "Persistence in variance, structural change, and the GARCH model," *Journal of Business and Economics Statistics*, 8, 225–234.

MEITZ, M., AND T. TERäSVIRTA (2006): "Evaluating Models of Autoregressive Conditional Duration," 24, 104–124.

O'HARA, M. (1995): *Market microstructure theory*. Oxford: Basil Blackwell Inc.

PACURAR, M. (2008): "Autoregressive condtional duration models in finance: a survey of the theoretical and empirical literature," *Journal of Economic Surveys*, 22(4), 711–751.

ZHANG, M., J. RUSSELL, AND R. TSAY (2001): "A nonlinear autoregressive conditional duration model with applications to financial transaction data," *Journal of Econometrics*, 104, 179–207.

Chapter 4

Value-at-Risk Estimation via Realized Higher Moments using High Frequency Data

4.1 Introduction

The financial crisis in 2008 has reinforced the importance of accurate risk management for financial institutions. Value-at-Risk (VaR) is the most widely used risk measurement tool in the banking sector, and consequently receives a great deal of attention. According to the 1996 market risk amendment to the Basel Accord by the Basel Committee on Banking Supervision at the Bank for International Settlements, banks and investment firms must meet the capital requirements to cover the losses on the trading portfolio over a 10-day holding period for 99% of the time. This 1% quantile of the return distribution, is referred to as the 99% VaR. More specifically, the $(1-\alpha)$ VaR for period $t+h$ of a portfolio, conditional on the information set given up to time t, is defined as the α-quantile of the return distribution. The formula is expressed as

$$VaR_{t+h}^{1-\alpha} = min\{x|P(-r_{t+h} \geq x|I_t) \leq \alpha\} \tag{4.1}$$

where r_t is the return on the asset or portfolio at time t, $1 - \alpha$ is the confidence level, and I_t denotes the information set up to time t. Typical values for $1 - \alpha$ are 95%, 97.5%, 99% and 99.5%. Banks often report daily VaR forecasts and choose their own internal VaR models, however, using inaccurate models can lead to a penalty of holding more capital, which hurts the bank's profitability as less money is available for use. Thus, having an accurate measurement of VaR is of great importance.

To calculate the VaR of a portfolio, one way is to aggregate the profit and loss of the portfolio and use a univariate forecasting model, while the other approach is to proceed with a multivariate model. In many situations, such as for the index-tracking funds, a complicated multivariate model is not able to outperform a simple univariate model using the aggregate data, according to Berkowitz and O'Brien (2002). In this chapter we focus on univariate VaR models to characterize risk at an aggregated level. There are three basic methods of calculating VaR, the parametric models, historical simulation which computes empirical quantiles based on past data, and Monte Carlo simulation which simulates market movements and evaluates the empirical distribution of portfolios priced along the movements. The most widely used full parametric approach is entitled as RiskMetrics, which has been shown to provide adequate out-of-sample forecasting performance. The central part in RiskMetrics method is to use an integrated GARCH model to compute the volatility of the return as a function of past daily squared returns. Furthermore, the availability of large databases and ultra-high frequency trading data nowadays has led to the notion of realized volatility, which is computed as the sum of squared intra-day returns for the given trading day. Applying realized volatility to compute Value-at-Risk, Giot and Laurent (2004) use a long-memory skewed student t model for the daily realized volatility and find it provides adequate one-day-ahead VaR forecasts for two stock indexes and exchange rates, which is equal to the performance of the skewed student APARCH model.

Many VaR models assume that the dependence structure of the return can be fully described by the mean and volatility, such as the GARCH type models and the realized volatility models. However, VaR is not only related to volatility for the following reasons. This specification has two major drawbacks. Firstly, financial data has a pattern of non-zero skewness and kurtosis larger than 3, which is related to higher moments and cannot be captured only by the mean and variance. Secondly, recent studies suggest that higher moments of returns are time varying and dependent. Harvey and Siddique (1999) propose a model with time-varying conditional skewness based on a non-central student t distribution and indicate evidence of autoregressive conditional skewness. Furthermore, Brooks *et al.* (2005) develop a model for conditional kurtosis and find the presence of autoregressive conditional kurtosis. Consequently, it is inadequate to assume the skewness and kurtosis to be constant or only dependent on the parameters of a prespecified distribution such as a student-t distribution.

As the mean and volatility cannot fully capture the characteristics of financial return series, in this chapter we incorporate higher moments through Cornish Fisher approximations. The Cornish Fisher expansion is a methodology to approximate the quantile of a random variable directly via correction for skewness and kurtosis and has been used in Jaschke (2001) and Knight, Satchell and Wang (2003) to calculate VaR, which is the α-quantile of the return distribution. However, only constant skewness and kurtosis of the full sample are used in the literature, which cannot capture the dynamics of higher moments. Consequently, in this chapter, the time-varying volatility, skewness and kurtosis are incorporated into the Cornish-Fisher approximations.

We investigate whether the incorporation of higher moments can capture the heavy tail of the return distribution and provide more accurate VaR forecasts before, during, and after the financial crisis in 2007 and 2008. Instead of assuming that higher moments rely heavily on the model specification, we use realized moments constructed from intra-day data. The realized moments are computed as the sum of powers of intraday returns, with the realized volatility being a special case of realized moments, i.e., the realized second moment. Firstly we investigate the performance of VaR estimates using *ex post* realized moments and compare it with VaR estimates using *ex post* realized volatility. Because realized moments are *ex post* measures of the population moments, we need to provide one-step-ahead forecasts of them in order to forecast VaR. We propose two new models of realized moments. The first model is the realized moments-exponentially weighted moving average (RM-EWMA) model, which extends the frequently used EWMA model in the VaR literature. We find that both the realized volatility and logarithmic realized fourth moment have some predicability that can be captured by the EWMA procedure. Moreover, similar to the characteristics of logarithmic realized volatility, logarithmic realized fourth moment is found to be an autoregressive long memory process. Therefore, we employ the autoregressive fractionally integrated moving average (ARFIMA) model for both the logarithmic realized volatility and logarithmic realized fourth moment as the second model, which is named the realized moments-ARFIMA (RM-ARFIMA) model. Then we use the Cornish Fisher approximation to compute VaR and compare the out-of-sample forecasting performance of these models.

The remainder of the chapter is organized in the following way. Section 4.2 presents the methodology of Cornish Fisher expansion. Section 4.3 presents the definition and theoretical results of the realized moments. In addition, section 4.3 also describes the data used for estimation, and investigates the empirical properties of the realized moments. Section 4.4 proposes the RM-EWMA model and the RM-ARFIMA model. In section 4.5 we discuss the evaluation procedure and the empirical results. Finally, section 4.6 concludes.

4.2 VaR via Cornish-Fisher Expansion

The unknown distribution function of return can be approximated by some known distribution with certain properties, such as the normal distribution or the Student's t distribution, or the Edgeworth expansion to approximate the density function after incorporating skewness and kurtosis. However, the Edgeworth expansion perform poorly in the tails of the distribution since the pdf is not guaranteed to be positive. Since VaR is a tail quantile of the distribution, the Edgeworth expansion is not suitable. Rather than approximating the distribution we can adjust the quantile via correction for skewness and kurtosis. This approximation is achieved by the use of a Cornish-Fisher expansion.

Cornish and Fisher (1937) derived an expansion for approximating the α-quantile of a random variable X based upon its cumulants or moments. Using the mean, variance, skewness and kurtosis, the corresponding VaR, given the significance level α, is expressed as

$$\hat{V_{\alpha,t}} = \mu_t + \sigma_t(z_\alpha + \frac{1}{6}(z_\alpha^2 - 1)k_{3,t} + \frac{1}{24}(z_\alpha^3 - 3z_\alpha)(k_{4,t} - 3) - \frac{1}{36}(2z_\alpha^3 - 5z_\alpha)k_{3,t}^2) \quad \textbf{(4.2)}$$

where z_α denotes the α-quantile of the standard normal distribution. μ_t, σ_t, $k_{3,t}$, $k_{4,t}$ are respectively the mean, standard deviation, skewness and kurtosis of the distribution of X_t. The Cornish-Fisher approximation avoids the complicated problem of estimating quantiles of a parametric distribution. Jaschke (2001) discusses the accuracy and computational efforts of the Cornish-Fisher expansion in the context of the Delta-Gamma approximations, concluding that this method leads to accurate results and is computationally faster than other methods like numerical Fourier inversion. Knight, Satchell and Wang (2003) use the Cornish-Fisher approximation for inclusion of skewness and kurtosis and propose a LINEX VaR procedure to adjust the forecasts when the loss functions of the forecaster are asymmetric. The sample skewness and kurtosis are used in the literature, which are constant in the full sample. As the skewness and kurtosis are also time-varying, in this chapter we use the time-varying realized moments instead of the sample moments in the approximation.

4.3 Realized Moments Measurement

4.3.1 Realized Moments Model to Calculate VaR

Over the last two decades, the rapid development in computer technology has led to the availability of high frequency data or intraday data in the financial market. Using intraday data, we are able to extract more information about the characteristics of the return series, such as the realized volatility developed by Andersen and Bollerslev (1998) and Barndorff-Nielsen and Shephard (2001). Constructed as the summation of squared intraday returns, realized volatility is a non-parametric measure of the unknown volatility, and it's a better proxy for the true volatility than the squared return.

While realized volatility is the summation of squared intra-day returns and converges to the quadratic variation as the sampling frequency increases, the sub-sampling methodology can also be applied to higher moments. As the distribution of return is well known to be skewed and leptokurtic, it is very important to take into consideration skewness and kurtosis, or the related third and fourth moments of returns. However, very little attention has been paid to realized higher moments in the previous literature to the best of our knowledge, except the work of Beine, Laurent and Palm (2009) and Amaya et. al (2011).

The third power and the fourth power of daily return are very noisy and thus are not efficient estimates of the daily third moment and fourth moment of return. Aiming to solve this problem, realized moments are constructed from intraday data such as 5-minute return. Let $P_{d,t}$ denote the dth intra-day price observation of the trading day t, for $d = 1, 2, ..., D$. Then the d-*th* return on day t, $r_{d,t}$, is computed as $r_{d,t} = 100(\ln P_{d,t} - \ln P_{d-1,t})$, for $d = 1, 2, ..., D$, and $t = 1, 2, ..., T$. The realized ith moment on day t can be defined by summing up the ith powers of intraday returns during the open-market period.

$$RM(i)_t = \sum_{d=1}^{D} r_{d,t}^i \tag{4.3}$$

for $i = 2, 3, 4, ...$. Realized variance is a special case of the realized moments when $i = 2$, which is widely discussed in the literature.

Amaya et. al (2011) shows that under realistic assumptions of an affine jump-diffusion process with stochastic volatility, the realized moments converges in mean square to the integrated moments for $i = 1, 2, 3, 4$.

Proposition 4.1. *(Amaya et. al (2011), Proposition 1) If the log price process* p_t

is defined by

$$dp_t = (\mu - V_t - \mu_J\lambda)dt + \sqrt{V_t}dW_t + JdN_t \tag{4.4}$$

$$dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dW_t \tag{4.5}$$

with μ as the drift, κ as the mean reversion speed to long-term volatility mean θ, and σ the diffusion coefficient of the volatility process V_t. N_t is an independent Poisson process with jump intensity λ and the distribution of jump size J is $N(\mu_J, \sigma_J^2)$. Then the realized moments converges in mean-square to

$$RM(3) \xrightarrow{m.s.} \lambda(\mu_J^3 + 3\mu_J\sigma_J^2)$$
$$RM(4) \xrightarrow{m.s.} \lambda(\mu_J^4 + 6\mu_J^2\sigma_J^2 + 3\sigma_J^4)$$

The proof is in Amaya et. al (2011). While the limit of realized volatility depends on both diffusion and jump parameters, the limits of realized third and fourth moments are solely relevant to jump parameters. Conducting a Monte Carlo simulation to allow for market microstructure noise, Amaya et. al (2011) find that although realized volatility is dominated by the variance of noise as sampling frequency increases, the realized third and fourth moments are not affected.

As the realized moments converge in mean-square to the integrated moments, we can compute the realized moments as estimators of the integrated moments, which contain a lot of information about the distribution of return, such as the skewness and kurtosis. As the realized moments are obtained, the realized skewness and kurtosis are calculated as:

$$k_{3,t} = \frac{\sqrt{D} RM(3)_t}{RM(2)_t^{\frac{3}{2}}} \tag{4.6}$$

$$k_{4,t} = \frac{D \times RM(4)_t}{RM(2)_t^2} \tag{4.7}$$

Using *ex post* realized second moment, skewness and kurtosis, we can calculate VaR via Cornish-Fisher expansion as given in equation (4.17). We call this model the realized moments (RM) model to calculate VaR. However, as these realized measures are obtained at day t instead of at day $t-1$, the RM model cannot be used to forecast VaR. In order to provide out-of-sample VaR forecasting using information conveyed by the realized moments, it is also necessary to provide out-of-sample one-day-ahead forecasts for daily realized moments $RM(i)$ first. In order to forecast the realized moments, we investigate the properties of the realized moments in the following subsection.

4.3.2 Data and Empirical Properties of Realized Moments

To study the properties of realized higher moments, we use one of the most actively traded U.S. equities, IBM stock prices, from Jan 3rd, 2005 to May 27th, 2011. 5-minute returns starting at 9:35 am and ending at 16:00 each weekday are used to construct the realized moments. In 6.5 trading hours we have 78 5-minute observations each day. The sample contains 1600 daily observations.

The trading days with less than 50 5-minute price observations are deleted because of low market activity. In order to investigate the performance of VaR models before, during and after the financial crisis, the sample is divided into 3 sub-periods. The first period is the pre-financial crisis period, from Jan 3rd, 2005 to November 30th, 2007, which contains 728 observations. The second period is the financial crisis period, from December 3rd, 2007 to March 31st, 2009, containing 330 observations. The beginning and end of the crisis period correspond roughly to the peak and trough of the U.S. stock indices, which covers the most volatile period of the financial market. The last period is the post-crisis period from April 1st, 2009 to May 27th, 2011, which contains 542 observations.

The daily return is defined as 100 times the log difference of daily closing price. Table 4.1 list the summary statistics for the daily return observations of the full sample and the three sub-samples. Figure 4.1 plots the daily return of IBM stock for the full sample period. The average return is negative during the crisis period, while it is positive in the other two periods. The standard deviation of return during the crisis period is about twice of the other periods, implying a much larger volatility during the financial crisis period. Interestingly, we find that the kurtosis in the crisis period is smaller than the kurtosis in the pre-crisis period. The null hypothesis of normality is strongly rejected for all sub-periods.

Table 4.1: Summary statistics of daily IBM return

	full sample	pre-crisis	crisis	post-crisis
mean	0.036	.012	-.025	.105
st.dev.	1.467	1.110	2.352	1.153
skewness	-.005	-.882	.253	.021
kurtosis	8.827	10.089	4.940	4.912
min	-8.719	-8.719	-6.485	-5.051
max	11.063	4.392	11.063	4.409
Jarque-bera(p-value)	0.000	0.000	0.000	0.000

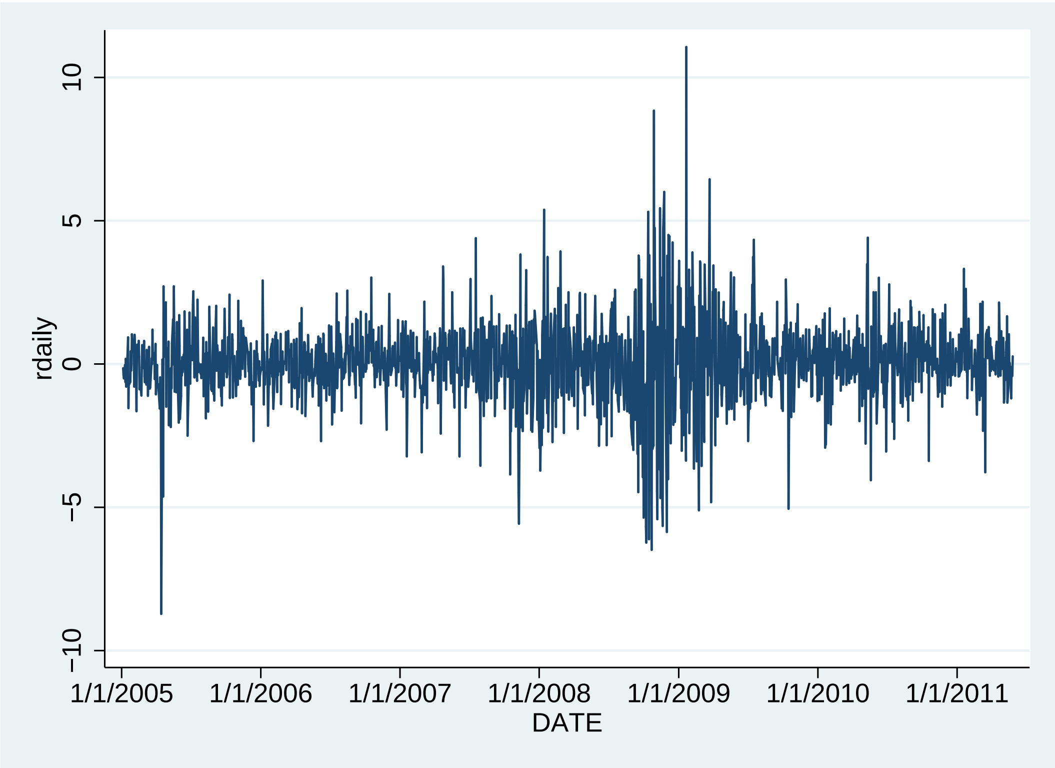

Figure 4.1: Time series of IBM daily return

For each day of each period we calculate the realized moments and Table 4.2 provides the summary statistics for the unconditional distribution of these moments for the full sample period and the three subsamples. The mean of realized volatility is much higher in the crisis period than in the other two periods. Similarly, the mean of realized third moment and the mean realized fourth moment are higher in the crisis period. Realized volatility and realized fourth moment are positively skewed and leptokurtic in all three sub-periods. Realized third moment is positively skewed in the crisis period, and negatively skewed in the other two sub-periods. The post-crisis period is relatively more volatile than the pre-crisis period. The Portmanteau (Q) test statistics for white noise is also reported, showing that the white noise assumption is strongly rejected for realized volatility in all sub-periods. The assumption of white noise cannot be rejected for realized third moments at 95% confidence interval. For realized fourth moment, the white noise assumption is rejected for the pre-crisis and crisis period, but cannot be rejected for the post-crisis period.

Figure 4.2, 4.3 and 4.4 plot the time series of daily realized volatility, realized third and fourth moments respectively. The realized volatility is quite tranquil before year 2007 and increases rapidly in year 2007, which is in correspondence with the bursting of the U.S. housing bubble. The realized volatility reaches its peak in September 2008, in correspondence with the most volatile period in the financial crisis. The peak happened in May 6, 2010 for realized

Table 4.2: Summary statistics for unconditional distributions of realized volatility, RM(3) and RM(4)

	RV	RM(3)	RM(4)
	full sample		
mean	1.860	-.085	2.072
st.dev.	4.182	3.207	32.125
skewness	8.261	-25.060	29.257
kurtosis	95.223	882.314	955.959
Q(20) stats (p-value)	0.000	0.997	0.082
	pre-crisis		
mean	.953	-.002	.116
st.dev.	.874	.227	.378
skewness	4.193	-2.323	9.883
kurtosis	28.657	72.751	130.661
Q(20) stats (p-value)	0.000	0.065	0.000
	crisis		
mean	5.241	-.107	6.228
st.dev.	7.755	3.592	34.483
skewness	4.118	.810	11.215
kurtosis	25.746	43.370	143.120
Q(20) stats (p-value)	0.000	0.454	0.000
	post-crisis		
mean	1.020	-.184	2.168
st.dev.	2.303	4.740	48.077
skewness	19.532	-23.109	23.214
kurtosis	426.453	536.751	539.933
Q(20) stats (p-value)	0.001	1.000	1.000

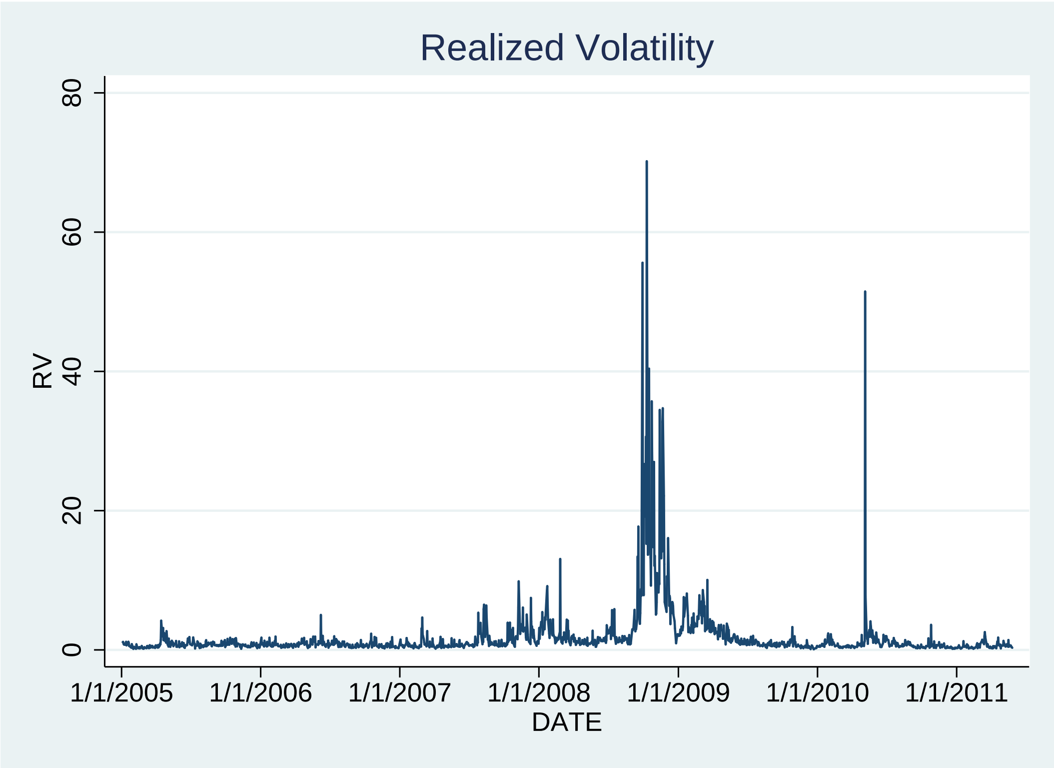

Figure 4.2: Time series of IBM realized volatility

volatility, realized third moment and fourth moment is the consequence of computerized trading executed based on an algorithm, which caused the Dow Jones Average to plummet nearly 10% of its total value in less than 20 minutes. The realized third and fourth moments are around zero for most of time, with some jumps happening infrequently. This phenomenon is consistent with the finding of Amaya et. al (2011), which show that the limits of realized third moment and fourth moment depend solely on jump parameters but not diffusion parameters.

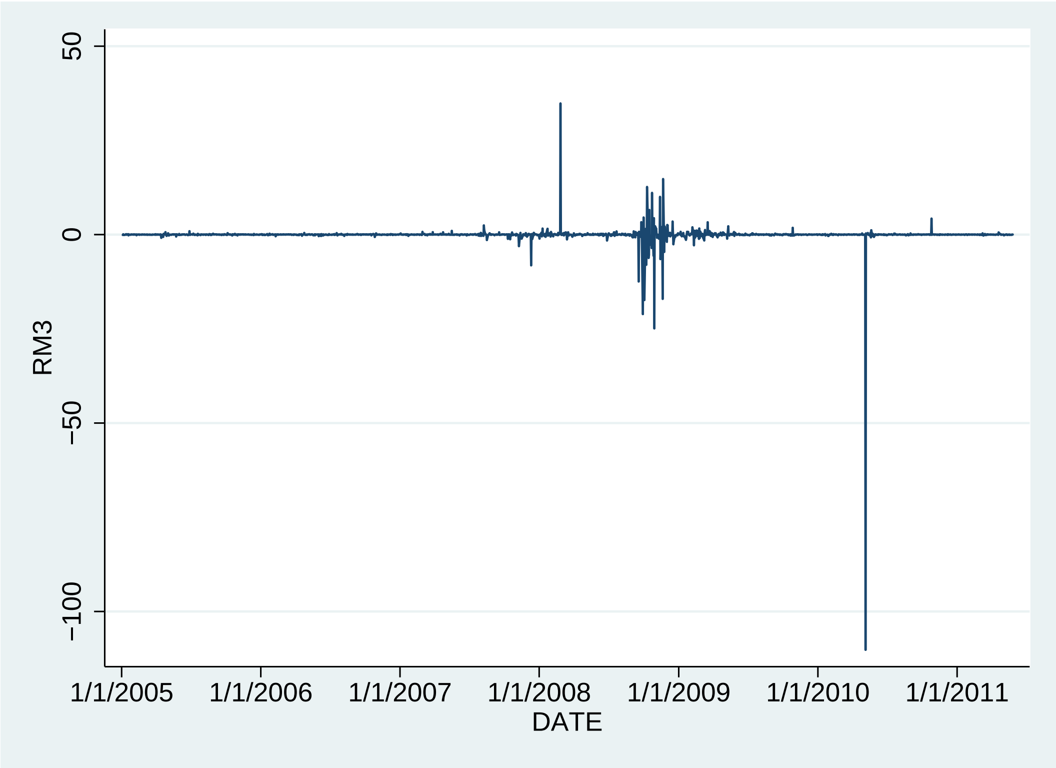

Figure 4.3: Time series of IBM realized third moment

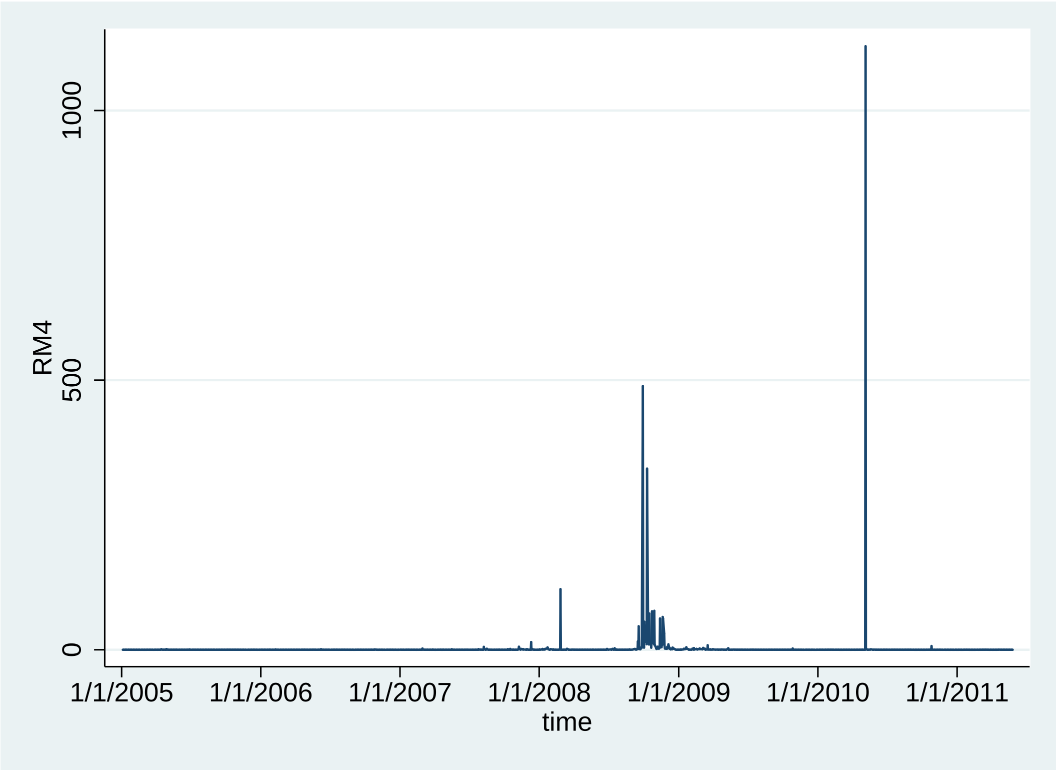

Figure 4.4: Time series of IBM realized fourth moment

In order to provide the forecasts for one-day-ahead realized third and fourth moments, we need to explore the properties of them. Figure 4.5 plots the autocorrelations of the realized fourth moments with 95% confidence intervals, indicating that the autocorrelation is insignificant for the full sample. As suggested by the Ljung-Box Q-statistics in Table 4.2, the realized fourth moment has some predictability during the pre-crisis period and post-crisis period. So we investigate characteristics of the logarithmic realized fourth moment. Table 4.3 summarizes the unconditional distribution of the logarithmic realized volatility and the logarithm of the realized fourth moments. Although the assumption of normality is still rejected, both of them have much smaller skewness and kurtosis, and is closer to a normal distribution, compared to corresponding realized moments. Figure 4.6 plots the time series of the logarithm of the realized fourth moment, which shows that the series is positively correlated. Furthermore, Figure 4.7, Figure 4.8, Figure 4.9 and Figure 4.10 plot the autocorrelations of the logarithm of the realized fourth moment for the full sample, the pre-crisis period, the crisis period and the post-crisis period respectively. The figures illustrate that there exists significant autocorrelations in all the three sub-samples. The autocorrelation is as large as 0.4 up to 40 lags. It is also evident that the series exhibits long memory property as the autocorrelations are above the conventional Bartlett 95% confidence band up to 80 lags for the full sample. It is noteworthy that the long memory property are shown in the sample which only spans six-and-a-half years.

To further investigate the long memory property, Table 4.4 reports the results of Ljung-Box portmanteau test and Augmented Dickey-Fuller test, and the estimate of the fractional integration parameter d following Geweke and Porter-Hudak (1983), hereafter denoted by GPH, for the logarithmic realized fourth moment. The white noise hypothesis is strongly rejected by the Ljung-Box portmanteau test for up to 20 order autocorrelation in all the three sub-samples, which is close to one month of trading days for stocks. The null hypothesis of a unit root process is rejected in the pre-crisis period, the post-crisis period and the full sample period by the augmented Dickey-Fuller test with 20 lag differences at the conventional 5% critical value -2.86. The unit root hypothesis cannot be rejected in the crisis period, either by the augmented Dickey-Fuller test with 20 lags, or by running the ADF test on quasi-differenced series, with the maximum lag determined by the Schwert criterion described in Schwert (1989), which equals the integer part of $12(T+1)/100^{0.25}$.

In literature, the long-run dependence in financial time series can be modeled by fractionally integrated processes, such as FIGARCH model proposed by Baillie et al. (1996) which accounts for long memory in financial volatility. Motivated by the slow decay rates of the autocorrelations, we use the GPH estimator to capture the fractal structure. In all the three sub-samples, the GPH fractional integration parameter estimate is highly significantly different from zero. The fractional integration parameter is less than 0.5 in the pre-crisis period and post-crisis period, but it is larger than 0.5 in the crisis

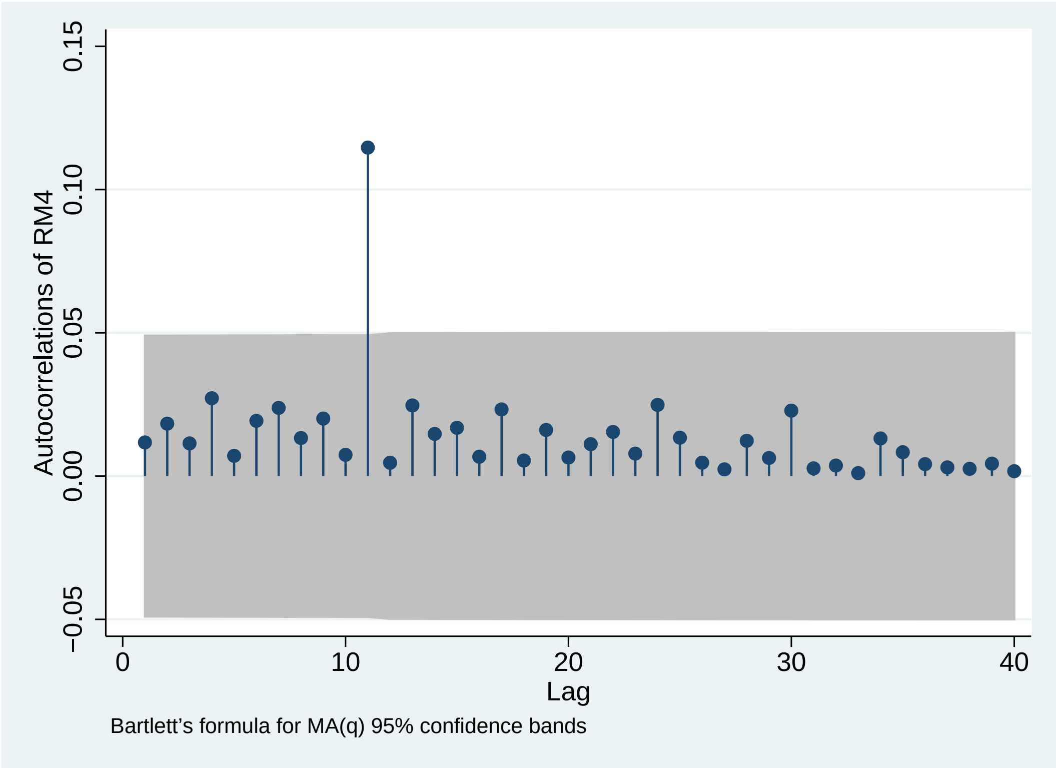

Figure 4.5: Autocorrelations of IBM realized fourth moment

Note: The figure plots autocorrelations of IBM realized fourth moment from Jan 2005 to May 2011 with 95% confidence intervals.

period and in the full sample period. Thus, it is possible that the crisis period is non-stationary.

Due to the overall long memory and autoregressive characteristics of the logarithmic realized fourth moment, we use the ARFIMA (p,d,q) model for estimation.

$$\Phi(L)(1-L)^d(\ln RM(4)_t - \mu_0) = \Theta(L)\epsilon_t \tag{4.8}$$

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} L^k \tag{4.9}$$

Table 4.3: Summary statistics for unconditional distributions of transformations of realized moments

	$\ln RV$	$\ln RM(4)$
mean	-.005	-2.870
st.dev.	.927	1.912
Skewness	1.092	1.042
kurtosis	4.797	4.791
Jarque-bera(p-value)	0.000	0.000

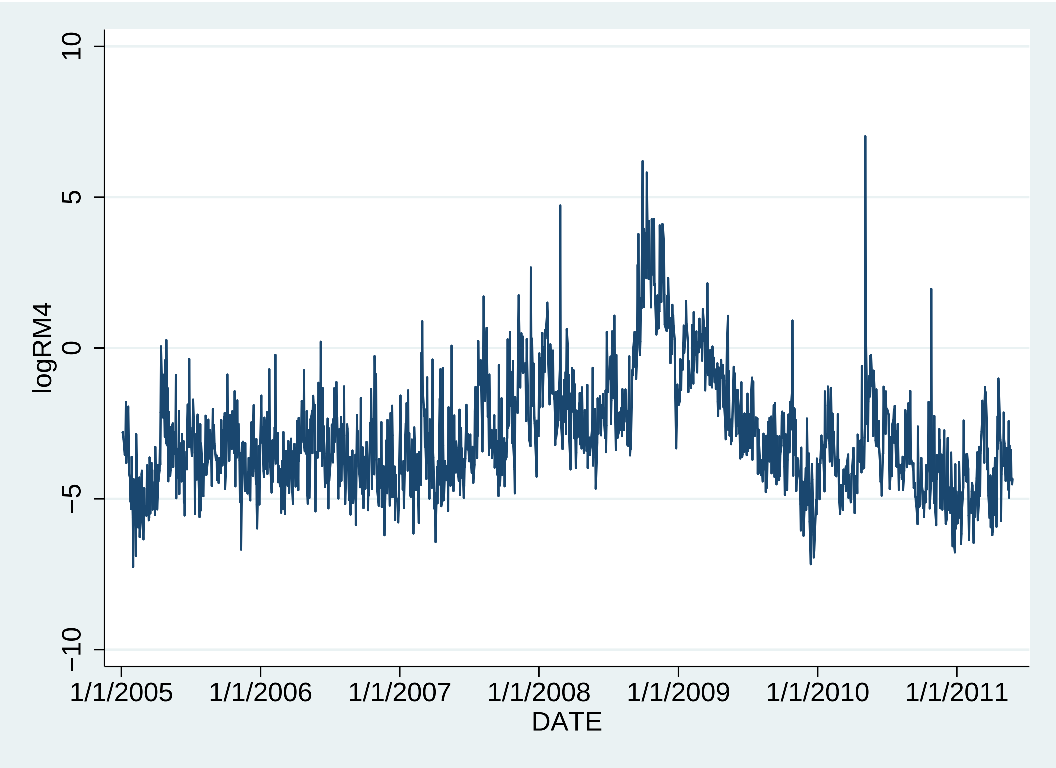

Figure 4.6: Time series of logarithmic realized fourth moment, IBM

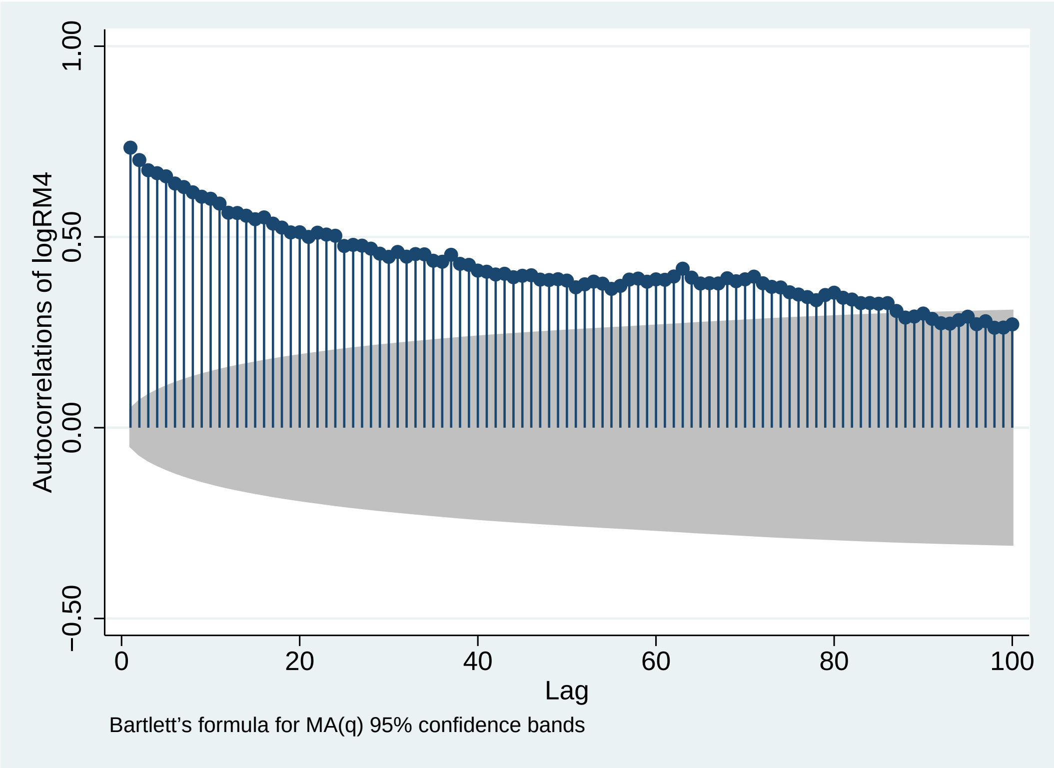

Figure 4.7: Autocorrelations of IBM logarithmic realized fourth moment in the full sample

Table 4.4: Test statistics of $\ln RM(4)$

$Ljung-BoxQ_{22}$	ADF	d_{GPH}
	full sample	
11694.341	-3.309	.750
	pre-crisis period	
1361.507	-3.100	.418
	crisis period	
2362.786	-1.814	.866
	post-crisis period	
1494.470	-3.684	.377

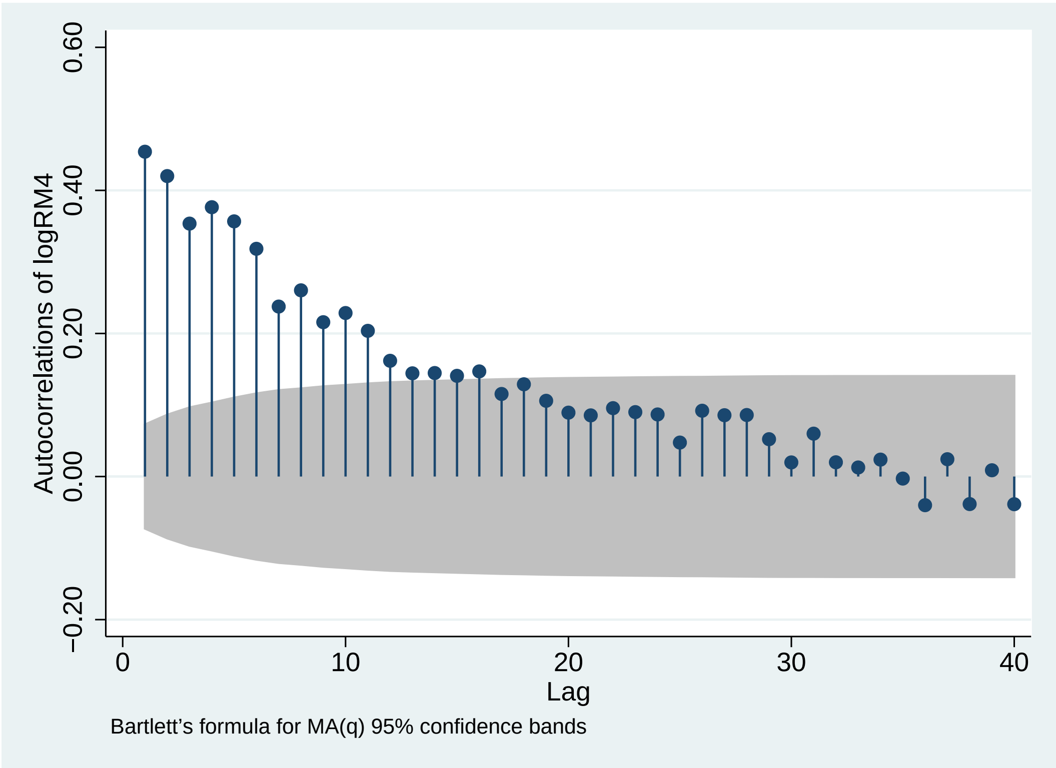

Figure 4.8: Autocorrelations of IBM logarithmic realized fourth moment in the pre-crisis period

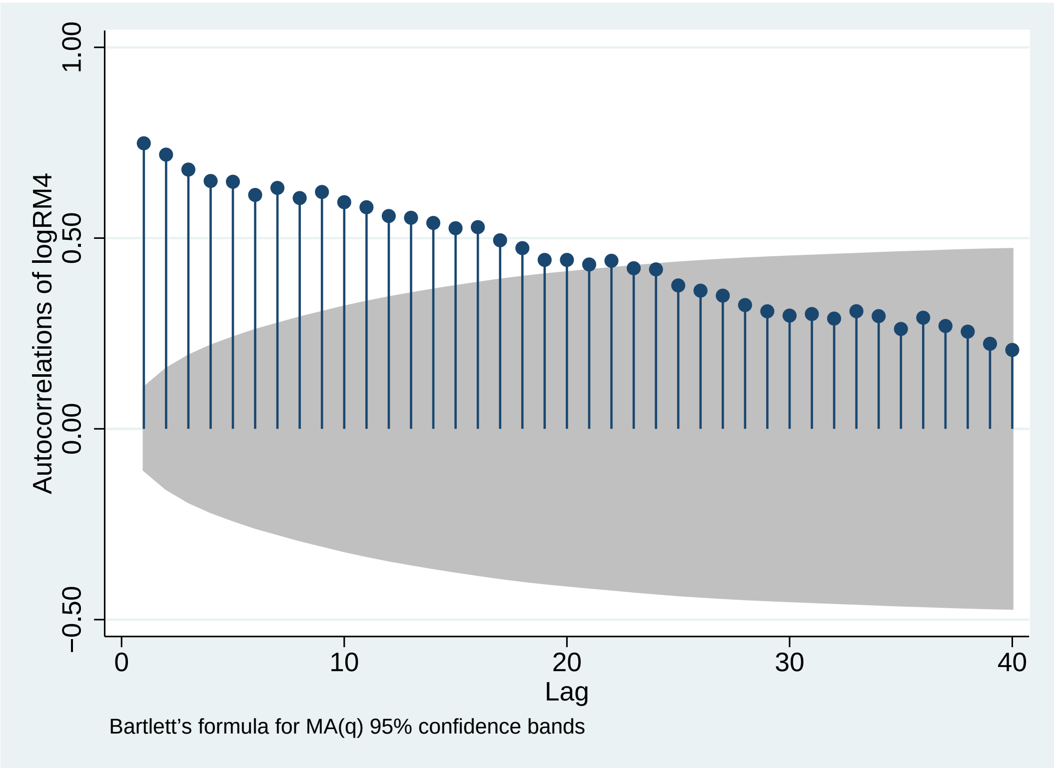

Figure 4.9: Autocorrelations of IBM logarithmic realized fourth moment in the crisis period

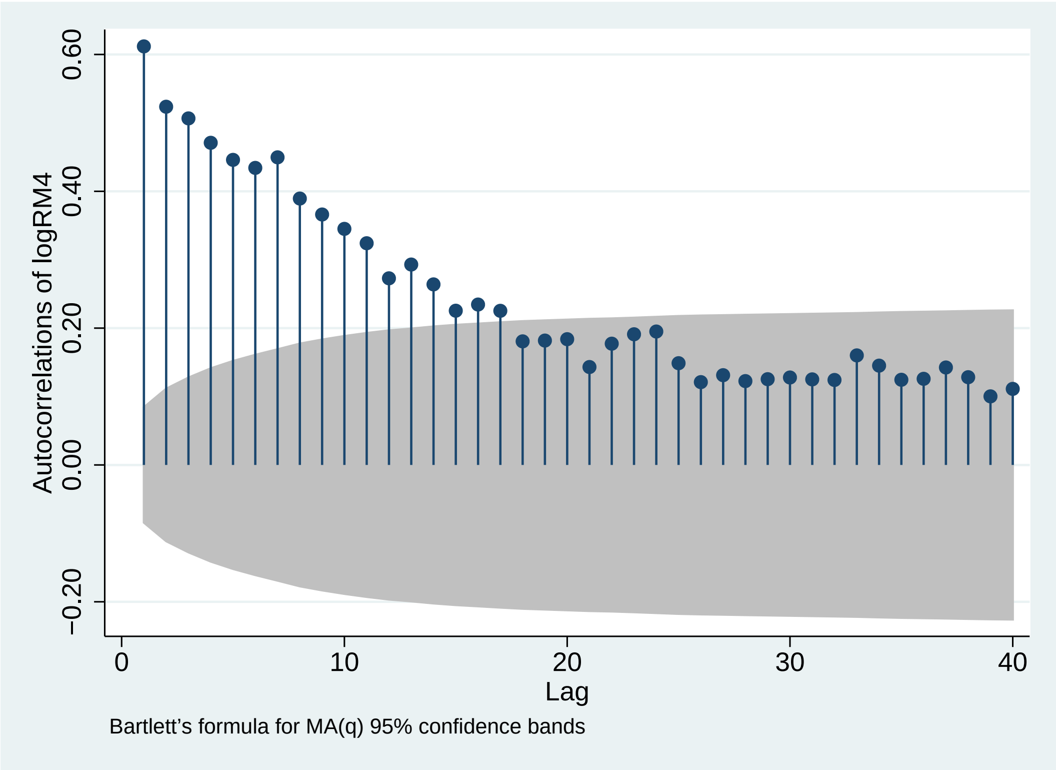

Figure 4.10: Autocorrelations of IBM logarithmic realized fourth moment in the post-crisis period

For estimation, the initial value of the fractional integration parameter is estimated from the log periodogram regression of Geweke and Porter-Hudak (1983) as reported in Table 4.4. Based on normality assumption of the error term, the estimation is carried out by exact maximum likelihood.

To investigate the process of the realized third moment, we plot the autocorrelation of realized third moment for IBM stock return in the full sample in Figure 4.11. The Ljung-Box portmanteau Q(20) test statistics has a p-value of 0.997, indicating that there is no serial correlation in the realized third moment series for the full sample period. Therefore, we only model the dependence in the realized fourth moment but not the realized third moment. After obtaining one-step-ahead predictions of the realized volatility and fourth moment, we can calculate the kurtosis and apply the Cornish-Fisher approximation to derive VaR.

4.4 Realized Moment Forecasting Models

4.4.1 RM-EWMA Model

According to the characteristics of realized moments discussed in the previous section, we propose two models to forecast realized moments. The first method for realized moments prediction is to extend the commonly used exponentially weighted moving average (EWMA) approach to realized higher moments, and thus we name it RM-EWMA model. It comes from the idea of the RiskMetrics

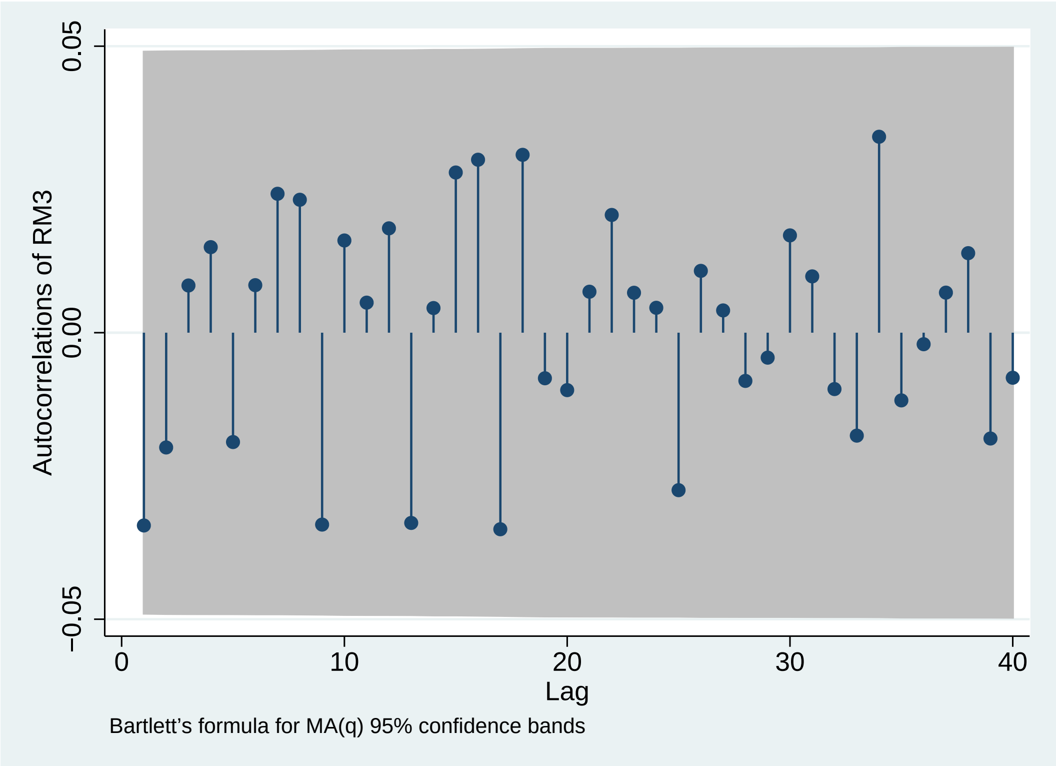

Figure 4.11: Autocorrelations of IBM realized third moment

model, which is made available by J.P Morgan and is adopted as an industry-wide approach to calculate market risk. The basic RiskMetrics model of EWMA procedure can be expressed as a normal integrated GARCH (1,1) model with prespecified autoregressive parameter. The data generation process of return is

$$r_t = \mu + \epsilon_t = \mu + \sigma_t z_t \tag{4.10}$$

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda)\epsilon_t^2 \tag{4.11}$$

where z_t follows an i.i.d standard normal distribution. The autoregressive parameter λ is pre-specified as 0.94. Although this specification is nested by GARCH model, it's widely used by practitioners as its out-of-sample performance is generally acceptable and the methodology is very easy to implement. Alexander and Leigh (1997) show that the predictions based on EWMA approach are favored over those based on the GARCH model for all but very short-term holding periods. It has the advantage to account for the unit-root property of the conditional variance of returns, which cannot be captured by a GARCH model as it's detrimental to the stationarity conditions.

However, the EWMA approach using only volatility may give an unacceptable higher number of outliers and underestimate VaR forecasts during volatile period, as it assumes normality of the return distribution and does not take into consideration the fat-tailed property of the return distribution. Aiming to solve this problem, we extend the EWMA methodology to forecast higher moments and then use the forecasts of higher moments to construct skewness and kurtosis. As autocorrelations are more significant in the logarithmic realized fourth moment instead of in the realized fourth moment, we use the EWMA procedure to capture the dependence in the logarithmic realized fourth moment. For realized second moment, we use the EWMA procedure to capture the dependence in the realized second moment directly.

Let $M(i)_t$ denote the ith moment at day t, for $i = 1, 2, 3$. The ith moment at the next period $M(i)_{t+1}$ is a weighted average of the current period realized moment $RM(i)_t$ and the current period $M(i)_t$:

$$M(i)_{t+1} = \lambda_i M(i)_t + (1 - \lambda_i) RM(i)_t \tag{4.12}$$

where λ_i is the decay factor. By recursive substitution, $M(i)_{t+1}$ can be expressed as the exponentially weighted moving average of past realized moments:

$$M(i)_{t+1} = (1 - \lambda_i) \sum_{j=0}^{\infty} \lambda_i^j RM(i)_{t-j} \tag{4.13}$$

The EWMA methodology gives old returns exponentially less weight. The optimal decay factor λ_i is obtained by minimizing the Mean Squared Error (MSE) between the EWMA ith moment forecast for period $t + 1$ and the *ex post* realized moment for period $t + 1$.

$$MSE(i) = \sum_{t=0}^{T} (RM(i)_{t+1} - M(i)_{t+1})^2 \tag{4.14}$$

We estimate the first moment to the third moment using this EWMA procedure. As mentioned previously, for the realized fourth moment, $log(M(i)_t)$

follows a EWMA process for $i = 4$. $M(\hat{4})_{t+1}$, the forecast of $M(4)_{t+1}$, is obtained by taking the exponent of the forecast of $log(M(4)_{t+1})$.

The forecasted moments are used to construct the forecasts of realized skewness and kurtosis as follows:

$$\hat{k_{3,t}} = \frac{\sqrt{D}M(\hat{3})_t}{M(\hat{2})_t^{\frac{3}{2}}} \tag{4.15}$$

$$\hat{k_{4,t}} = \frac{D \times M(\hat{4})_t}{M(\hat{2})_t^{2}} \tag{4.16}$$

where $\hat{k_{4,t}}$ denotes the forecast of realized kurtosis. After obtaining the forecasts of the realized skewness and kurtosis, we use the Cornish Fisher approximation to incorporate them to provide VaR forecasts:

$$\hat{V_{\alpha,t}} = \mu_t + \hat{\sigma_t}(z_\alpha + \frac{1}{6}(z_\alpha^2 - 1)\hat{k_{3,t}} + \frac{1}{24}(z_\alpha^3 - 3z_\alpha)(\hat{k_{4,t}} - 3) - \frac{1}{36}(2z_\alpha^3 - 5z_\alpha)\hat{k_{3,t}}^2) \tag{4.17}$$

where z_α denotes the α-quantile of the standard normal distribution, $\hat{\sigma_t}$ denotes the forecast for realized volatility respectively.

4.4.2 RM-ARFIMA Process

The second approach is the RM-ARFIMA process, in which we use the autoregressive fractionally integrated moving average (ARFIMA) process to forecast realized moments. In the literature, the realized volatility is found to be autoregressive and exhibit long memory property, which can be captured using either an ARFIMA process or the Heterogenous Autoregressive Realized Volatility model (HAR) proposed in Corsi (2009).

The HAR model given in 4.18 is a simple and parsimonious process that takes into consideration a daily, a weekly and a monthly component of realized volatility. It captures the long memory and autoregressive characteristics of realized volatility and has good forecasting performance.

$$\sqrt{RV_t} = \beta_0 + \beta_1\sqrt{RV_{t-1}} + \beta_2\sqrt{RV^w_{t-1}} + \beta_3\sqrt{RV^m_{t-1}} + \epsilon_t \tag{4.18}$$

where RV_t denotes the daily realized variance, $RV^w_{t-1} = \frac{1}{5}\sum_{i=1}^{5} RV_{t-i}$ denotes the weekly realized variance and $RV^w_{t-1} = \frac{1}{22}\sum_{i=1}^{22} RV_{t-i}$ denotes the monthly realized variance. The error term is a Gaussian white noise process.

The other commonly used realized volatility model is the autoregressive fractionally integrated moving average (ARFIMA) model. Initially developed by Granger (1980), the process is long memory when $0 < d < 0.5$ holds. The ARFIMA(p,d,q) model is:

$$\Phi(L)(1-L)^d(\ln RV_t - \mu_0) = \Theta(L)\epsilon_t \tag{4.19}$$

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} L^k \tag{4.20}$$

with L as the lag operator, $\Phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$ the AR lag polynomial and $\Theta(L) = 1 + \theta_1 L + ... + \theta_p L^p$ the MA lag polynomial. Following Andersen et al. (2001) and Giot and Laurent (2004), we apply an ARFIMAX model to account for the correlation between past negative shocks and logarithmic realized volatility.

As there is no literature about the data generating process for realized third moment and fourth moment, we analyze the time series of IBM stock price to investigate the characteristics of them.

4.5 Empirical Results

4.5.1 Data Sampling

We use the IBM stock data to compare the forecasting performance of RiskMetrics, RV-EWMA, RM-ARFIMA, RM-EWMA and realized moments(RM) models. In the RM model, the *ex post* realized moments are directly used to provide VaR estimate, and thus can only be used to examine whether the realized moments can provide information for VaR estimation. All the other models can be used to forecast VaR.

The full sample is split into an in-sample period for the first 500 observations, and an out-of-sample period that contains observation 501 to observation 1600 to evaluate 1100 one-day-ahead VaR forecasts. A rolling-window forecasting scheme with a fixed window size of 500 is adopted, which is the size of the in-sample period. The models are re-estimated every trading day to incorporate new information, as many risk managers update their models at daily frequency. In order to investigate the models' performance during the financial crisis, the out-of-sample period is divided into three sub-samples, as mentioned in previous sections.

4.5.2 Realized Volatility Incorporating Overnight Information

As the US stock market is only open from 9:30 am to 4:00 pm eastern time, the IBM intra-day price data are unavailable for a large fraction of the day. There are several ways to incorporate the overnight information into realized volatility. Hansen and Lunde (2005) argue that the overnight information should not be ignored and propose a bias correction mechanism. In their paper, the daily close-to-close return is divided into two parts, the close-to-open return $r_{1,t}$ and the open-to-close return $r_{2,t}$, such that $r_t = r_{1,t} + r_{2,t}$. The realized variance for the full day, RV_t^*, given in equation 4.21, is a weighted average of the realized variance for the active part of the day, RV_t, and the squared return over the inactive period, $r_{1,t}^2$, with the weights chosen to minimize the mean squared error.

$$RV_t^* = \omega_1 r_{1,t}^2 + \omega_2 RV_t \tag{4.21}$$

The weights are respectively $\omega_1 = (1-\phi)\mu_0/\mu_1$ and $\omega_2 = \phi\mu_0/\mu_2$, with $\phi = (\mu_2^2\eta_1^2 - \mu_1\mu_2\eta_{12})/(\mu_2^2\eta_1^2 + \mu_1^2\eta_2^2 - 2\mu_1\mu_2\eta_{12})$. For the parameters, μ_0, μ_1, μ_2 are calculated as the mean of $r_{1,t}^2 + RV_t$, $r_{1,t}^2$ and RV_t respectively, and $\eta_1^2 = var(r_{1,t}^2)$, $\eta_2^2 = var(RV_t)$, $\eta_{12} = cov(r_{1,t}^2, RV_t)$. Xu and Li (2012) provide supporting empirical evidence for this consistent-scaling and optimal-weighted measure of realized volatility compared to other scaling methods.

An alternative and easier way to remedy realized volatility from the presence of non-trading hours is proposed in Marten (2002), where a scaled estimator is introduced to accommodate the "close-to-open" and "open-to-close" effects. The realized variance for the whole day, RV_t^*, named as the scaled realized variance, is now the multiplication of the realized variance for the active part of the day, RV_t, and a scaling factor ξ.

$$RV_t^* = \xi RV_t \tag{4.22}$$

$$\xi = \frac{\sigma_{oc}^2 + \sigma_{co}^2}{\sigma_{oc}^2} \tag{4.23}$$

where σ_{oc}^2 is the "open-to-close" variance, and σ_{co}^2 is the "close-to-open" variance, which can be derived from the logarithmic open price and logarithmic close price of the trading days as follows:

$$\sigma_{oc}^2 = \sum_{t=1}^{T} 10000[log(P_{D,t}) - log(P_{0,t})]^2/T \tag{4.24}$$

$$\sigma_{co}^2 = \sum_{t=1}^{T} 10000[log(P_{0,t}) - log(P_{D,t-1})]^2/T \tag{4.25}$$

We use both ways to construct realized volatility incorporating overnight information. We find that the first approach by Hansen and Lunde (2005) has a much higher mean of realized volatility for IBM stock return than the second approach. The mean of realized volatility using the first method is 2.99, and the mean of realized volatility is 2.07 using the second method. As the mean of realized volatility before incorporating overnight information is 1.79, the second method seems to be more reliable, as the overnight information is generally considered to have a much smaller impact on volatility than the trading hours information. Therefore, we follow Marten (2002) to derive realized volatility for the whole day.

4.5.3 Realized Moments Forecasting Results

Firstly we estimate the RM-ARFIMA model. Figure 4.12 plots the time series of logarithmic realized volatility and Figure 4.13 plots the autocorrelation of logarithmic realized volatility of IBM stock return. The slowing decaying autocorrelation implies the long-memory property of the logarithmic realized volatility, which is in correspondence with previous literature.

In order to determine the lag order of the ARFIMA (p,d,q) model , we estimate different specifications with $p, q = 1, 2, 3, 4, 5$. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are minimized to choose the optimal bundle of (p, q). Then we apply the likelihood ratio test for model specification. Table 4.5 reports the estimates of ARFIMA model for both the logarithmic realized volatility and the logarithmic realized fourth moments

Table 4.5: Estimation Results for ARFIMA models of $\ln RV$ and $\ln RM(4)$

parameters	$\ln RV$	$\ln RM(4)$
ϕ_1	0.957(0.000)	
θ_1	-0.857(0.000)	-0.169(0.000)
σ	0.229(0.000)	1.325(0.000)
d	0.311(0.000)	0.495(0.000)
AIC	2191.45	5001.224
BIC	2212.961	5017.735

Note: P-values are reported in the parentheses.

during the sample period with a fixed window of 1600, with the corresponding p-values reported in the parentheses. For the RV-ARFIMA (p,d,q) model of the logarithmic realized volatility, we find that an ARFIMA (1,d,1) model without constant term fits the data best. The 95% confidence interval for d is (.217,.405), indicating that fractional integration parameter is significantly different from 0.5. For the logarithmic realized fourth moment, the ARFIMA (0,d,1) model provides the best estimation. The parameters are all significantly different from zero. The 95% confidence interval for d is (.482,.508), with the upper bound of the fractional integration parameter slightly larger than 0.5, indicating that the series may be non-stationary. Thus, we test whether the series of logarithmic realized fourth moment is generated by an I(1) process by testing whether the first difference is overdifferenced. Applying the ARFIMA model to the first difference of the logarithmic realized moment, the 95% confidence interval of the fractional integration parameter is (-.460,-.200), which is strictly less than zero. Therefore, the difference series is overdifferenced, and the logarithmic realized fourth moment is stationary in the full sample period, which is in accordance with the results in Table 4.4.

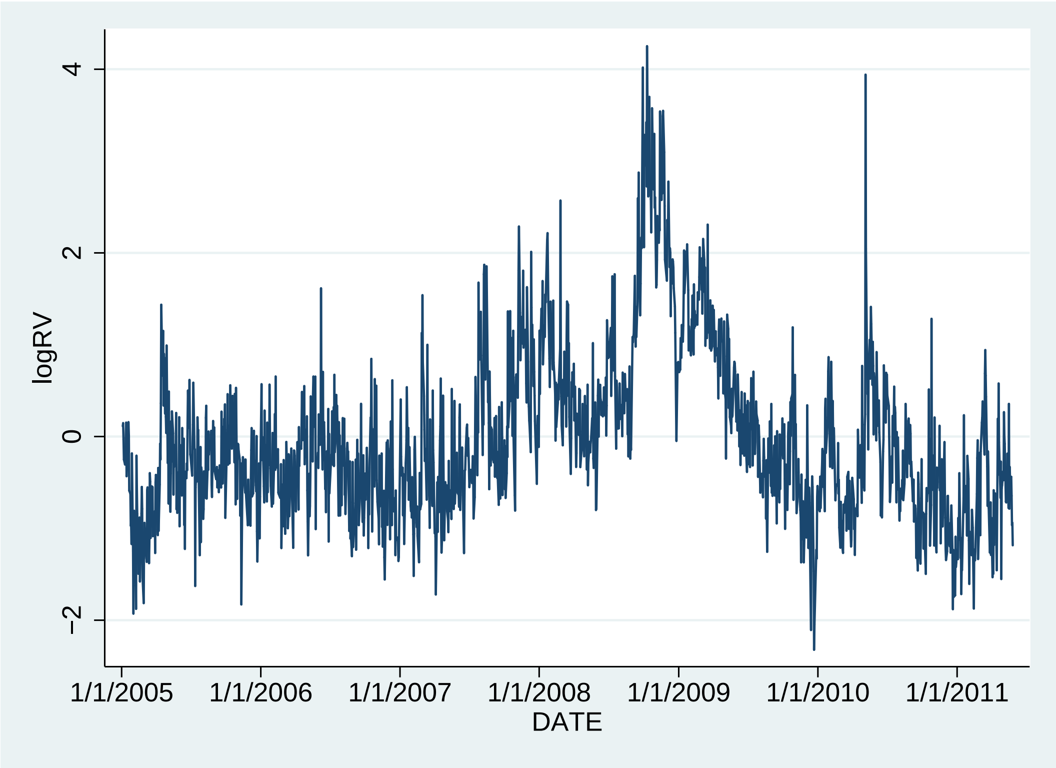

Figure 4.12: Time series of logarithmic realized volatility, IBM

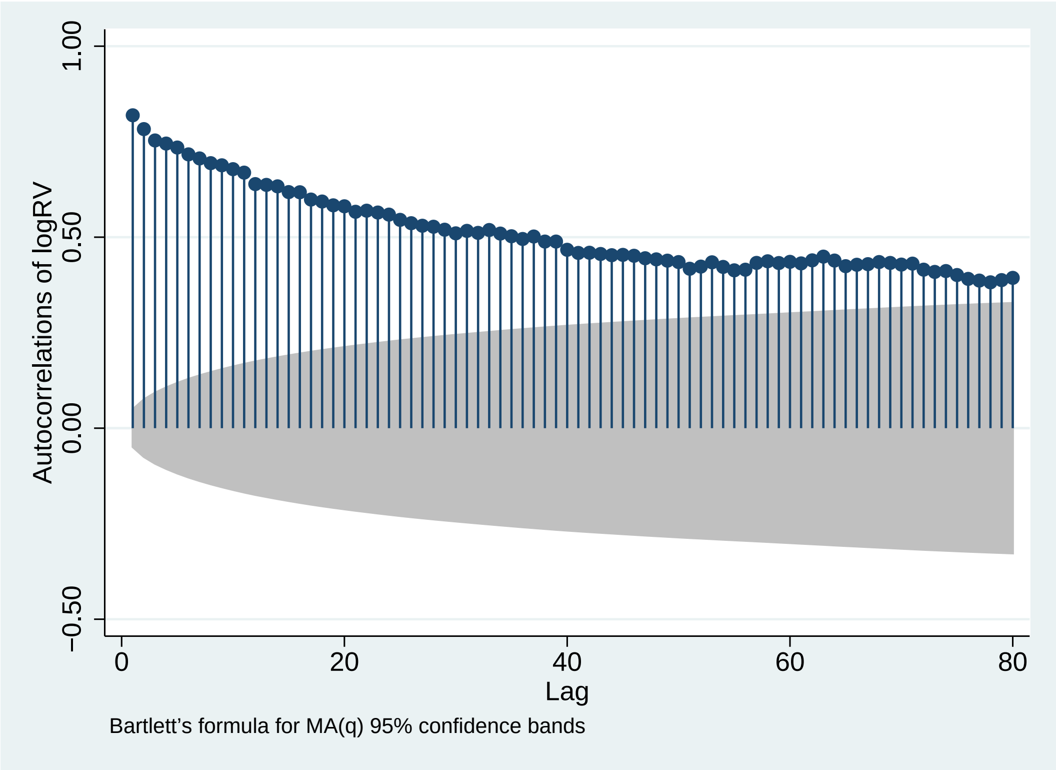

Figure 4.13: Autocorrelations of logarithmic realized volatility, IBM

For the RM-EWMA model, Table 4.6 reports the EWMA optimal values of the decay factor λ_i for $i = 1, 2, 3, 4$, and the correlation coefficients between the estimated ith moment and the corresponding realized moment for the full sample period. As mentioned previously, for the fourth moment, the table reports the correlation between the logarithmic realized fourth moment and its forecast. From Table 4.6, we find that for even numbers of i, λ_2 and λ_4 are less than 1, which indicates predictability in the realized volatility and fourth moment. For odd numbers of i, λ_1 and λ_3 are equal to 1, which means that the one-day-ahead forecast for the mean and third moment are equal to the unconditional mean and third moment, and there is no predictability. In addition, the correlation coefficients between the *ex post* realized moments and the forecasted moments are much larger than zero for $i = 2, 4$, and are equal to zero for $i = 1, 3$. For the realized fourth moment, the correlation is as high as 0.798, showing that the model provides very good forecast.

Using the RM-ARFIMA model and the RM-EWMA model, we can forecast the next period's realized volatility and realized fourth moment. It is worth noting that for the RM-ARFIMA model, the 1100 one-day-ahead realized volatility and realized fourth moment forecasts are obtained by fitting the ARFIMA (p,q) model with the optimal bundle of (p,q) chosen by minimizing the AIC and BIC. The forecasted kurtosis can be derived using the forecasts of the realized volatility and realized fourth moment. Incorporating the volatility and kurtosis into the Cornish Fisher expansion, the VaR of the return series can be

Table 4.6: Lambdas and correlations of EWMA procedure

λ_1	1	corr(RM(1),$M(1)$)	0.000
λ_2	0.784	corr(RM(2),$M(2)$)	0.724
λ_3	1	corr(RM(3),$M(3)$)	0.000
λ_4	0.721	corr(log(RM(4)),$log(M(4))$)	0.798

The table reports the optimal values of λ_k and the correlation coefficients between the realized moments and its forecast. For the fourth moment it reports the correlation between the logarithmic realized fourth moment and its forecast. The estimates are averaged over all windows (1600 daily observations).

forecasted.

4.5.4 Statistical Evaluation

Kupiec (1995) proposes a likelihood ratio test to evaluate VaR forecast in terms of unconditional accuracy, by determining if the rate of violation is statistically compatible with the significance level. It implies that, for all possible significance level α, the VaR estimate should only be exceeded α of the time. The null hypothesis is $H_0 : p = \alpha$. The VaR forecast is accurate if the null hypothesis is accepted. The test statistics of unconditional coverage is given by:

$$LR_{uc} = -2\ln{(1-p)^{T-N}p^N} + 2\ln{(1-\frac{N}{T})^{T-N}(\frac{N}{T})^N} \tag{4.26}$$

where p is one minus the specified confidence level, T is the number of observations, and N is the number of exceptions that the return is larger than the VaR forecast. The test statistics LR_{uc} conforms to $\chi^2_{1,\alpha}$ distribution under the null hypothesis.

In addition to the unconditional coverage test of Kupiec (1995), Christoffersen (1998) proposed a conditional accuracy test. In contrast to the Kupiec test which only focuses on the frequency of exceptions and ignores the time dynamics, Christoffersen (1998) tests the independence of exceptions. Conditional accuracy indicates that the current violation of VaR should not have influence on the future violation of VaR in the next period. Let I_t denote the indicator function for whether VaR is exceeded or not: $I_t = 1$ if VaR is exceeded, and the return is in state 1. $I_t = 0$ if VaR is not exceeded and the return is in state 0. N_{ij} is defined as the number of times that state j follows state i, for $i, j = 0, 1$. The likelihood ratio test for conditional accuracy is given by

$$LR_c = 2(\ln L_A - \ln L_0) \tag{4.27}$$

where

$$L_A = (1 - \pi_{01})^{N_{00}} \pi_{01}^{N_{01}} (1 - \pi_{11})^{N_{10}} \pi_{11}^{N_{11}} \tag{4.28}$$

$$L_0 = (1 - \pi)^{N_{00}+N_{10}} \pi^{N_{01}+N_{11}} \tag{4.29}$$

For the parameters, $\pi_{ij} = N_{ij}/(N_{i0} + N_{i1})$, and $\pi = (N_{01} + N_{11})/(N_{00} + N_{01} + N_{10}+N_{11})$. The statistic follows an asymptotic chi-squared distribution with one degree of freedom under the null hypothesis that the sequence is independent.

4.5.5 VaR Forecasting Results

We use both the unconditional accuracy test and the conditional accuracy test to evaluate the VaR forecasts for five models, including the RiskMetrics model, RV-EWMA model, RM-EWMA model, RM-ARFIMA model and RM model. The RiskMetrics model only extract information from the mean and volatility of the daily return. The RV-EWMA model adopts the mean-variance framework, using realized volatility as the variance to forecast VaR. Applying Cornish-Fisher expansion, the RM model directly uses realized moments to estimate VaR, which cannot be used for VaR forecasting. The RM-EWMA and RM-ARFIMA model incorporate forecasts of realized volatility and realized fourth moment to provide VaR forecasts. Table 4.7 reports the VaR expected number of violations and estimated number of violations by different models for different significance level α, which varies from 1% to 5%. By definition, the failure rate $x\%$ is the percentage of the number of times that the negative return exceeds the VaR forecast in the left tail of the distribution. If the VaR model is correctly specified, the estimated number of violation should be equal to the expected number of violation. If the estimated number of violation is close to the target number, the model provides accurate VaR forecasts.

From the Table we find that the RM model passs both tests for all α in any sub-sample. The RM model uses *ex post* realized moments given in equation 4.3. [1] It provides very accurate VaR estimation results as the number of violation is very close to the expected number for all significance levels. Although the RM model cannot be used for forecasting, it shows that the realized

[1]The forecasting result of the RM model without considering the impact of overnight information on the realized volatility is much better than the result of the RM model with scaled realized volatility. For the RM-ARFIMA model and RM-EWMA model the scaled realized volatility has better forecasting performance. One possible reason is that the realized third moment is not incorporated in the RM-EWMA and RM-ARFIMA model, thus a higher realized volatility leads to better result.

Table 4.7: Number of VaR violations

significance level	95%	97.5%	98%	99%
		pre-crisis period (228 days)		
Target	11.4	5.7	4.56	2.28
RiskMetrics	13*	8*	8*	8
RV-EWMA	13*	8*	8*	7
RM-EWMA	12*	8*	8*	7
RM-ARFIMA	15*	8*	8*	8
RM	13*	5*	3*	2*
		crisis period (330 days)		
Target	16.5	8.25	6.6	3.3
RiskMetrics	22*	12	11	5*
RV-EWMA	19*	8*	8*	4*
RM-EWMA	19*	8*	8*	2*
RM-ARFIMA	18*	10*	7*	5*
RM	18*	9*	6*	2*
		post-crisis period (542 days)		
Target	27.1	13.55	10.84	5.42
RiskMetrics	22*	13*	13*	10*
RV-EWMA	31*	17*	14*	12
RM-EWMA	29*	14*	13*	7*
RM-ARFIMA	30*	16*	14*	9*
RM	23*	10*	9*	5*

Note: The table summarizes out-of-sample performance of 5 models' VaR violations in the three sub-samples after excluding the first 500 in-sample observations. ”Target” refers to the expected number of VaR violations in the time period. The asterisk shows that the model passes both the Kupiec unconditional accuracy test and the Christoffersen conditional accuracy test at 5% significance level.

moments contain useful information of the return distribution. It is worth noting that only the RM model passes both tests for 99% VaR in the pre-crisis period, as all the other models have higher expected number of violations at 1% significance level.

The RiskMetrics forecasts are acceptable in the post-crisis period. However, in the crisis period, it performs poorly for 97.5% VaR and 98% VaR, with a significantly higher number of violations than the expected number. The RiskMetrics model tends to underestimate the risk by only considering the mean and variance, which is in accordance with previous literature. Compared to the RiskMetrics model which uses the EWMA procedure with a pre-specified weighting parameter, the RV-EWMA model, which also follows the EWMA approach but uses the realized volatility instead of an integrated GARCH process, generally provides number of violations that's closer to the expected number for all α. However, similar to the RiskMetrics model, it still underestimates the risk in all the three sub-samples. Although the RV-EWMA model provides acceptable VaR estimates in the crisis period, it does not pass the unconditional test for 99% VaR in the post-crisis period.

The RM-EWMA model performs very well for all significance levels in the crisis period and the post-crisis period. It passes both tests and the number of violations is very close to the expected number in the crisis period. For all sub-samples, the RM-EWMA model yields a significant improvement upon the RiskMetrics model and the RV-EWMA model, as it takes into consideration the skewness and kurtosis which cannot be captured in the volatility model. The performance of the RM-EWMA model and the RM-ARFIMA model is quite similar, and the RM-EWMA model usually provides slightly closer number of violations to the target number than the RM-ARFIMA model. In the pre-crisis period, all the forecasting models have higher number of violations than the target number for 99% VaR, indicating that the associated 99% market risk is underestimated in the pre-crisis period. We find that although the *ex post* logarithmic realized volatility and fourth moment is close to Gaussian, they are still skewed and thus the error term in the ARFIMA model might not follow an Gaussian distribution. Figure 4.14 and Figure 4.15 show the normal quantile plots of the residuals of the logarithmic realized volatility and logarithmic realized fourth moment respectively, indicating there are still excess skewness and kurtosis that cannot be captured by a normal distribution. Therefore, other distributions of the error term could be considered for future study.

4.6 Conclusion

In this chapter, we have derived two models incorporating realized moments through Cornish Fisher approximation to provide Value-at-risk forecasts. Constructed as the summation of the powers of the intraday returns, the realized moments are consistent estimators of the corresponding population moments

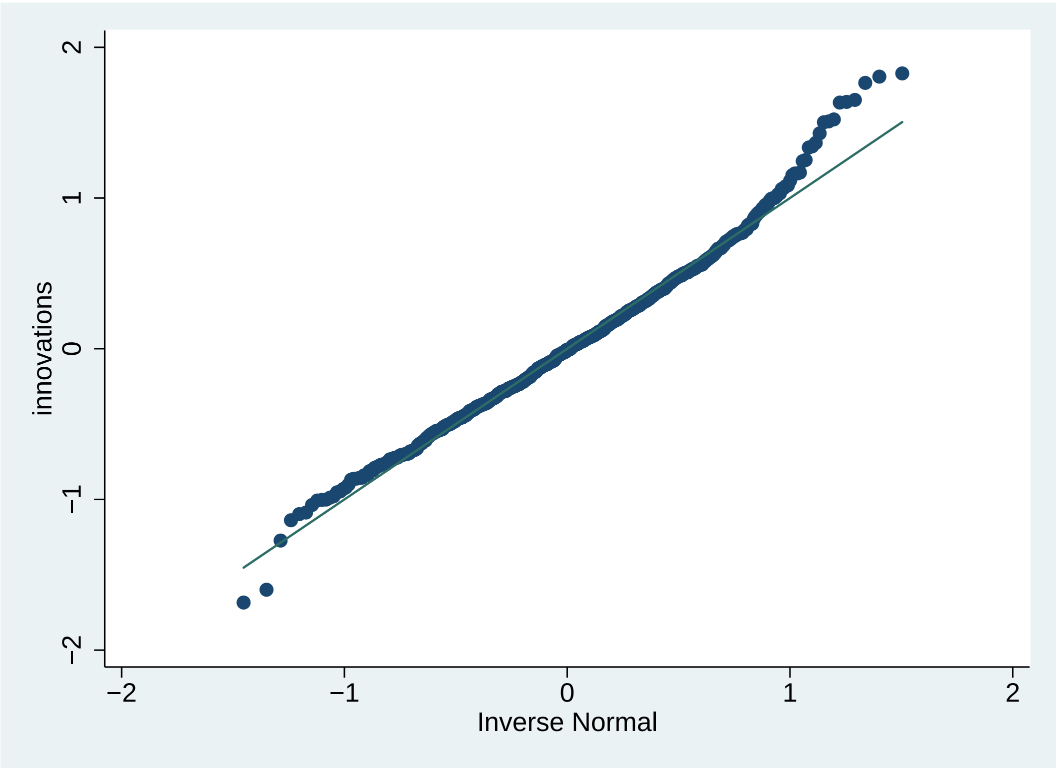

Figure 4.14: Normal quantile plot of residuals of ARFIMA model for IBM logarithmic realized volatility

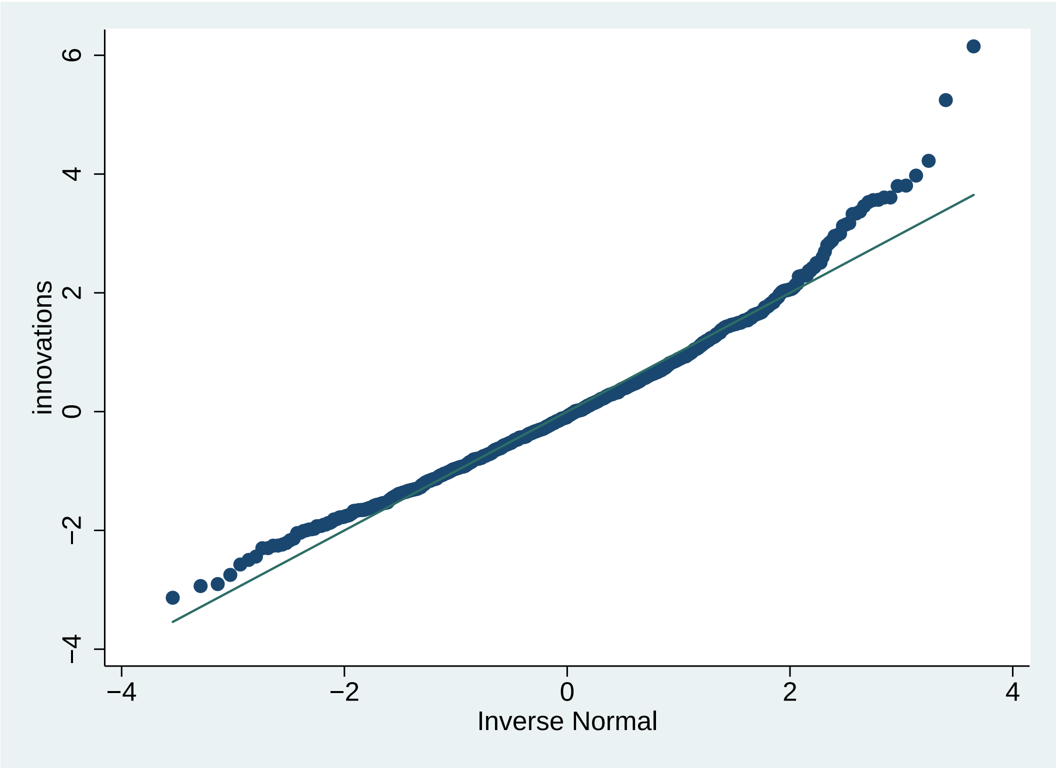

Figure 4.15: Normal quantile plot of residuals of ARFIMA model for IBM logarithmic realized fourth moment

under a typical jump diffusion setting. The fat-tail property of the return distribution, which hinders the accuracy of VaR forecasts, can be captured by the skewness and kurtosis constructed from realized moments.

The first model is the RM-EWMA model, which extends the EWMA procedure up to realized fourth moments. In this model, the next period's ith moment is specified as the weighted average of the current moment and the realized ith moment for $i = 1, 2, 3, 4$. The optimal weights are determined by minimizing the mean squared error. The second model is the RM-ARFIMA model, in which both the logarithmic realized volatility and fourth moment take an ARFIMA model. We find that the logarithmic realized fourth moment is a autoregressive process whose autocorrelations decline at a slow rate, similar to the logarithmic realized volatility. The one-step-ahead forecasts of the logarithmic realized fourth moment can be used to calculate the one-step-ahead forecast for realized kurtosis. Then we employ the Cornish Fisher approximation of the standard normal distribution in terms of the mean, variance, time-varying skewness and kurtosis estimated from the models to provide one-day-ahead VaR.

We compare the RM-EWMA model and the RM-ARFIMA model with the RiskMetrics model, the benchmark RM model and the RV-EWMA model using IBM stock price. To evaluate the performance of the models at different time periods of the business cycle, the full sample is divided into three subsamples, the pre-crisis period, the crisis period around 2008, and the post-crisis period. Kupiec unconditional and Christoffersen conditional accuracy tests reveal that the RM-EWMA model and the RM-ARFIMA model outperform the RiskMetrics model and the RV-EWMA model in all three subsamples. In the crisis period, the RiskMetrics model underestimate the VaR as it only considers the mean and variance. can remedy the problem by considering the fat-tail property. In summation, the time-varying realized moments models provide considerable improvement over the realized volatility models and the RiskMetrics model.

4.7 Bibliography

AMAYA, D., P. CHRISTOFFERSEN, K. JACOBS, AND A. VASQUEZ (2011): "Does Realized Skewness and Kurtosis Predict the Cross-section of Equity Returns?," *working paper*.

ANDERSEN, T., AND T. BOLLERSLEV (1998): "Answering skeptics:Yes,standard volatility models do provide accurate forecasts," *International Economic Reviews*, 39, 115–158.

BAO, Y., T. LEE, AND B. SALTOGLU (2006): "Evaluating predictive performance of value-at-risk models in emerging markets: a reality check," *Journal of Forecasting*, 25, 101–128.

BARNDOFF-NIELSEN, O., AND N. SHEPHARD (2001): "Non-Gaussian Ornstein-Unlenbeck models and some of their uses in financial economics," *The Royal Statistical Society*, B63, 167–241.

BEINE, M., S. LAURENT, AND F. PALM (2009): "Central Bank Forex Interventions Assessed Using Realized Moments," *Journal of International Financial Markets, Institutions and Money*, 19, 112–127.

BERKOWITZ, J., AND J. OBRIEN (2002): "How Accurate Are Value-at-Risk Models at Commercial Banks?," *Journal of Finance*, 57, 1093–1112.

BROOKS, C., S. BURKE, S. HERAVI, AND G. PERSAND (2005): "Autoregressive Conditional Kurtosis," *Journal of Financial Econometrics*, 3, 399–421.

CHRISTOFFERSEN, P. (1998): "Evaluating Interval Forecasts," *International Economic Review*, 39, 841–862.

CORNISH, E. A., AND R. A. FISHER (1937): "Moments and Cumulants in the Specification of Distributions," *Extrait de la Revue de l'Institute International de Statistique*, 4, 1–14.

CORSI, F. (2009): "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 7(2), 174–196.

ENGLE, R., AND S. MANGANELLI (2004): "CAViaR: Conditional autoregressive value at risk by regression quantiles," *Journal of Business & Economic Statistics*, 22, 367–381.

GEWEKE, J., AND S. PORTER-HUDAK (1983): "The Estimation and Application of Long Memory Time-series Models," *Journal of Time Series Analysis*, 4, 221–238.

GIOT, P., AND S. LAURENT (2004): "Modelling Daily Value-at-risk using Realized Volatility and ARCH Type Models," *Journal of Empirical Finance*, 11.

HANSEN, P. R., AND A. LUNDE (2005): "A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data," *Journal of Financial Econometrics*, 3, 525–554.

HARVEY, C., AND A. SIDDIQUE (1999): "Autoregressive Conditional Skewness," *Journal of Financial and Quantitative Analysis*, 34, 465–487.

KNIGHT, J., S. SATCHELL, AND G. WANG (2003): "Value-at-Risk Linear Exponent (VARLINEX) Forecasts," *Quantitative Finance*, 3, 332–344.

KUESTER, K., S. MITTNIK, AND M. PAOLELLA (2006): "Value-at-risk prediction: a comparison of alternative strategies," *Jounral of Finanicial Econometrics*, 4, 53–89.

KUPIEC, P. (1995): "Technique for Verifying the Accuracy of Risk Measurement Models," *Journal of Derivatives*, 3, 73–84.

LOPEZ, J. (2001): "Evaluating the Predictive Accuracy of Volatility Models," *Journal of Forecasting*, 20.

MARTEN, M. (2002): "Measuring and Forecasting S&P 500 Index Futures Volatility Using High-Frequency Data," 22, 497–518.

MCNEIL, A., AND R. FREY (2000): "Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach," *Journal of Empirical Finance*, 7, 271–300.

PERIGNON, C., AND D. SMITH (2010): "The level and quality of value-at-risk diclosure by commercial banks," *Journal of Banking&Finance*, 34, 362–377.

POLANSKI, A., AND E. STOJA (2010): "Incorporating Higher Moments into Value-at-Risk Forecasting," *Journal of Forecasting*, 29, 523–535.

SCHWERT, G. W. (1989): "Tests for unit roots: A Monte Carlo investigation," 2, 147–159.

XU, D., AND Y. LI (2012): "Empirical Evidence of Leverage Effect in a Stochastic Volatility Model: a Realized Volatility Approach," *Forthcoming in Frontiers of Economics in China*.

ZHANG, L., P. MYKLAND, AND Y. AÏT-SAHALIA (2010): "Edgeworth Expansions for Realized Volatility and Related Estimators," *Journal of Econometrics*, 160, 190–203.

Chapter 5

Conclusion

This thesis explores various topics in the financial econometrics literature, about durations between adjacent trades in tick-by-tick data, structural changes in the volatility of the return series, and Value-at-Risk. The chapter on volatility presents two types of regime-switching GARCH-jump models with autoregressive jump intensity to model the non-linearity in return series. The first model is a Markov regime-switching model which generalizes the GARCH model by distinguishing two regimes with different GARCH volatility and jump intensity levels. As the regimes are unknown to the econometrician in Markov regime-switching models, which leads to difficulty in forecasting, a threshold GARCH-jump model with an exogenous threshold variable is also proposed. The stationarity conditions and moments of returns are derived for the threshold GARCH-jump model. Using Japanese YEN-US Dollar exchange rate, it's shown that both types of regime-switching models have better performance than the traditional GARCH model for the in-sample period. The threshold

GARCH-jump model outperforms the single-regime GARCH-jump model for the out-of-sample period.

The chapter on trade durations deals with structural changes in the autoregressive conditional duration series. Monte Carlo experiments are conducted to show that both permanent changes and temporary changes for a relatively short time period in the parameters of the autoregressive conditional duration (ACD) model can lead to a big bias in the estimates of the autoregressive parameters, which converge to one as jump size increases. The sample mean of the conditional expected duration is derived for the ACD model with unaccounted structural changes.

The chapter on Value-at-Risk presents two models to provide Value-at-risk forecasts by incorporating realized moments through the Cornish Fisher approximation. Constructed from high frequancy data, the realized moments are consistent estimators of the corresponding population moments in return series under a typical jump diffusion setting. The traditional mean-variance VaR models tend to underestimate the risk especially during volatile periods, such as the RiskMetrics model which is widely used in the financial industry. The fat-tail property of the return distribution can be captured by the skewness and kurtosis constructed from realized moments. The first model proposed in this chapter is the RM-EWMA model, which extends the EWMA procedure up to realized fourth moments. As we find that the logarithmic realized fourth moment is autoregressive and often exhibits long memory property, we also propose an RM-ARFIMA model to forecast realized moments. Comparing the RM-EWMA model and the RM-ARFIMA model with the RiskMetrics model, the benchmark RM model and the RV-EWMA model using IBM stock price, we find that the models using realized moments outperform the RiskMetrics model in all subsamples. During the financial crisis period around 2008, the realized moment models accurately predict the market risk.

5.1 Bibliography

AMAYA, D., P. CHRISTOFFERSEN, K. JACOBS, AND A. VASQUEZ (2011): "Does Realized Skewness and Kurtosis Predict the Cross-section of Equity Returns?," *working paper*.

CHAN, W., AND J. M. MAHEU (2002): "Conditional jump dynamics in stock market returns," *Journal of Business & Economic Statistics*, 20(3), 377–389.

ENGLE, R., AND J. RUSSELL (1998): "Autoregressive conditional duration: a new model for irregularly spaced transaction data," *Econometrica*, 66, 1127–1162.

Appendix A

Proof of Proposition 2.1

Proof: As $\lambda_t = E[N(t)|\Phi_{t-1}, s_t]$, applying the law of iterated expectations to equation (2.17), we have

$$\begin{aligned} E[N(t)|s_t] &= E[\lambda_t|s_t] \\ &= \alpha_{s_t} + \rho_{s_t} E[E[N(t-1)|\Phi_{t-2}, s_{t-1}]|s_t] + \gamma_{s_t} E[\nu_t|s_t] \\ &= \alpha_{s_t} + \rho_{s_t} E[E[N(t-1)|\Phi_{t-2}, s_{t-1}, s_t]|s_t] + \gamma_{s_t} E[\nu_t] \\ &= \alpha_{s_t} + \rho_{s_t} E[E[N(t-1)|s_{t-1}]|s_t] \end{aligned}$$

$$\begin{aligned} E[E[N(t-1)|s_{t-1}]|s_t = 1] &= E[N(t-1)|s_{t-1} = 1]P(s_{t-1} = 1|s_t = 1) \\ &\quad + E[N(t-1)|s_{t-1} = 2]P(s_{t-1} = 1|s_t = 1) \end{aligned}$$

$$\begin{aligned} E[E[N(t-1)|s_{t-1}]|s_t = 2] &= E[N(t-1)|s_{t-1} = 1]P(s_{t-1} = 1|s_t = 2) \\ &\quad + E[N(t-1)|s_{t-1} = 2]P(s_{t-1} = 1|s_t = 2) \end{aligned}$$

Substituting the latter two equations into the first equation leads to following equations.

$$\begin{aligned} \lambda_1 &= E[N(t)|s_t = 1] \\ &= \alpha_1 + \rho_1 P(s_{t-1} = 1|s_t = 1)\lambda_1 + \rho_1 P(s_{t-1} = 2|s_t = 1)\lambda_2 \qquad \textbf{(A.1)} \end{aligned}$$

$$\begin{aligned}\lambda_2 &= E[N(t)|s_t = 2] \\ &= \alpha_2 + \rho_2 P(s_{t-1} = 1|s_t = 2)\lambda_1 + \rho_2 P(s_{t-1} = 2|s_t = 2)\lambda_2\end{aligned} \tag{A.2}$$

Organizing the above expressions for λ_1 and λ_2 leads to

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = B^{-1} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \tag{A.3}$$

$$B = \begin{bmatrix} 1 - \rho_1 P(s_{t-1} = 1|s_t = 1) & -\rho_1 P(s_{t-1} = 2|s_t = 1) \\ -\rho_2 P(s_{t-1} = 1|s_t = 2) & 1 - \rho_2 P(s_{t-1} = 2|s_t = 2) \end{bmatrix} \tag{A.4}$$

By the Bayes' rule, the conditional probability of s_{t-1} given s_t is easy to compute.

$$P(s_{t-1}|s_t) = \frac{P(s_{t-1})P(s_t|s_{t-1})}{\sum_{s_{t-1}=1,2} P(s_{t-1})P(s_t|s_{t-1})} \tag{A.5}$$

By calculation, $P(s_{t-1} = 1|s_t = 1) = p_{11}$, $P(s_{t-1} = 1|s_t = 2) = 1 - p_{22}$, $P(s_{t-1} = 2|s_t = 1) = 1 - p_{11}$, and $P(s_{t-1} = 2|s_t = 2) = p_{22}$. So the unconditional intensities are given by equations (2.10) and (2.11). □

Appendix B

Proof of Proposition 2.2

Proof: The return at time t can be divided into 3 parts, the mean μ, and $\epsilon_{1,t}$, whose conditional variance is σ_t^2, and the jump part, whose conditional variance is $(\theta^2 + \delta^2)\lambda_t$.

The conditional jump intensity equation can be rewritten as

$$\lambda_t = \alpha_1 + (\alpha_2 - \alpha_1)s_{t-1} + \rho_1\lambda_{t-1} + (\rho_2 - \rho_1)s_{t-1}\lambda_{t-1} + \gamma_1\xi_{t-1} + (\gamma_2 - \gamma_1)s_{t-1}\xi_{t-1} \quad \text{(B.1)}$$

Knight and Satchell (2011) present a TAR(1) model with constant intercept coefficient across regimes, and the above equation is a variation of their model with different coefficients in both intercept and error term.

Let

$$c_0 = \alpha_1(1-\pi) + \alpha_2\pi \; ; \; c_1 = \alpha_2 - \alpha_1$$

$$d_0 = \rho_1(1-\pi) + \rho_2\pi \; ; \; d_1 = \rho_2 - \rho_1$$

$$e_0 = \gamma_1(1-\pi) + \gamma_2\pi \; ; \; e_1 = \gamma_2 - \gamma_1$$

$B_{t-1} = s_{t-1} - \pi$

Then

$$\lambda_t = c_0 + c_1 B_{t-1} + (d_0 + d_1 B_{t-1})\lambda_{t-1} + (e_0 + e_1 B_{t-1})\xi_{t-1} \quad \text{(B.2)}$$

We have $P(B_{t-1} = -\pi) = 1 - \pi$, $P(B_{t-1} = 1 - \pi) = \pi$.

Backward substitution in (B.2) leads to

$$\begin{aligned}\lambda_t &= c_0 + c_1 B_{t-1} + \sum_{n=1}^{k-1}(c_0 + c_1 B_{t-n-1})[\prod_{m=1}^{n}(d_0 + d_1 B_{t-m})] + (e_0 + e_1 B_{t-1})\xi_{t-1} \\ &+ \sum_{n=1}^{k-1}(e_0 + e_1 B_{t-n-1})\xi_{t-n-1}[\prod_{m=1}^{n}(d_0 + d_1 B_{t-m})] + \lambda_{t-k}\prod_{m=1}^{k}(d_0 + d_1 B_{t-m})\end{aligned}$$

Following Quinn (1982) and Knight and Satchell (2011), and letting $G_n(t) = \prod_{m=1}^{n}(d_0 + d_1 B_{t-m})$,

$$\ln G_n(t) = \sum_{m=1}^{n} \ln(d_0 + d_1 B_{t-m})$$

and $\frac{1}{n}\ln|G_n(t)| \longrightarrow E(\ln|d_0 + d_1 B_{t-m}|)$

Then $G_n(t)\xi_{t-n-1}$ are geometrically bounded if $E(\ln|d_0 + d_1 B_{t-m}|) < 0$,i.e,
$\pi \ln|d_0 + d_1(1-\pi)| + (1-\pi)\ln|d_0 - d_1\pi| < 0$

or

$\pi \ln|\rho_2| + (1-\pi)\ln|\rho_1| < 0$

Then equation (B.2) has the solution that

$$\begin{aligned}\lambda_t &= c_0 + c_1 B_{t-1} + \sum_{n=1}^{\infty} G_n(t)(c_0 + c_1 B_{t-n-1}) \\ &+ (e_0 + e_1 B_{t-1})\xi_{t-1} + \sum_{n=1}^{\infty}(e_0 + e_1 B_{t-n-1})\xi_{t-n-1}G_n(t)\end{aligned} \quad \text{(B.3)}$$

From Quinn (1982) and Feigin and Tweedie (1985), the mean of the stationary distribution will exist, i.e., $E(|\lambda_t|) < \infty$, if $E(|d_0 + d_1 B_{t-1}|) < 1$.
That is, $|\rho_1|(1-\pi) + |\rho_2|\pi < 1$

Consequently, given $\rho_1(1-\pi) + \rho_2\pi < 1$, and $\alpha_1 < \infty$,$\alpha_2 < \infty$, the unconditional mean of λ_t exists: $E(\lambda_t) = c_0 + c_0(\sum_{n=1}^{\infty} d_0^n) = \frac{c_0}{1-d_0} = \frac{\alpha_1(1-\pi)+\alpha_2\pi}{1-\rho_1(1-\pi)-\rho_2\pi}$

To find the unconditional expectation of σ_t, the GARCH type conditional variance is written as

$$\sigma_t^2 = \omega_1 + (\omega_2 - \omega_1)s_{t-1} + a_1\epsilon_{t-1}^2 + (a_2 - a_1)s_{t-1}\epsilon_{t-1}^2 + b_1\sigma_{t-1}^2 + (b_2 - b_1)s_{t-1}\epsilon_{t-1}^2 \quad \text{(B.4)}$$

in which $\epsilon_{t-1}^2 = \epsilon_{1,t-1}^2 + \epsilon_{2,t-1}^2 + 2\epsilon_{1,t-1}\epsilon_{2,t-1}$

Let

$$f_0 = \omega_1(1-\pi) + \omega_2\pi \; ; \; f_1 = \omega_2 - \omega_1$$

$$g_0 = a_1(1-\pi) + a_2\pi \ ; \ g_1 = a_2 - a_1$$

$$h_0 = b_1(1-\pi) + b_2\pi \ ; \ h_1 = b_2 - b_1$$

$B_{t-1} = s_{t-1} - \pi$

Then

$$\begin{aligned}
\sigma_t^2 &= f_0 + f_1 B_{t-1} + g_0 \epsilon_{t-1}^2 + g_1 \epsilon_{t-1}^2 B_{t-1} + h_0 \sigma_{t-1}^2 + h_1 \sigma_{t-1}^2 B_{t-1} \\
&= f_0 + f_1 B_{t-1} + (g_0 + g_1 B_{t-1})(\epsilon_{2,t-1}^2 + 2\epsilon_{1,t-1}\epsilon_{2,t-1}) \\
&+ [(g_0 + g_1 B_{t-1}) z_{t-1}^2 + h_0 + h_1 B_{t-1}] \sigma_{t-1}^2 \qquad \text{(B.5)}
\end{aligned}$$

Backward substitution in (B.5) leads to

$$\begin{aligned}
\sigma_t &= f_0 + f_1 B_{t-1} + \sum_{n=1}^{k-1} \prod_{m=1}^{n} [(g_0 + g_1 B_{t-m}) z_{t-1}^2 + h_0 + h_1 B_{t-m}](f_0 + f_1 B_{t-m}) \\
&+ (g_0 + g_1 B_{t-1})(\epsilon_{2,t-1}^2 + 2\epsilon_{1,t-1}\epsilon_{2,t-1}) + \sum_{n=1}^{k-1} (g_0 + g_1 B_{t-n-1}) \\
&(\epsilon_{2,t-n-1}^2 + 2\epsilon_{1,t-n-1}\epsilon_{2,t-n-1}) \prod_{m=1}^{n} [(g_0 + g_1 B_{t-m}) z_{t-m}^2 + h_0 + h_1 B_{t-m}] \\
&+ \sigma_{t-k}^2 \prod_{m=1}^{k} [(g_0 + g_1 B_{t-m}) z_{t-m}^2 + h_0 + h_1 B_{t-m}]
\end{aligned}$$

Similarly, letting $Q_n(t) = \prod_{m=1}^{n} [(g_0 + g_1 B_{t-m}) z_{t-m}^2 + h_0 + h_1 B_{t-m}]$, if $E(\ln[(g_0 + g_1 B_{t-m}) z_{t-m}^2 + h_0 + h_1 B_{t-m}]) < 0$, that is, $(a_1 + b_1)(1-\pi) + (a_2 + b_2)\pi < 1$, equation (B.5) has the solution that

$$\begin{aligned}
\sigma_t^2 &= f_0 + f_1 B_{t-1} + \sum_{n=1}^{\infty} Q_n(t)(f_0 + f_1 B_{t-n-1}) + (g_0 + g_1 B_{t-1})(\epsilon_{2,t-1}^2 + 2\epsilon_{1,t-1}\epsilon_{2,t-1}) \\
&+ \sum_{n=1}^{\infty} (g_0 + g_1 B_{t-n-1})(\epsilon_{2,t-n-1}^2 + 2\epsilon_{1,t-n-1}\epsilon_{2,t-n-1}) Q_n(t)
\end{aligned}$$

Then the unconditional expectation can be computed as

$$E(\sigma_t^2) = f_0 + \sum_{n=1}^{\infty} E(Q_n(t)) f_0 + g_0 E(\epsilon_{2,t-1}^2) + \sum_{n=1}^{\infty} g_0 E(\epsilon_{2,t-1}^2) E(Q_n(t)) \qquad \text{(B.6)}$$

We have $E(\epsilon^2_{2,t-1}) = E(E(\epsilon_{2,t-1}|\lambda_{t-1})) = E((\theta^2+\delta^2)\lambda_{t-1}) = (\theta^2+\delta^2)E(\lambda(t-1))$
$E(Q_n(t)) = E(\prod_{m=1}^{n}((g_0+g_1B_{t-m})z^2_{t-m}+h_0+h_1B_{t-m})) = (g_0+h_0)^n$
Taking $E(Q_n(t))$ and $E(\epsilon^2_{2,t-1})$ into equation (B.6),

$$E(\sigma_t^2) \quad = \quad \frac{f_0}{1-g_0-h_0} + (\theta^2+\delta^2)\frac{(\alpha_1(1-\pi)+\alpha_2\pi)g_0}{(1-\rho_1(1-\pi)-\rho_2\pi)(1-g_0-h_0)} \qquad \text{(B.7)}$$

Thus,given $\omega_1 < \infty$, $\omega_2 < \infty$, $\alpha_1 < \infty$,$\alpha_2 < \infty$, and $\rho_1(1-\pi)+\rho_2\pi < 1$, $(a_1+b_1)(1-\pi)+(a_2+b_2)\pi < 1$, $Var(R_t) = E(\sigma_t^2)+(\theta^2+\delta^2)E(\lambda_t)$ is used to calculate the value of $Var(R_t)$. $\square$

Appendix C

Proof of Proposition 3.1

Proof: For a stationary single-regime ACD (1,1) model, let $E_j(\psi_i)$ denote the expected value of ψ_i conditional on the information set at transaction j. Then $E_0(\psi_i)$ is the expected value of ψ_i conditional on the initial value ψ_0.

Step 1. In order to prove the proposition, firstly we show that for the stationary single regime ACD (1,1) model, $E_0(\psi_i) = E(\psi_i) + o(1)_I$, for $i = 1, 2, ...I$, where the unconditional expectation is $E(\psi_i) = \omega/(1 - \alpha - \beta)$ and $o(1)_I \to 0$ as $I \to \infty$:

As $E(\epsilon_i) = 1$ and ϵ_i is i.i.d and independent of x_i, the expected value of ψ_i conditional on the information set at transaction $i - 2$ expressed as

$$\begin{aligned} E_{i-2}(\psi_i) &= \omega + \alpha E_{i-2}(x_{i-1}) + \beta E_{i-2}(\psi_{i-1}) \\ &= \omega + \alpha E_{i-2}(\psi_{i-1}\epsilon_{i-1}) + \beta E_{i-2}(\psi_{i-1}) \\ &= \omega + (\alpha + \beta) E_{i-2}(\psi_{i-1}) \end{aligned} \tag{C.1}$$

By iterative substitution,

$$\begin{aligned} E_0(\psi_i) &= E_0(\omega + \alpha x_{i-1} + \beta \psi_{i-1}) \\ &= \omega + \alpha E_0(\psi_{i-1}) + \beta E_0(\psi_{i-1}) \\ &= \omega + (\alpha + \beta) E_0(\psi_{i-1}) \\ &= ... = \omega \frac{1 - (\alpha + \beta)^i}{1 - \alpha - \beta} + (\alpha + \beta)^i \psi_0 \end{aligned} \tag{C.2}$$

Thus, for a stationary process with $\alpha+\beta<1$,

$$\begin{aligned} E_0(\psi_i)-E(\psi_i) &= \omega\frac{1-(\alpha+\beta)^i}{1-\alpha-\beta}+(\alpha+\beta)^i\psi_0-\frac{\omega}{1-\alpha-\beta} \\ &= (\alpha+\beta)^i(\psi_0-\frac{\omega}{1-\alpha-\beta})=o(1)_I \end{aligned} \tag{C.3}$$

Step 2. In step 2 we prove that the proposition holds for a ACD (1,1) model with segments of different parameters.
For the k*th* segment where $i=I_{k-1},...,I_k$, let $d_i=\psi_i-E_k(\psi_i)$ be the deviation of ψ_i from the expected value of it conditional on the initial value of the k*th* segment. Applying the Law of Large Numbers, within the k*th* segment,

$$\sum_{i=I_{k-1}+1}^{I_k} d_i/(I_k-I_{k-1})=o(1)_{I_k-I_{k-1}} \tag{C.4}$$

where I_k-I_{k-1} is the length of the k*th* segment. Then the sample mean of ψ_i on the entire sample can be expressed in terms of the summation of the conditional mean on the initial value of each segment and the deviations.

$$\begin{aligned} \overline{\psi} &= \sum_{i=1}^{I}\psi_i/I \\ &= \frac{1}{I}(\sum_{i=1}^{I_1}E_1(\psi_i)+\sum_{i=1}^{I_1}d_i+...+\sum_{i=I_{K-1}}^{I}E_K(\psi_i)+\sum_{i=I_{K-1}}^{I}d_i) \\ &= \frac{1}{I}(\sum_{i=1}^{I_1}E_1(\psi_i)+...+\sum_{i=I_{K-1}}^{I}E_K(\psi_i))+\sum_{k=1}^{K}o(1)_{I_k-I_{k-1}} \end{aligned} \tag{C.5}$$

Using the result obtained in step 1, we have

$$\begin{aligned} \overline{\psi} &= \frac{1}{I}(\sum_{i=1}^{I_1}(E(\psi_i)+o(1)_{I_1})+...+\sum_{i=I_{K-1}}^{I}(E(\psi_i)+o(1)_{I-I_{K-1}}))+\sum_{i=1}^{I}o(1)_{I_i-I_{i-1}} \\ &= \frac{1}{I}\sum_{k=1}^{K}(I_k-I_{k-1})E(\psi_i(k))+\frac{1}{I}(\sum_{i=1}^{I_1}o(1)_{I_1}+...+\sum_{i=I_{K-1}}^{I}o(1)_{I-I_{K-1}}) \\ &\quad +\sum_{k=1}^{K}o(1)_{I_k-I_{k-1}} \end{aligned} \tag{C.6}$$

where $E(\psi_i(k))$ is the unconditional expectation of the conditional expected duration within the k*th* segment. With the assumption that the segment length $I_k - I_{k-1} \to \infty$ as $I \to \infty$, the last two terms in the above equation can be written as $o(1)_I$. Therefore,

$$\overline{\psi} = \frac{1}{I} \sum_{k=1}^{K} (I_k - I_{k-1}) E(\psi_i(k)) + o(1)_I \tag{C.7}$$

□

Currirculum Vitae

Name:	Pujun Liu
Place of Birth:	Hubei, China
Year of Birth:	1985
Post-Secondary Education and Degrees:	Huazhong University of Science and Technology Wuhan, China 2001-2005 B.A. (Economics) and B.S. (Mathematics).
	The University of Western Ontario London, Ontario 2005-2006 M.A. (Economics)
	The University of Western Ontario London, Ontario 2006-2012 Ph.D. (Economics)
Honors and Awards:	Western Graduate Research Scholarship The University of Western Ontario 2005-2010
	Student Excellence Scholarship Huazhong University of Science and Technology 2003-2005
Related Work Experience:	*Teaching Assistant* The University of Western Ontario 2005-2010
	Instructor The University of Western Ontario 2008, 2011
	Research Assistant The University of Western Ontario 2008, 2011