



Statistical Inference for Poisson Time Series  
Models

Abdullah Maedh Almarashi

Department of Mathematics and Statistics

University of Strathclyde

Glasgow, UK

This thesis is submitted to the University of Strathclyde for the  
degree of Doctor of Philosophy in the Faculty of Science

2014

# Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

©The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

Date:

# Acknowledgements

First of all, I would like to express my sincere thanks and gratitude to Allah that gave me the patience, strength and the ability to finish this work.

Secondly, I would like to thank the Ministry of Higher Education of Saudi government and University of King Abdulaziz for their financial support.

I am extremely thankful to my supervisor Dr Jiazhu Pan. He has helped, supported and encouraged me in everything during the course of my studies.

I would also like to thank Prof Kurt Brännäs and Prof William Dunsmuir for providing me with the data sets that were used in Chapter 3 and Chapter 5 respectively.

Special thanks go to my friend Alexios Prelorendjos, and I am grateful for helping me in many aspects of my thesis during my time in Glasgow.

Many thanks for all people staff and students in the Department of Mathematics and Statistics in University of Strathclyde, who have helped me directly or indirectly during my research period.

Next, I would like to thank my parents, brothers and sisters for their support and encouragement and special thanks to my wife (Amina) and my children (Shada, Ziyad, Nada, Tareq and Jenan) for all love, supporting and encouragement as well as for the patience for all that time away from their friends. So I am grateful to them.

Last but not least, many thanks for people who gave me suggestions, constructive comments and some various comments during my study period.

# Abstract

There are many nonlinear econometric models which are useful in analysis of financial time series. In this thesis, we consider two kinds of nonlinear autoregressive models for nonnegative integer-valued time series: threshold autoregressive models and Markov switching models, in which the conditional distribution given historical information is the Poisson distribution. The link between the conditional variance (i.e. the conditional mean for the Poisson distribution) and its past values as well as the observed values of the Poisson process may be different according to the threshold variable in threshold autoregressive models, and to an unobservable state variable in Markov switching models in different regimes. We give a condition on parameters under which the Poisson generalized threshold autoregressive heteroscedastic (PTGARCH) process can be approximated by a geometrically ergodic process. Under this condition, we discuss statistical inference (estimation and tests) for PTGARCH models, and give the asymptotic theory on the inference. The complete structure of the threshold autoregressive model is not exactly specific in economic theory for the most financial applications of the model. In particular, the number of regimes, the value of threshold and the delay parameter are often unknown and cannot be assumed known. Therefore, in this research, the performance of various information criteria for choosing the number of regimes, the threshold value and the delay parameters for different sample sizes is investigated. Tests for threshold nonlinearity are applied. The characteristics of Markovian switching Poisson generalized autoregressive hetero-

stochastic (MS-PGARCH) models are given, and the maximum likelihood estimation of parameters is discussed. Simulation studies and applications to modelling financial counting time series are presented to support our methodology for both the PTGARCH model and the MS-PGARCH model.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	1
1.2	Overview of the study . . . . .	6
<b>2</b>	<b>The PTGARCH Model and its Statistical Inference</b>	<b>8</b>
2.1	The Model . . . . .	8
2.1.1	Definition of the model . . . . .	8
2.1.2	Ergodicity of a perturbed model . . . . .	9
2.2	Statistical Inference . . . . .	25
2.2.1	Parameter estimation and information criteria . . . . .	25
2.2.1.1	Conditional maximum likelihood inference . . . . .	25
2.2.1.2	Akaike information criterion (AIC) . . . . .	44
2.2.1.3	Bayesian information criterion (BIC) . . . . .	44
2.2.1.4	Hannan-Quinn information criterion (HQIC) . . . . .	45
2.2.2	Testing nonlinearity . . . . .	45
2.2.2.1	Lagrange multiplier (LM) test . . . . .	46
2.2.2.2	Likelihood ratio (LR) test . . . . .	47
2.3	Summary . . . . .	48
<b>3</b>	<b>Simulation and Data Analysis</b>	<b>49</b>
3.1	Numerical results: simulation and real data example . . . . .	50
3.1.1	The performance of Maximum Likelihood Estimators (MLE) . . . . .	50

3.1.2	Determination of the threshold value $r$ . . . . .	63
3.1.3	Test for threshold nonlinearity (Lagrange Multiplier Test)	66
3.1.4	A real data example . . . . .	67
3.2	Summary . . . . .	69
<b>4</b>	<b>Determination of the Number of Regimes</b>	<b>72</b>
4.1	Introduction . . . . .	72
4.2	The PTGARCH(1,1; $s$ ,1) Model . . . . .	73
4.3	Determination of the number of regimes with different information criteria . . . . .	76
4.4	Simulations . . . . .	77
4.4.1	Experiment study 1: Linearity model against nonlinearity	78
4.4.2	Experiment study 2: Linear PGARCH (1,1) against PTG- ARCH(1,1:2,1) . . . . .	82
4.4.3	Experiment study 3: Nonlinear PTGARCH (1,1:2,1) with different coefficients, and different order . . . . .	86
4.4.4	Experiment study 4: Nonlinear PTGARCH (1,1:3,1) with different coefficients . . . . .	94
4.5	A real data example . . . . .	97
4.6	Summary . . . . .	112
<b>5</b>	<b>Simultaneous Determination of the Time Delay and the Threshold Value</b>	<b>114</b>
5.1	Introduction . . . . .	114
5.2	Determination of the delay parameter $d$ . . . . .	115
5.3	Simultaneous determination of the delay parameter $d$ and the value of threshold $r$ . . . . .	119
5.4	A real data example . . . . .	122
5.5	Summary . . . . .	123

<b>6</b>	<b>Poisson Markov Switching Model</b>	<b>128</b>
6.1	Introduction . . . . .	128
6.2	The Markov Switching Model of Conditional Variance (Mean) . .	129
6.2.1	Definition of the model . . . . .	129
6.2.2	Ergodicity of a perturbed model . . . . .	130
6.2.3	A simple model . . . . .	147
6.3	Statistical Inference . . . . .	149
6.3.1	Maximum likelihood estimation of MS-PGARCH model . .	150
6.3.2	Simulation . . . . .	153
6.3.3	A real data example . . . . .	186
6.4	Summary . . . . .	187
<b>7</b>	<b>Conclusions and Further Work</b>	<b>190</b>
7.1	Conclusions . . . . .	190
7.2	Further work . . . . .	192
	<b>Bibliography</b>	<b>194</b>



# List of Tables

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The PTGARCH Model and its Statistical Inference</b>	<b>8</b>
<b>3</b>	<b>Simulation and Data Analysis</b>	<b>49</b>
3.1	Maximum likelihood estimates and least-squares estimates with their mean square errors . . . . .	53
3.2	Comparison of simulated standard error and standard error from $G(\theta_0)$ when initial values for parameters by the ARIMA(1,1) fit to the data . . . . .	54
3.3	Maximum likelihood estimates with their mean square errors . . .	55
3.4	Comparison of simulated standard error and standard error from $G(\theta_0)$ when the initial values follow uniform distribution . . . . .	56
3.5	Average of MSPR for two nonlinear models . . . . .	61
3.6	The average of AIC, BIC and HQIC values of the estimated threshold value . . . . .	64
3.7	The percentage of correct choices and average of AIC, BIC and HQIC values of the estimated the value of threshold . . . . .	65
3.8	Achieved significance level of the $LM^*$ test for threshold nonlinearity when $\theta_{10} \neq \theta_{20}$ . . . . .	67
3.9	Achieved significance level of the $LM^*$ test for threshold nonlinearity when $\theta_{10} = \theta_{20}$ . . . . .	67

3.10	Possible threshold values for the number of transactions . . . . .	68
3.11	The estimated parameters of the data example . . . . .	68
<b>4</b>	<b>Determination of the Number of Regimes</b>	<b>72</b>
4.1	The percentage of correct model detecting for a nonlinear PTGARCH when compared to a linear PGARCH (1,1) model using the LR and the LM test . . . . .	79
4.2	The percentage of correct model detecting for a linear PGARCH when compared to a nonlinear PTGARCH model using the LR and the LM test . . . . .	81
4.3	The percentage of correct model detecting for a linear PGARCH when compared to a nonlinear PTGARCH model using various information criteria . . . . .	82
4.4	The percentage of correct model detection when a linear PTGARCH is compared to a linear PGARCH model using the LR test	83
4.5	The percentage of correct model detection for a nonlinear PTGARCH model when compared to a linear PGARCH model for different threshold values using different information criteria . . .	84
4.6	The percentage of correct model detection for a nonlinear PTGARCH model when compared to a linear PGARCH model for different threshold values using the LR test . . . . .	85
4.7	Model selection for a linear GARCH model and two different PTGARCH models using different information criteria . . . . .	87
4.8	Model selection for a linear GARCH model and two different PTGARCH models using different information criteria . . . . .	90
4.9	Model selection for a linear GARCH model and two different PTGARCH models using different information criteria . . . . .	91
4.10	Model selection for a linear GARCH model and two different PTGARCH models using different information criteria . . . . .	92

4.11	Model selection percentage for a different nonlinear PTGARCH models using various information criteria . . . . .	93
4.12	The percentage of correct model detection for a nonlinear PTGARCH model when compared to a nonlinear PTGARCH model for different threshold values and $s = 3$ . . . . .	95
4.13	Model selection percentages for a nonlinear PTGARCH model ( $s = 3$ ) when compared with linear PGARCH model ( $s = 1$ ) and nonlinear PTGARCH model ( $s = 2$ ) . . . . .	96
4.14	Descriptive statistics for transaction data. . . . .	98
4.15	The estimated parameters for the TGARCH model . . . . .	98
4.16	The estimated parameters for the TGARCH model . . . . .	99
4.17	The estimated parameters for the TGARCH model . . . . .	100
4.18	The estimated parameters for the TAR model . . . . .	101
4.19	The estimated parameters for the TAR model . . . . .	102
4.20	The estimated parameters for the TGARCH model . . . . .	103
4.21	Linear and nonlinear autoregressive models . . . . .	105
4.22	Nonlinear autoregressive model . . . . .	107
4.23	Nonlinear autoregressive model . . . . .	108
4.24	The estimated parameters of the data example . . . . .	109
4.25	Summary of experiment studies . . . . .	113

## 5 Simultaneous Determination of the Time Delay and the Threshold

<b>Value</b>	<b>114</b>	
5.1	The relative rates of correct choice of threshold delay using various information criteria . . . . .	117
5.2	The relative rates of correct choice of possible threshold variables $d$	118
5.3	The relative rates of correct choice of threshold delay and threshold value together . . . . .	121

5.4	Suggested threshold values and possible time delay for the number of daily admissions for asthma to a single hospital . . . . .	125
5.5	The estimated threshold value and time delay for nonlinear model for real data . . . . .	125
5.6	The estimated parameters of the data example . . . . .	126
<b>6</b>	<b>Poisson Markov Switching Model</b>	<b>128</b>
6.1	Maximum likelihood estimates with their mean square errors . . .	154
6.2	The simulated standard error for model (6.9) and (6.10) . . . . .	154
6.3	Distribution of the maximum likelihood estimates of model (6.9) and (6.10) . . . . .	156
6.4	Maximum likelihood estimates with their mean square errors (in parentheses) . . . . .	157
6.5	The simulated standard errors for two different models for different sample sizes . . . . .	158
6.6	The variance of average absolute error for two different models for different sample sizes . . . . .	158
6.7	Summary of average absolute error of two various models for different sample sizes . . . . .	159
6.8	Distribution of the maximum likelihood estimates of model (6.9) and (6.10) . . . . .	160
6.9	Distribution of the maximum likelihood estimates of model (6.9) and (6.10) . . . . .	161
6.10	Maximum likelihood estimates with their mean square errors (in parentheses) . . . . .	167
6.11	The simulated standard errors for model (6.10) . . . . .	168
6.12	The variance of average absolute error for model (6.9) and (6.10) for different sample sizes . . . . .	168

6.13	Summary of average absolute error for model (6.10) for different sample sizes . . . . .	168
6.14	Distribution of the maximum likelihood estimates of model (6.9) and (6.10) . . . . .	169
6.15	Maximum likelihood estimates with their mean square errors (in parentheses) . . . . .	174
6.16	The simulated standard error for two different MS-GARCH (2;1,1) models for different sample sizes for different sample sizes . . . . .	175
6.17	The variance of average absolute error for two different MS-GARCH (2;1,1) models for different sample sizes for different sample sizes . . . . .	175
6.18	Summary of average absolute errors for two different MS-GARCH (2;1,1) models for different sample sizes for different sample sizes . . . . .	176
6.19	Distribution of the maximum likelihood estimates of model (6.9) and (6.10) . . . . .	177
6.20	Maximum likelihood estimates with their mean square errors (in parentheses) . . . . .	180
6.21	The simulated standard error for model (6.18) . . . . .	181
6.22	Distribution of the maximum likelihood estimates of model (6.18) . . . . .	182
6.23	The estimated parameters for the Markov switching model . . . . .	186
<b>7</b>	<b>Conclusions and Further Work</b>	<b>190</b>

# List of Figures

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The PTGARCH Model and its Statistical Inference</b>	<b>8</b>
<b>3</b>	<b>Simulation and Data Analysis</b>	<b>49</b>
3.1	Histograms and QQ-plots of the sampling distribution of the MLE and LSE estimator . . . . .	57
3.2	Boxplots of average absolute errors (AAE) for maximum likelihood estimator and the conditional least squares estimator . . . . .	58
3.3	Boxplot of average absolute error of MLE for the first nonlinear model . . . . .	59
3.4	Boxplot of average absolute error of MLE for the second nonlinear model . . . . .	60
3.5	Histograms and QQ-plot of the sampling distribution of the estimators for the second model . . . . .	61
3.6	Boxplot of MSPR for two different models . . . . .	62
3.7	Time plot, histogram and autocorrelation function of real data . .	70
3.8	Pearson residuals . . . . .	71
<b>4</b>	<b>Determination of the Number of Regimes</b>	<b>72</b>
4.1	Time series and their autocorrelation functions . . . . .	75

4.2	Time plot, histogram and autocorrelation function of the number of transactions per minute for the stock Ericsson B . . . . .	97
4.3	Time plot of the first and the second regime separately for real data	110
4.4	Pearson residuals . . . . .	111
<b>5</b>	<b>Simultaneous Determination of the Time Delay and the Threshold</b>	
	<b>Value</b>	<b>114</b>
5.1	Time series of the asthma data . . . . .	124
5.2	Time series of asthma counts for the first and the second regime separately . . . . .	126
5.3	Pearson residuals and their cumulative periodogram plot . . . . .	127
<b>6</b>	<b>Poisson Markov Switching Model</b>	<b>128</b>
6.1	Boxplot of AAE for maximum likelihood estimator of model (6.9) and (6.10) for different sample sizes . . . . .	155
6.2	Boxplot of the values of estimated parameters of model (6.9) and (6.10) . . . . .	162
6.3	Boxplot of AAE for maximum likelihood estimator of model (6.9) and (6.10) . . . . .	163
6.4	Histogram of AAE for maximum likelihood estimator of model (6.9) and (6.10) . . . . .	164
6.5	Boxplot of AAE for maximum likelihood estimator of model (6.9) and (6.10) . . . . .	165
6.6	Histogram of AAE for maximum likelihood estimator of model (6.9) and (6.10) . . . . .	166
6.7	Boxplot of AAE for maximum likelihood estimator of model (6.9) and (6.10) . . . . .	170
6.8	Boxplot of AAE for maximum likelihood estimator of model (6.9) and (6.10) . . . . .	171

6.9	Histogram of AAE for maximum likelihood estimator of model (6.9) and (6.10) . . . . .	172
6.10	Boxplot of AAE for maximum likelihood estimator of model (6.9) and (6.10) . . . . .	176
6.11	Boxplot of AAE for maximum likelihood estimator of model (6.9) and (6.10) . . . . .	178
6.12	Histogram of AAE for maximum likelihood estimator of model (6.9) and (6.10) . . . . .	179
6.13	Boxplot of AAE for maximum likelihood estimator of model (6.18)	183
6.14	Boxplot of AAE for maximum likelihood estimator of model (6.9) and (6.10) . . . . .	184
6.15	Histogram of AAE for maximum likelihood estimator of model (6.18)	185
6.16	Filtered and smoothed probabilities . . . . .	188
<b>7</b>	<b>Conclusions and Further Work</b>	<b>190</b>



# Abbreviations

<b>ACF</b>	Autocorrelation Function
<b>AIC</b>	Akaike Information Criterion
<b>BIC</b>	Bayesian Information Criterion
<b>HQIC</b>	Hannan-Quinn Information Criterion
<b>MLE</b>	Maximum Likelihood Estimation
<b>MSE</b>	Mean Square Error
<b>CLSE</b>	Conditional Least Square Estimation
<b>LM</b>	Lagrange Multiplier Test
<b>LR</b>	Likelihood Ratio Test
<b>AAE</b>	Average Absolute Error
<b>ARMA</b>	Autoregressive Moving Average
<b>ARIMA</b>	Autoregressive Integrated Moving Average
<b>GARCH</b>	Generalized Autoregressive Conditional Heteroscedastic
<b>PGARCH</b>	Poisson Generalized Autoregressive Conditional Heteroscedastic
<b>PTGARCH</b>	Poisson Threshold Generalized Autoregressive Conditional Heteroscedastic
<b>AR</b>	Autoregressive Process
<b>TAR</b>	Threshold Autoregressive Process
<b>MSAR</b>	Markov Switching Autoregressive Process
<b>MS-PGARCH</b>	Markov Switching Poisson Generalized Autoregressive Conditional Heteroscedastic
<b>MSPR</b>	Mean Square of the Pearson Residuals

# Chapter 1

## Introduction

### 1.1 Background

A time series can be discrete or continuous valued and may be modelled linearly or nonlinearly. For example, AR models, MA models and ARMA models are linear models. These models are usually used for modelling means. These kinds of models have become more popular nowadays, partly because they have been integrated into much economics and statistics software. In contrast, they have some limitations, for example, they cannot handle many dynamic patterns of variance, e.g. volatility clustering (the variance of the returns is high for extended periods of time, and then becomes low for other extended periods) and asymmetry (rising and declining patterns). Many applications in economic time series exhibit periods of unusually large volatility followed by periods of relative tranquility (*i.e* the variance does not stay constant over time). In such circumstances, if we plan to pay the price of assets at time  $t$  and sell at time  $t + 1$ , then the conditional variance would be important, but the unconditional variance would not. Therefore, it is informative to introduce an explanatory variable which helps to predict the volatility. If the explanatory variables are constant over time then the conditional variance is constant. If the explanatory variables change over

time, then the conditional variance is a function of time (it is a stochastic process!). Actually, in time series analysis, there are many volatility models, which are used to forecast volatility. In widespread use, there are two different classes of volatility models. First, the conditional variance is considered directly as a function of observed variables. The second kind formulates a volatility model that is not a function purely of observed variables (latent volatility). This research is concerned with the first type of volatility models. The simplest examples here are autoregressive conditional heteroscedastic (ARCH) models (Engle (1982)), CHARMA models (Tsay (1987)), and generalized autoregressive conditional heteroscedastic (GARCH) models (Bollerslev (1986)). A GARCH  $(p,q)$  model is defined as a process  $\{Y_t\}_{t \in Z}$  which satisfies

$$\begin{cases} Y_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, \sigma_t^2), \forall t \in Z, \\ \sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i Y_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \end{cases}$$

where  $\alpha_0 > 0, \alpha_i \geq 0, i = 1, \dots, q, \beta_j \geq 0, j = 1, \dots, p$ . But, up to recent years, most of the literature on this kind of model has been continuous-valued time series.

In the past few decades, many different nonlinear models have been proposed (see Tong and Series (1990) and Granger and Terasvirta (1993) for example). TAR models (Tong and Lim (1980)), STAR models (Chan and Tong (1986)), MSA models (Hamilton (1989)) and SETAR models (Chan and Tsay (1998)) are among the popular ones. Regime-switching models such as the threshold autoregressive (TAR) and the Markov switching autoregressive (MSAR) model have become increasingly popular for nonlinear time series and state regime switching phenomena in various fields such as science and social science studies. There are many studies concerning these kinds of models, (see, Tong (1978), Tong and Lim (1980), Caner and Hansen (2001), Tsay (2005), Enders (2010), Hansen (2011), Medeiros et al. (2002), Engel and Hamilton (1989), Huang (1999) and Kuan (2002)). These models consist of several regimes which can describe the time

series pattern in various states. The switching between states is according to threshold variable ( $Y_{t-d}$ ) in the TAR model and unobservable (hidden) state ( $S_t$ ) in the Markov switching model.

TAR models are used for capturing nonlinearities such as asymmetric features and jump phenomena of the conditional mean of a time series. This kind of models is characterized by piecewise linear processes separated according to the magnitude of a threshold variable. More recently, people have applied the threshold idea to modelling the conditional variance (i.e. volatility) of a time series. In the management of risk, the volatility is considered as a measure of risk of an economic position. Threshold GARCH (TGARCH) models, have been proposed to describe the asymmetrical features of the time-varying volatility of time series (see, Higgins and Bera (1992), Li and Li (1998), Hwang and Kim (2004), Hwang and Basawa (2004) and Pan et al. (2008)). However, most of the research on TAR and TGARCH models in the literature is about continuous-valued time series, and this project will focus on the statistical work on threshold models for discrete-valued time series.

Recently, there has been growing interest in modeling discrete-valued time series, such as count time series, that arise in various fields such as environmental, medical or financial and economic applications. Examples include the monthly cases of rare infectious diseases reported in a specified area, the number of customers waiting to be served at a counter recorded at discrete points in time, the daily number of absent workers in a firm, the monthly number of claimants collecting wage loss benefit for injuries in the workplace, the number of transaction times for the shares of a company in the markets and so on. There has been some research on discrete-valued time series analysis. McKenzie (1985) is a survey paper of early work on linear models for discrete-valued time series such as discrete AR, MA and ARMA. Chapter 4 of Kedem and Fokianos (2002) gives a comprehensive account for models for time series of counts. The most popu-

lar model among authors is the log-linear model. But recently published studies focus on a specified discrete conditional distribution with a conditional mean as a linear function of historical observations (see, Ferland et al. (2006), Fokianos et al. (2009), Fokianos and Fried (2010), and Weiss and Testik (2009)). They have discussed the Poisson autoregressive models, which assume the conditional distribution given the history is a Poisson distribution with mean as a linear function of previous mean and previous states. Because the mean and variance of a Poisson distribution are the same, the model for conditional mean and conditional variance coincides. Therefore, the Poisson autoregressive model is called Poisson GARCH model too. All previously mentioned models assume that the conditional mean (and variance) is a linear function of the past observations so that we call them linear Poisson AR models or linear Poisson GARCH models.

In this research, we study a type of threshold autoregressive model for non-negative integer-valued time series in which the conditional distribution given the history is a Poisson distribution, but the conditional mean (i.e. variance) is a nonlinear threshold function of the past observations as well as the past conditional means (and variances), which may be different when the threshold variable is in different regimes. It will be shown that this kind of models have good probabilistic properties and are good at capturing the asymmetrical features of the count time series. This is an attempt on discrete valued threshold AR (or GARCH) models. We discuss the conditions for geometric ergodicity of the proposed Poisson threshold AR (or GARCH) models, which are denoted by PTAR (or PTGARCH). We derive the maximum likelihood estimation for PTAR (or PTGARCH) and prove the consistency and asymptotic normality of the proposed estimators.

In forecasting and analyzing time series observations, there are surely tradeoffs between different stochastic time series models. Therefore, it is very important to look for a suitable test of the capability of selecting a practitioner with respect to

whether a linear or a nonlinear model is better to describe the time series under study. A test for threshold nonlinearity has been discussed by many authors. For example, Tsay (1989) suggests some predictive residuals and computes the simple statistic for an arranged autoregression, Cryer and Chan (2008) apply a Likelihood Ratio test with threshold model as the alternative hypothesis, and Al-Awadhi and Jolliffe (1998) use the value of Normalized Akaike's Information Criterion (NAIC), as well as the value of mean square error (MSE). In this project, we also discuss the threshold nonlinearity test of PTAR (or PTGARCH). Nonlinearity is caused by the threshold variable switching between regimes and by the model having more than one regime with different linear PGARCH models.

In many time series applications of the TAR model, both the value of threshold and the time delay often cannot be assumed known. Furthermore, the number of regimes in the model is often unknown as well. Some proposals exist on how to select the threshold value, the threshold delay and the number of regimes in TAR models. Tsay (1989) proposed to select the threshold value using a graphical method (in arranged autoregression). To choose the number of regimes he used scatterplots of predictive residuals and recursive  $t$ -ratios of an AR model's parameters against the threshold variable. Aide et al. (1988) suggested that for determining the value of threshold and the delay parameter, the AIC criterion and the normalized value of AIC can be used. They used a grid search: the original data series is arranged from small to large so that the sample of the threshold variable is divided into  $\alpha$  intervals with  $\beta$  grid points corresponding to  $\nu \times 100th$  percentile point to the  $\nu \times 100th$  percentile point of the observations, say between the 30th and 70th percentile points, the corresponding values of the AIC are calculated, and the estimated value of the threshold is the one which minimizes the value of the AIC. Strikholm and Teräsvirta (2006) used smooth transition autoregression to determine the number of regimes in TAR models. Hamaker (2009) suggested an information criteria to estimate the number of re-

gimes in threshold autoregressive models in small to moderate sample sizes. To determine the appropriate value of threshold and the delay of parameter as well as the number of regimes in the PTGARCH model, we use the different information criteria, which includes  $-2$  times the log-likelihood function and the number of unknown parameters.

We carry out much simulation work on the proposed methods for estimation of parameters and determination of threshold value, delay parameter and the number of regimes. We also give a real data example to show that the proposed model can fit to data better than the single regime linear Poisson GARCH model. Regarding the MS-PGARCH model many experiments are carried out for the maximum likelihood estimation for different cases. An application of the MS-GARCH model to a real time series shows that the suggested MS-GARCH models can fit to data well.

## **1.2 Overview of the study**

In this thesis, we study the statistical inference for threshold autoregressive models and Markov switching autoregressive models for nonnegative integer-valued time series. The conditional distribution given a history in such models is the Poisson distribution. The research studies the probabilistic and statistical properties of these models as well as introducing a huge number of simulation experiments to examine the suggested methods.

In Chapter 1, we review the history and background of nonlinear time series. Chapter 2 introduces the Poisson threshold GARCH (PTGARCH) model and discusses its probabilistic properties such as geometric ergodicity. We use the "perturbation" approach introduced by Fokianos et al. (2009) to obtain a geometrically ergodic approximation to the model. Statistical tools including the maximum likelihood estimators of parameters, their asymptotic theory and information criteria are given. Furthermore, we discuss two types of nonlinearity

tests: the Lagrange multiplier and the Likelihood Ratio test. Chapter 3 presents many simulation results to illustrate the parameter estimation for the PTGARCH model such as autoregressive coefficients and the threshold value as well as non-linearity tests. At the end of Chapter 3 a real data example is given, to support the proposed model by numerical evidence. Chapter 4 focuses on the determination of the number of regimes in PTGARCH models, using various information criteria. In Chapter 5, we illustrate the determination of the time delay and the simultaneous determination of the delay parameter and the value of threshold for the PTGARCH model. It presents several simulations and some real data examples to evaluate the model performance using various information criteria (AIC, BIC and HQIC). The general framework in this chapter comes from Aide et al. (1988), Tong and Lim (1980) and Al-Awadhi and Jolliffe (1998). Chapter 6 introduces another kind of nonlinear time series model for count time series. The switches between the regimes are according to an unobservable state variable. This model is called a Poisson Markov switching model for count time series. Statistical inference, a few simulation results on the maximum likelihood estimation for the model and real data analysis are discussed. Chapter 7, the last chapter of the thesis, consists of a summary of our findings and some problems for further work related to this area.



# Chapter 2

## The PTGARCH Model and its Statistical Inference

The current chapter provides an introduction to unperturbed and perturbed models, as well as some basic concepts in time series analysis and statistics. The purpose of this chapter is to present a kind of threshold GARCH model for non-negative integer valued time series and to discuss the link between unperturbed and perturbed models and the probabilistic aspect of time series such as geometric ergodicity (and stationarity), which will be used in the following chapters of the thesis.

### 2.1 The Model

#### 2.1.1 Definition of the model

Suppose that  $\{Y_t\}_{t \geq 1}$  is a nonnegative integer-valued time series. Consider the model defined by

$$\begin{cases} Y_t | \mathcal{F}_{t-1}^{Y, \lambda} \sim \text{Poisson}(\lambda_t), \\ \lambda_t = \phi_{j0} + \phi_{j1} Y_{t-1} + \phi_{j2} \lambda_{t-1}, \quad \text{when } r_{j-1} \leq Y_{t-d} < r_j, \quad j = 1, \dots, s, \end{cases} \quad (2.1)$$

for  $t \geq d$ , where  $Y_0, \lambda_0, \dots, Y_{d-1}, \lambda_{d-1}$  are fixed, the parameters  $\phi_{j0}, \phi_{j2}$  are assumed to be positive real numbers,  $\phi_{j1}$  are assumed to be nonnegative real numbers,  $0 = r_0 < r_1 < \dots < r_{s-1} < r_s = \infty$  are positive integers,  $Y_{t-d}$  is called the threshold variable with delay  $d$ ,  $r_j$  are called threshold values, and  $\mathcal{F}_{t-1}$  denotes the information set available at time  $t - 1$  which is a  $\sigma$ -field generated by  $(Y_{t-1}, Y_{t-2}, \dots, Y_0, \lambda_0)$ . One of the properties of the Poisson distribution is the first moment and the second moments are the same, that is,  $E[Y_t | \mathcal{F}_{t-1}^{Y, \lambda}] = Var[Y_t | \mathcal{F}_{t-1}^{Y, \lambda}] = \lambda_t$ . Therefore, we can call the model defined by (2.1) a Poisson threshold generalized autoregressive (PTGAR) model with  $s$  regimes, order  $(1, 1)$  and delay  $d$ , denoted by  $PTGAR(1, 1; s, d)$ . Because this is also a model for conditional variance, similar to a GARCH model (Bollerslev (1986)), we can call it a Poisson threshold generalized autoregressive conditional heteroscedastic (PTGARCH) model with  $s$  regimes, denoted by  $PTGARCH(1, 1; s; d)$ .

The model (2.1) can be rewritten in a form of Poisson process by assuming that the time series of counts  $\{Y_t\}$  is equal to the number of events of  $N_t(\cdot)$  in the time interval  $[0, \lambda_t]$  as follows:

$$\begin{cases} Y_t = N_t(\lambda_t) \\ \lambda_t = \phi_{j0} + \phi_{j1}Y_{t-1} + \phi_{j2}\lambda_{t-1}, \quad \text{when } r_{j-1} \leq Y_{t-d} < r_j, \quad j = 1, \dots, s. \end{cases} \quad (2.2)$$

for  $t \geq d$  and both  $Y_0$  and  $\lambda_0$  are fixed. Here  $N_t(\lambda_t)$  denotes a Poisson process with intensity  $\lambda_t$ , and  $\lambda_t$  is defined by (2.1).

## 2.1.2 Ergodicity of a perturbed model

To create a PTGAR model, we need to estimate the parameters. To this end, we would like to discuss stationarity and geometrical ergodicity for  $\{Y_t\}$ . Stationarity and ergodicity are important concepts in time series analysis. It is informative to present definitions of these concepts.

**Definition 1.** If  $(Y_{t_1}, Y_{t_2}, Y_{t_3}, \dots, Y_{t_n})$  and  $(Y_{t_1+l}, Y_{t_2+l}, Y_{t_3+l}, \dots, Y_{t_n+l})$

have the same joint distribution for all  $t_1 < t_2 < \dots < t_n$ ,  $n$ , and  $l$ , then we say that the process  $\{Y_t\}$  is strictly stationary. That is,  $\{Y_t\}$  is strictly stationary if and only if

$$F_{t_1, t_2, \dots, t_n}(y_1, y_2, y_3, \dots, y_n) = F_{t_1+l, t_2+l, \dots, t_n+l}(y_1, y_2, y_3, \dots, y_n)$$

where  $F_{t_1, t_2, \dots, t_n}(y_1, y_2, y_3, \dots, y_n)$  denotes the joint distribution function of  $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ .

**Definition 2.** The process  $\{Y_t\}$  is called geometrically ergodic, if there exists a probability vector  $\nu = (\nu_1, \dots, \nu_s)$  with  $\nu_j \geq 0$ ,  $\sum_{j=1}^s \nu_j = 1$ , and a number  $0 < A < 1$  such that

$$|P_{ij}^n - \nu_j| \leq M_{ij} A^n \quad \text{for all } i, j, n$$

where  $P_{ij}^n$  is the  $n$ -step transition probability, and  $M_{ij}$  is a constant depending on  $i, j$ .

To show geometric ergodicity, we need  $\psi$ -irreducibility. Here  $\psi$  represents the Lebesgue measure with support interval  $[k, \infty)$  for some  $k \geq \lambda^*$  where  $\lambda^* = \frac{\phi_{10}}{1-\phi_{12}}$  is the solution of the skeleton of the model defined by (2.2). Establishing  $\psi$ -irreducibility and finding a small set for the model defined by (2.2) is quite complicated. We are going to resort a perturbed chain  $(Y_t^m, \lambda_t^m)$  to avoid this issue:

$$\begin{cases} Y_t^m = N_t(\lambda_t^m) \\ \lambda_t^m = \phi_{j0} + \phi_{j1} Y_{t-1}^m + \phi_{j2} \lambda_{t-1}^m + \varepsilon_{t,m}, \quad \text{when } r_{j-1} \leq Y_{t-d}^m < r_j, \quad j = 1, \dots, s \end{cases} \quad (2.3)$$

where

$$\varepsilon_{t,m} = C_m U_t,$$

with  $U_t$  being a sequence of iid random variables uniformly distributed on  $[0, 1]$  such that  $U_t$  is independent of  $\{N_t(\cdot)\}$ , and  $C_m > 0$  is a sequence of positive numbers which tends to zero as  $m \rightarrow \infty$ .

In the following, we provide a simple sufficient condition for the geometric ergodicity of the model defined by (2.3).

**Theorem 1** . If  $\max_{1 \leq j \leq s}(\phi_{j1} + \phi_{j2}) < 1$ , then  $(Y_t^m, \lambda_t^m)$  defined by (2.3) is geometrically ergodic.

To prove Theorem 1, we need three lemmas.

**Lemma 2.1.** Let  $\{\lambda_t\}$  be a Markov chain defined by (2.1) or (2.3). If there is a  $\phi_{12}$  such that  $0 < \phi_{12} < 1$ , then every point in  $[\lambda^*, \infty)$  is reachable, where  $\lambda^* = \frac{\phi_{10}}{1-\phi_{12}}$  (for the definition of reachable, see Meyn, S.P. and Tweedie, R.L. (1993), p. 135).

**Proof.** Without loss of generality, assume  $0 < \phi_{12} < 1$ . For a point  $x \in [\lambda^*, \infty)$ , if  $\lambda_d > x$ , we start from  $\lambda_n = \sum_{j=1}^s(\phi_{j0} + \phi_{j2}\lambda_{n-1} + \phi_{j1}Y_{n-1})I(r_{j-1} \leq Y_{n-d} < r_j), n = d+1, d+2, \dots$ . Consider the path  $Y_k = N \in [r_{j-1}, r_j), k = 1, \dots, d$  and  $Y_k = 0, d+1 \leq k \leq n$ .

We can prove that  $\lambda_t$  could be arbitrarily close to  $x$ . For  $n > d$ ,

$$\begin{aligned}
 \lambda_{d+1} &= \phi_{j0} + \phi_{j2}\lambda_d + \phi_{j1}Y_d \\
 &= \phi_{j0} + \phi_{j2}\lambda_d + \phi_{j1}N \\
 \lambda_{d+2} &= \phi_{j0} + \phi_{j2}\lambda_{d+1} + \phi_{j1}Y_{d+1} \\
 &= \phi_{j0} + \phi_{j2}[\phi_{j0} + \phi_{j2}\lambda_d + \phi_{j1}N] \\
 &= \phi_{j0} + \phi_{j0}\phi_{j2} + \phi_{j2}^2\lambda_d + \phi_{j1}\phi_{j2}N \\
 \lambda_{d+3} &= \phi_{j0} + \phi_{j0}\phi_{j2} + \phi_{j0}\phi_{j2}^2 + \phi_{j2}^3\lambda_d + \phi_{j1}\phi_{j2}^2N \\
 &\dots\dots \\
 \lambda_{2d} &= \phi_{j0} + \phi_{j0}\phi_{j2} + \phi_{j0}\phi_{j2}^2 + \phi_{j0}\phi_{j2}^3 + \dots + \phi_{j0}\phi_{j2}^{d-1} + \phi_{j2}^d\lambda_d + \phi_{j1}\phi_{j2}^{d-1}N \\
 \lambda_{2d+1} &= \phi_{10} + \phi_{12}\lambda_{2d} \\
 &\dots\dots
 \end{aligned}$$

$$\begin{aligned}
 \lambda_{2d+k} &= \phi_{10} + \phi_{10}\phi_{12} + \phi_{10}\phi_{12}^2 + \cdots + \phi_{10}\phi_{12}^{k-1} + \phi_{12}^k\lambda_{2d} \\
 &= \phi_{10} \left[ \frac{1 - \phi_{12}^k}{1 - \phi_{12}} \right] + \phi_{12}^k\lambda_{2d} \\
 &= \phi_{10} \left[ \frac{1 - \phi_{12}^k}{1 - \phi_{12}} \right] + \phi_{12}^k\phi_{j0} \left[ \frac{1 - \phi_{j2}^d}{1 - \phi_{j2}} \right] + \phi_{12}^k\phi_{j2}^d\lambda_d + \phi_{j1}\phi_{12}^k\phi_{j2}^{d-1}N.
 \end{aligned}$$

Note that  $\lambda_{2d+k}$  is increasing with  $N$ . Let  $N =: N_n$  be the least integer such that  $\lambda_{2d+k}(N) > x$ ,  $\lambda_{2d+k}(N-1) \leq x$ . Then

$$\lambda_{2d+k}(N) - \lambda_{2d+k}(N-1) = \phi_{12}^k\phi_{j2}^{d-1}\phi_{j1}$$

and for any  $\delta > 0$ , because  $0 < \phi_{12} < 1$ , we can select a  $k$  such that  $\phi_{12}^k\phi_{j2}^{d-1}\phi_{j1} < \delta$ . Therefore,  $x \in \left[ \lambda_{2d+k}(N-1), \lambda_{2d+k}(N) \right)$ , where the distance between the minimum and maximum of the interval is less than  $\delta$ . Because  $\delta$  is arbitrary,  $x$  can be approximated arbitrarily closely, and so  $x$  is reachable if  $x \in [\lambda^*, \infty)$ . (Note that the reasoning is exactly the same for the model defined in (2.3)).

**Lemma 2.2.** Let  $\{\lambda_t^m\}$  be a Markov chain defined by (2.3). Under the same conditions as Lemma 2.1, for any constant  $A > \lambda^*$ , the interval  $[\lambda^*, A]$  is a small set (for the definition of a small set, see Meyn, S.P. and Tweedie, R.L. (1993)).

**Proof.** This lemma can be proved by the same technique as that for Lemma 2.1 with a slight modification. Without loss of generality, we assume  $d = 1$ . Let  $C = [\lambda^*, A]$  for a finite  $A > \lambda^*$ . Because  $0 < \phi_{12} < 1$ , for given  $\epsilon > 0$ , there exists a positive integer  $n$  such that  $\phi_{12}^{n-1}|\phi_{j2}\lambda_1^m - \lambda^*| < \epsilon$ . Consider a path  $Y_1^m = N, Y_2^m = Y_3^m = \dots Y_n^m = 0, \lambda_1^m = \lambda$ .

Then for  $n \geq 1$

$$\begin{aligned}
 \lambda_2^m &= \phi_{j0} + \phi_{j2}\lambda_1^m + \phi_{j1}Y_1^m + \varepsilon_{2,m} \\
 &= \phi_{j0} + \phi_{j2}\lambda_1^m + \phi_{j1}N + \varepsilon_{2,m}
 \end{aligned}$$

$$\begin{aligned}
\lambda_3^m &= \phi_{10} + \phi_{12}\lambda_2^m + \varepsilon_{3,m} \\
&= \phi_{10} + \phi_{j_0}\phi_{12} + \phi_{12}\phi_{j_2}\lambda_1^m + \phi_{j_1}\phi_{12}N + \phi_{12}\varepsilon_{2,m} + \varepsilon_{3,m} \\
\lambda_4^m &= \phi_{10} + \phi_{10}\phi_{12} + \phi_{j_0}\phi_{12}^2 + \phi_{12}^2\phi_{j_2}\lambda_1^m + \phi_{j_1}\phi_{12}^2N + \phi_{12}^2\varepsilon_{2,m} + \phi_{12}\varepsilon_{3,m} + \varepsilon_{4,m} \\
&\dots\dots\dots \\
\lambda_n^m &= \phi_{10} + \phi_{10}\phi_{12} + \phi_{10}\phi_{12}^2 + \dots + \phi_{10}\phi_{12}^{n-3} + \phi_{j_0}\phi_{12}^{n-2} + \phi_{12}^{n-2}\phi_{j_2}\lambda_1^m \\
&+ \phi_{j_1}\phi_{12}^{n-2}N + \phi_{12}^{n-2}\varepsilon_{2,m} + \phi_{12}^{n-3}\varepsilon_{3,m} + \dots + \varepsilon_{n,m} \\
\lambda_{n+1}^m &= \phi_{10} + \phi_{10}\phi_{12} + \phi_{10}\phi_{12}^2 + \dots + \phi_{10}\phi_{12}^{n-2} + \phi_{j_0}\phi_{12}^{n-1} + \phi_{12}^{n-1}\phi_{j_2}\lambda_1^m \\
&+ \phi_{j_1}\phi_{12}^{n-1}N + \phi_{12}^{n-1}\varepsilon_{2,m} + \phi_{12}^{n-2}\varepsilon_{3,m} + \dots + \varepsilon_{n+1,m} \\
&= \phi_{10} \left[ \frac{1 - \phi_{12}^{n-1}}{1 - \phi_{12}} \right] + \phi_{j_0}\phi_{12}^{n-1} + \phi_{j_2}\phi_{12}^{n-1}\lambda_1^m + \phi_{j_1}\phi_{12}^{n-1}N \\
&+ \sum_{i=0}^{n-1} (\phi_{12})^i \varepsilon_{n+1-i,m} \\
&= \lambda^*(1 - \phi_{12}^{n-1}) + \phi_{j_2}\phi_{12}^{n-1}\lambda_1^m + \phi_{j_0}\phi_{12}^{n-1} + \phi_{j_1}\phi_{12}^{n-1}N + \sum_{i=0}^{n-1} (\phi_{12})^i \varepsilon_{n+1-i,m} \\
&= \phi_{12}^{n-1}(\phi_{j_2}\lambda_1^m - \lambda^*) + \lambda^* + \phi_{j_0}\phi_{12}^{n-1} + \phi_{j_1}\phi_{12}^{n-1}N + \sum_{i=0}^{n-1} (\phi_{12})^i \varepsilon_{n+1-i,m}.
\end{aligned}$$

Consider an open interval  $(a_1, a_2)$  with  $a_1 > \lambda^*$ , and let  $N$  be the least integer such that

$$\psi_{n+1}(N) \equiv \lambda^* + \phi_{j_0}\phi_{12}^{n-1} + \phi_{j_1}\phi_{12}^{n-1}N > a_2,$$

without loss of generality, assume that  $N > 2$ , and for any given  $\delta > 0$ , choose  $n$  large enough such that  $\phi_{j_1}\phi_{12}^{n-1} < \delta$ . Therefore,

$$a_2 - \delta < \psi_{n+1}(N - 1) < a_2.$$

Note that

$$\lambda_{n+1}^m(N - 1) = \psi_{n+1}(N - 1) + \phi_{12}^{n-1}(\phi_{j_2}\lambda_1^m - \lambda^*) + \sum_{i=0}^{n-1} (\phi_{12})^i \varepsilon_{n+1-i,m}$$

and  $P\left(\sum_{i=0}^{n-1} (\phi_{12})^i \varepsilon_{n+1-i,m} < \epsilon - \phi_{12}^{n-1}(\phi_{j_2}\lambda_1^m - \lambda^*)\right) > 0$ . Choose  $\epsilon$  small enough

such that

$$a_2 - \delta < \psi_{n+1}(N-1) < \lambda_{n+1}^m(N-1) < a_2,$$

if

$$0 < \sum_{i=0}^{n-1} (\phi_{12})^i \varepsilon_{n+1-i,m} + \phi_{12}^{n-1} (\phi_{j2} \lambda - \lambda^*) < \epsilon.$$

So these choices of  $n, N, \lambda_{n+1}^m(N-1) \in (a_1, a_2)$ , for all  $\lambda \in C$ , and therefore,

$$\inf_{\lambda \in C} P^{n+1} \left( \lambda, (a_1, a_2) \right) \geq P \left( Y_1^m = N, Y_2^m = Y_3^m = \dots Y_n^m = 0 \right) > 0.$$

This means that the open interval  $(a_1, a_2)$  is uniformly reachable from all  $\lambda \in [\lambda^*, A]$ .

**Lemma 2.3.** Let  $\{\lambda_t^m\}$  be a Markov chain defined by (2.3). Under the same conditions as Lemma 2.1,  $\{\lambda_t^m\}$  is aperiodic.

**Proof.** Suppose we have a small set  $C = (\lambda^*, A]$ . Note that  $\psi(C) > 0$ , let  $\lambda_{t-1}^m = \lambda \in C$ . Then  $\lambda_t^m = \phi_{j0} + \phi_{j1} Y_{t-1}^m + \phi_{j2} \lambda_{t-1}^m + \varepsilon_{t,m}$ , when  $r_{j-1} < Y_{t-1}^m \leq r_j$ . For the path where  $Y_{t-1}^m \in [0, r_s)$  is zero, we have  $\lambda_t^m = \phi_{10} + \phi_{12} \lambda + \varepsilon_{t,m} = \lambda^*(1 - \phi_{12}) + \phi_{12} \lambda + \varepsilon_{t,m} = \lambda^* + \phi_{12}(\lambda - \lambda^*) + \varepsilon_{t,m} \geq \lambda^*$  because  $\phi_{12} \geq 0$  and  $\varepsilon_{t,m} \geq 0$ . On the other hand,  $\lambda_t^m - \lambda = \lambda^*(1 - \phi_{12}) - \lambda(1 - \phi_{12}) + \varepsilon_{t,m} = -(1 - \phi_{12})(\lambda - \lambda^*) + \varepsilon_{t,m} \leq 0$  because  $\phi_{12} < 1$  and  $\lambda > \lambda^*$ . This means that, if  $0 < \phi_{12} < 1$ , then  $\lambda \in C, Y_{t-1}^m = 0$  and  $\varepsilon_{t,m}$ , small enough, imply  $\lambda_t^m \in C$ . Then  $P(\lambda, C) \geq P(Y_{t-1}^m = 0 \text{ and } \varepsilon_{t,m} \text{ is small enough} | \lambda_{t-1}^m = \lambda) = P(N_t(\lambda) = 0) > 0$ . Similarly,  $P^2(\lambda, C) \geq P(Y_t^m = Y_{t-1}^m = 0 \text{ and } \varepsilon_{t,m}, \varepsilon_{t+1,m} \text{ are small enough} | \lambda_{t-1}^m = \lambda) > 0$ , and for any positive integer  $k \geq 2$ ,  $P^k(\lambda, C) \geq P(Y_{t-2+k}^m = Y_{t-3+k}^m = 0, \text{ and } \varepsilon_{t,m}, \varepsilon_{t+1,m}, \dots, \varepsilon_{t-1+k,m} \text{ are small enough} | \lambda_{t-1}^m = \lambda) > 0$ . Therefore,  $\{\lambda_t^m\}$  is aperiodic.

**Proof of Theorem 1.** Denote

$$c_0 = \max_{1 \leq j \leq s} \phi_{j0}, \quad c_1 = \max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}),$$

and

$$\varepsilon_{Y_{t-d}^m} = \sum_{j=1}^s I(r_{j-1} \leq Y_{t-d}^m < r_j) \varepsilon_{t,m}.$$

Note that

$$\begin{aligned} \lambda_t^m &= \sum_{j=1}^s \{\phi_{j0} + \phi_{j1} Y_{t-1}^m + \phi_{j2} \lambda_{t-1}^m\} I(r_{j-1} \leq Y_{t-d}^m < r_j) + \varepsilon_{Y_{t-d}^m} \\ &= \sum_{j=1}^s \phi_{j0} I(r_{j-1} \leq Y_{t-d}^m < r_j) + Y_{t-1}^m \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \\ &\quad + \lambda_{t-1}^m \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) + \varepsilon_{Y_{t-d}^m} \\ &\leq c_0 + Y_{t-1}^m \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \\ &\quad + \lambda_{t-1}^m \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) + \varepsilon_{Y_{t-d}^m} =: \Lambda_t^m \end{aligned}$$

and

$$\begin{aligned} E(\varepsilon_{Y_{t-d}^m}) &= \sum_{j=1}^s E \left[ \varepsilon_{t,m} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right] \\ &\leq \sum_{j=1}^s E(\varepsilon_{t,m}) \leq C_m \rightarrow 0 \end{aligned}$$

as  $m \rightarrow \infty$ .

Take the test function as  $V(x) = 1 + x^k$ . We will show that  $\{X_t^m\}$  is  $V_x$ -geometrically ergodic Markovian Chain by showing that  $\{X_t^m\}$  is irreducible and



aperiodic, and there exist constants  $0 < \kappa_1 < 1, 0 < \kappa_2 < \infty$  such that

$$E[V(\lambda_t^m) | \lambda_{t-1}^m = \lambda] \leq (1 - \kappa_1)V(\lambda) + \kappa_2 I(\lambda \in C) \quad (2.4)$$

for a small set  $C$ . Note that

$$\begin{aligned} & E[V(\lambda_t^m) | \lambda_{t-1}^m = \lambda] \\ = & E[1 + (\lambda_t^m)^k | \lambda_{t-1}^m = \lambda] \\ \leq & 1 + E[(\Lambda_t^m)^k | \lambda_{t-1}^m = \lambda] \\ = & 1 + \sum_{i=0}^k \binom{k}{i} \left( \lambda \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right)^i \\ & \cdot E \left[ \left( N_{t-1}(\lambda_{t-1}^m) \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) + c_0 + \varepsilon_{Y_{t-1}^m} \right)^{k-i} \middle| \lambda_{t-1}^m = \lambda \right]. \end{aligned}$$

We then use a simple property of a Poisson random variable: The  $k$ -th moment of a Poisson random variable with parameter  $\lambda$  is a polynomial of  $\lambda$  with degree  $k$ . This property is an application of the binomial theorem plus the fact that  $E(\varepsilon_{Y_{t-1}^m}) = 0$  as  $m \rightarrow \infty$ .

Hence,

$$\begin{aligned} E[V(\lambda_t^m) | \lambda_{t-1}^m = \lambda] & \leq 1 + \sum_{i=0}^k \binom{k}{i} \left( \lambda \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right)^i \\ & \cdot \left( \lambda \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right)^{k-i} \\ & \cdot \left( 1 + \frac{c}{\lambda \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j)} \right)^{k-i}. \end{aligned}$$

if  $c$  is a large enough constant. Therefore,

$$\begin{aligned}
 E[V(\lambda_t^m) | \lambda_{t-1}^m = \lambda] &\leq 1 + \sum_{i=0}^k \binom{k}{i} \left( \lambda \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right)^i \\
 &\quad \cdot \left( \lambda \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right)^{k-i} \\
 &\quad \cdot \left[ \sum_{m=1}^{k-i} \binom{k-i}{m} \left( \frac{c}{\lambda \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j)} \right)^m + 1 \right] \\
 &= 1 + \sum_{i=0}^k \binom{k}{i} \left( \lambda \sum_{m=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right)^i \\
 &\quad \cdot \left( \lambda \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right)^{k-i} \\
 &\quad + \sum_{i=0}^k \binom{k}{i} \left( \lambda \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right)^i \\
 &\quad \cdot \left( \lambda \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right)^{k-i} \\
 &\quad \cdot \sum_{m=1}^{k-i} \binom{k-i}{m} \left( \frac{c}{\lambda \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j)} \right)^m.
 \end{aligned}$$

The last term in the above equation is

$$\sum_{i=0}^k \lambda^k \sum_{m=1}^{k-i} F_m \lambda^{-m}$$

for some constants  $F_m$ , depending on  $\sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \leq \max_{1 \leq j \leq s} \phi_{j1}$ ,  $\sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) < \max_{1 \leq j \leq s} \phi_{j2}$ , and  $c$ . Note that

$$\sum_{i=0}^k \sum_{m=1}^{k-i} \lambda^{k-m} = \sum_{m=0}^{k-1} (m+1) \lambda^m = \sum_{m=0}^{k-1} c_m \lambda^m.$$

Hence

$$\begin{aligned}
 E[V(\lambda_t^m) | \lambda_{t-1}^m = \lambda] &\leq 1 + \sum_{i=0}^k \binom{k}{i} \left( \lambda \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right)^i \\
 &\quad \cdot \left( \lambda \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right)^{k-i} + \sum_{m=0}^{k-1} c_m \lambda^m \\
 &= 1 + \left( \sum_{j=1}^s (\phi_{j1} + \phi_{j2}) I(r_{j-1} \leq Y_{t-d}^m < r_j) \right)^k \lambda^k + \sum_{m=0}^{k-1} c_m \lambda^m \\
 &= 1 + \left( \max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}) \right)^k \lambda^k + \sum_{m=0}^{k-1} c_m \lambda^m
 \end{aligned}$$

for some constants  $c_m$  depending on  $F_m$ .

Take  $C = [\lambda^*, K]$  as the small set, where  $\lambda^* = \frac{\phi_{j0}}{1-\phi_{j2}}$  is a fixed point of the skeleton  $\lambda_t = \phi_{j0} + \phi_{j2}\lambda_{t-1}$  of (2.2) and (2.3).

Note that

$$\begin{aligned}
 1 + \left( \max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}) \right)^k \lambda^k &= \left\{ 1 - \frac{\lambda^k \left[ 1 - \left( \max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}) \right)^k \right]}{1 + \lambda^k} \right\} \\
 &\quad \cdot (1 + \lambda^k) [I(\lambda \in C) + I(\lambda \in C^c)].
 \end{aligned}$$

If  $\lambda \geq 1$ , then  $\frac{\lambda^k}{1+\lambda^k} \geq \frac{1}{2}$ . Since  $\max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}) < 1$ ,

$$\begin{aligned}
 1 + \left( \max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}) \right)^k \lambda^k &\leq \left\{ 1 - \frac{1}{2} \left[ 1 - \left( \max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}) \right)^k \right] \right. \\
 &\quad \left. \cdot (1 + \lambda^k) \right\} [I(\lambda \in C) + I(\lambda \in C^c)].
 \end{aligned}$$

Defining  $\kappa_1 = \frac{1}{2} \left[ 1 - \left( \max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}) \right)^k \right]$ , we will show that (2.4) holds.

For  $\lambda \in C^c$ , we obtain

$$\sup_{\lambda \in C^c} \left\{ 1 - \frac{\lambda^k \left[ 1 - \left( \max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}) \right)^k \right]}{1 + \lambda^k} \right\} \rightarrow \left( \max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}) \right)^k$$

as  $K$  increases to infinity. Take  $K \geq 1$  large enough such that

$$\sup_{\lambda \in C^c} \frac{\sum_{m=1}^{k-1} c_m \lambda^m}{1 + \lambda^k} < \delta,$$

where  $0 < \delta < \frac{1}{2} \left[ 1 - \left( \max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}) \right)^k \right]$ . But on  $C$ ,  $1 + \left( \max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}) \right)^k \lambda^k + \sum_{m=1}^{k-1} c_m \lambda^m$  is bounded and  $(1 - \kappa_1)V(\lambda)$  is always positive. This shows that (2.4) holds for all  $\lambda \in C$ . This completes the proof of Theorem 1. By Lemmas 2.1, 2.2, 2.3 and Theorem 1 (drift-criteria for the geometric ergodicity of a Markov chain (Isaacson (1979))), we conclude that the chain  $\{\lambda_t^m, t \geq 0\}$  is geometrically ergodic.

The theorem below shows that  $\{(Y_t^m, \lambda_t^m)\}$  approximates to  $\{(Y_t, \lambda_t)\}$  when  $m \rightarrow \infty$ .

**Theorem 2.** Assume that  $\{(Y_t, \lambda_t)\}$  and  $\{(Y_t^m, \lambda_t^m)\}$  are defined by (2.2) and (2.3) respectively. If the condition  $\max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}) < 1$  holds, then as  $c_m \rightarrow 0$  as  $m \rightarrow \infty$ , if  $\lambda_0^m = \lambda_0$ , the following results hold true:

$$(i) |E(\lambda_t^m - \lambda_t)| = |E(Y_t^m - Y_t)| \leq \alpha_{1,m};$$

$$(ii) E(\lambda_t^m - \lambda_t)^2 \leq \alpha_{2,m};$$

$$(iii) E(Y_t^m - Y_t)^2 \leq \alpha_{3,m},$$

where  $\alpha_{i,m} \rightarrow 0$  as  $m \rightarrow \infty$  for  $i = 1, 2, 3$ . The sequences  $\alpha_{i,m}$  can be chosen to be independent of  $t$ .

**Proof of Theorem 2.** Denote  $c_1 = \max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}) < 1$ . We first prove

(i). By subtracting model defined (2.2) from model defined (2.3), we obtain that

$$\begin{aligned}
 \lambda_t^m - \lambda_t &= \sum_{j=1}^s \phi_{j0} \left[ \left( I(r_{j-1} \leq Y_{t-d}^m < r_j) - I(r_{j-1} \leq Y_{t-d} < r_j) \right) \right] \\
 &\quad + \sum_{j=1}^s \phi_{j1} \left[ (Y_{t-1}^m - Y_{t-1}) I(r_{j-1} \leq Y_{t-d}^m < r_j) \right. \\
 &\quad \left. + Y_{t-1} \left( I(r_{j-1} \leq Y_{t-d}^m < r_j) - I(r_{j-1} \leq Y_{t-d} < r_j) \right) \right] \\
 &\quad + \sum_{j=1}^s \phi_{j2} \left[ (\lambda_{t-1}^m - \lambda_{t-1}) I(r_{j-1} \leq Y_{t-d}^m < r_j) \right. \\
 &\quad \left. + \lambda_{t-1} \left( I(r_{j-1} \leq Y_{t-d}^m < r_j) - I(r_{j-1} \leq Y_{t-d} < r_j) \right) \right] + \varepsilon(Y_{t-d}^m) \\
 &= (Y_{t-1}^m - Y_{t-1}) \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \\
 &\quad + (\lambda_{t-1}^m - \lambda_{t-1}) \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) + R_t^m,
 \end{aligned} \tag{2.5}$$

where

$$\begin{aligned}
 R_t^m &= \sum_{j=1}^s \phi_{j0} \left[ \left( I(r_{j-1} \leq Y_{t-d}^m < r_j) - I(r_{j-1} \leq Y_{t-d} < r_j) \right) \right] \\
 &\quad + \left( \sum_{j=1}^s \phi_{j1} Y_{t-1} + \sum_{j=1}^s \phi_{j2} \lambda_{t-1} \right) \left[ I(r_{j-1} \leq Y_{t-d}^m < r_j) - I(r_{j-1} \leq Y_{t-d} < r_j) \right] \\
 &\quad + \varepsilon(Y_{t-d}^m).
 \end{aligned}$$

It can be easily seen that

$$E(R_t^m) \leq C_m + D_m,$$

where  $D_m \rightarrow 0$  as  $m \rightarrow \infty$ .

Hence

$$\begin{aligned}
 E(\lambda_t^m - \lambda_t) &= E\left[(Y_{t-1}^m - Y_{t-1}) \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right] \\
 &\quad + E\left[(\lambda_{t-1}^m - \lambda_{t-1}) \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right] \\
 &\quad + E(R_t^m).
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 \left|E(\lambda_t^m - \lambda_t)\right| &\leq \left|E\left[(Y_{t-1}^m - Y_{t-1}) \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right]\right| \\
 &\quad + \left|E\left[(\lambda_{t-1}^m - \lambda_{t-1}) \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right]\right| + \left|E(R_t^m)\right| \\
 &\leq \left|E\left[E\left((Y_{t-1}^m - Y_{t-1}) \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right) \middle| \mathcal{F}_{t-2}\right]\right| \\
 &\quad + \left|E\left[(\lambda_{t-1}^m - \lambda_{t-1}) \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right]\right| + \left|E(R_t^m)\right| \\
 &\leq \left|E\left[(\lambda_{t-1}^m - \lambda_{t-1}) \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right]\right| \\
 &\quad + \left|E\left[(\lambda_{t-1}^m - \lambda_{t-1}) \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right]\right| + \left|E(R_t^m)\right| \\
 &\leq \left|E\left[(\lambda_{t-1}^m - \lambda_{t-1}) \sum_{j=1}^s (\phi_{j1} + \phi_{j2}) I(r_{j-1} \leq Y_{t-d}^m < r_j)\right]\right| \\
 &\quad + \left|E(R_t^m)\right| \\
 &\leq \max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}) \left|E(\lambda_{t-1}^m - \lambda_{t-1})\right| + \left|E(R_t^m)\right|.
 \end{aligned}$$

By using simple recursion, we get

$$\left|E(\lambda_t^m - \lambda_t)\right| \leq \sum_{i=0}^{t-1} \left(\max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2})\right)^i \left|E(R_t^m)\right|.$$

Since  $\max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}) < 1$ , and  $E(R_t^m) \rightarrow 0$  as  $m \rightarrow \infty$ , we have

$$\left| E(\lambda_t^m - \lambda_t) \right| \leq \frac{C_m}{1 - c_1} =: \alpha_{1,m}.$$

By using equation (2.5) again, the second statement becomes

$$\begin{aligned} E(\lambda_t^m - \lambda_t)^2 &= E \left[ (Y_{t-1}^m - Y_{t-1}) \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right. \\ &\quad \left. + (\lambda_{t-1}^m - \lambda_{t-1}) \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) + R_t^m \right]^2 \\ &= E \left[ (Y_{t-1}^m - Y_{t-1}) \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right. \\ &\quad \left. + (\lambda_{t-1}^m - \lambda_{t-1}) \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right]^2 + E(R_t^m)^2 \\ &\quad + 2E \left\{ \left[ (Y_{t-1}^m - Y_{t-1}) \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right. \right. \\ &\quad \left. \left. + (\lambda_{t-1}^m - \lambda_{t-1}) \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right] R_t^m \right\} \\ &= E \left[ (Y_{t-1}^m - Y_{t-1})^2 \left( \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right)^2 \right. \\ &\quad \left. + (\lambda_{t-1}^m - \lambda_{t-1})^2 \left( \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right)^2 \right. \\ &\quad \left. + 2(Y_{t-1}^m - Y_{t-1}) \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right. \\ &\quad \left. \cdot (\lambda_{t-1}^m - \lambda_{t-1}) \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right] \\ &\quad + E(R_t^m)^2 + 2E \left\{ \left[ (Y_{t-1}^m - Y_{t-1}) \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right. \right. \\ &\quad \left. \left. \cdot (\lambda_{t-1}^m - \lambda_{t-1}) \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right] R_t^m \right\} \end{aligned}$$

However, for  $\lambda_t^m \geq \lambda_t$  (if  $\lambda_t > \lambda_t^m$ , the method is the same), using the properties

of the Poisson process, we get that

$$\begin{aligned}
 & E \left[ E \left( (Y_{t-1}^m - Y_{t-1}) \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right. \right. \\
 & \left. \left. \cdot (\lambda_{t-1}^m - \lambda_{t-1}) \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \middle| \mathcal{F}_{t-2} \right) \right] \\
 &= E \left[ \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right. \\
 & \left. \cdot (\lambda_{t-1}^m - \lambda_{t-1}) \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) E(\lambda_{t-1}^m - \lambda_{t-1} | \mathcal{F}_{t-2}) \right] \\
 &= E \left[ (\lambda_{t-1}^m - \lambda_{t-1})^2 \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \cdot \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right]
 \end{aligned}$$

and

$$\begin{aligned}
 & E \left\{ E \left[ (Y_{t-1}^m - Y_{t-1})^2 \left( \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right)^2 \middle| \mathcal{F}_{t-2} \right] \right\} \\
 & \leq E(\lambda_{t-1}^m - \lambda_{t-1})^2 \left( \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right)^2 \\
 & \quad + 2|E(\lambda_{t-1}^m - \lambda_{t-1})| \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j)
 \end{aligned} \tag{2.6}$$

where  $\mathcal{F}_{t-2} = \sigma(\lambda_{t-2}, Y_{t-2}, \lambda_{t-2}^m, Y_{t-2}^m)$ , we used the fact that  $Y_t^m = N_t(\lambda_t^m)$  is the event numbers for  $N_t(\cdot)$  in the stochastic time interval  $[0, \lambda_t^m]$  and similarly for  $Y_t = N_t(\lambda_t)$ . In other word,  $\{Y_t^m\}$  and  $\{Y_t\}$  are generated by the same sequences of independent Poisson processes  $\{N_t(\cdot)\}$  of unit intensity.

Finally, with  $K$  positive constant,

$$\begin{aligned}
 & E(R_t^m)^2 + 2E \left[ \left( (Y_{t-1}^m - Y_{t-1}) \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right. \right. \\
 & \left. \left. \times (\lambda_{t-1}^m - \lambda_{t-1}) \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \right) R_t^m \right] \leq KC_m^2
 \end{aligned}$$



Hence

$$\begin{aligned}
 E(\lambda_t^m - \lambda_t)^2 &\leq E\left[(Y_{t-1}^m - Y_{t-1})^2 \left(\sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right)^2\right. \\
 &\quad \left.+ (\lambda_{t-1}^m - \lambda_{t-1})^2 \left(\sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right)^2\right. \\
 &\quad \left.+ 2(\lambda_{t-1}^m - \lambda_{t-1})^2 \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right. \\
 &\quad \left. \times \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right] + KC_m^2 \\
 &\leq E\left\{(\lambda_{t-1}^m - \lambda_{t-1})^2 \left[\left(\sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right)^2\right. \right. \\
 &\quad \left. \left.+ 2 \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right. \right. \\
 &\quad \left. \left.+ \left(\sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right)^2\right]\right\} \\
 &\quad \left. + 2(\lambda_{t-1}^m - \lambda_{t-1}) \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right\} + KC_m^2 \\
 &\leq E\left\{\left[\sum_{j=1}^s (\phi_{j1} + \phi_{j2}) I(r_{j-1} \leq Y_{t-d}^m < r_j)\right]^2 (\lambda_{t-1}^m - \lambda_{t-1})^2\right. \\
 &\quad \left.+ 2|(\lambda_{t-1}^m - \lambda_{t-1})| \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right\} + KC_m^2 \\
 &\leq \left[\max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2})\right]^2 E(\lambda_{t-1}^m - \lambda_{t-1})^2 \\
 &\quad + 2|E(\lambda_{t-1}^m - \lambda_{t-1})| \sum_{j=1}^s \phi_{j1} I(r_{j-1} \leq Y_{t-d}^m < r_j) + KC_m^2 \\
 &\leq \left[\max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2})\right]^2 E(\lambda_{t-1}^m - \lambda_t)^2 + 2c_2 |E(\lambda_{t-1}^m - \lambda_{t-1})| \\
 &\quad + KC_m^2 \\
 &\leq \left[\max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2})\right]^2 E(\lambda_{t-1}^m - \lambda_{t-1})^2 + 2c_2 \alpha_{1,m} \\
 &\quad + KC_m^2 \\
 &\leq \alpha_{2,m}
 \end{aligned}$$

where  $c_2 = \max_{1 \leq j \leq s} \phi_{j1}$

The third statement, from equation (2.6), and the second statement of the Theorem 2, we conclude that

$$E(Y_t^m - Y_t)^2 \leq \alpha_{3,m}.$$

## 2.2 Statistical Inference

We introduce some statistical methods in time series first. The goal of this section is to estimate the parameters of models using Maximum likelihood estimation (MLE) and to study the asymptotic properties of the MLE. We discuss the main tool for linking between the estimated parameters of the unperturbed and the perturbed model. We present two different nonlinear tests to detect the threshold nonlinearity for the model: the Lagrange Multiplier test and the Likelihood Ratio test.

### 2.2.1 Parameter estimation and information criteria

The parameters to be estimated for the stochastic TAR model are the number of regimes, the threshold value, the time delay, and the intercepts and the autoregressive coefficients. The determination and selection of these parameters is required to fit an observed stationary time series. We use maximum likelihood and information criteria to estimate these parameters.

#### 2.2.1.1 Conditional maximum likelihood inference

We consider conditional maximum likelihood inference about Poisson TGARCH models.

Let  $\theta = (\phi_{j0}, \phi_{j1}, \phi_{j2}, j = 1, \dots, s)'$  denote a  $3s$ -dimensional parameter vector which has true value  $\theta_0 = (\phi_{j0}^0, \phi_{j1}^0, \phi_{j2}^0, j = 1, \dots, s)'$ . The estimate of  $\theta$  can

be found by maximizing the conditional likelihood function of the observations  $Y_1, Y_2, Y_3, \dots, Y_n$  given by

$$L(\theta) = \prod_{t=1}^n \frac{\lambda_t(\theta)^{Y_t} \exp(-\lambda_t(\theta))}{Y_t!},$$

where  $\lambda_t(\theta)$  is defined by (2.2). The log-likelihood function is equal to

$$\ell(\theta) = \sum_{t=1}^n (Y_t \log \lambda_t(\theta) - \lambda_t(\theta) - \log(Y_t!)) = \sum_{t=1}^n \ell_t(\theta). \quad (2.7)$$

Clearly, the term  $\log(Y_t!)$  can be neglected, because the derivative of this term is equal to zero. Note that the derivative of  $\ell_t(\theta)$  is

$$\frac{\partial \ell_t}{\partial \theta} = \left( \frac{Y_t}{\lambda_t(\theta)} - 1 \right) \frac{\partial \lambda_t(\theta)}{\partial \theta}.$$

The score function is then defined by

$$S_n(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = \sum_{t=1}^n \frac{\partial \ell_t(\theta)}{\partial \theta} = \sum_{t=1}^n \left( \frac{Y_t}{\lambda_t(\theta)} - 1 \right) \frac{\partial \lambda_t(\theta)}{\partial \theta} \quad (2.8)$$

where

$$\begin{aligned} \frac{\partial \lambda_t}{\partial \phi_{i0}} &= I\left(r_{i-1} \leq Y_{t-d} < r_i\right) + \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}}{\partial \phi_{i0}} I\left(r_{j-1} \leq Y_{t-d} < r_j\right) \\ \frac{\partial \lambda_t}{\partial \phi_{i1}} &= Y_{t-1} I\left(r_{i-1} \leq Y_{t-d} < r_i\right) + \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}}{\partial \phi_{i1}} I\left(r_{j-1} \leq Y_{t-d} < r_j\right) \\ \frac{\partial \lambda_t}{\partial \phi_{i2}} &= \lambda_{t-1} I\left(r_{i-1} \leq Y_{t-d} < r_i\right) + \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}}{\partial \phi_{i2}} I\left(r_{j-1} \leq Y_{t-d} < r_j\right). \end{aligned} \quad (2.9)$$

If the solution to the equation  $S_n(\theta) = 0$  exists, then we can find the conditional maximum likelihood estimate  $\hat{\theta}$  of  $\theta$ . In addition, the conditional information

matrix for  $\theta$  is equal to

$$\begin{aligned} G_n(\theta) &= \sum_{t=1}^n \text{Var} \left[ \frac{\partial \ell(\theta)}{\partial \theta} \middle| \mathcal{F}_{t-1} \right] = \sum_{t=1}^n E \left[ \left( \frac{Y_t}{\lambda_t(\theta)} - 1 \right)^2 \left( \frac{\partial \lambda_t(\theta)}{\partial \theta} \right) \left( \frac{\partial \lambda_t(\theta)}{\partial \theta} \right)' \middle| \mathcal{F}_{t-1} \right] \\ &= \sum_{t=1}^n \frac{1}{\lambda_t(\theta)} \left( \frac{\partial \lambda_t(\theta)}{\partial \theta} \right) \left( \frac{\partial \lambda_t(\theta)}{\partial \theta} \right)'. \end{aligned} \quad (2.10)$$

Because the region of ergodicity for the model defined by (2.1) is unknown, we use the approximate model defined by (2.3). Theorem 2 shows that  $\{(Y_t^m, \lambda_t^m)\}$  converges in distribution to the  $\{(Y_t, \lambda_t)\}$  model. Hence, it is informative for studying asymptotic properties of the maximum likelihood estimate using the process  $(Y_t^m, \lambda_t^m)$ . The likelihood function for model (2.3), including the observations  $U_1, U_2, U_3, \dots, U_n$  is defined as

$$L^m(\theta) = \prod_{t=1}^n \frac{\lambda_t^m(\theta)^{Y_t^m} \exp(-\lambda_t^m(\theta))}{Y_t^m!} \prod_{t=1}^n f_u(U_t),$$

where  $f_u(\cdot)$  is the density function of  $U_t$  which is an i.i.d sequence of uniformly distributed random variables on  $(0, 1)$ , and  $\lambda_t^m(\theta)$  is defined by (2.3). The log likelihood function is given, up to a constant, by

$$\ell^m(\theta) = \sum_{t=1}^n (Y_t^m \log \lambda_t^m(\theta) - \lambda_t^m(\theta) - \log(Y_t^m!)) + \sum_{t=1}^n \log f_u(U_t) = \sum_{t=1}^n \ell_t(\theta).$$

Taking the derivatives, we get the score function

$$S_n^m(\theta) = \frac{\partial \ell^m(\theta)}{\partial \theta} = \sum_{t=1}^n \frac{\partial \ell_t^m(\theta)}{\partial \theta} = \sum_{t=1}^n \left( \frac{Y_t^m}{\lambda_t^m(\theta)} - 1 \right) \frac{\partial \lambda_t^m(\theta)}{\partial \theta}, \quad (2.11)$$

where

$$\begin{aligned}
 \frac{\partial \lambda_t^m}{\partial \phi_{i0}} &= I\left(r_{i-1} \leq Y_{t-d}^m < r_i\right) + \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}^m}{\partial \phi_{i0}} I\left(r_{j-1} \leq Y_{t-d}^m < r_j\right) \\
 \frac{\partial \lambda_t^m}{\partial \phi_{i1}} &= Y_{t-1}^m I\left(r_{i-1} \leq Y_{t-d}^m < r_i\right) + \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}^m}{\partial \phi_{i1}} I\left(r_{j-1} \leq Y_{t-d}^m < r_j\right) \\
 \frac{\partial \lambda_t^m}{\partial \phi_{i2}} &= \lambda_{t-1}^m I\left(r_{i-1} \leq Y_{t-d}^m < r_i\right) + \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}^m}{\partial \phi_{i2}} I\left(r_{j-1} \leq Y_{t-d}^m < r_j\right).
 \end{aligned} \tag{2.12}$$

We can find the conditional maximum estimate  $\hat{\theta}^m$  of  $\theta$  by solving the equation  $S_n^m(\theta) = \frac{\partial \ell^m(\theta)}{\partial \theta} = 0$ . In addition, the conditional information matrix for  $\theta^m$  is equal to

$$G_n^m(\theta) = \sum_{t=1}^n \frac{1}{\lambda_t^m(\theta)} \left( \frac{\partial \lambda_t^m(\theta)}{\partial \theta} \right) \left( \frac{\partial \lambda_t^m(\theta)}{\partial \theta} \right)'. \tag{2.13}$$

**Theorem 3.** Assume that  $\{(Y_t, \lambda_t)\}$  and  $\{(Y_t^m, \lambda_t^m)\}$  are defined by (2.2) and (2.3) respectively. If  $0 < \max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}) < 1$  holds, and at  $\theta = \theta_0$ , where  $\theta_0$  is the true value of  $\theta$ , then the following statements hold:

- (1).  $\left| E \left( \frac{\partial \lambda_t^m}{\partial \phi_{i0}} - \frac{\partial \lambda_t}{\partial \phi_{i0}} \right) \right| \leq \beta_{1,m},$
- (2).  $\left| E \left( \frac{\partial \lambda_t^m}{\partial \phi_{i2}} - \frac{\partial \lambda_t}{\partial \phi_{i2}} \right) \right| \leq \beta_{2,m},$
- (3).  $\left| E \left( \frac{\partial \lambda_t^m}{\partial \phi_{i1}} - \frac{\partial \lambda_t}{\partial \phi_{i1}} \right) \right| \leq \beta_{3,m},$
- (4).  $E \left( \frac{\partial \lambda_t^m}{\partial \phi_{i2}} - \frac{\partial \lambda_t}{\partial \phi_{i2}} \right)^2 \leq \beta_{4,m}.$

The sequences  $\beta_{k,m}$  can be chosen to be independent of  $t$  and  $\beta_{k,m} \rightarrow 0$  as  $m \rightarrow \infty$  for  $i = 1, 2, \dots, s$ , and  $k = 1, 2, 3, 4$ .

### Proof of Theorem 3

Note that  $\theta_0 = (\phi_{i0}^0, \phi_{i1}^0, \phi_{i2}^0, i = 1, \dots, s)'$ . The differences of derivatives with

respect to  $\phi_{im}, m = 0, 1, 2$ , can be derived by subtracting equation (2.9) from equation (2.12) as follows:

$$\begin{aligned}
 \frac{\partial \lambda_t^m}{\partial \phi_{i0}} - \frac{\partial \lambda_t}{\partial \phi_{i0}} &= [I(r_{i-1} \leq Y_{t-d}^m < r_i) - I(r_{i-1} \leq Y_{t-d} < r_i)] \\
 &\quad + \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}^m}{\partial \phi_{i0}} I(r_{j-1} \leq Y_{t-d}^m < r_j) \\
 &\quad - \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}}{\partial \phi_{i0}} I(r_{j-1} \leq Y_{t-d} < r_j) \\
 &= [I(r_{i-1} \leq Y_{t-d}^m < r_i) - I(r_{i-1} \leq Y_{t-d} < r_i)] \\
 &\quad + \sum_{j=1}^s \phi_{j2} \left[ \left( \frac{\partial \lambda_{t-1}^m}{\partial \phi_{i0}} - \frac{\partial \lambda_{t-1}}{\partial \phi_{i0}} \right) I(r_{j-1} \leq Y_{t-d}^m < r_j) \right] \\
 &\quad + \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}}{\partial \phi_{i0}} \left[ I(r_{j-1} \leq Y_{t-d}^m < r_j) - I(r_{j-1} \leq Y_{t-d} < r_j) \right].
 \end{aligned}$$

Hence,

$$\begin{aligned}
 E \left( \frac{\partial \lambda_t^m}{\partial \phi_{i0}} - \frac{\partial \lambda_t}{\partial \phi_{i0}} \right) &= E \left( I(r_{i-1} \leq Y_{t-d}^m < r_i) - I(r_{i-1} \leq Y_{t-d} < r_i) \right) \\
 &\quad + E \left[ \sum_{j=1}^s \phi_{j2} (I(r_{j-1} \leq Y_{t-d}^m < r_j) \left( \frac{\partial \lambda_{t-1}^m}{\partial \phi_{i0}} - \frac{\partial \lambda_{t-1}}{\partial \phi_{i0}} \right)) \right] \\
 &\quad + E \left[ \sum_{j=1}^s \phi_{j2} \left( \frac{\partial \lambda_{t-1}}{\partial \phi_{i0}} \right) \right. \\
 &\quad \left. \times \left( I(r_{j-1} \leq Y_{t-d}^m < r_j) - I(r_{j-1} \leq Y_{t-d} < r_j) \right) \right] \\
 \left| E \left( \frac{\partial \lambda_t^m}{\partial \phi_{i0}} - \frac{\partial \lambda_t}{\partial \phi_{i0}} \right) \right| &\leq \left| E \left( I(r_{i-1} \leq Y_{t-d}^m < r_i) - I(r_{i-1} \leq Y_{t-d} < r_i) \right) \right| \\
 &\quad + \left| E \left[ \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \left( \frac{\partial \lambda_{t-1}^m}{\partial \phi_{i0}} - \frac{\partial \lambda_{t-1}}{\partial \phi_{i0}} \right) \right] \right| \\
 &\quad + \left| E \left[ \sum_{j=1}^s \phi_{j2} \left( \frac{\partial \lambda_{t-1}}{\partial \phi_{i0}} \right) \right. \right. \\
 &\quad \left. \left. \times \left( I(r_{j-1} \leq Y_{t-d}^m < r_j) - I(r_{j-1} \leq Y_{t-d} < r_j) \right) \right] \right| \\
 &\leq \max_{1 \leq j \leq s} \phi_{j2} \left| E \left( \frac{\partial \lambda_{t-1}^m}{\partial \phi_{i0}} - \frac{\partial \lambda_{t-1}}{\partial \phi_{i0}} \right) \right| + \left| E(\hat{R}_{1,t}^m) \right| \\
 &\leq \beta_{1,m},
 \end{aligned}$$

where  $\beta_{1,m} \rightarrow 0$  as  $m \rightarrow \infty$ , and

$$\begin{aligned} \hat{R}_{1,t}^m &= \left( I(r_{i-1} \leq Y_{t-d}^m < r_i) - I(r_{i-1} \leq Y_{t-d} < r_i) \right) \\ &\quad + \left[ \sum_{j=1}^s \phi_{j2} \left( \frac{\partial \lambda_{t-1}}{\partial \phi_{i0}} \right) \left( I(r_{j-1} \leq Y_{t-d}^m < r_j) - I(r_{j-1} \leq Y_{t-d} < r_j) \right) \right] \end{aligned}$$

and  $E(\hat{R}_{1,t}^m) \rightarrow 0$  as  $m \rightarrow \infty$ .

For  $\phi_{i2}^0$ , and by using the same technique, we have

$$\begin{aligned} \frac{\partial \lambda_t^m}{\partial \phi_{i2}} - \frac{\partial \lambda_t}{\partial \phi_{i2}} &= [\lambda_{t-1}^m I(r_{i-1} \leq Y_{t-d}^m < r_i) - \lambda_{t-1} I(r_{i-1} \leq Y_{t-d} < r_i)] \\ &\quad + \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}^m}{\partial \phi_{i2}} I(r_{j-1} \leq Y_{t-d}^m < r_j) \\ &\quad - \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}}{\partial \phi_{i2}} I(r_{j-1} \leq Y_{t-d} < r_j) \\ &= \left[ \lambda_{t-1}^m - \lambda_{t-1} \right] I(r_{i-1} \leq Y_{t-d}^m < r_i) \\ &\quad + \lambda_{t-1} \left[ I(r_{i-1} \leq Y_{t-d}^m < r_i) - I(r_{i-1} \leq Y_{t-d} < r_i) \right] \\ &\quad + \left[ \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \left( \frac{\partial \lambda_{t-1}^m}{\partial \phi_{i2}} - \frac{\partial \lambda_{t-1}}{\partial \phi_{i2}} \right) \right] \\ &\quad + \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}}{\partial \phi_{i2}} \left[ I(r_{j-1} \leq Y_{t-d}^m < r_j) - I(r_{j-1} \leq Y_{t-d} < r_j) \right] \\ E \left( \frac{\partial \lambda_t^m}{\partial \phi_{i2}} - \frac{\partial \lambda_t}{\partial \phi_{i2}} \right) &= E \left[ (\lambda_{t-1}^m - \lambda_{t-1}) I(r_{i-1} \leq Y_{t-d}^m < r_i) \right] \\ &\quad + E \left[ \lambda_{t-1} \left( I(r_{i-1} \leq Y_{t-d}^m < r_i) - I(r_{i-1} \leq Y_{t-d} < r_i) \right) \right] \\ &\quad + E \left[ \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \left( \frac{\partial \lambda_{t-1}^m}{\partial \phi_{i2}} - \frac{\partial \lambda_{t-1}}{\partial \phi_{i2}} \right) \right] \\ &\quad + E \left\{ \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}}{\partial \phi_{i2}} \right. \\ &\quad \left. \times \left[ I(r_{i-1} \leq Y_{t-d}^m < r_i) - I(r_{i-1} \leq Y_{t-d} < r_i) \right] \right\} \end{aligned}$$

$$\begin{aligned}
\left| E\left(\frac{\partial \lambda_t^m}{\partial \phi_{i2}} - \frac{\partial \lambda_t}{\partial \phi_{i2}}\right) \right| &\leq \left| E\left(\lambda_{t-1}^m - \lambda_{t-1}\right) I(r_{i-1} \leq Y_{t-d}^m < r_i) \right| \\
&+ \left| E\left\{ \lambda_{t-1} \left[ I(r_{i-1} \leq Y_{t-d}^m < r_i) - I(r_{i-1} \leq Y_{t-d} < r_i) \right] \right\} \right| \\
&+ \left| E\left[ \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \left( \frac{\partial \lambda_{t-1}^m}{\partial \phi_{i2}} - \frac{\partial \lambda_{t-1}}{\partial \phi_{i2}} \right) \right] \right| \\
&+ \left| E\left\{ \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}}{\partial \phi_{i2}} \right. \right. \\
&\quad \left. \left. \times \left[ I(r_{i-1} \leq Y_{t-d}^m < r_i) - I(r_{i-1} \leq Y_{t-d} < r_i) \right] \right\} \right| \\
&\leq \left| E\left(\lambda_{t-1}^m - \lambda_{t-1}\right) I(r_{i-1} \leq Y_{t-d}^m < r_i) \right| \\
&+ \left| E\left[ \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \left( \frac{\partial \lambda_{t-1}^m}{\partial \phi_{i2}} - \frac{\partial \lambda_{t-1}}{\partial \phi_{i2}} \right) \right] \right| \\
&+ |E(\hat{R}_{2,t}^m)| \\
&\leq \alpha_{1,m} + \max_{1 \leq j \leq s} \phi_{j2} \left| E\left(\frac{\partial \lambda_{t-1}^m}{\partial \phi_{i2}} - \frac{\partial \lambda_{t-1}}{\partial \phi_{i2}}\right) \right| + \left| E(\hat{R}_{2,t}^m) \right| \\
&\leq \beta_{2,m},
\end{aligned}$$

where  $\beta_{2,m} \rightarrow 0$  as  $m \rightarrow \infty$ , and

$$\begin{aligned}
\hat{R}_{2,t}^m &= \lambda_{t-1} \left[ I(r_{i-1} \leq Y_{t-d}^m < r_i) - I(r_{i-1} \leq Y_{t-d} < r_i) \right] \\
&+ \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}}{\partial \phi_{i2}} \left[ I(r_{j-1} \leq Y_{t-d}^m < r_j) - I(r_{j-1} \leq Y_{t-d} < r_j) \right].
\end{aligned}$$

Note that  $E(\hat{R}_{2,t}^m) \rightarrow 0$  as  $m \rightarrow \infty$



For  $\phi_{i1}^0$ , we get

$$\begin{aligned}
 \frac{\partial \lambda_t^m}{\partial \phi_{i1}} - \frac{\partial \lambda_t}{\partial \phi_{i1}} &= [Y_{t-1}^m I(r_{i-1} \leq Y_{t-d}^m < r_i) - Y_{t-1} I(r_{i-1} \leq Y_{t-d} < r_i)] \\
 &\quad + \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}^m}{\partial \phi_{i1}} I(r_{j-1} \leq Y_{t-d}^m < r_j) \\
 &\quad - \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}}{\partial \phi_{i1}} I(r_{j-1} \leq Y_{t-d} < r_j) \\
 &= \left[ Y_{t-1}^m - Y_{t-1} \right] I(r_{i-1} \leq Y_{t-d}^m < r_i) \\
 &\quad + Y_{t-1} \left[ I(r_{i-1} \leq Y_{t-d}^m < r_i) - I(r_{i-1} \leq Y_{t-d} < r_i) \right] \\
 &\quad + \left[ \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \left( \frac{\partial \lambda_{t-1}^m}{\partial \phi_{i1}} - \frac{\partial \lambda_{t-1}}{\partial \phi_{i1}} \right) \right] \\
 &\quad + \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}}{\partial \phi_{i1}} \left[ I(r_{j-1} \leq Y_{t-d}^m < r_j) - I(r_{j-1} \leq Y_{t-d} < r_j) \right] \\
 \left| E \left( \frac{\partial \lambda_t^m}{\partial \phi_{i1}} - \frac{\partial \lambda_t}{\partial \phi_{i1}} \right) \right| &\leq \left| E \left[ \left( Y_{t-1}^m - Y_{t-1} \right) \left( I(r_{i-1} \leq Y_{t-d}^m < r_i) \right) \right] \right| \\
 &\quad + \max_{1 \leq j \leq s} \phi_{j2} \left| E \left( \frac{\partial \lambda_{t-1}^m}{\partial \phi_{i1}} - \frac{\partial \lambda_{t-1}}{\partial \phi_{i1}} \right) \right| + \left| E(\hat{R}_{3,t}^m) \right|,
 \end{aligned}$$

where

$$\begin{aligned}
 \hat{R}_{3,t}^m &= Y_{t-1} \left[ I(r_{i-1} \leq Y_{t-d}^m < r_i) - I(r_{i-1} \leq Y_{t-d} < r_i) \right] \\
 &\quad + \sum_{j=1}^s \phi_{j2} \frac{\partial \lambda_{t-1}}{\partial \phi_{i1}} \left[ I(r_{j-1} \leq Y_{t-d}^m < r_j) - I(r_{j-1} \leq Y_{t-d} < r_j) \right],
 \end{aligned}$$

and  $E(\hat{R}_{3,t}^m) \rightarrow 0$  as  $m \rightarrow \infty$ .

Hence,

$$\left| E \left( \frac{\partial \lambda_t^m}{\partial \phi_{i1}} - \frac{\partial \lambda_t}{\partial \phi_{i1}} \right) \right| \leq \alpha_{1,m} + \alpha_{4,m} + \max_{1 \leq j \leq s} \phi_{j2} \left| E \left( \frac{\partial \lambda_{t-1}^m}{\partial \phi_{i1}} - \frac{\partial \lambda_{t-1}}{\partial \phi_{i1}} \right) \right| \leq \beta_{3,m},$$

where  $\alpha_{4,m}, \beta_{3,m} \rightarrow 0$  as  $m \rightarrow \infty$ .

For the last assertion,

$$\begin{aligned}
& E\left(\frac{\partial\lambda_t^m}{\partial\phi_{i2}} - \frac{\partial\lambda_t}{\partial\phi_{i2}}\right)^2 \\
= & E\left\{\left(\lambda_{t-1}^m - \lambda_{t-1}\right)^2 \left(I(r_{i-1} \leq Y_{t-d}^m < r_i)\right)^2\right\} \\
& + \left[\left(\sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right)^2 \left(\frac{\partial\lambda_{t-1}^m}{\partial\phi_{i2}} - \frac{\partial\lambda_{t-1}}{\partial\phi_{i2}}\right)^2\right] + (\hat{R}_{2,t}^m)^2 \\
& + 2 \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \left(\lambda_{t-1}^m - \lambda_{t-1}\right) \left(\frac{\partial\lambda_{t-1}^m}{\partial\phi_{i2}} - \frac{\partial\lambda_{t-1}}{\partial\phi_{i2}}\right) \\
& + 2 \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \left(\lambda_{t-1}^m - \lambda_{t-1}\right) \hat{R}_{2,t}^m \Big\} \\
& + 2 \sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \left(\frac{\partial\lambda_{t-1}^m}{\partial\phi_{i2}} - \frac{\partial\lambda_{t-1}}{\partial\phi_{i2}}\right) \hat{R}_{2,t}^m \Big\} \\
= & E\left[\left(\lambda_{t-1}^m - \lambda_{t-1}\right)^2 \left(I(r_{i-1} \leq Y_{t-d}^m < r_i)\right)^2\right] \\
& + E\left[\left(\sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j)\right)^2 \left(\frac{\partial\lambda_{t-1}^m}{\partial\phi_{i2}} - \frac{\partial\lambda_{t-1}}{\partial\phi_{i2}}\right)^2\right] + E(\hat{R}_{2,t}^m)^2 \\
& + 2E\left[\sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \left(\lambda_{t-1}^m - \lambda_{t-1}\right) \left(\frac{\partial\lambda_{t-1}^m}{\partial\phi_{i2}} - \frac{\partial\lambda_{t-1}}{\partial\phi_{i2}}\right)\right] \\
& + 2E\left[I(r_{i-1} \leq Y_{t-d}^m < r_i) \left(\lambda_{t-1}^m - \lambda_{t-1}\right) \hat{R}_{2,t}^m\right] \\
& + 2E\left[\sum_{j=1}^s \phi_{j2} I(r_{j-1} \leq Y_{t-d}^m < r_j) \left(\frac{\partial\lambda_{t-1}^m}{\partial\phi_{i2}} - \frac{\partial\lambda_{t-1}}{\partial\phi_{i2}}\right) \hat{R}_{2,t}^m\right] \\
\leq & \alpha_{2,m} + c_{20}^2 E\left(\frac{\partial\lambda_{t-1}^m}{\partial\phi_{i2}} - \frac{\partial\lambda_{t-1}}{\partial\phi_{i2}}\right)^2 + 2c_{20}\beta_{2,m}\sqrt{\alpha_{2,m}} + \gamma_m \\
\leq & \beta_{4,m},
\end{aligned}$$

where  $\beta_{4,m} \rightarrow 0$ , and  $\gamma_m \rightarrow 0$  as  $m \rightarrow \infty$ , and  $c_{20} = \max_{1 \leq j \leq s} \phi_{j2}$ .

Furthermore, it is easy to see that ( the idea of proof is same as the last assertion in Theorem 3).

$$\begin{aligned}
& E\left(\frac{\partial\lambda_t^m}{\partial\theta_{i0}} - \frac{\partial\lambda_t}{\partial\theta_{i0}}\right)^2 \leq \beta_{5,m} \\
\text{and} \quad & E\left(\frac{\partial\lambda_t^m}{\partial\theta_{i1}} - \frac{\partial\lambda_t}{\partial\theta_{i1}}\right)^2 \leq \beta_{6,m},
\end{aligned}$$

where  $\beta_{5,m}, \beta_{6,m}$  are independent of  $t$  and  $\beta_{5,m}, \beta_{6,m} \rightarrow 0$ , as  $t \rightarrow \infty$ .

**Definition 3.** Define the matrices

$$\begin{aligned} G^m(\theta) &= E\pi^m \left[ \frac{1}{\lambda_t^m} \left( \frac{\partial \lambda_t^m}{\partial \theta} \right) \left( \frac{\partial \lambda_t^m}{\partial \theta} \right)' \right] \quad \text{and} \\ G(\theta) &= E\pi \left[ \frac{1}{\lambda_t} \left( \frac{\partial \lambda_t}{\partial \theta} \right) \left( \frac{\partial \lambda_t}{\partial \theta} \right)' \right], \end{aligned}$$

where  $\pi^m$  and  $\pi$  are the equilibrium distribution of the markov chains  $\lambda_t^m$  and  $\lambda_t$  respectively.

**Theorem 4.** Under the assumption  $\max_{1 \leq j \leq s} (\phi_{j1} + \phi_{j2}) < 1$ , the Rao score statistic satisfies

$$\Omega_n(\theta) \rightarrow \chi_D^2,$$

in distribution at  $\theta = \theta_0$ , as  $n \rightarrow \infty$ , where

$$\Omega_n(\theta) = S_n(\theta)' G_n^{-1}(\theta) S_n(\theta).$$

Here  $S_n(\theta)$  is the score function defined by (2.8),  $G_n(\theta)$  is the conditional information matrix defined by (2.10), and  $\chi_D^2$  is the Chi-square distribution with degree of freedom  $D$ , where  $D$  is the number of parameters.

**Proof of Theorem 4.** To prove this theorem, we need to prove two assertions: Stability and Asymptotic Normality (Basawa (1991)):

(i) Stability,

$$\frac{1}{n} G_n(\theta) \rightarrow G(\theta)$$

in probability at  $\theta = \theta_0$ , as  $n \rightarrow \infty$ , where  $G(\theta)$  is a  $s \times s$  positive conditional information matrix.

To prove the stability assertion, we use a proposition in Brockwell and Davis (2009, Proposition 6.3.9) which is the main tool for linking between  $\hat{\theta}$  and  $\hat{\theta}^m$ . This means, the following have to be proved:

- (1)  $\frac{1}{n}G_n^m(\theta) \xrightarrow{P} G^m(\theta)$ , as  $n \rightarrow \infty$ , for each  $m = 1, 2, \dots$ ,
- (2)  $G^m(\theta) \rightarrow G(\theta)$  as  $m \rightarrow \infty$ ,
- (3)  $\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P[|G_n^m(\theta) - G_n(\theta)| > \epsilon n] = 0$  for any  $\epsilon > 0$ .

The first assertion follows by Jensen and Rahbek (2007). They present that  $\frac{1}{n}G_n^m(\theta)$  converges in probability to  $G^m(\theta)$ , by using the Law of Large Number (LLN) for a geometrically ergodic process, that is

$$\frac{1}{n} \sum_{t=1}^n \left[ \frac{1}{\lambda_t^m} \left( \frac{\partial \lambda_t^m}{\partial \theta} \right) \left( \frac{\partial \lambda_t^m}{\partial \theta} \right)' \right] \rightarrow E\pi^m \left[ \frac{1}{\lambda_t^m} \left( \frac{\partial \lambda_t^m}{\partial \theta} \right) \left( \frac{\partial \lambda_t^m}{\partial \theta} \right)' \right]$$

in probability, as  $n \rightarrow \infty$ .

For the second statement, the limiting conditional information matrix converges to another conditional information matrix of PGARCH (1,1), as has been discussed by Fokianos et al. (2009). We need to evaluate the differences between  $\frac{1}{\lambda_t^m} \left( \frac{\partial \lambda_t^m}{\partial \theta} \right)^2$  and  $\frac{1}{\lambda_t} \left( \frac{\partial \lambda_t}{\partial \theta} \right)^2$  where  $\theta_0 = (\phi_{i0}^0, \phi_{i1}^0, \phi_{i2}^0)$ , i.e. we need to evaluate the following difference  $\frac{1}{\lambda_t^m} \left( \frac{\partial \lambda_t^m}{\partial \phi_{i2}} \right)^2 - \frac{1}{\lambda_t} \left( \frac{\partial \lambda_t}{\partial \phi_{i2}} \right)^2$ ,  $\frac{1}{\lambda_t^m} \left( \frac{\partial \lambda_t^m}{\partial \phi_{i1}} \right)^2 - \frac{1}{\lambda_t} \left( \frac{\partial \lambda_t}{\partial \phi_{i1}} \right)^2$  and  $\frac{1}{\lambda_t^m} \left( \frac{\partial \lambda_t^m}{\partial \phi_{i0}} \right)^2 - \frac{1}{\lambda_t} \left( \frac{\partial \lambda_t}{\partial \phi_{i0}} \right)^2$ .

Firstly, with  $\theta_0 = \phi_{i2}^0$ ,

$$\begin{aligned}
 & E \left| \frac{1}{\lambda_t^m} \left( \frac{\partial \lambda_t^m}{\partial \phi_{i2}} \right)^2 - \frac{1}{\lambda_t} \left( \frac{\partial \lambda_t}{\partial \phi_{i2}} \right)^2 \right| \\
 = & E \left| \left( \lambda_t \left( \frac{\partial \lambda_t^m}{\partial \phi_{i2}} \right)^2 - \lambda_t^m \left( \frac{\partial \lambda_t}{\partial \phi_{i2}} \right)^2 \right) / \lambda_t \lambda_t^m \right| \\
 = & E \left| \left( \lambda_t \left( \frac{\partial \lambda_t^m}{\partial \phi_{i2}} \right)^2 - \lambda_t \left( \frac{\partial \lambda_t}{\partial \phi_{i2}} \right)^2 + \lambda_t \left( \frac{\partial \lambda_t}{\partial \phi_{i2}} \right)^2 - \lambda_t^m \left( \frac{\partial \lambda_t}{\partial \phi_{i2}} \right)^2 \right) / \lambda_t \lambda_t^m \right| \\
 = & E \left| \left( \left( \frac{\partial \lambda_t^m}{\partial \phi_{i2}} \right)^2 - \left( \frac{\partial \lambda_t}{\partial \phi_{i2}} \right)^2 \right) / \lambda_t^m + \left( \left( \frac{\partial \lambda_t}{\partial \phi_{i2}} \right)^2 (\lambda_t - \lambda_t^m) \right) / \lambda_t \lambda_t^m \right| \\
 \leq & \frac{1}{c_0} E \left[ \left| \left( \frac{\partial \lambda_t^m}{\partial \phi_{i2}} \right) - \left( \frac{\partial \lambda_t}{\partial \phi_{i2}} \right) \right|^2 \right]^{\frac{1}{2}} E \left[ \left| \left( \frac{\partial \lambda_t^m}{\partial \phi_{i2}} \right) + \left( \frac{\partial \lambda_t}{\partial \phi_{i2}} \right) \right|^2 \right]^{\frac{1}{2}} \\
 + & \frac{1}{c_0^2} \left[ E(\lambda_t - \lambda_t^m)^2 E \left( \frac{\partial \lambda_t}{\partial \phi_{i2}} \right)^4 \right]^{\frac{1}{2}} \\
 \leq & \frac{c_1}{c_0} \sqrt{\beta_{4,m}} + \frac{c_2}{c_0^2} \sqrt{\alpha_{2,m}} \leq \beta,
 \end{aligned}$$

where  $\beta$  can be chosen arbitrarily small, and  $c_i$  are some constants. Here, because of the facts that  $\lambda_t \geq \min_{1 \leq j \leq s} \phi_{j0}$ , and  $\lambda_t^m \geq \min_{1 \leq j \leq s} \phi_{j0}$ , the first part of Theorem 3, the second conclusion of Theorem 2 and finite moments of  $\left( \frac{\partial \lambda_t^m}{\partial \phi_{i2}} \right)^4$ ,  $\left( \frac{\partial \lambda_t^m}{\partial \phi_{i2}} \right)$  and  $\left( \frac{\partial \lambda_t}{\partial \phi_{i2}} \right)$ .

Secondly, with  $\theta_0 = \phi_{i0}^0$ , similar to the above discussion

$$\begin{aligned}
 & E \left| \frac{1}{\lambda_t^m} \left( \frac{\partial \lambda_t^m}{\partial \phi_{i0}} \right)^2 - \frac{1}{\lambda_t} \left( \frac{\partial \lambda_t}{\partial \phi_{i0}} \right)^2 \right| \\
 \leq & \frac{1}{c_0} E \left| \left( \frac{\partial \lambda_t^m}{\partial \phi_{i0}} \right)^2 - \left( \frac{\partial \lambda_t}{\partial \phi_{i0}} \right)^2 \right| + \frac{1}{c_0^2} E \left| \left( \frac{\partial \lambda_t}{\partial \phi_{i0}} \right)^2 (\lambda_t - \lambda_t^m) \right| \\
 \leq & \frac{1}{c_0} E \left| \left( \left( \frac{\partial \lambda_t^m}{\partial \phi_{i0}} \right) - \left( \frac{\partial \lambda_t}{\partial \phi_{i0}} \right) \right) \left( \left( \frac{\partial \lambda_t^m}{\partial \phi_{i0}} \right) + \left( \frac{\partial \lambda_t}{\partial \phi_{i0}} \right) \right) \right| \\
 + & \frac{1}{c_0^2} E \left| \left( \frac{\partial \lambda_t}{\partial \phi_{i0}} \right)^2 (\lambda_t - \lambda_t^m) \right|.
 \end{aligned}$$

But, the first expectation can be bounded by  $c_1 \beta_{1,m}$ , and the second expectation,

$$\begin{aligned}
 E \left| \left( \frac{\partial \lambda_t}{\partial \phi_{i0}} \right)^2 (\lambda_t - \lambda_t^m) \right| &= \left[ E(\lambda_t - \lambda_t^m)^2 E \left( \frac{\partial \lambda_t}{\partial \phi_{i0}} \right)^4 \right]^{\frac{1}{2}} \\
 &\leq c_2 \sqrt{\alpha_{2,m}}.
 \end{aligned}$$

Hence,

$$\begin{aligned} E \left| \frac{1}{\lambda_t^m} \left( \frac{\partial \lambda_t^m}{\partial \phi_{i0}} \right)^2 - \frac{1}{\lambda_t} \left( \frac{\partial \lambda_t}{\partial \phi_{i0}} \right)^2 \right| &\leq \frac{c_1}{c_0} \beta_{1,m} + \frac{c_2}{c_0^2} \sqrt{\alpha_{2,m}} \\ &\leq \beta_1 \end{aligned}$$

where  $\beta_1$  also can be chosen arbitrarily small.

Finally, with  $\theta_0 = \phi_{i1}^0$ , it follows that

$$\begin{aligned} &E \left| \frac{1}{\lambda_t^m} \left( \frac{\partial \lambda_t^m}{\partial \phi_{i1}} \right)^2 - \frac{1}{\lambda_t} \left( \frac{\partial \lambda_t}{\partial \phi_{i1}} \right)^2 \right| \\ &\leq \frac{1}{c_0} E \left| \left( \frac{\partial \lambda_t^m}{\partial \phi_{i1}} \right)^2 - \left( \frac{\partial \lambda_t}{\partial \phi_{i1}} \right)^2 \right| + \frac{1}{c_0^2} E \left| \left( \frac{\partial \lambda_t}{\partial \phi_{i1}} \right)^2 (\lambda_t - \lambda_t^m) \right| \\ &\leq \frac{1}{c_0} E \left| \left( \left( \frac{\partial \lambda_t^m}{\partial \phi_{i1}} \right) - \left( \frac{\partial \lambda_t}{\partial \phi_{i1}} \right) \right) \left( \left( \frac{\partial \lambda_t^m}{\partial \phi_{i1}} \right) + \left( \frac{\partial \lambda_t}{\partial \phi_{i1}} \right) \right) \right| \\ &\quad + \frac{1}{c_0^2} E \left[ \left( \frac{\partial \lambda_t}{\partial \phi_{i1}} \right)^4 (\lambda_t - \lambda_t^m)^2 \right]^{\frac{1}{2}} \\ &\leq \frac{c_1}{c_0} \beta_{3,m} + \frac{c_2}{c_0^2} \sqrt{\alpha_{2,m}} \leq \beta_2, \end{aligned}$$

where  $\beta_2$  can be chosen arbitrarily small.

Then

$$\begin{aligned}
 & \frac{1}{n} \left[ G_n^m(\theta) - G_n(\theta) \right] \\
 = & \frac{1}{n} \sum_{t=1}^n \left[ \frac{1}{\lambda_t^m} \left( \frac{\partial \lambda_t^m}{\partial \theta} \right) \left( \frac{\partial \lambda_t^m}{\partial \theta} \right)' - \frac{1}{\lambda_t} \left( \frac{\partial \lambda_t}{\partial \theta} \right) \left( \frac{\partial \lambda_t}{\partial \theta} \right)' \right] \\
 = & \frac{1}{n} \sum_{t=1}^n \left[ \left( \lambda_t \left( \left( \frac{\partial \lambda_t^m}{\partial \theta} \right) \left( \frac{\partial \lambda_t^m}{\partial \theta} \right)' \right) - \lambda_t^m \left( \left( \frac{\partial \lambda_t}{\partial \theta} \right) \left( \frac{\partial \lambda_t}{\partial \theta} \right)' \right) \right) / \lambda_t \lambda_t^m \right] \\
 = & \frac{1}{n} \sum_{t=1}^n \left[ \left( \lambda_t \left( \left( \frac{\partial \lambda_t^m}{\partial \theta} \right) \left( \frac{\partial \lambda_t^m}{\partial \theta} \right)' \right) - \lambda_t \left( \left( \frac{\partial \lambda_t}{\partial \theta} \right) \left( \frac{\partial \lambda_t}{\partial \theta} \right)' \right) \right) \right. \\
 & \left. + \lambda_t \left( \left( \frac{\partial \lambda_t}{\partial \theta} \right) \left( \frac{\partial \lambda_t}{\partial \theta} \right)' \right) - \lambda_t^m \left( \left( \frac{\partial \lambda_t}{\partial \theta} \right) \left( \frac{\partial \lambda_t}{\partial \theta} \right)' \right) \right) / \lambda_t \lambda_t^m \right] \\
 \leq & \frac{1}{n} \sum_{t=1}^n \left[ \frac{1}{c_0} \left\| \left( \frac{\partial \lambda_t^m}{\partial \theta} \right) \left( \frac{\partial \lambda_t^m}{\partial \theta} \right)' - \left( \frac{\partial \lambda_t}{\partial \theta} \right) \left( \frac{\partial \lambda_t}{\partial \theta} \right)' \right\| \right. \\
 & \left. + \frac{1}{c_0^2} \left\| \left( \frac{\partial \lambda_t}{\partial \theta} \right) \left( \frac{\partial \lambda_t}{\partial \theta} \right)' (\lambda_t - \lambda_t^m) \right\| \right].
 \end{aligned}$$

For the first summand

$$E \left\| \left( \frac{\partial \lambda_t^m}{\partial \theta} \right) \left( \frac{\partial \lambda_t^m}{\partial \theta} \right)' - \left( \frac{\partial \lambda_t}{\partial \theta} \right) \left( \frac{\partial \lambda_t}{\partial \theta} \right)' \right\| \leq k \beta_{c,m} \rightarrow 0$$

as  $m \rightarrow \infty$  and  $k$  is constant. To see this

$$\begin{aligned}
 & E \left\| \left( \frac{\partial \lambda_t^m}{\partial \theta} \right) \left( \frac{\partial \lambda_t^m}{\partial \theta} \right)' - \left( \frac{\partial \lambda_t}{\partial \theta} \right) \left( \frac{\partial \lambda_t}{\partial \theta} \right)' \right\| \\
 = & E \left\| \left[ \left( \frac{\partial \lambda_t^m}{\partial \theta} \right) \left( \frac{\partial \lambda_t^m}{\partial \theta} \right)' - \left( \frac{\partial \lambda_t}{\partial \theta} \right) \left( \frac{\partial \lambda_t^m}{\partial \theta} \right)' \right. \right. \\
 & \left. \left. + \left( \frac{\partial \lambda_t}{\partial \theta} \right) \left( \frac{\partial \lambda_t^m}{\partial \theta} \right)' - \left( \frac{\partial \lambda_t}{\partial \theta} \right) \left( \frac{\partial \lambda_t}{\partial \theta} \right)' \right] \right\| \\
 \leq & E \left\| \left[ \left( \frac{\partial \lambda_t^m}{\partial \theta} \right) - \left( \frac{\partial \lambda_t}{\partial \theta} \right) \right] \left( \frac{\partial \lambda_t^m}{\partial \theta} \right)' \right\| \\
 & + E \left\| \left( \frac{\partial \lambda_t}{\partial \theta} \right) \left[ \left( \frac{\partial \lambda_t^m}{\partial \theta} \right)' - \left( \frac{\partial \lambda_t}{\partial \theta} \right)' \right] \right\| \\
 \leq & k_1 \beta_{c,m} + k_1 \beta_{a,m} \\
 \leq & \beta_3,
 \end{aligned}$$

where  $\beta_3$  can be chosen arbitrarily small and because of Theorem 3 and the finite moments of  $(\frac{\partial \lambda_t^m}{\partial \phi})'$  and  $(\frac{\partial \lambda_t}{\partial \phi})'$ .

For the second summand, the same idea applies because of  $E \left\| \left( \frac{\partial \lambda_t}{\partial \theta} \right) \left( \frac{\partial \lambda_t}{\partial \theta} \right)' \right\| < \infty$ , and the fact, for  $\frac{1}{c_0^2} E |\lambda_t - \lambda_t^m| \leq \frac{1}{c_0^2} (\alpha_{2,m})^{1/2}$ .

This leads to

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P[|G_n^m(\theta) - G_n(\theta)| > \epsilon n] = 0.$$

This completes the proof of the stability.

(ii) Asymptotic Normality

$$\frac{1}{\sqrt{n}} S_n(\theta) \rightarrow \mathcal{N}_D(0, G(\theta)),$$

in distribution at  $\theta = \theta_0$ , as  $n \rightarrow \infty$ , where  $\mathcal{N}_D$  denotes a  $D$ -variate normal distribution, and  $G(\theta)$  is a non singular  $D \times D$  matrix.

To prove the Asymptotic Normality assertion, we use the same proposition as in the proof of stability, but in this case, in the following way:

- (1)  $\frac{1}{\sqrt{n}} S_n^m \xrightarrow{D} S^m := \mathcal{N}(0, G^m)$  as  $n \rightarrow \infty$  for each  $m = 1, 2, \dots$ ,
- (2)  $S^m \xrightarrow{D} \mathcal{N}(0, G)$  as  $m \rightarrow \infty$ ,
- (3)  $\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P(\|S_n^m - S_n\| > \delta \sqrt{n}) = 0$ , for any  $\delta > 0$ .

For assertion (1), having evaluated the true value  $\theta = \theta_0$ , the equation (2.8) illustrates that the score is defined as

$$\begin{aligned} S_n^m(\theta) &= \frac{\partial \ell^m(\theta)}{\partial \theta} = \sum_{t=1}^n \frac{\partial \ell_t^m(\theta)}{\partial \theta} \\ &= \sum_{t=1}^n \left( \frac{Y_t^m}{\lambda_t^m(\theta)} - 1 \right) \frac{\partial \lambda_t^m(\theta)}{\partial \theta} \end{aligned}$$



where  $\frac{\partial \lambda_t^m(\theta)}{\partial \theta}$ , is given by (2.12). Furthermore, from (2.3) and (2.12), we find,

$$\begin{aligned}\frac{\partial \lambda_t^m}{\partial \phi_{i2}} &\leq \sum_{k=0}^{t-1} c_{2,0}^k \lambda_{t-k-1}^m I(r_{i-1} \leq Y_{t-d-k}^m < r_i), \\ \frac{\partial \lambda_t^m}{\partial \phi_{i1}} &\leq \sum_{k=0}^{t-1} c_{2,0}^k Y_{t-k-1}^m I(r_{i-1} \leq Y_{t-d-k}^m < r_i), \\ \frac{\partial \lambda_t^m}{\partial \phi_{i0}} &\leq \sum_{k=0}^{t-1} c_{2,0}^k I(r_{i-1} \leq Y_{t-d-k}^m < r_i),\end{aligned}$$

where  $c_{2,0} = \sum_{j=1}^s \phi_{j2}$ . Now, we must find the conditional expected value of our score function such that

$$\begin{aligned}E\left(\frac{\partial \ell^m(\theta)}{\partial \theta} \middle| \mathcal{F}_{t-1}\right) &= \sum_{t=1}^n E\left[\left(\frac{Y_t^m}{\lambda_t^m} - 1\right) \frac{\partial \lambda_t^m}{\partial \theta} \middle| \mathcal{F}_{t-1}\right] \\ &= \sum_{t=1}^n E\left[\frac{Y_t^m}{\lambda_t^m} \frac{\partial \lambda_t^m}{\partial \theta} \middle| \mathcal{F}_{t-1} - \frac{\partial \lambda_t^m}{\partial \theta} \middle| \mathcal{F}_{t-1}\right] \\ &= \sum_{t=1}^n E\left[\frac{\partial \lambda_t^m}{\partial \theta} \middle| \mathcal{F}_{t-1} - \frac{\partial \lambda_t^m}{\partial \theta} \middle| \mathcal{F}_{t-1}\right] = 0\end{aligned}$$

where  $E(Y_t^m | \mathcal{F}_{t-1}) = \lambda_t^m$ , and  $\mathcal{F}_{t-1}$  is the  $\sigma$ - algebra generated by  $\{U_{k+1}, N_k, k \leq t-1\}$ . By using the central limit theorem for martingale differences in Brown and Eagleson (1971), we have the variance of the score function as

$$\begin{aligned}Var\left(\frac{\partial \ell^m(\theta)}{\partial \theta} \middle| \mathcal{F}_{t-1}\right) &= \sum_{t=1}^n E\left[\left(\left(\frac{Y_t^m}{\lambda_t^m} - 1\right) \frac{\partial \lambda_t^m}{\partial \theta}\right) \left(\left(\frac{Y_t^m}{\lambda_t^m} - 1\right) \frac{\partial \lambda_t^m}{\partial \theta}\right)' \middle| \mathcal{F}_{t-1}\right] \\ &= \sum_{t=1}^n E\left[\left(\frac{Y_t^m}{\lambda_t^m} - 1\right)^2 \left(\frac{\partial \lambda_t^m}{\partial \theta}\right) \left(\frac{\partial \lambda_t^m}{\partial \theta}\right)' \middle| \mathcal{F}_{t-1}\right] \\ &= \sum_{t=1}^n \left[\left(\frac{\partial \lambda_t^m}{\partial \theta}\right) \left(\frac{\partial \lambda_t^m}{\partial \theta}\right)' E\left(\frac{Y_t^m}{\lambda_t^m} - 1\right)^2 \middle| \mathcal{F}_{t-1}\right].\end{aligned}$$

But

$$\begin{aligned} E\left[\left(\frac{Y_t^m}{\lambda_t^m} - 1\right)^2 \middle| \mathcal{F}_{t-1}\right] &= E\left[\left(\frac{(Y_t^m)^2}{(\lambda_t^m)^2} - 2\frac{Y_t^m}{\lambda_t^m} + 1\right) \middle| \mathcal{F}_{t-1}\right] \\ &= \frac{1}{\lambda_t^m}. \end{aligned}$$

Hence,

$$\begin{aligned} \text{Var}\left(\frac{\partial \ell^m(\theta)}{\partial \theta} \middle| \mathcal{F}_{t-1}\right) &= \sum_{t=1}^n \frac{1}{\lambda_t^m} \left(\frac{\partial \lambda_t^m}{\partial \theta}\right) \left(\frac{\partial \lambda_t^m}{\partial \theta}\right)' \\ &= G_n^m. \end{aligned}$$

From the above,  $E\left[\left(\frac{Y_t^m}{\lambda_t^m} - 1\right) \middle| \mathcal{F}_{t-1}\right] = 0$ . In addition, from Hölder's inequality, we conclude that the expectation of  $\left\|\frac{\partial \ell^m(\theta)}{\partial \theta}\right\|$  is finite. Therefore,  $\frac{\partial \ell^m(\theta)}{\partial \theta}$  is a martingale difference sequence with respect to the past information  $(\mathcal{F}_{t-1})$ . An application of the CLT, gives  $\frac{1}{\sqrt{n}}S_n^m$  is asymptotically Normality with variance given by the limit,

$$\frac{1}{n} \sum_{t=1}^n E\left((N_t^m)^2 \left(\frac{\partial \lambda_t^m}{\partial \theta}\right) \left(\frac{\partial \lambda_t^m}{\partial \theta}\right)' \middle| \mathcal{F}_{t-1}\right) \xrightarrow{P} G^m,$$

where  $N_t^m = \frac{Y_t^m}{\lambda_t^m} - 1$  (Hall and Heyde (1980)) and by the LLN for geometrically ergodic process (Jensen and Rahbek (2007)). Hence the conditional Lindeberg's condition which guarantee that the variance is arbitrarily small for the contribution of any random variable. This implies that

$$\frac{1}{n} \sum_{t=1}^n E\left(\left\|\frac{\partial \ell_t^m(\theta)}{\partial \theta}\right\|^2 I\left(\left\|\frac{\partial \ell_t^m(\theta)}{\partial \theta}\right\| > \sqrt{n}\delta\right) \middle| \mathcal{F}_{t-1}\right) \leq \frac{1}{n^2\delta^2} \sum_{t=1}^n E\left(\left\|\frac{\partial \ell_t^m(\theta)}{\partial \theta}\right\|^4 \middle| \mathcal{F}_{t-1}\right) \rightarrow 0,$$

as  $n \rightarrow \infty$  and  $E\left\|\frac{\partial \ell_t^m(\theta)}{\partial \theta}\right\|^4 < \infty$ . This prove the first part of the Asymptotic Normality.

The second part follows from assertion (2) of the stability.

In the last assertion, for convenience, we write

$$\frac{\partial \ell^m(\theta)}{\partial \theta} = \left( \frac{Y_t^m}{\lambda_t^m} - 1 \right) \frac{\partial \lambda_t^m}{\partial \theta} = N_t^m \frac{\partial \lambda_t^m}{\partial \theta},$$

and

$$\frac{\partial \ell(\theta)}{\partial \theta} = \left( \frac{Y_t}{\lambda_t} - 1 \right) \frac{\partial \lambda_t}{\partial \theta} = N_t \frac{\partial \lambda_t}{\partial \theta},$$

where  $\frac{\partial \lambda_t^m}{\partial \theta}$  and  $\frac{\partial \lambda_t}{\partial \theta}$  are defined by (2.12) and (2.9) respectively. Then

$$\begin{aligned} \frac{1}{\sqrt{n}}(S_n^m - S_n) &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \left( \frac{\partial \ell^m(\theta)}{\partial \theta} - \frac{\partial \ell(\theta)}{\partial \theta} \right) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \left( N_t^m \frac{\partial \lambda_t^m}{\partial \theta} - N_t \frac{\partial \lambda_t}{\partial \theta} \right) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \left[ N_t^m \left( \frac{\partial \lambda_t^m}{\partial \theta} - \frac{\partial \lambda_t}{\partial \theta} \right) + (N_t^m - N_t) \frac{\partial \lambda_t}{\partial \theta} \right]. \end{aligned}$$

For the first summand in the above expression,

$$\begin{aligned} P \left( \left\| \sum_{t=1}^n N_t^m \left( \frac{\partial \lambda_t^m}{\partial \theta} - \frac{\partial \lambda_t}{\partial \theta} \right) \right\| > \delta \sqrt{n} \right) &\leq P \left( \beta_m \left| \sum_{t=1}^n N_t^m \right| > \delta \sqrt{n} \right) \\ &\leq \frac{\beta_m^2}{\delta^2 n} \sum_{t=1}^n E \left| N_t^m \right|^2 \leq K \beta_m^2 \rightarrow 0 \end{aligned}$$

as  $m \rightarrow \infty$ . The second summand,  $E \left\| \frac{\partial \lambda_t}{\partial \theta} \right\|^2$  is finite and  $|N_t^m - N_t|$  is bounded above by  $\beta_{a,m}$ , and  $\beta_{a,m} \rightarrow 0$  as  $m \rightarrow \infty$ . To see this, we have

$$\begin{aligned} N_t^m - N_t &= \left( \frac{Y_t^m}{\lambda_t^m} - 1 \right) - \left( \frac{Y_t}{\lambda_t} - 1 \right) \\ &= \frac{\lambda_t(Y_t^m - Y_t) - Y_t(\lambda_t^m - \lambda_t)}{\lambda_t \lambda_t^m}, \end{aligned}$$

$E \left| \frac{Y_t^m - Y_t}{\lambda_t^m} \right| \leq \frac{1}{c_0} \sqrt{\alpha_{3,m}}$ . Using Theorem 2 and the fact that  $E|Y_t|^2 < \infty$ , we find

$$E \left| \frac{Y_t(\lambda_t^m - \lambda_t)}{\lambda_t \lambda_t^m} \right| \leq \frac{\sqrt{\alpha_{2,m} E|Y_t|^2}}{c_0^2} \leq k\alpha_m.$$

Having proved assertions (1), (2) and (3), we get

$$\frac{1}{\sqrt{n}} S_n(\theta) \rightarrow \mathcal{N}_D(0, G(\theta)),$$

in distribution, as  $n \rightarrow \infty$ . This completes the proof of Asymptotic Normality.

From Stability assertion, we have

$$\frac{1}{n} G_n(\theta) \xrightarrow{P} G(\theta) \text{ as } n \rightarrow \infty,$$

hence

$$[\frac{1}{n} G_n(\theta)]^{-1} = [G(\theta)]^{-1} (1 + o_p(1))$$

Therefore, from Theorem 4, we get

$$\begin{aligned} \Omega_n(\theta) &= S_n(\theta)' [G_n(\theta)]^{-1} S_n(\theta) \\ &= \left[ \frac{1}{\sqrt{n}} S_n(\theta) \right]' \left[ n^{-1} G_n(\theta) \right]^{-1} \left[ \frac{1}{\sqrt{n}} S_n(\theta) \right] \\ &= \left[ \frac{1}{\sqrt{n}} S_n(\theta) \right]' \left[ G(\theta) \right]^{-1} \left[ \frac{1}{\sqrt{n}} S_n(\theta) \right] (1 + o_p(1)) \\ &= \left[ \frac{G^{-1/2}(\theta) S_n(\theta)}{\sqrt{n}} \right]' \left[ \frac{G^{-1/2}(\theta) S_n(\theta)}{\sqrt{n}} \right] \left( 1 + o_p(1) \right). \end{aligned}$$

But from Asymptotic Normality, we have

$$\frac{1}{\sqrt{n}} G^{-1/2}(\theta) S_n(\theta) \rightarrow \mathcal{N}(0, I_D),$$

where  $I_D$  is the identity matrix of size  $D$ .

Hence,

$$\Omega_n(\theta) \rightarrow \chi_D^2, \quad \text{as } n \rightarrow \infty.$$

In the next part, we will present the most popular information criteria in time series analysis used to identify the suitable model. In particular, we discuss the AIC (Akaike (1973)), BIC (Schwarz (1978)) and HQIC (Hannan and Quinn (1979)). The aim of these criteria is to find the model with the lowest value of the selected information criteria. We will use these criteria to estimate the parameters of PTGARCH (1,1) model such as the value of threshold, time delay and the number of regimes in the next chapters.

### 2.2.1.2 Akaike information criterion (AIC)

The Akaike Information criterion (AIC) is considered to be one of the most important criteria used to determine the parameters of a statistical model. The AIC is a measure of the relative goodness of fit of a statistical model and was introduced by Akaike (1973). Assuming a stationary time series  $Y_t, t = 1, 2, \dots, N$ , the AIC statistic is defined as two parts; first part is maximizing the conditional log-likelihood function and the second part is a penalty for model complexity.

$$\text{AIC} = -\frac{2}{N}\ln(L) + \frac{2}{N}k,$$

where  $L$  is likelihood function,  $k$  is the number of parameters to be estimated and  $N$  is the number of observations.

### 2.2.1.3 Bayesian information criterion (BIC)

The Bayesian information criterion (BIC) is another criterion for model selection among a finite set of models, which was introduced by Schwarz (1978). It is given as

$$\text{BIC} = -\frac{2}{N}\ln(L) + \frac{k}{N}\ln(N),$$

#### 2.2.1.4 Hannan-Quinn information criterion (HQIC)

The Hannan-Quinn information criterion (HQIC) is one objective measure of model suitability that balances model complexity and model fit. The HQIC, introduced by Hannan and Quinn (1979) is defined as

$$\text{HQIC} = \frac{-2}{N} \ln(L) + \frac{2k}{N} \ln(\ln(N)),$$

where  $L$  is the likelihood function,  $k$  is the number of parameters to be estimated and  $N$  is the number of observations. The best model is the model that corresponds to the minimum (*i.e.* minHQIC).

### 2.2.2 Testing nonlinearity

The model defined by (2.2) has  $s$  regimes. Each regime follows a linear GARCH model, but the whole process follows a nonlinear GARCH model. The nonlinearity is caused by the endogenous switching between the regimes and there are  $s$  regimes with different linear models. In this part, we discuss tests for the threshold nonlinearity. The majority of tests for the threshold nonlinearity are built for detecting quadratic nonlinearity; for example, Keenan's test and Tasy's test. We will focus on the Lagrange Multiplier Test (LM) and the log likelihood Ratio test (LR) with the threshold autoregressive model as the alternative hypothesis. The main problem for the threshold nonlinearity test is that the nuisance parameters are not identifiable under the null hypothesis. Because the log-likelihood function of these tests for the model defined by (2.2) are quadratic, then the sampling distribution of these tests is no longer approximately  $\chi_k^2$  distribution with  $k$  degrees of freedom under the null hypothesis for all sample sizes, see Buse (1982) and Cryer and Chan (2008). These test have a nonstandard sampling distribution (Chan (1991), Tong and Series (1990) and Hansen (1996)). These tests are widely used in nonlinear time series.

### 2.2.2.1 Lagrange multiplier (LM) test

In 1959, David Silvey used the name Lagrange multiplier test. This test has become increasingly popular under the title of Lagrange multiplier in econometrics, although it was proposed approximately a decade earlier by Rao (1948). The LM test measures the distance between composite hypotheses  $H_0$  and  $H_1$  in terms of the score function  $S(\theta_0)$ . The null hypothesis  $H_0 : \lambda_t$  has one regime (linear GARCH (1,1) model), whereas the alternative hypothesis is  $H_1$  : that the process follows the model defined by (2.1). Suppose that  $\hat{\theta}_0$  is the maximum likelihood estimate of  $\theta$  under the null hypothesis  $H_0$ . Then, the LM test depends on the score function at  $\theta_0$ , where

$$LM = (S_n(\hat{\theta}_0))' (G_n(\hat{\theta}_0))^{-1} (S_n(\hat{\theta}_0)) \rightarrow \chi_k^2.$$

Here  $S_n(\hat{\theta}_0)$  and  $G_n(\hat{\theta}_0)$  are the score function and conditional information matrix respectively. The  $LM^*$  test statistic function was calculated by using the standard Lagrange Multiplier test (LM). This test statistic function uses the maximum (max) and average (ave) to generate the test statistic  $LM^*$ . That is,

$$LM^* = \max_{u \leq r \leq v} LM \tag{2.14}$$

and

$$LM^* = \text{ave}_{u \leq r \leq v} LM. \tag{2.15}$$

In general, simulation is used to calculate empirical critical values of the test statistic  $LM^*$  which depends upon the selection of the lower and upper threshold value  $r$ . The maximum and average LM test is considered by Hansen (1996) as a test statistic.

### 2.2.2.2 Likelihood ratio (LR) test

The LR test is used to compare the fit of two models, and one of them is the null hypothesis which is a special case of the other (the alternative hypothesis model). This test is based on the Likelihood ratio. The LR test depends on the difference between the maximum likelihood estimates of the parameters ( $\theta_0$ ) under the null hypothesis and the alternative hypothesis. The null hypotheses is  $H_0$ : the process follows a linear GARCH (1,1) model

$$\lambda_t = \phi_0 + \phi_{01}\lambda_{t-1} + \phi_{02}Y_{t-1},$$

whereas the alternative hypothesis is  $H_1$ : the process follows a nonlinear TAR model

$$\lambda_t = \phi_{j0} + \phi_{j1}Y_{t-1} + \phi_{j2}\lambda_{t-1}, \quad \text{when } r_{j-1} < Y_{t-d} \leq r_j, \quad j = 1, \dots, s.$$

The log likelihood ratio  $\ell(r)$  is defined, as a function of the threshold value  $r$ , by

$$\ell(r) = \ell_1(r; \hat{\theta}) - \ell_0(\hat{\theta}),$$

where  $r$  is the vector of the selected threshold values  $(r_1, r_2, \dots, r_s)$ . The parameter  $r$  is referred to as a noise parameter under the null hypotheses. Usually, we use  $LR_{\max} = \max_{u \leq r \leq v} |\ell(r)|$  as a test statistic, where  $u$  and  $v$  are lower and upper bounds of the threshold value respectively. Andrews and Ploberger (1994) used the average of  $\ell(r)$  as a test statistic. Under the null hypotheses, the LR test is no longer approximately  $\chi_k^2$  with  $k$  degrees of freedom. When selecting  $u$  and  $v$ , we should ensure that enough data falls into each regime. Both the LM test and the LR test are a suitable measure of the distance between the null hypothesis and the alternative hypothesis.

The decision rule is to reject the null hypothesis if both tests,  $LM^*$  which is defined by (2.14) or (2.15) and LR, are greater than  $\chi_k^2$ , where  $\chi_k$  denotes the 90th, 95th and 99th percentiles of the chi-square distribution with degree of freedom  $\alpha$ . In general, large values of these tests lead to the rejection of the null



hypothesis. Most statistical programmes will provide the  $p$ -value for the LM test. Hence, the decision rule is to reject  $H_0$  if the  $p$ -value is less than or equal to the significance level  $\alpha$ .

## **2.3 Summary**

This chapter introduces a kind of threshold autoregressive model for nonlinear count time series. We have illustrated a simple sufficient condition for geometric ergodicity of a perturbed model. Theorem 2 shows the difference between the unperturbed model and the perturbed model. In addition, we have discussed maximum likelihood estimation (MLE) for the model. We also provided the asymptotic properties for the nonlinear model defined by (2.2) by deriving the asymptotic properties of the nonlinear model (2.3). The estimation method used in this chapter enable us to fit a threshold model to given series of data. We have introduced some criteria for model selection. Our model is a nonlinear model, and therefore two different kinds of nonlinearity tests were presented. The performance of different information criteria for choosing the rest of the parameters such as the number of regimes, the threshold value and the delay parameter for the Poisson TAR model will be investigated in the next chapters.

# Chapter 3

## Simulation and Data Analysis

In this chapter, we carry out a few simulations to show the asymptotic distribution of the maximum likelihood estimator (MLE) for the model defined by (2.2). We also use real data to illustrate our methodology. The values of the maximum likelihood estimators are computed by optimizing the log-likelihood function defined by (2.7). The analytical solution of the equation (2.8) does not exist. Therefore, we use a quasi-Newton method which is an iterative method for finding roots of equations (2.8). This method is used for solving nonlinear optimization problems, and specifically, to find steady-state of the score function.

An important subject for PTGARCH models is how to determine the threshold value  $r$ . Therefore, we also carry out some simulations for selecting the appropriate threshold value using various information criteria (AIC, BIC, HQIC) which are presented in the previous chapter.

Our model includes at least two different autoregressive regimes. It is useful for us to study the test for threshold nonlinearity. To that end, we use the Lagrange Multiplier test (LM) which is defined by (2.14) and (2.15). We give a few simulation results to illustrate asymptotic normality of the MLE for model (2.2) and the threshold nonlinearity first, then apply the model defined by (2.2) to the real data example.

## 3.1 Numerical results: simulation and real data example

### 3.1.1 The performance of Maximum Likelihood Estimators (MLE)

In this section, we carry out some simulation studies to show the finite sample properties of the MLE for the PTGARCH models. To make the analysis simple, we consider a two-regime model defined as follows

$$\begin{aligned}
 Y_t | \mathcal{F}_{t-1}^{Y, \lambda} &\sim \text{Poisson}(\lambda_t), \\
 \lambda_t &= \begin{cases} d_1 + b_1 Y_{t-1} + a_1 \lambda_{t-1}, & \text{if } Y_{t-d} \leq r, \\ d_2 + b_2 Y_{t-1} + a_2 \lambda_{t-1}, & \text{if } Y_{t-d} > r, \end{cases} \quad (3.1)
 \end{aligned}$$

where  $t \geq 1$ , the initial values  $Y_0, \lambda_0$  are fixed, the intercept coefficients  $d_i, i = 1, 2$  are assumed to be positive real numbers, and the autoregression coefficients  $a_i, b_i, i = 1, 2$  are assumed to be nonnegative real numbers. Theorem 1 and Theorem 2 in Chapter 2 show that the process  $\lambda_t$  is approximated by a geometrically ergodic process  $\lambda_t^m$ . This process consists of two regimes and it switches between them according to the threshold variable  $Y_{t-d}$ . If the threshold variable does not exceed the threshold value then the process  $\lambda_t$  falls in the first regime, otherwise the process  $\lambda_t$  follows the second regime. The systems may differ from each other with respect to the constants (i.e.,  $d_i, i = 1, 2$ ) and the GARCH coefficients (i.e.,  $a_i, b_i, i = 1, 2$ ). Each regime follows a linear PGARCH(1,1) model. Ferland et al. (2006) showed that the PGARCH (1,1) model has the first and the second moments identical to those of an ARMA (1,1) model, and it is stationary, provided that the sum of the different autoregression coefficients is less than one. In contrast, we find out that the PTGARCH (1,1; 2;1) process can be approximated by a stationary process provided that  $\max(a_1 + b_1, a_2 + b_2) < 1$ . The log

likelihood function and the score function are defined by equation (2.7) and (2.8) respectively, where  $\frac{\partial \lambda_t(\theta)}{\partial(\theta)}$  is a six-dimensional vector with components given as follows:

$$\begin{aligned}
 \frac{\partial \lambda_t}{\partial d_1} &= I\left(Y_{t-d} \leq r\right) + \frac{\partial \lambda_{t-1}}{\partial d_1} \left[ a_1 I(Y_{t-d} \leq r) + a_2 I(Y_{t-d} > r) \right], \\
 \frac{\partial \lambda_t}{\partial d_2} &= I\left(Y_{t-d} > r\right) + \frac{\partial \lambda_{t-1}}{\partial d_2} \left[ a_1 I(Y_{t-d} \leq r) + a_2 I(Y_{t-d} > r) \right], \\
 \frac{\partial \lambda_t}{\partial a_1} &= \lambda_{t-1} I\left(Y_{t-d} \leq r\right) + \frac{\partial \lambda_{t-1}}{\partial a_1} \left[ a_1 I(Y_{t-d} \leq r) + a_2 I(Y_{t-d} > r) \right], \\
 \frac{\partial \lambda_t}{\partial a_2} &= \lambda_{t-1} I\left(Y_{t-d} > r\right) + \frac{\partial \lambda_{t-1}}{\partial a_2} \left[ a_1 I(Y_{t-d} \leq r) + a_2 I(Y_{t-d} > r) \right], \\
 \frac{\partial \lambda_t}{\partial b_1} &= Y_{t-1} I\left(Y_{t-d} \leq r\right) + \frac{\partial \lambda_{t-1}}{\partial b_1} \left[ a_1 I(Y_{t-d} \leq r) + a_2 I(Y_{t-d} > r) \right], \\
 \frac{\partial \lambda_t}{\partial b_2} &= Y_{t-1} I\left(Y_{t-d} > r\right) + \frac{\partial \lambda_{t-1}}{\partial b_2} \left[ a_1 I(Y_{t-d} \leq r) + a_2 I(Y_{t-d} > r) \right].
 \end{aligned} \tag{3.2}$$

We calculate the MLEs by optimising equation (2.8) using the quasi-Newton method. Simulated time series with different sample sizes were generated from two PTGARCH  $(1, 1; r; d)$  models. The parameters of the first model are  $(d_1, a_1, b_1; d_2, a_2, b_2, r, d) = (0.4, 0.5, 0.3; 0.5, 0.3, 0.4, 2, 1)$ , while the second model is with  $(d_1, a_1, b_1; d_2, a_2, b_2, r, d) = (0.6, 0.7, 0.2; 0.3, 0.4, 0.5, 3, 2)$ . Note that, for these choices of coefficient values,  $\max(0.5 + 0.3, 0.3 + 0.4) = 0.8$  and  $\max(0.7 + 0.2, 0.4 + 0.5) = 0.9$  for the first and the second model respectively. The rest of the results for different sets of parameter values were analogous, the condition  $\max(a_1 + b_1, a_2 + b_2) < 1$  was taken into account, and hence they are omitted here. Ferland et al. (2006) showed that any PGARCH (1,1) model has first and second moments identical to those of an ARMA (1,1) model

$$(Y_t - \mu) - (a + b)(Y_{t-1} - \mu) = e_t - ae_{t-1},$$

where the sequence  $e_t$  is a white noise with constant first and second moments,

that is,  $E(e_t) = \text{Var}(e_t) = \mu$ . As mentioned previously, the model defined by (3.1) has two different sub-models, each sub-model may follow a linear PGARCH (1,1) model and has moment up to second order. This means we have two different ARMA (1,1) models, one for each PGARCH (1,1) model. This result indicates that the least-squares method may be used for estimating the parameters. We use this property to compare the least-squares estimators (LSE) with maximum likelihood estimators (MLE). To do this, the algorithm for optimisation of equation (2.7) is done by using the recursions of the equations (2.9). In what follows, we carry out simulations and real data analysis.

Table 3.1 compares the maximum likelihood estimators (MLE) and the least-squares estimators (LSE). In Table 3.1 the results are given for the data generated with the first model.

Table 3.1 reports the estimates of the parameters obtained by averaging out the results from all simulations. It also reports the average absolute error (AAE) of both estimators (MLE and LSE) for different sample sizes. The mean square error (MSE) has been computed using the simulation output. Note that the MSE of both estimators decreases as the sample size increases. In all cases, the least-squares estimators have greater errors than the maximum likelihood estimators and as the sample size increases, the average absolute error (AAE) of both types of estimator decreases. The MSE is defined as  $MSE(\hat{\theta}) = \frac{1}{N} \sum (\hat{\theta} - \theta)^2$  while the AAE is defined as  $AAE(\hat{\theta}) = \frac{1}{N} \sum |\hat{\theta} - \theta|$ . As expected, the MLE performs better than the LSE. Both estimators are close to the true parameter value because as the sample size increases, the mean values of the estimates approach the true parameter values. The MLE is more consistent with smaller MSE and more focused around true parameters when the sample size increases from small to large. The initial values for parameters are obtained by the ARIMA (1,1) fit to the data for MLE.

Table 3.2 supports the results for the standard error found by standard devi-

Table 3.1: Maximum likelihood estimates and least-squares estimates with their mean square errors (in parentheses) for model (3.1) when  $(d_1, a_1, b_1; d_2, a_2, b_2) = (0.4, 0.5, 0.3; 0.5, 0.3, 0.4)$ , and threshold value and time delay are 2 and 1 respectively, for different sample sizes. The initial values for parameters were obtained by the ARIMA (1,1,1) fit to the data. Results are based on 1000 simulations.

sizes	Maximum Likelihood Estimates						Least Square Estimates						AAE	
	$d_1$	$a_1$	$b_1$	$d_2$	$a_2$	$b_2$	$d_1$	$a_1$	$b_1$	$d_2$	$a_2$	$b_2$	MLE	LSE
2500	0.4100 (0.0043)	0.4954 (0.0019)	0.2996 (0.0013)	0.5154 (0.0439)	0.2949 (0.0056)	0.3981 (0.0030)	0.4045 (0.0134)	0.5097 (0.0127)	0.1285 (0.0298)	1.9883 (2.5024)	0.2590 (0.0233)	0.2283 (0.0319)	0.0062	0.3144
3000	0.4077 (0.0038)	0.4960 (0.0017)	0.3000 (0.0011)	0.5146 (0.0362)	0.2947 (0.0045)	0.3985 (0.0025)	0.3982 (0.0106)	0.5142 (0.0102)	0.1293 (0.0295)	1.9789 (2.4516)	0.25973 (0.0213)	0.2302 (0.0308)	0.0055	0.3126
5000	0.4066 (0.0024)	0.4965 (0.0010)	0.3002 (0.0006)	0.5159 (0.0186)	0.2957 (0.0025)	0.3983 (0.0013)	0.3915 (0.0061)	0.5213 (0.0064)	0.1286 (0.0297)	1.9681 (2.2758)	0.2632 (0.0118)	0.2295 (0.0295)	0.0053	0.3127
10000	0.4028 (0.0010)	0.4985 (0.0004)	0.3001 (0.0003)	0.5017 (0.0109)	0.2991 (0.0013)	0.3997 (0.0007)	0.3822 (0.0031)	0.5294 (0.0036)	0.1285 (0.0295)	1.9235 (2.0954)	0.2719 (0.0059)	0.2327 (0.0285)	0.0011	0.3062

ation (SD) of the simulated estimators (MLE) and obtained by the asymptotic theory. The SD is defined as  $SD(\hat{\theta}) = \sqrt{\frac{1}{N} \sum (\hat{\theta} - \theta)^2}$ .

Table 3.2: Comparison of standard errors for model (3.1) with  $(d_1, a_1, b_1; d_2, a_2, b_2) = (0.4, 0.5, 0.3; 0.5, 0.3, 0.4)$ , for different sample sizes. Threshold value is 2. Results are based on 1000 simulations.

Sizes		Simulated standard error						Standard error from $G(\theta_0)$					
$n$	$d_1$	$a_1$	$b_1$	$d_2$	$a_2$	$b_2$	$d_1$	$a_1$	$b_1$	$d_2$	$a_2$	$b_2$	
2500	0.0663	0.0443	0.0364	0.2096	0.0754	0.0554	0.0647	0.0434	0.0359	0.1995	0.0708	0.0531	
3000	0.0570	0.0380	0.0343	0.1876	0.0657	0.0502	0.0592	0.0396	0.0326	0.1853	0.0656	0.0492	
5000	0.0473	0.0316	0.0252	0.1510	0.0523	0.0387	0.0459	0.0308	0.0252	0.1431	0.0507	0.0380	
10000	0.0321	0.0219	0.0174	0.1046	0.0373	0.0271	0.03250	0.0218	0.0178	0.1008	0.0357	0.0268	

The results show that the square error obtained by simulation decreases as the sample size increases and similarly for the standard error obtained after computing the inverse of matrix  $G$ . In all cases considered, the standard errors obtained after calculating the matrix inverse  $G$  are smaller than the standard error based on the simulations.

Tables 3.3 shows maximum likelihood estimators with their mean square errors (in parentheses), and average absolute errors of two different nonlinear models for different sample sizes. The results have been calculated by using the simulation output. The table reports that the values of MSE and AAE decrease when the sample size increases in both nonlinear models. The starting values for parameters are obtained by a uniform distribution as follows  $d_1 \sim \text{unif}(0.1, \text{mean}(\text{data}))$ ,  $a_1 \sim \text{unif}(0.1, 0.9)$ ,  $b_1 \sim \text{unif}(0.1, 0.9)$ ,  $d_2 \sim \text{unif}(0.1, \text{mean}(\text{data}))$ ,  $a_2 \sim \text{unif}(0.1, 0.9)$  and  $b_2 \sim \text{unif}(0.1, 0.9)$ . Tables 3.3 shows that the MLE estimates become more efficient with smaller MSEs and more closed to the true parameters  $\theta_0 = (d_1, a_1, b_1; d_2, a_2, b_2)$  as well as the AAE of estimates decreases gradually when the sample sizes increases from 200 to 10000 in both nonlinear models.

Table 3.4 compares the standard errors of the estimators obtained by the standard deviation of the simulated estimators and by the standard errors obtained after calculating the matrix inverse  $G$ . Clearly, in both nonlinear models, the

Table 3.3: Maximum likelihood estimates with their mean square errors (in parentheses), and the average absolute error of estimates for two different nonlinear model and different threshold value and time delay for different sample sizes.

Parameters		Sample sizes	Maximum likelihood estimators (MLE)						Average absolute errors
Regime 1	Regime 2		$d_1$	$a_1$	$b_1$	$d_2$	$a_2$	$b_2$	AAE
$d_1=0.4$	$d_2=0.5$	200	0.5617	0.4212	0.2973	0.7920	0.2061	0.3718	0.2778
$a_1=0.5$	$a_2=0.3$		(0.1611)	(0.0513)	(0.0214)	(0.9627)	(0.1261)	(0.0478)	
$b_1=0.3$	$b_2=0.4$	500	0.4476	0.4795	0.2953	0.5986	0.2673	0.3907	
	$r=2$		(0.0277)	(0.0114)	(0.0068)	(0.2548)	(0.0317)	(0.0157)	0.1502
	$d=1$	1000	0.4311	0.485	0.2991	0.5255	0.2858	0.3999	
			(0.0127)	(0.0052)	(0.0034)	(0.1094)	(0.0147)	(0.0077)	
		5000	0.4053	0.4979	0.2993	0.5169	0.2933	0.3996	0.0455
			(0.0023)	(0.0010)	(0.0007)	(0.0215)	(0.0028)	(0.0014)	
		10000	0.4023	0.4988	0.3005	0.5052	0.2991	0.3989	
			(0.0010)	(0.0004)	(0.0003)	(0.0105)	(0.0013)	(0.0008)	0.0316
$d_1=0.6$	$d_2=0.3$	200	1.1025	0.5683	0.1730	0.4514	0.3754	0.4874	0.2742
$a_1=0.7$	$a_2=0.4$		(0.9628)	(0.1151)	(0.0158)	(0.5814)	(0.0381)	(0.0097)	
$b_1=0.2$	$b_2=0.5$	500	0.8835	0.6129	0.1968	0.3680	0.3881	0.4963	
	$r=3$		(0.2963)	(0.0351)	(0.0045)	(0.1393)	(0.0102)	(0.0033)	0.1590
	$d=2$	1000	0.7497	0.6551	0.1986	0.3448	0.3898	0.4998	
			(0.1158)	(0.0138)	(0.0023)	(0.0565)	(0.0041)	(0.0015)	
		5000	0.6456	0.6862	0.1992	0.3058	0.3996	0.4990	0.0478
			(0.0213)	(0.0026)	(0.0004)	(0.0109)	(0.0007)	(0.0003)	
		10000	0.6241	0.6924	0.1998	0.3055	0.3986	0.5000	
			(0.0106)	(0.0013)	(0.0002)	(0.0053)	(0.0004)	(0.0001)	0.0341

NOTE: The starting values for parameters follow uniform distributions. Results are based on 1000 simulations.

result shows that the standard errors based on the simulations are larger than the ones obtained after calculating and inverting the matrix  $G$ . Both measures decrease gradually as the sample size increases.

Figure 3.1 shows both histograms and QQ-plots of the sampling distribution of the estimators for the first model (3.1) when there are 5000 and for 1000 simulation runs. The QQ-plots support the claim of asymptotic normality of the parameters  $\theta = (d_1, a_1, b_1; b_2, a_2, b_2)$ . The figures show that the sampling distribution of the estimator of the intercept is slightly skewed (the right tail). It has less accuracy compared to the autoregressive coefficients. These results are in agreement with those reported by Fokianos et al. (2009) for the linear model. In addition, the figures indicate that the histograms look like the normal distribution.



Table 3.4: Comparison of standard errors for two nonlinear models with different thresholds and time delays for different sample sizes. Result based on 1000 simulations.

Parameters		Sizes	Simulated standard error						Standard error from $G(\theta_0)$					
Regime 1	Regime 2	$n$	$d_1$	$a_1$	$b_1$	$d_2$	$a_2$	$b_2$	$d_1$	$a_1$	$b_1$	$d_2$	$a_2$	$b_2$
$d_1=0.4$	$d_2=0.5$	200	0.4014	0.2264	0.1464	0.9812	0.3551	0.2185	0.2212	0.1460	0.1284	0.7977	0.2709	0.2084
$a_1=0.5$	$a_2=0.3$	500	0.1664	0.1068	0.0826	0.5048	0.1780	0.1254	0.1431	0.0955	0.0803	0.4706	0.1646	0.1246
$b_1=0.3$	$b_2=0.4$	1000	0.1126	0.0718	0.0580	0.3307	0.1214	0.0876	0.1021	0.0682	0.0566	0.3246	0.1144	0.0862
	$r=2$	5000	0.0476	0.0318	0.0259	0.1466	0.0527	0.0371	0.0460	0.0308	0.0253	0.1430	0.0506	0.0380
	$d=1$	10000	0.0322	0.0212	0.0173	0.1026	0.0362	0.0277	0.0325	0.0218	0.0179	0.1009	0.0358	0.0269
$d_1=0.6$	$d_2=0.3$	200	1.2419	0.2890	0.1320	1.1365	0.1759	0.1630	0.3851	0.1454	0.1023	0.5785	0.1491	0.08701
$a_1=0.7$	$a_2=0.4$	500	0.5443	0.1874	0.0685	0.3733	0.1008	0.0574	0.3136	0.1130	0.0634	0.3375	0.0882	0.0535
$b_1=0.2$	$b_2=0.5$	1000	0.3403	0.1175	0.0477	0.2377	0.0644	0.0386	0.2530	0.0900	0.0449	0.2321	0.0608	0.0375
	$r=3$	5000	0.1461	0.0513	0.0209	0.1046	0.0266	0.0165	0.1316	0.0462	0.0200	0.1014	0.0267	0.0167
	$d=2$	10000	0.1029	0.0359	0.0144	0.0727	0.0193	0.0119	0.0953	0.0334	0.0142	0.0716	0.0188	0.0118

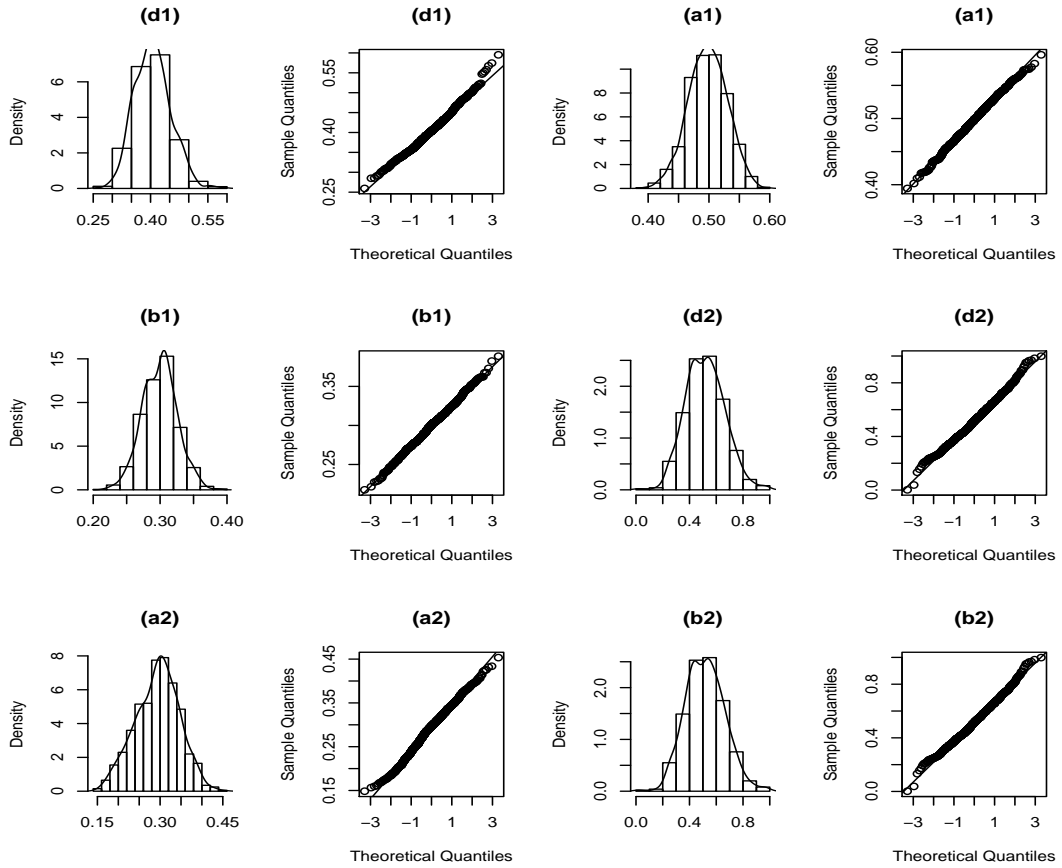


Figure 3.1: Histograms and QQ-plots of the sampling distribution of the estimators of  $\hat{\theta} = (\hat{d}_1, \hat{a}_1, \hat{b}_1; \hat{d}_2, \hat{a}_2, \hat{b}_2)$  for model (3.1), when the true values are  $(0.4, 0.5, 0.3; 0.5, 0.3, 0.4)$ . The initial values follow a special uniform distribution. The results are based on 5000 data points and 1000 simulations.

Figure 3.2 illustrates the boxplot of the average absolute errors of both estimators of the first model for different sample sizes. The figure reports the comparison between the average absolute errors of the maximum likelihood estimators and the least-square estimators. The plot reports in all cases considered, that the maximum likelihood estimators have lower average absolute errors than the least-squares estimators and the average absolute errors of both estimators decrease as sample size increases. The average of the outliers of LSE is far larger than that of MLE in all sample size cases. In general, MLE estimation is much more accurate than that of LSE.

Figure 3.3 provides the boxplot of the average absolute errors of the maximum

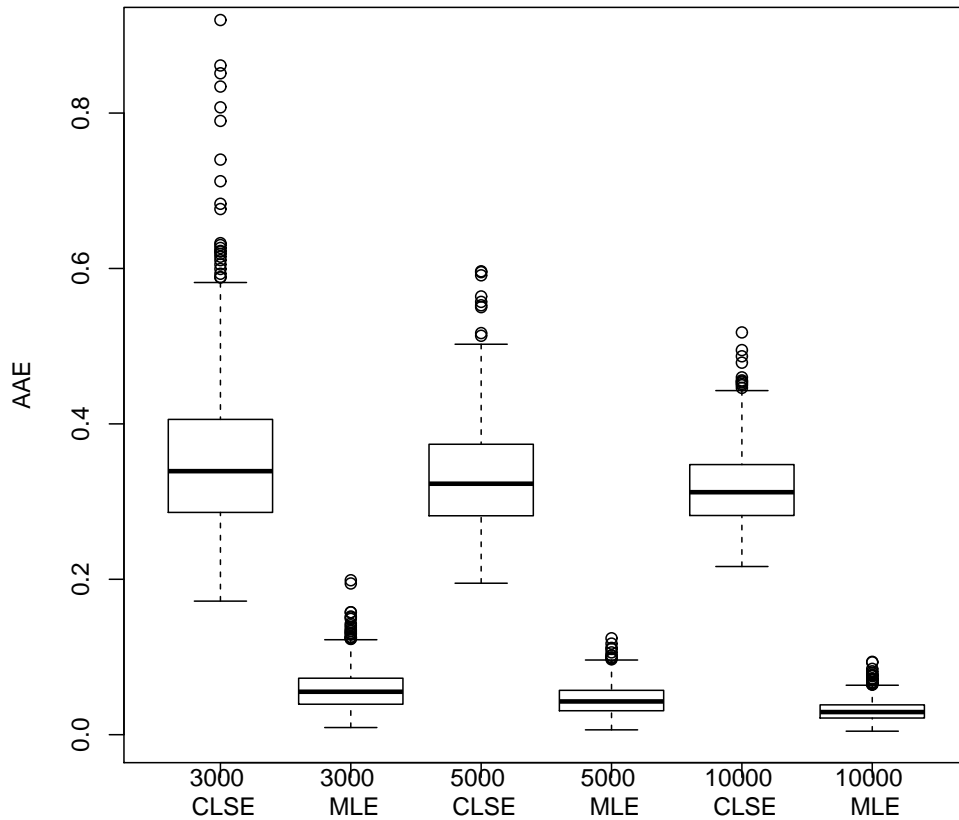


Figure 3.2: Boxplots of average absolute error (AAE) for maximum likelihood estimator (bottom boxplots) and the conditional least square estimator (top boxplots) for different sample sizes. The initial values of parameters of the MLE follow ARIMA (1,1) fit data model. The plots are based on 1000 simulations.

likelihood estimator of model (3.1) when  $(d_1, a_1, b_1; d_2, a_2, b_2) = (0.4, 0.5, 0.3; 0.5, 0.3, 0.4)$  for different sample sizes. Figure 3.3 shows that the larger the sample size is, the better the estimation of the MLE, because the smaller the AAE of estimation is. There are outliers in all five different sample size cases and the smaller the sample size is, the larger the range of values of the outliers is. Overall, the plot shows the boxplot size decreases gradually as the sample size increases in both estimators.

Figure 3.4 presents the boxplot of the average absolute errors of the maximum likelihood estimator for model (3.1) when  $(d_1, a_1, b_1; d_2, a_2, b_2) = (0.6, 0.7, 0.2; 0.3, 0.4, 0.5)$ , and threshold value and time delay are 3 and 2 respectively for different sample sizes. It reports that the values of AAE reduce gradually when the sample

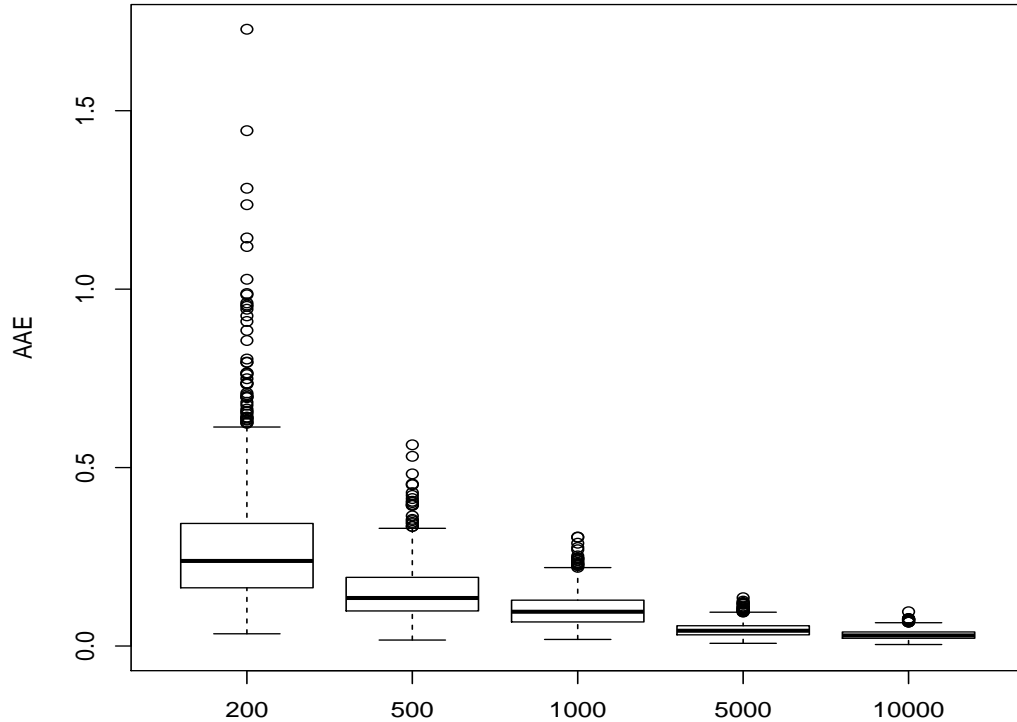


Figure 3.3: Boxplot of average absolute errors (AAE) for maximum likelihood estimator for model (3.1) when  $(d_1, a_1, b_1; d_2, a_2, b_2) = (0.4, 0.5, 0.3; 0.5, 0.3, 0.4)$ , and threshold value and time delay are 2 and 1 respectively for different sample sizes. The initial values of parameters follow uniform distributions. The plots are based on 1000 simulations.

size increases. From the plot, it can be seen that the range of outliers is larger for small sample size but this range gradually becomes smaller as the sample size increases. Figure 3.5 illustrates the histograms and QQ-plots for the sampling distribution of the standardized maximum likelihood estimators of model (3.1) when  $(d_1, a_1, b_1; d_2, a_2, b_2) = (0.6, 0.7, 0.2; 0.3, 0.4, 0.5)$ , and threshold value and time delay are 3 and 2 respectively for 5000 data points and 1000 simulations.

Table 3.5 reports the mean square Pearson residuals (MSPR) which is defined by  $\sum_{t=1}^N \frac{e_t^2}{N-p}$ , where  $e_t = \frac{Y_t - \lambda_t}{\sqrt{\lambda_t}}$  and  $p$  denotes the number of estimated parameters for two different models (Kedem and Fokianos (2002)). The MSPR of both models

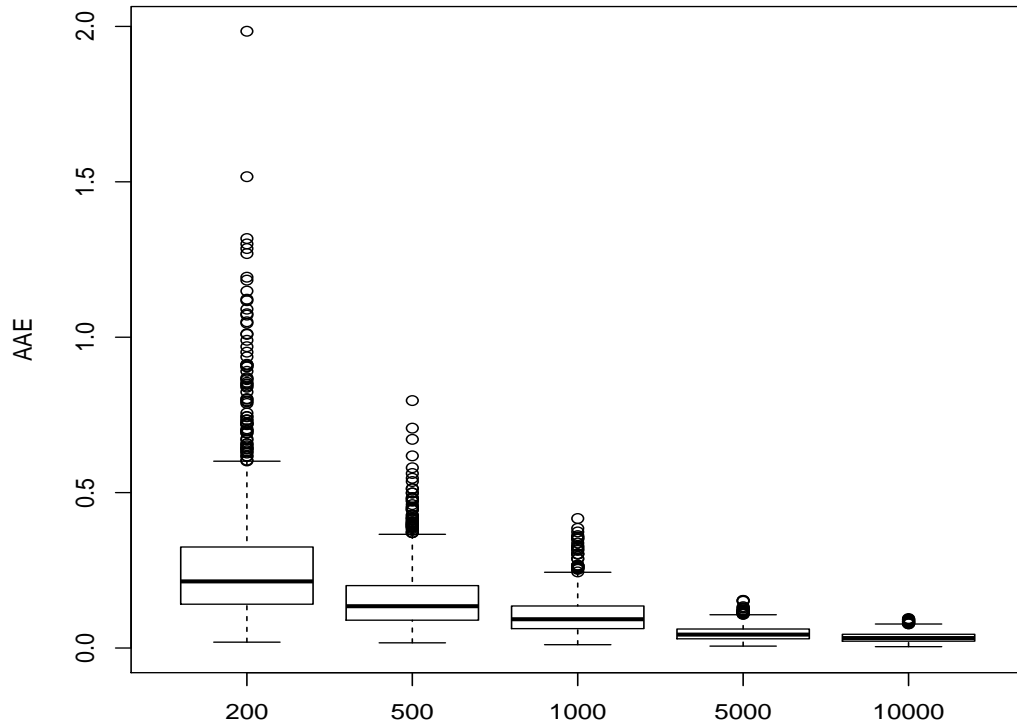


Figure 3.4: Boxplot of average absolute errors (AAE) for maximum likelihood estimator for model (3.1) when  $(d_1, a_1, b_1; d_2, a_2, b_2) = (0.6, 0.7, 0.2; 0.3, 0.4, 0.5)$ , and threshold value and time delay are 3 and 2 respectively for different sample sizes. The initial values of parameters follow uniform distributions. The plots are based on 1000 simulations.

was obtained by averaging out the results from all simulations. In all cases, the values of MSPR of both models decrease gradually when the sample sizes increase. This means that the models become more adequate with larger sample sizes. Figure 3.6 illustrates boxplot of MSPR for two nonlinear models with different sample sizes. The plot shows that the MSPR of both models clearly reduces gradually as the sample size increases from 200 to 2000.

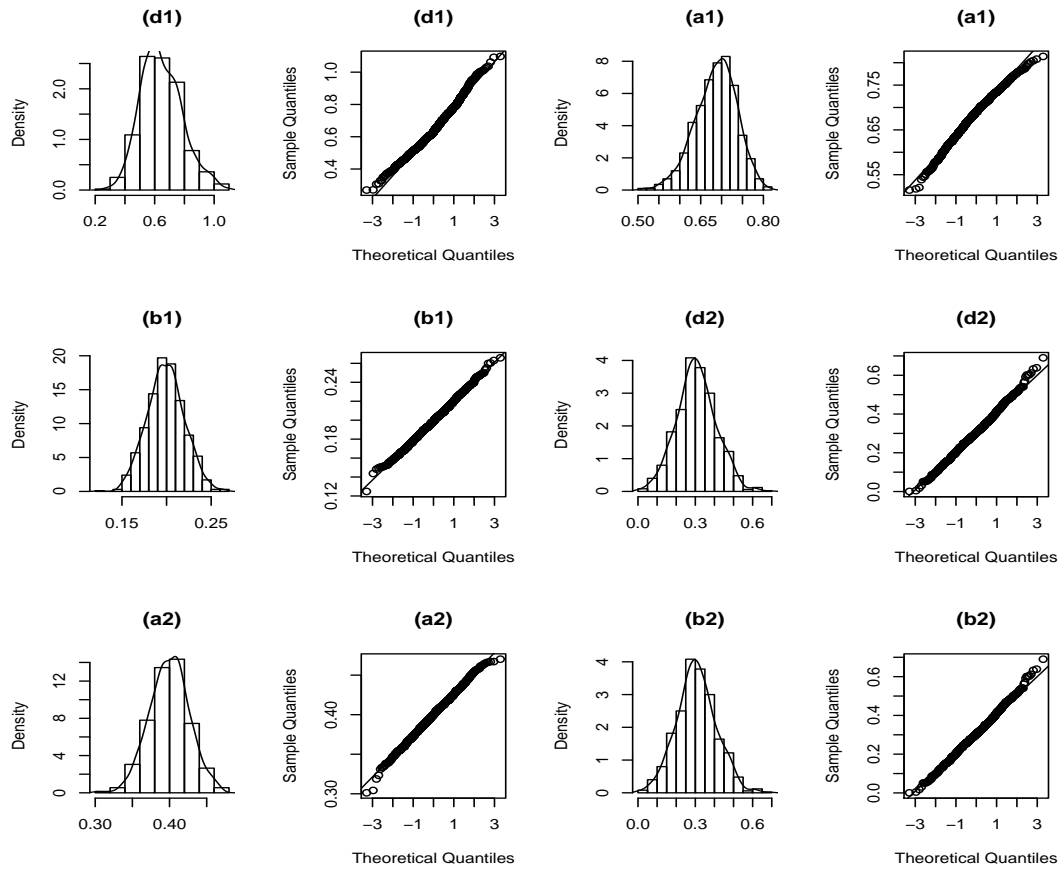


Figure 3.5: Histograms and QQ-plots of the sampling distribution of the estimators of  $\hat{\theta} = (\hat{d}_1, \hat{a}_1, \hat{b}_1; \hat{d}_2, \hat{a}_2, \hat{b}_2)$  for model (3.1), when the true value are  $(0.6, 0.7, 0.2; 0.3, 0.4, 0.5)$ . The initial values follow a uniform distribution. The results are based on 5000 data points and 1000 simulations.

Table 3.5: The MSPR calculated for different models. The first model was generated with  $d_1 = 0.4, a_1 = 0.5, b_1 = 0.3; d_2 = 0.5, a_2 = 0.3, b_2 = 0.4; r = 2$  and  $d = 1$ , while the second model was generated with  $d_1 = 0.6, a_1 = 0.7, b_1 = 0.2; d_2 = 0.3, a_2 = 0.4, b_2 = 0.5; r = 3$  and  $d = 2$ . The results are based on 1000 simulations.

Sample Sizes	Average of MSPR	
	First model	Second model
200	0.00515	0.00517
500	0.00203	0.00202
1000	0.00101	0.00100
2000	0.00050	0.00050

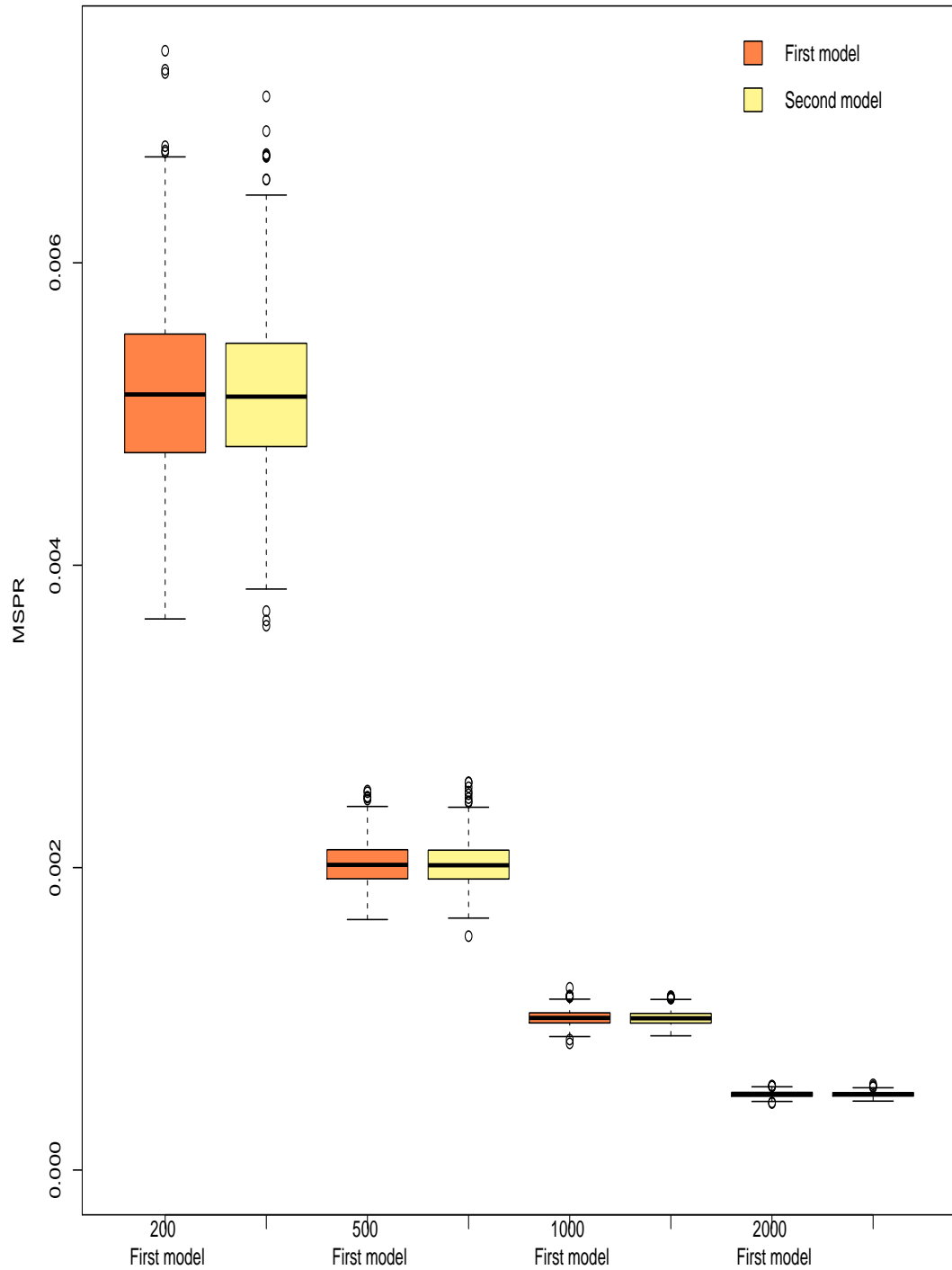


Figure 3.6: Boxplot of MSPR for different models. The first model generated with  $(d_1 = 0.4, a_1 = 0.5, b_1 = 0.3; d_2 = 0.5, a_2 = 0.3, b_2 = 0.4; r = 2$  and  $d = 1)$ , while the second model generated with  $(d_1 = 0.6, a_1 = 0.7, b_1 = 0.2; d_2 = 0.3, a_2 = 0.4, b_2 = 0.5; r = 3$  and  $d = 2)$ . The results are based on 1000 simulations.

### 3.1.2 Determination of the threshold value $r$

In this part, we present a few simulations to determine the threshold value when the delay parameter  $d$  and the number of regimes  $s$  are given. To estimate the threshold value, we arrange the original data in increasing order, then use a lower bound of at least 20% – 25% of the total number of observations in each state distribution (regime). Let the corresponding data in the arranged series be the threshold candidate value  $\{r_c\}$ , for each of which the  $AIC(r_c, d)$  is calculated, then

$$AIC(\hat{r}_c, d) = \min_{\{r_c\}} AIC(r_c, d)$$

and the corresponding  $\hat{r}$  is the threshold value ( $r$ ).

Table 3.6 displays the AIC, BIC and HQIC values of the estimated threshold parameter  $r$  for four different PTGARCH (1,1;2,1) models and three different sample sizes. In the first and the second model, the table shows the information criteria is smallest when  $\hat{r} = 2$  at a sample size of 2000, so we estimate the threshold value to be 2. This is true also for sample sizes 5000 and 10000, where the AIC, BIC and HQIC are smallest when  $\hat{r} = 2$ , therefore the corresponding  $\hat{r} = 2$ . Similar results, for  $\hat{r} = 3$ , are obtained for the rest of the models in all different sample sizes. Table 3.6 reports the estimate of the threshold parameter obtained by averaging out the results of the AIC, BIC and HQIC for all simulations. From the results in Table 3.6, it can be seen that the maximum of possible threshold values  $r$  for the first three models is 4 for all sample sizes, whereas it is 6 for the sample size 2000 and 5 for the sample sizes 5000 and 10000 in the last model respectively.



Table 3.6: The average of AIC, BIC and HQIC values of model (3.1) for four different PTGARCH (1,1;2,1) models and three different sample sizes.

Regime 1	Regime 2	Sample size	Possible values	AIC ( $\hat{r}, 1$ )	BIC ( $\hat{r}, 1$ )	HQIC ( $\hat{r}, 1$ )	Percentage of correct choices	Estimate of ( $\hat{r}$ )	
$d_1=0.5$ $a_1=0.4$ $b_1=0.3$	$d_2=0.4$ $a_2=0.5$ $b_2=0.3$	2000	1	1.2939	1.3107	1.2940	14.9	2	
			2	<b>1.2920</b>	<b>1.3088</b>	<b>1.2921</b>	<b>64</b>	<b>2</b>	
			3	1.2933	1.3101	1.2934	20.2	2	
			4	1.2968	1.3136	1.2969	0.9	2	
		$r = 2$	5000	1	1.2894	1.2972	1.2895	5.1	2
				2	<b>1.2874</b>	<b>1.2952</b>	<b>1.2876</b>	<b>82.8</b>	<b>2</b>
				3	1.2887	1.2965	1.2889	12	2
				4	1.2853	1.2931	1.2954	0.1	2
			10000	1	1.2872	1.2915	1.2873	0.9	2
				2	<b>1.2853</b>	<b>1.2896</b>	<b>1.2854</b>	<b>94.7</b>	<b>2</b>
				3	1.2866	1.2909	1.2867	4.3	2
				4	1.2863	1.2907	1.2865	0.1	2
$d_1=0.4$ $a_1=0.5$ $b_1=0.3$	$d_2=0.5$ $a_2=0.3$ $b_2=0.4$	2000	1	0.9774	0.9942	0.9775	8.4	2	
			2	<b>0.9734</b>	<b>0.9902</b>	<b>0.9735</b>	<b>72.3</b>	<b>2</b>	
			3	0.9752	0.9920	0.9753	12.2	2	
			4	0.9754	0.9922	0.9755	7.1	2	
		$r = 2$	5000	1	0.9733	0.9811	0.9735	3.1	2
				2	<b>0.9695</b>	<b>0.9773</b>	<b>0.9696</b>	<b>89.5</b>	<b>2</b>
				3	0.97125	0.9790	0.9714	6.1	2
				4	0.9723	0.9802	0.9725	1.3	2
			10000	1	0.9696	0.9739	0.9697	0.2	2
				2	<b>0.9657</b>	<b>0.9700</b>	<b>0.9658</b>	<b>98.6</b>	<b>2</b>
				3	0.9674	0.9718	0.9676	1.0	2
				4	0.9685	0.9728	0.9686	0.2	2
$d_1=0.4$ $a_1=0.5$ $b_1=0.3$	$d_2=0.5$ $a_2=0.4$ $b_2=0.3$	2000	1	1.1861	1.2029	1.1862	6.2	3	
			2	1.1831	1.2000	1.1832	16	3	
			3	<b>1.1816</b>	<b>1.1984</b>	<b>1.1817</b>	<b>54.7</b>	<b>3</b>	
			4	1.1836	1.2004	1.1837	23.1	3	
		$r = 3$	5000	1	1.1776	1.1854	1.1777	0.2	3
				2	1.1744	0.1.1822	0.1.1746	7.7	3
				3	<b>1.1728</b>	<b>1.1806</b>	<b>1.1730</b>	<b>81.1</b>	<b>3</b>
				4	1.1749	1.1827	1.1751	11	3
			10000	1	1.1763	1.1806	1.1764	0.0	3
				2	1.1732	0.1.1775	0.1.1733	1.9	3
				3	<b>1.1716</b>	<b>1.1759</b>	<b>1.1717</b>	<b>92.9</b>	<b>3</b>
				4	1.1723	1.1767	1.1725	5.2	3
$d_1=0.9$ $a_1=0.5$ $b_1=0.2$	$d_2=0.5$ $a_2=0.3$ $b_2=0.4$	2000	2	-0.5823	-0.5654	-0.5822	0.1	3	
			3	<b>-0.6003</b>	<b>-0.5835</b>	<b>-0.6002</b>	<b>93.2</b>	<b>3</b>	
			4	-0.5961	-0.5793	-0.5960	4	3	
			5	-0.5953	-0.5785	-0.5952	2.5	3	
			6	-0.5804	-0.5636	-0.5803	0.2	3	
			3	<b>-0.6070</b>	<b>-0.5992</b>	<b>-0.6069</b>	<b>99</b>	<b>3</b>	
		$r = 3$	5000	4	-0.9029	-0.5828	-0.6027	0.9	3
				5	-0.6020	-0.5942	-0.6019	0.1	3
				1	-0.5331	-0.5288	-0.5330	0.0	3
			10000	2	-0.5871	-0.5828	-0.5870	0.0	3
				3	<b>-0.6051</b>	<b>-0.6008</b>	<b>-0.6050</b>	<b>100</b>	<b>3</b>
				4	-0.6010	-0.5967	-0.6009	0.0	3
5	-0.6002	-0.5959	-0.6001	0.0	3				

NOTE: The fourth column shows the possible values of threshold parameters. The fifth, sixth and seventh column report the averages of AIC, BIC and HQIC respectively. The eighth column represents the percentages of correct choices of corresponding threshold values to those in the fourth column. Simulation results are given for four different PTGARCH (1,1;2,1) models and different threshold values. Correct threshold value detection percentages are printed bold. Results are based on 1000 simulations.

Table 3.7 reports the relative rates of correct choices of the threshold parameter for different PTGARCH models, as well as the average of information criteria for model (3.1). Obviously the results report that the relative rate of correct choices

Table 3.7: The percentage of correct choices and average of AIC, BIC and HQIC values for four PTGARCH models (3.1) and for three different sample sizes.

Regime 1	Regime 2	Sample sizes	Percentage of correct choices	Average of $AIC(\hat{r}, 1)$	Average of $BIC(\hat{r}, 1)$	Average of $HQIC(\hat{r}, 1)$
$d_1=0.5$	$d_2=0.4$	2000	64	1.2920	1.3088	1.2921
$a_1=0.4$	$a_2=0.5$	5000	82.8	1.2874	1.2952	1.2876
$b_1=0.3$	$b_2=0.3$	10000	94.7	1.2853	1.2896	1.2854
$d_1=0.4$	$d_2=0.5$	2000	72.3	0.9734	0.9902	0.9735
$a_1=0.5$	$a_2=0.3$	5000	89.5	0.9695	0.9773	0.9696
$b_1=0.4$	$b_2=0.4$	10000	98.6	0.9657	0.9700	0.9658
$d_1=0.4$	$d_2=0.5$	2000	54.7	1.1816	1.1984	1.1817
$a_1=0.5$	$a_2=0.4$	5000	81.1	1.1728	1.1806	1.1730
$b_1=0.3$	$b_2=0.3$	10000	92.9	1.1716	1.1759	1.1717
$d_1=0.9$	$d_2=0.5$	2000	93.2	-0.6003	-0.5835	-0.6002
$a_1=0.5$	$a_2=0.3$	5000	99	-0.6070	-0.5992	-0.6069
$b_1=0.2$	$b_2=0.4$	10000	100	-0.6051	-0.6008	-0.6050

NOTE: The fourth column reports the relative rates of correct choices of threshold value for model (3.1). The threshold value of the first and the second model is 2 while the value of threshold in the third and fourth model were 3. Results are based on 1000 simulations.

increases as the sample size increases from 2000 to 10000 for all different models. In addition, the table shows that the average of the information criteria decreases as the sample size increases.

### 3.1.3 Test for threshold nonlinearity (Lagrange Multiplier Test)

In the previous chapter, we mentioned the Lagrange Multiplier (LM) and Likelihood Ratio (LR) test as a test for threshold nonlinearity. Here we do some simulations to support the threshold nonlinearity test using the  $LM^*$  test which is defined by (2.14) and (2.15).

Table 3.8 shows a few simulation results examining the adequacy of the  $LM^*$  test for different sample sizes when  $(d_1, a_1, b_1; d_2, a_2, b_2) = (0.4, 0.5, 0.3; 0.5, 0.3, 0.4)$  and the value of the threshold is 2. The outcomes are given for the data generated with a PTGARCH (1,1;2,1) process. In addition, the values of max  $LM$  and *ave LM* test for a PTGARCH (1,1;2,1) model with their  $p$ -values (at 95% level of significance) for the  $\chi_k^2$  test which are obtained using  $1 - pchisq(LM^*, 3)$  can be seen. This test is an one-tail (right-tail) test with 5% level of significance. From the outcomes in Table 3.8, it can be seen that the values of max  $LM$  are larger than the critical value of 7.815. In other words, the  $p$ -values are less than the significant level for different sample sizes. That is, there is clearly significant benefit at significant level  $\alpha = 0.05$  in using the nonlinear model for all sample size cases. In contrast, the values of *ave LM* for the sample sizes (500 and 1000) indicate that there is no evidence for us to reject the null hypothesis, but when the sample size increases, there is significant improvement at the 5% level of significance in using the nonlinear model. In general, the values of  $LM^*$  improve gradually as the sample size increases.

Table 3.9 contains the results for the data that were generated with a specific PTGARCH (1,1;2,1) model. This model has specific parameters which were fixed across the regimes. That is, the specific model has the same autoregressive coefficients and the intercept in both regimes. Table 3.9 reports that the  $p$ -values of tests are greater than the level of significance  $\alpha = 0.05$  for all sample sizes. Therefore, the decision is to accept the null hypothesis. In other words, there is

Table 3.8: Achieved significance level of the  $LM^*$  test for different sample sizes. The data were generated by PTGARCH (1,1;2,1) process.

Sample sizes		$LM^*$		
$n$	max $LM$	$p$ -value	ave $LM$	$p$ -value
500	8.32*	0.039	6.75	0.0801
1000	9.19*	0.026	7.21	0.065
1500	10.25*	0.014	8.02*	0.045
2000	11.54*	0.009	8.53*	0.036
2500	12.46*	0.005	9.07*	0.028
5000	18.74*	0.0003	12.28*	0.006

NOTE: The parameters used are  $(\theta_{10}; \theta_{20}) = (0.4, 0.5, 0.3; 0.5, 0.3, 0.4)$ , and the LM denotes the Lagrange multiplier test. \* denotes reject  $H_0$  with significant level 0.05. Results are based on 1000 simulations.

not sufficient evidence to reject the null hypothesis at the  $\alpha = 0.05$  level in using the linear model.

Table 3.9: Achieved significance level of the  $LM^*$  test for different sample sizes. The data were generated by PTGARCH (1,1;2,1) process.

Sample sizes		$LM^*$		
$n$	max $LM$	$p$ -value	ave $LM$	$p$ -value
500	7.17	0.066	6.20	0.102
1000	6.77	0.079	5.93	0.115
1500	6.87	0.076	6.01	0.111
2000	6.90	0.075	6.04	0.109
2500	6.99	0.071	6.075	0.107
5000	6.68	0.083	5.81	0.121

NOTE: The parameters used are  $(\theta_{10}; \theta_{20}) = (0.4, 0.5, 0.3; 0.4, 0.5, 0.3)$ . The third column reports the  $p$ -values of the  $\chi^2$  test statistic. Results are based on 1000 simulations.

### 3.1.4 A real data example

To apply our methodology to model (3.1), we use as data the number of transactions per minute for the stock Ericsson  $B$ , in the period 2 – 3 July, 2002. There are approximately 7000 observations in the data set which includes all of the transactions per minutes for the stock Ericsson B for the period 2-22 July, 2002. For the purpose of the analysis in the thesis, only the first 650 observations of the original data set, conveying around 11 hours of transactions were used. Note that the first and the last minute were neglected. The sample mean of the particular realization was 10.08, with a standard deviation of 5.62072. Figure 3.7 presents the observation numbers, the histogram and the sample autocorrelation function.

From Figure 3.7, it can be seen that the time series does not exhibit any trending behaviour. The plot of the sample autocorrelation function of the number of transactions suggests significant serial correlation between transactions.

To model these data, we considered the model (3.1). Table 3.10 reports possible threshold values ( $r$ ) and their information criteria (AIC,BIC and HQIC) for the PTGARCH (1,1,2,1) model fitted to the number of transactions when the number of regimes ( $s=2$ ) and the time delay ( $d=1$ ) are given. From Table 3.10, it can be

Table 3.10: Information criteria for the PTGARCH (1,1,2,1) model fitted to the number of transactions.

Possible threshold values ( $r$ )	Information criteria		
	<i>AIC</i>	<i>BIC</i>	<i>HQIC</i>
5	-27.0960	-27.0547	-27.0800
6	-27.0728	-27.0315	-27.0568
7	-27.1124	-27.0711	-27.0964
8	-27.1184	-27.0771	-27.1024
9	-27.1177	-27.0764	-27.1017
10	-27.1085	-27.0672	-27.0925
11	-27.1118	-27.0705	-27.0958
12	-24.7331	-24.6918	-24.7171
13	-27.0742	-27.0329	-27.0582

seen that the smallest values of information criteria are reached when the possible threshold value is 8. Therefore, the estimated threshold value is 8.

The Table 3.11 shows the maximum likelihood estimates of parameters ( $d_1, a_1, b_1; d_2, a_2, b_2$ ) and the standard error of the estimators obtained by inversion of the information matrix (see Theorem 4) .

Table 3.11: The estimated parameters for model (3.1) and their standard errors.

Model	First regime, $n_1 = 240$			Second regime, $n_2 = 410$		
Parameters	$d_1$	$a_1$	$b_1$	$d_2$	$a_2$	$b_2$
Estimated value	0.3914	0.8915	0.0123	0.5157	0.7721	0.1750
Standard error	0.2217	0.0278	0.0480	0.2621	0.0276	0.0260

The ave LM test of the model is 33.5, so there is clearly significant benefit at the 5% level of significance in using the nonlinear model. In addition, the mean square of the Pearson residuals (MSPR) for nonlinear models, is equal to 2.3295 which

is slightly smaller than the corresponding mean square of the Pearson residuals obtained from the linear model fit (2.3523) (Kedem and Fokianos (2002)). The estimated model can be expressed as

$$\hat{\lambda}_t = \begin{cases} 0.3914 + 0.01234Y_{t-1} + 0.8915\hat{\lambda}_{t-1}, & \text{if } Y_{t-1} \leq 8, \\ 0.5157 + 0.1750Y_{t-1} + 0.7721\hat{\lambda}_{t-1}, & \text{if } Y_{t-1} > 8. \end{cases}$$

The plot 3.8 shows the sequence of residuals and their cumulative periodogram plot, as well as their autocorrelation function of the transaction data. The cumulative periodogram plot obviously indicates that the Pearson residuals obtained after the application of the nonlinear model ( $s = 2, d = 1$ ) do not deviate from a white noise sequence. The plot of the autocorrelation function reveals no significant serial correlation in the sequence of residuals.

## 3.2 Summary

In this chapter, we presented our methodology on how to estimate the model (3.1) and how to detect a nonlinearity and the value of threshold as well. We carried out a few simulation results to estimate the parameters of the model (3.1) using MLE and LSE, and used the average of the AIC criterion to determine the threshold value. The MLE estimator is more efficient than the CLSE estimator. Simulated standard error was compared with standard error defined by (2.10). Testing the threshold nonlinearity is investigated using  $LM^*$  defined by (2.14) and (2.15). Boxplot of average absolute errors (AAE) for both estimators MLE and LSE are illustrated. In addition, histograms and QQ-plots of the sampling distribution of the MLE are presented. As a real data example, we applied the model (3.1) to the number of transactions per minute for stock Ericsson B when the delay parameter  $d$  and the number of regimes ( $s$ ) are given. The simple sufficient condition for the geometric ergodicity is taken into account.

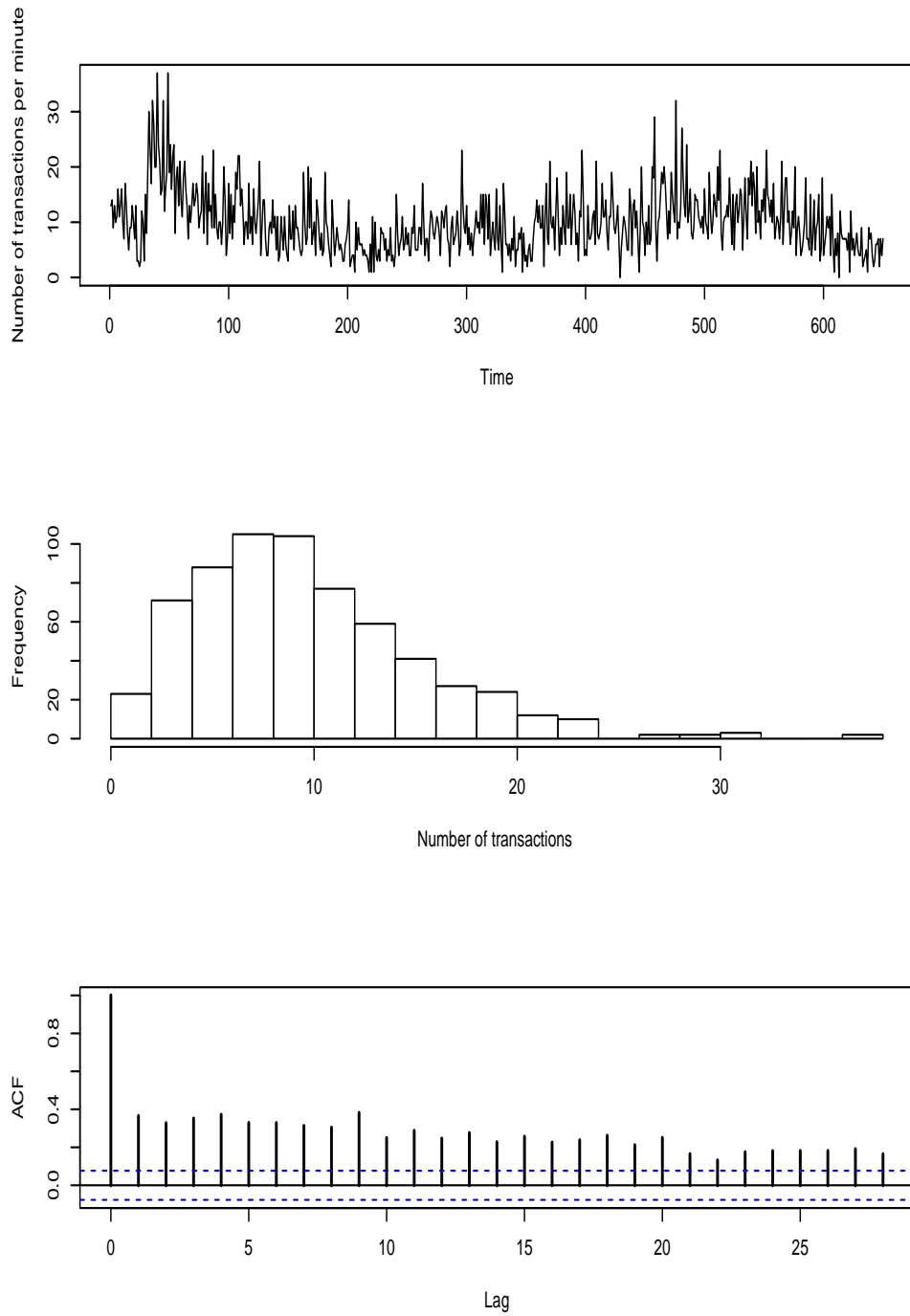


Figure 3.7: From top to bottom: Time plot of the number of transactions per minute for the stock Ericsson B in the period July 2-22,2002, histogram of the observation number and the sample autocorrelation function.

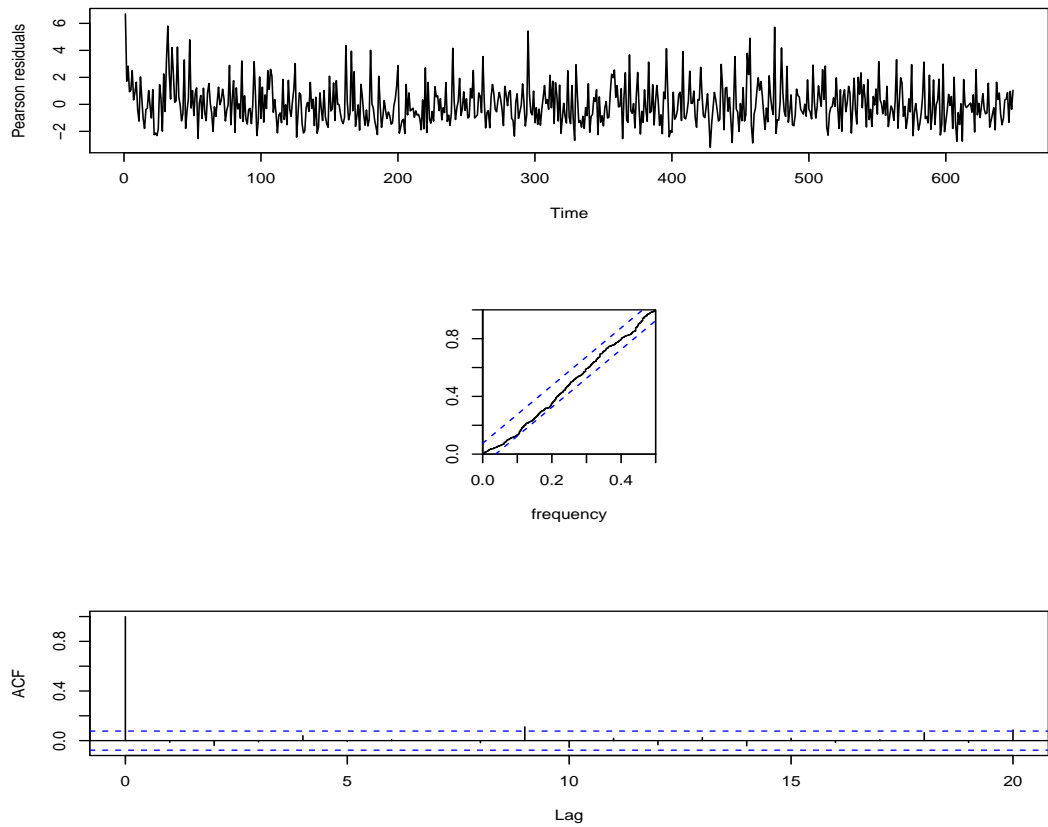


Figure 3.8: Pearson residuals (top) and their cumulative periodogram plot (middle), and the sample autocorrelation function of Pearson residuals (bottom) for the nonlinear model fitted to the number of transactions.



# Chapter 4

## Determination of the Number of Regimes

### 4.1 Introduction

The economic theory in the majority of financial applications concerning the TAR model is not specific about the general framework of the model. For example, the appropriate number of regimes usually cannot be fixed, as well as the value of threshold and the time delay (which determines the threshold variable in the TAR model). Some studies have been published that concentrate on how to choose a suitable number of regimes in TAR models. Tsay (1989) proposed a scatter-plot for locating the threshold value in an arranged autoregressive model. He used a lagged variable against the recursive estimates' t-ratios of an AR model's parameters to determine both the number and location of the threshold.

Recently, Hamaker (2009) suggested determining the number of regimes in a TAR model using six different information criteria with data sets containing small to moderate sample sizes ( $N = 50, 100, 200$ ). Gonzalo and Pitarakis (2002) suggested selecting the number of regimes ( $s$ ) or the value of threshold sequentially starting from a single sub-model ( $s=1$ ) and using model selection criteria for

determining between model with one or more threshold values, in moderate to large sample sizes ( $N = 200, 400, 800$ ).

The research on TAR models in the literature mostly is about continuous-valued time series, and this project will work on threshold models for discrete time series. In recent times, modeling count time series has gained significant interest with various applications in many different areas. For example, in medicine, the number of cases of campylobacteriosis infectious diseases reported in a specified area, in economy and finance, the number of transaction time series for the stock in a company, and in the environment, the daily temperature in a country and the number of accidents in a city. A few studies have been focused on a specified discrete conditional distribution with a conditional mean as a linear function of historical observations (Ferland et al. (2006), Fokianos et al. (2009), and Fokianos and Fried (2010)). Two studies have discussed Poisson autoregressive models (Fokianos et al. (2009), and Weiss and Testik (2009)) and both assume that the conditional distribution, given the history, follows a Poisson distribution with mean and variance being the same. The conditional mean and variance are linear functions of their previous values and past observations. Therefore, we call them linear Poisson GARCH models.

One of the main goal in TAR modeling is how to determine the suitable number of regimes if both the time delay and the values of threshold are given. The objective of this chapter is to discuss such an issue.

## 4.2 The PTGARCH(1,1; $s$ ,1) Model

Consider the model defined by (3.1). In the PTGARCH process the conditional variance for Poisson distribution is response, and its past values as well as the observed values are the covariants. A PTGARCH process consists of two or more regimes. Each regime follows a linear PGARCH process. PTGARCH processes

are characterized by frequent switches between the regimes. They switch between the regimes according to the threshold variable. If the threshold variable  $Y_{t-d}$  exceeded the value of threshold  $r_1$  then the processes fall into the first regime, and if  $r_1 < Y_{t-d} \leq r_2$ , then the processes switches to the second regime, and so on. Figure 4.1 shows six different models, a linear PGARCH process, and five different nonlinear PTGARCH processes in which regimes are created by first order linear PGARCH processes.

The linear PGARCH process is illustrated at the top of the graph, and was generated with ( $d = 0.3, a = 0.5, b = 0.4$ ). The first PTGARCH process is

$$\lambda_t = \begin{cases} 0.9 + 0.2Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-1} \leq 2, \\ 0.4 + 0.4Y_{t-1} + 0.3\lambda_{t-1}, & \text{if } Y_{t-1} > 2, \end{cases} \quad (4.1)$$

the second PTGARCH model is

$$\lambda_t = \begin{cases} 1.0 + 0.3Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-1} \leq 3, \\ 0.5 + 0.3Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-1} > 3, \end{cases} \quad (4.2)$$

the third PTGARCH model is

$$\lambda_t = \begin{cases} 1.2 + 0.45\lambda_{t-1}, & \text{if } Y_{t-1} \leq 2, \\ 0.6 + 0.5Y_{t-1} + 0.2\lambda_{t-1}, & \text{if } Y_{t-1} > 2, \end{cases} \quad (4.3)$$

the fourth PTGARCH model is given

$$\lambda_t = \begin{cases} 0.6 + 0.2Y_{t-1} + 0.46\lambda_{t-1}, & \text{if } Y_{t-1} \leq 3, \\ 1.0 + 0.6Y_{t-1}, & \text{if } Y_{t-1} > 3, \end{cases} \quad (4.4)$$

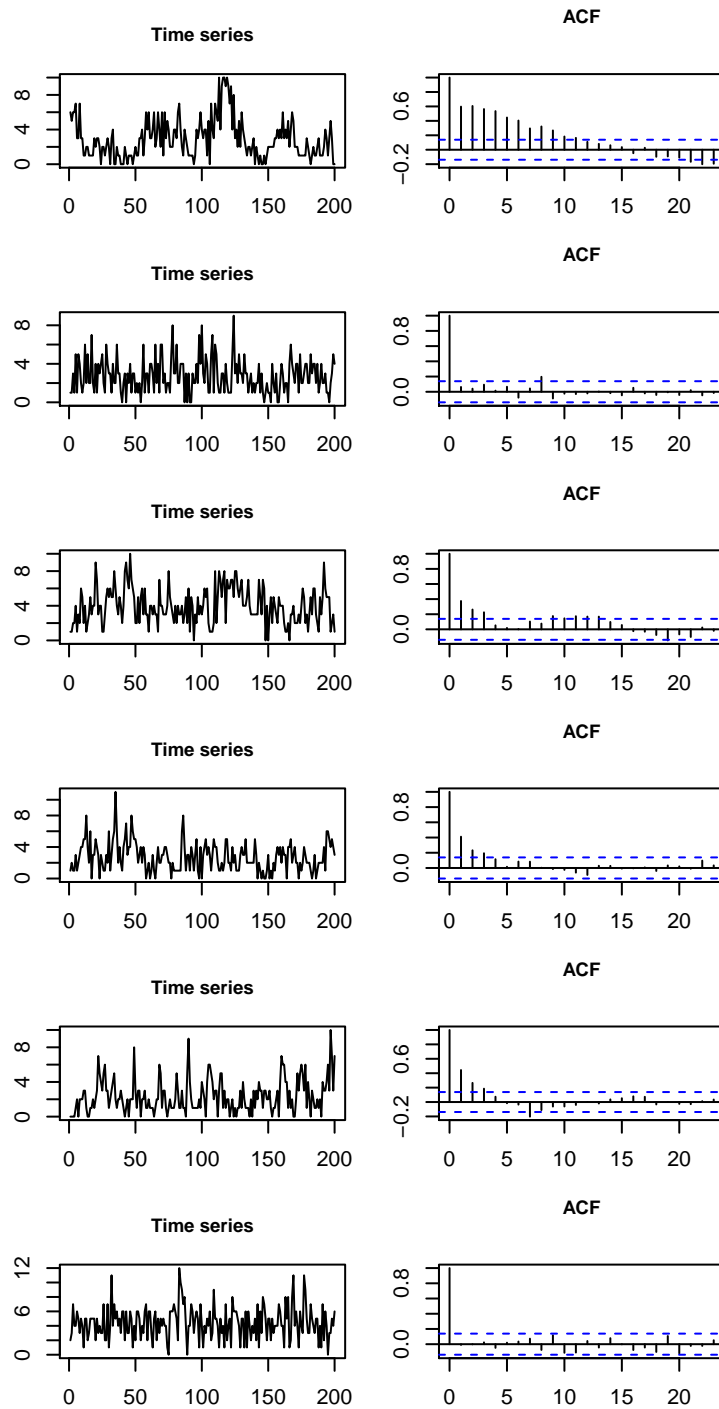


Figure 4.1: From top to bottom: Linear PGARCH process and five different PT-GARCH processes with different threshold values. The left column contains a plot of the time series and the right column contains their autocorrelation functions.

and the last PTGARCH process is

$$\lambda_t = \begin{cases} 3.0 + 0.2Y_{t-1} + 0.4\lambda_{t-1}, & \text{if } Y_{t-1} \leq 3, \\ 1.5 + 0.3Y_{t-1} + 0.2\lambda_{t-1}, & \text{if } 3 < Y_{t-1} \leq 7, \\ 0.7 + 0.4Y_{t-1} + 0.3\lambda_{t-1}, & \text{if } Y_{t-1} > 7. \end{cases} \quad (4.5)$$

In Figure 4.1, the plots on the left are the time series plotted against time for different models and the plots in the right are the autocorrelation function (ACF). The figure shows that the different PTGARCH models have low autocorrelation functions, and highly random appearance. In contrast, the plot of the autocorrelation function of linear model reveals a high dependence between observations.

### 4.3 Determination of the number of regimes with different information criteria

Recently, a few authors have concentrated on the study of our main problem, the determination of the number of regimes in a TAR model, using various information criteria in small to moderate sample sizes ( $N = 50, 100, 200$ ) (Hamaker (2009)), and in moderate to large sample sizes ( $N = 200, 400, 600, 800$ ) (Gonzalo and Pitarakis (2002); Strikholm and Teräsvirta (2006)). Peña and Rodriguez (2005) have studied the same problem by using different statistical tests. All studies discussed the determination number of regimes in a threshold autoregressive model only for continuous-valued time series. However, to the best of our knowledge, no one has discussed the same problem for count time series. Therefore, this chapter as far as we know, is the first statistical work to identify the number of regimes in a threshold autoregressive model for discrete-valued time series using different information criteria. The different information criteria used in the following experiments are the AIC, BIC, BIC2 and HQIC which were mentioned previously in the third chapter. These criteria can be expressed in a

unified formula as

$$IC = \frac{-2}{N} * \log(\text{likelihood}) + \frac{K}{N} * \lambda_{IC}. \quad (4.6)$$

The information criteria consist of two parts, first part is -2 time the log-likelihood function, and the second term corresponds to penalty terms  $\lambda_{IC} = 2$ ,  $\lambda_{IC} = \log(N)$ ,  $\lambda_{IC} = 2\log(N)$ , and  $\lambda_{IC} = 2\log(\log(N))$  respectively for AIC, BIC, BIC2, HQIC.  $K$  is the number of parameters in each regime and  $N$  is a sample size. The Loglikelihood function is given in (2.7). Actually, we have three parameters: an intercept  $\phi_{j0}$ , and autoregressive coefficients  $\phi_{j1}, \phi_{j2}$  in each regime. Hence the total number of regular coefficients across regimes is  $3s$ . In addition, there are  $s - 1$  threshold values.

## 4.4 Simulations

This section will concentrate on the performance of information criteria in moderate to large samples. The selected sample sizes are 300, 500 and 800 observations. All results are based on 1000 simulations. Four simulation studies are examined and their results are presented below.

In the first experiment study we use two kinds of nonlinearity tests, the Likelihood ratio (LR) and the Lagrange multiplier (max LM) tests, to determine the type of model. LR and LM tests are used to distinguish between the linearity and nonlinearity. The second experiment study compares two models. A PGARCH (1,1) model as a linear PGARCH, and PTGARCH (1,1;2,1) as a nonlinear PTGARCH. The performance of various information criteria is evaluated according to the true model rate detection and LR test is given for both linear GARCH and nonlinear GARCH. The third experiment studies the performance of four information criteria calculated when the data is generated from a nonlinear model (the true model) with different orders. In the last experiment, we examine whether

the four information criteria can be used to determine a higher number of regimes ( $s=3$ ). To study the performance of the different information criteria and compare their behaviour with results obtained from various model selection criteria, we assume the various models, such as PGARCH (1,1) model, PTGARCH (1,1;2,1) model, and PTGARCH (1,1;3,1) model.

#### 4.4.1 Experiment study 1: Linearity model against nonlinearity

The first study creates two types of linearity test: LR test and LM test, to determine whether the selected model is a linear PGARCH (1,1) model or a PTGARCH (1,1;2,1) model. It should be noted that a PTGARCH (1,1;2,1) model has two regimes and each regime is formed by a linear PGARCH (1,1) process. We use a different parameter value in simulating the data. Hence, both models are fitted with the simulated observations. Both types of model have a different intercept, and autoregressive coefficient values are used in simulating data for different sample sizes ( $N = 500, 800$  and  $1100$ ). The threshold value  $r$  and the time delay  $d$  of the models are given. In Table 4.1 the data were generated with five different PTGARCH (1,1;2,1) processes.

The different coefficients are used to discover the difference in the coefficients' effect on the model selection results. For example, the first model has two different parameters in each regime and this model can be written as

$$\lambda_t = \begin{cases} 0.9 + 0.2Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-1} \leq 2, \\ 0.4 + 0.4Y_{t-1} + 0.3\lambda_{t-1}, & \text{if } Y_{t-1} > 2. \end{cases} \quad (4.7)$$

In the second model, the intercept and the coefficient of  $Y_{t-1}$  are equal in each regime, while the autoregressive coefficient of  $\lambda_{t-1}$  differs with parameter values

Table 4.1: The percentage of correct model detecting for a nonlinear PTGARCH (1,1;2,1) model when compared to a linear PGARCH (1,1) model using LR test and LM test for different significance levels, for different threshold values, coefficients and sample sizes.

Threshold $r$	Regime		Sample size $n$	L-R test				L-M test			
	1	2		One regime		Two-regime		One regime		Two-regime	
				$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
$r=2$	$d_1=0.9$	$d_2=0.4$	500	13.3	39.3	86.7	60.7	0.0	0.0	100	100
	$a_1=0.5$	$a_2=0.3$	800	1.4	8.3	98.6	91.7	0.0	0.0	100	100
	$b_1=0.2$	$b_2=0.4$	1100	0.0	1.5	100	98.5	0.0	0.0	100	100
$r=2$	$d_1=0.5$	$d_2=0.5$	500	21.2	66.7	78.8	33.3	0.1	0.7	99.9	99.3
	$a_1=0.5$	$a_2=0.3$	800	1.5	14.2	98.5	85.8	0.0	0.0	100	100
	$b_1=0.4$	$b_2=0.4$	1100	0.2	0.8	99.8	99.2	0.0	0.0	1000	100
$r=2$	$d_1=0.6$	$d_2=0.6$	500	16.3	56.2	83.7	43.8	0.2	1.6	99.8	98.4
	$a_1=0.5$	$a_2=0.5$	800	1.4	11.0	98.6	89.0	0.0	0.2	100	99.8
	$b_1=0.4$	$b_2=0.2$	1100	0.1	0.7	99.9	99.3	0.0	0.0	100	100
$r=3$	$d_1=1.0$	$d_2=0.5$	500	46.6	86.6	53.4	13.4	0.0	0.0	100	100
	$a_1=0.5$	$a_2=0.5$	800	10.4	49.7	89.6	50.3	0.0	0.0	100	100
	$b_1=0.3$	$b_2=0.3$	1100	1.2	19.2	98.8	80.8	0.0	0.0	100	100
$r=4$	$d_1=1.0$	$d_2=0.5$	500	4.4	24.1	95.6	75.9	0.0	0.1	100	99.9
	$a_1=0.5$	$a_2=0.5$	800	0.0	1.6	100	98.4	0.0	0.0	100	100
	$b_1=0.3$	$b_2=0.3$	1100	0.0	0.1	100	99.9	0.0	0.0	100	100

NOTE: Percentage of correctly selecting a nonlinear (two-regime)

PTGARCH(1,1;2,1) model with  $p = q = 1$  based on 1000 simulations. LR test and LM test denote likelihood ratio test and Lagrange multiplier test respectively.

$a_1 = 0.5$  and  $a_2 = 0.3$ . The model is defined as

$$\lambda_t = \begin{cases} 0.5 + 0.4Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-1} \leq 2, \\ 0.5 + 0.4Y_{t-1} + 0.3\lambda_{t-1}, & \text{if } Y_{t-1} > 2. \end{cases} \quad (4.8)$$

In the third model, all parameters in each regime are the same except for the autoregressive coefficient of  $Y_{t-1}$  which differs with parameter values  $b_1 = 0.4$  and  $b_2 = 0.2$  :

$$\lambda_t = \begin{cases} 0.6 + 0.4Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-1} \leq 2, \\ 0.6 + 0.2Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-1} > 2. \end{cases} \quad (4.9)$$

The fourth and the fifth model have the same autoregressive coefficients in both regimes, while the intercept and the threshold value differ, with the threshold value being 3 in the fourth and 4 in the fifth model. The fourth model is

$$\lambda_t = \begin{cases} 1.0 + 0.3Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-1} \leq 3, \\ 0.5 + 0.3Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-1} > 3, \end{cases} \quad (4.10)$$



and the fifth model is

$$\lambda_t = \begin{cases} 1.0 + 0.3Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-1} \leq 4, \\ 0.5 + 0.3Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-1} > 4. \end{cases} \quad (4.11)$$

The reason for choosing the fourth and fifth models with these characteristics is to study the effect of the threshold value on the models.

In Table 4.1, the percentages of correctly identified models are given. The percentages of the LR test correctly selected two regimes between 53.4 and 100, and between 13.4 and 99.9 for  $\alpha = 0.05, 0.01$  respectively. In contrast, the percentages are between 0.0 and 46.6, and 0.1 and 46.6, for falsely selecting the one regime model (linear model). On the other hand, the percentages of LM test correctly selected models lie between 99.8 and 100, and between 98.4 and 100 for  $\alpha = 0.05, 0.01$  in all models respectively. It should be mentioned the fact that the LR and the LM tests are specifically designed for detecting quadratic nonlinearity. The percentages of the values of both LR and LM tests increase as the sample size increases. The distribution of the LR test and LM test is no longer approximately  $\chi_3^2$ . Note that  $\chi_{3,0.05}^2 = 7.82$  and  $\chi_{3,0.01}^2 = 11.35$ . (see Cryer and Chan (2008)).

In Table 4.2 the data were generated from two regimes. It reports six different models with different threshold values. Each sub-model has the same intercepts and the parameter coefficients. That is, in general,

$$\lambda_t = \begin{cases} d_1 + b_1Y_{t-1} + a_1\lambda_{t-1}, & \text{if } Y_{t-1} \leq r, \\ d_1 + b_1Y_{t-1} + a_1\lambda_{t-1}, & \text{if } Y_{t-1} > r, \end{cases} \quad (4.12)$$

Table 4.2: The percentage of correct model detecting for a linear PGARCH(1,1) model when compared to a nonlinear PTGARCH(1,1;2,1) model using test statistic (LR test and LM test) for different significance levels, for different threshold values, coefficients and sample sizes.

Threshold $r$	Regime1= $\theta_1=\theta_2$ Regime2	Sample size $n$	L-R test				L-M test			
			One regime		Two-regime		One regime		Two-regime	
			$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
$r=2$	$d_1 = d_2=0.9$	500	100	100	0.0	0.0	53.7	81.7	46.3	18.3
	$a_1 = a_2=0.5$	800	99.8	100	0.2	0.0	57.5	84.0	42.5	16.0
	$b_1 = b_2=0.2$	1100	99.9	100	0.1	0.0	59.5	85.8	40.5	14.2
$r=2$	$d_1 = d_2=0.5$	500	99.8	100	0.2	0.0	68.6	89.8	31.4	10.2
	$a_1 = a_2=0.3$	800	99.7	100	0.3	0.0	68.3	90.9	31.7	9.1
	$b_1 = b_2=0.4$	1100	99.8	100	0.2	0.0	66.1	88.2	33.9	11.8
$r=2$	$d_1 = d_2=0.5$	500	99.9	99.9	0.1	0.1	52.0	80.9	48.0	19.1
	$a_1 = a_2=0.5$	800	99.8	99.9	0.2	0.1	51.3	81.2	48.7	18.8
	$b_1 = b_2=0.4$	1100	99.9	100	0.1	0.0	53.8	82.3	46.2	17.7
$r=3$	$d_1 = d_2=0.6$	500	99.6	100	0.4	0.0	61.6	85.7	38.4	14.3
	$a_1 = a_2=0.5$	800	100	100	0.0	0.0	62.8	86.7	37.4	13.3
	$b_1 = b_2=0.2$	1100	100	100	0.0	0.0	64.7	87.3	35.3	12.3
$r=3$	$d_1 = d_2=1.0$	500	99.7	99.9	0.3	0.1	56.6	81.7	43.4	18.3
	$a_1 = a_2=0.5$	800	99.8	100	0.2	0.0	54.3	81.4	45.7	18.6
	$b_1 = b_2=0.3$	1100	99.9	100	0.1	0.0	55.1	82.2	44.9	17.8
$r=4$	$d_1 = d_2=1.0$	500	99.9	100	0.1	0.0	53.6	79.7	46.4	20.3
	$a_1 = a_2=0.5$	800	99.9	100	0.1	0.0	56.6	83.0	43.4	17.0
	$b_1 = b_2=0.3$	1100	100	100	0.0	0.0	56.8	82.1	43.2	17.9

NOTE: Percentage of correctly selecting a linear model (one regime) PGARCH(1,1) model with  $p = q = 1$  based on 1000 simulations. LR test and LM test denote likelihood ratio test and Lagrange multiplier test respectively.

which can be rewritten as follows

$$\begin{aligned}
 \lambda_t &= (d_1 + b_1 Y_{t-1} + a_1 \lambda_{t-1}) I(Y_{t-1} \leq r) + (d_1 + b_1 Y_{t-1} + a_1 \lambda_{t-1}) I(Y_{t-1} > r) \\
 &= (d_1 + b_1 Y_{t-1} + a_1 \lambda_{t-1}) \left( I(Y_{t-1} \leq r) + I(Y_{t-1} > r) \right) \\
 &= (d_1 + b_1 Y_{t-1} + a_1 \lambda_{t-1}).
 \end{aligned}$$

Table 4.2 shows the percentages of LR test of model (one regime) between 99.6 and 100, and between 99.9 and 100 for  $\alpha = 0.05, 0.01$  in all models respectively, while the percentages of LM test were between 51.3 and 68.6 and between 79.7 and 90.9 for  $\alpha = 0.05, 0.01$  respectively. In contrast, the results indicate the false selection of the two regime models with different parameters in between 31.4 and 48.7 and between 9.1 and 20.3 of the cases for  $\alpha = 0.05$  and  $\alpha = 0.01$  respectively, in the LM values. The LR test and LM test results improve as sample size increases.

### 4.4.2 Experiment study 2: Linear PGARCH (1,1) against PTGARCH(1,1;2,1)

The purpose of this study is to compare the performance of four information criteria in case of a linear PGARCH (1,1) model and a nonlinear PTGARCH (1,1;2,1) model, which has two sub-models formed by PGARCH (1,1) processes. Different parameters are used in each regime in simulating the data for different sample sizes ( $N = 300, 500, 800$ ). The information criteria are used to distinguish between a linear and a nonlinear model. LR tests of both models are calculated. Table 4.3 shows the percentages of correct decisions for a linear PGARCH (1,1) when compared to a nonlinear PTGARCH (1,1;2,1) model. The AIC, BIC, BIC2,

Table 4.3: The percentage of correct model detection for a linear PGARCH (1,1) model when compared to a PTGARCH (1,1;2,1) model for different coefficients and sample sizes.

PGARCH Parameter	Size	<i>AIC</i>	<i>BIC</i>	<i>BIC2</i>	<i>HQIC</i>
$d_1=0.9$	300	85.3	99.6	100	97.5
$a_1=0.5$	500	85.6	99.9	100	97.8
$b_1=0.2$	800	88.8	100	100	99.0
$d_1=0.8$	300	84.9	99.9	100	96.8
$a_1=0.6$	500	85.1	99.8	100	97.8
$b_1=0.3$	800	84.2	100	100	97.4
$d_1=0.6$	300	86.8	99.6	100	97.9
$a_1=0.5$	500	85.3	99.8	100	98.0
$b_1=0.4$	800	89.1	100	100	98.5

NOTE: Percentage of correctly selecting a linear model with  $p = q = 1$  based on 1000 simulations.

and HQIC refer to the model choosing criteria with different penalty terms. In Tables 4.3 and 4.4, the data were generated from a linear PTGARCH (1,1) model.

The BIC, BIC2 and HQIC perform very well above chance level (i.e, 50%) (percentages between 96.8 and 100). Clearly, the HQIC and BIC values improve as sample size increases. The BIC and its variants (BIC2) indicate the correct model with probability close to 1. The BIC2 results are in agreement with those reported by Hamaker (2009) for different sample sizes. In light of the above,

Table 4.4: The percentage of correct model detection for a linear PGARCH (1,1) model when compared to a nonlinear PTGARCH (1,1;2,1) model for different threshold values, coefficients and sample sizes.

Parameters Regime1	Size $n$	L-R Test			Average of LR
		$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	ALR
$d_1=0.9$	300	98.8	99.5	99.9	1.710
$a_1=0.5$	500	99.0	99.6	99.9	1.657
$b_1=0.2$	800	99.2	99.7	100	1.518
$d_1=0.8$	300	98.0	99.0	99.4	1.99
$a_1=0.6$	500	98.7	99.4	99.6	2.024
$b_1=0.3$	800	97.8	99.2	99.5	2.200
$d_1=0.6$	300	99.3	99.9	100	1.651
$a_1=0.5$	500	98.8	99.7	99.9	1.662
$b_1=0.4$	800	99.0	99.6	100.0	1.564

NOTE: Percentage of correctly selecting a linear PGARCH model with  $p = q = 1$  based on 1000 simulations.

we can conclude that the BIC and its variants have a strong performance for the linear model for different sample sizes. Table 4.4 shows LR tests of different significance levels ( $\alpha = 0.1, 0.05, 0.01$ ) for different sample sizes. The average of the LR tests is given. The table reports that the LR test improves as sample size increases for different significance levels. All results support a correctly selected model.

The results in Tables 4.5 and 4.6 are created for the data generated with six different PTGARCH (1,1;2,1) processes, when comparing a linear PGARCH (1,1) to a nonlinear PTGARCH (1,1;2,1) model. Specific coefficients were selected across the regimes to see how differences in coefficients affect the model selection results. The six models are expressed respectively as

$$\lambda_t = \begin{cases} 0.9 + 0.3Y_{t-1} + 0.2\lambda_{t-1}, & \text{if } Y_{t-1} \leq 2, \\ 0.6 + 0.4Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-1} > 2, \end{cases} \quad (4.13)$$

$$\lambda_t = \begin{cases} 0.8 + 0.2Y_{t-1} + 0.7\lambda_{t-1}, & \text{if } Y_{t-1} \leq 2, \\ 0.5 + 0.2Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-1} > 2, \end{cases} \quad (4.14)$$

Table 4.5: The percentage of correct model detection for a nonlinear PTGAR-ARCH(1,1:2,1) model when compared to a linear PGARCH(1,1) model for different threshold values, coefficients and sample sizes.

Threshold	Regime 1	Regime 2	Size	AIC	BIC	BIC2	HQIC
$r=2$	$d_1=0.9$	$d_2=0.6$	300	90.0	30.7	0.5	66.5
	$a_1=0.2$	$a_2=0.5$	500	98.3	59.2	4.4	90.8
	$b_1=0.3$	$b_2=0.4$	800	100	88.3	24.8	99.1
$r=2$	$d_1=0.8$	$d_2=0.5$	300	89.9	31.8	1.7	69.1
	$a_1=0.7$	$a_2=0.5$	500	97.8	58.2	5.7	87.5
	$b_1=0.2$	$b_2=0.2$	800	99.4	84.9	22.5	97.8
$r=3$	$d_1=0.5$	$d_2=1.5$	300	83.2	21.8	0.4	56.4
	$a_1=0.5$	$a_2=0.5$	500	97.1	47.4	2.3	85.4
	$b_1=0.3$	$b_2=0.3$	800	100	79.9	14.3	97.5
$r=3$	$d_1=0.9$	$d_2=0.9$	300	96.9	53.4	4.8	86.8
	$a_1=0.3$	$a_2=0.6$	500	99.8	88.6	20.6	98.5
	$b_1=0.3$	$b_2=0.3$	800	100	99.5	66.0	100
$r=3$	$d_1=0.9$	$d_2=0.9$	300	76.6	17.7	0.3	47.8
	$a_1=0.3$	$a_2=0.3$	500	91.5	30.1	1.3	68.7
	$b_1=0.3$	$b_2=0.6$	800	98.6	58.0	5.0	89.1
$r=4$	$d_1=1.2$	$d_2=1.2$	300	94.8	44.7	3.0	79.6
	$a_1=0.6$	$a_2=0.3$	500	98.5	74.9	12.6	64.0
	$b_1=0.3$	$b_2=0.3$	800	99.9	95.3	45.1	99.6

NOTE: Percentage of correctly selecting two-regime PTGAR-ARCH model with  $p = q = 1$  based on 1000 simulations.

$$\lambda_t = \begin{cases} 0.5 + 0.3Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-1} \leq 3, \\ 1.5 + 0.3Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-1} > 3, \end{cases} \quad (4.15)$$

$$\lambda_t = \begin{cases} 0.9 + 0.3Y_{t-1} + 0.3\lambda_{t-1}, & \text{if } Y_{t-1} \leq 3, \\ 0.9 + 0.3Y_{t-1} + 0.6\lambda_{t-1}, & \text{if } Y_{t-1} > 3, \end{cases} \quad (4.16)$$

$$\lambda_t = \begin{cases} 0.9 + 0.3Y_{t-1} + 0.3\lambda_{t-1}, & \text{if } Y_{t-1} \leq 3, \\ 0.9 + 0.6Y_{t-1} + 0.3\lambda_{t-1}, & \text{if } Y_{t-1} > 3, \end{cases} \quad (4.17)$$

$$\lambda_t = \begin{cases} 1.2 + 0.3Y_{t-1} + 0.6\lambda_{t-1}, & \text{if } Y_{t-1} \leq 4, \\ 1.2 + 0.3Y_{t-1} + 0.3\lambda_{t-1}, & \text{if } Y_{t-1} > 4. \end{cases} \quad (4.18)$$

The first and second models are models for which the parameters across the two regimes differ. The third model has the same autoregressive coefficients,

Table 4.6: The percentage of correct model detection for a nonlinear PTGARCH (1,1:2,1) model when compared to a linear PGARCH (1,1) model for different threshold values, coefficients and sample sizes.

Threshold	Parameters		Size		L-R Test			Average of LR
	Regime 1	Regime 2	$n$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	ALR	
$r=2$	$d_1=0.9$	$d_2=0.6$	300	54.5	38.2	11.9	7.09	
	$a_1=0.2$	$a_2=0.5$	500	86.3	73.0	41.0	10.74	
	$b_1=0.3$	$b_2=0.4$	800	98.5	96.2	82.0	16.52	
$r=2$	$d_1=0.8$	$d_2=0.5$	300	55.7	39.2	14.2	7.33	
	$a_1=0.7$	$a_2=0.5$	500	82.3	70.6	38.8	10.59	
	$b_1=0.2$	$b_2=0.2$	800	97.2	94.1	78.6	15.91	
$r=2$	$d_1=0.5$	$d_2=1.5$	300	43.9	27.2	6.6	6.16	
	$a_1=0.5$	$a_2=0.5$	500	78.6	65.0	28.9	9.50	
	$b_1=0.3$	$b_2=0.3$	800	96.4	91.9	71.4	14.44	
$r=3$	$d_1=0.9$	$d_2=0.9$	300	76.7	61.3	28.3	9.36	
	$a_1=0.3$	$a_2=0.6$	500	97.5	94.3	75.4	14.99	
	$b_1=0.3$	$b_2=0.3$	800	100	100	98.3	23.06	
$r=3$	$d_1=0.9$	$d_2=0.9$	300	37.0	22.7	5.1	5.55	
	$a_1=0.3$	$a_2=0.3$	500	60.4	43.3	16.8	7.76	
	$b_1=0.3$	$b_2=0.6$	800	85.9	74.7	47.0	11.35	
$r=4$	$d_1=1.2$	$d_2=1.2$	300	70.6	53.3	22.2	8.57	
	$a_1=0.6$	$a_2=0.3$	500	91.8	83.8	59.1	12.80	
	$b_1=0.3$	$b_2=0.3$	800	99.4	98.2	93.1	19.74	

NOTE: Percentage of correctly choice two-regime PTGARCH model with  $p = q = 1$  based on 1000 simulations.

whereas the intercepts differ across regimes. In the fourth model, the intercepts and autoregressive coefficients of the past observation ( $Y_{t-1}$ ) are fixed across the regimes, while the parameters of the past value of conditional variance ( $\lambda_{t-1}$ ) are different, and the value of threshold is 3. The coefficients of the autoregressive coefficients of the observed values ( $Y_{t-1}$ ) differ in the fifth model, whereas the intercepts and coefficients of the conditional variance ( $\lambda_{t-1}$ ) are fixed across the regimes. The latter model has different parameters of the past value of conditional variance ( $\lambda_{t-1}$ ), whereas the intercepts and autoregressive coefficients of the observed values ( $Y_{t-1}$ ) are fixed across the regimes, and the threshold value is 4.

Table 4.5 reports the percentage of correctly identified models. For all sample sizes, the AIC performs best. The performance of HQIC seems good when sample size increases. The Table 4.5 shows that when the intercepts and autoregressive coefficients of the observed values ( $Y_{t-1}$ ) are fixed across the regimes, the BIC criterion is good for large sample sizes. The BIC2 criterion has the worst per-

formance among all criteria for the complex models.

In Table 4.6, the percentages of LR tests for the threshold nonlinearity are given for different significance levels ( $\alpha = 0.1, 0.05, 0.01$ ). The percentages have been compared using critical values of  $\chi_{3,\alpha}^2$  which are 6.25, 7.82, and 11.35 of significance levels ( $\alpha = 0.1, 0.05, 0.01$ ) respectively. Table 4.6 indicates that the LR test supports a linearity model when the intercepts and autoregressive coefficients of the past values of conditional variance ( $\lambda_{t-1}$ ) are fixed across the regimes at  $\alpha = 0.01$ , although the data were generated from a nonlinear model, and the rest of the results support the correct models at the different significance levels. The average of LR tests is given.

### 4.4.3 Experiment study 3: Nonlinear PTGARCH (1,1;2,1) with different coefficients, and different order

The aim of the third experiment study is to define whether the information criteria can distinguish among three different models, namely a linear PGARCH (1,1) process, a nonlinear PTGARCH (1,1;2,1) process with  $p = q = 1$  and a nonlinear PTGARCH (1,1;2,1) process with different coefficients. The first (linear) PGARCH (1,1) model was simulated with  $d_1 = 0.7, a_1 = 0.5, b_1 = 0.3$ . That is,  $\lambda_t = 0.7 + 0.3Y_{t-1} + 0.5\lambda_{t-1}$ . The first nonlinear model was simulated with ( $d_1 = 1.2, a_1 = 0.2, b_1 = 0.3; d_2 = 0.6, a_2 = 0.4, b_2 = 0.5$ )

$$\lambda_t = \begin{cases} 1.2 + 0.3Y_{t-1} + 0.2\lambda_{t-1}, & \text{if } Y_{t-1} \leq 3, \\ 0.6 + 0.5Y_{t-1} + 0.4\lambda_{t-1}, & \text{if } Y_{t-1} > 3, \end{cases} \quad (4.19)$$

and the second nonlinear model has two regimes, but the autoregressive coefficient of  $Y_{t-1}$  of the first regime disappears. That is,

$$\lambda_t = \begin{cases} 1.2 + 0.45\lambda_{t-1}, & \text{if } Y_{t-1} \leq 2, \\ 0.6 + 0.5Y_{t-1} + 0.2\lambda_{t-1}, & \text{if } Y_{t-1} > 2. \end{cases} \quad (4.20)$$

From the results in Table 4.7, it can be seen that the AIC performs best when looking at all three models with correct model detection percentages of 83.9 for a linear model, 74.8 for the first nonlinear PTGARCH (1,1;2,1) model and 53.1 for the second nonlinear PTGARCH (1,1;2,1) model when the sample size is 300. Obviously, the HQIC performs quite well focusing on the results for all three

Table 4.7: Model selection for a linear GARCH(1,1) model and two different PTGARCH(1,1;2,1) models for different sample sizes.

Threshold	Regime 1	Regime 2	Size	Model	<i>AIC</i>	<i>BIC</i>	<i>BIC2</i>	<i>HQIC</i>	
<i>r</i> =2	<i>d</i> <sub>1</sub> =0.7		300	<b>a</b>	<b>83.9</b>	<b>99.5</b>	<b>100</b>	<b>96.2</b>	
				b	8.6	0.2	0.0	0.7	
				c	7.5	0.3	0.0	3.1	
		<i>a</i> <sub>1</sub> =0.5		500	<b>a</b>	<b>83.3</b>	<b>99.7</b>	<b>100</b>	<b>96.6</b>
					b	9.6	0.1	0.0	1.5
					c	7.1	0.2	0.0	1.9
		<i>b</i> <sub>1</sub> =0.3		800	<b>a</b>	<b>87.4</b>	<b>99.9</b>	<b>100</b>	<b>98.4</b>
					b	8.7	0.1	0.0	0.5
					c	3.9	0.0	0.0	1.1
<i>r</i> =2	<i>d</i> <sub>1</sub> =1.2	<i>d</i> <sub>2</sub> =0.6	300	a	12.1	67.7	95.8	36.0	
				<b>b</b>	<b>74.8</b>	<b>11.3</b>	<b>0.1</b>	<b>42.1</b>	
				c	13.1	21.0	4.1	21.9	
		<i>a</i> <sub>1</sub> =0.2	<i>a</i> <sub>2</sub> =0.45	500	a	2.0	45.4	90.3	12.6
					<b>b</b>	<b>94.2</b>	<b>37.5</b>	<b>1.3</b>	<b>78.5</b>
					c	3.8	17.1	8.4	8.9
		<i>b</i> <sub>1</sub> =0.3	<i>b</i> <sub>2</sub> =0.25	800	a	0.1	15.7	79.1	1.8
					<b>b</b>	<b>99.3</b>	<b>78.5</b>	<b>9.2</b>	<b>96.6</b>
					c	0.6	5.8	11.7	1.6
<i>r</i> =2	<i>d</i> <sub>1</sub> =1.2	<i>d</i> <sub>2</sub> =0.6	300	a	20.8	77.4	98.8	46.7	
				b	26.1	2.8	0.0	12.2	
				<b>c</b>	<b>53.1</b>	<b>19.8</b>	<b>1.2</b>	<b>41.1</b>	
		<i>a</i> <sub>1</sub> =0.45	<i>a</i> <sub>2</sub> =0.2	500	a	7.8	55.1	95.3	24.3
					b	30.4	5.6	0.1	15.2
					<b>c</b>	<b>61.8</b>	<b>39.3</b>	<b>4.6</b>	<b>60.5</b>
		<i>b</i> <sub>1</sub> =0.5	<i>b</i> <sub>2</sub> =0.5	800	a	1.6	31.6	82.3	8.1
					b	34.8	9.2	0.4	20.3
					<b>c</b>	<b>63.6</b>	<b>59.2</b>	<b>17.3</b>	<b>71.6</b>

NOTE: Simulation results for a linear GARCH (1,1) model and two different PTGARCH (1,1;2,1) models. All three models were fitted to the data. The correct model detection percentages are printed in bold. Results are based on 1000 simulations.

models for different sample sizes. In particular, when the sample size is 800, the percentage of correct model detection for a linear model is 98.4, and 96.6 and 71.6 for the first and second nonlinear PTGARCH (1,1;2,1) model respectively. The BIC2 does not perform well, as it cannot distinguish between the last two



nonlinear PTGARCH (1,1;2,1) models, whereas the BIC improves fairly well when the sample size increases to 800.

Table 4.8 contains three different models, a linear model and two various non-linear models. The difference between this table and Table 4.7 is in the last model in which the autoregression coefficient of the past observation ( $Y_{t-1}$ ) disappeared in the second regime. That is, the last model in each table contains two regimes with different order and parameters. These models in Table 4.8 can be expressed respectively as  $\lambda_t = 0.9 + 0.3Y_{t-1} + 0.2\lambda_{t-1}$ , and

$$\lambda_t = \begin{cases} 0.9 + 0.3Y_{t-1} + 0.2\lambda_{t-1}, & \text{if } Y_{t-1} \leq 2, \\ 0.6 + 0.4Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-1} > 2, \end{cases} \quad (4.21)$$

$$\lambda_t = \begin{cases} 0.9 + 0.3Y_{t-1} + 0.2\lambda_{t-1}, & \text{if } Y_{t-1} \leq 2, \\ 0.6 + 0.2\lambda_{t-1}, & \text{if } Y_{t-1} > 2. \end{cases} \quad (4.22)$$

Tables 4.9 and 4.10 contain also three different models, linear and two different nonlinear models. The difference between these tables is in the last model in each table. In Table 4.9 and 4.10, the parameter of the past value of conditional variance ( $\lambda_{t-1}$ ) has disappeared in the first regime and the second regime respectively. The last model in Tables 4.9 and 4.10 are expressed respectively as:

$$\lambda_t = \begin{cases} 1.2 + 0.4Y_{t-1}, & \text{if } Y_{t-1} \leq 3, \\ 0.6 + 0.6Y_{t-1} + 0.3\lambda_{t-1}, & \text{if } Y_{t-1} > 3, \end{cases} \quad (4.23)$$

and

$$\lambda_t = \begin{cases} 1.5 + 0.4Y_{t-1} + 0.3\lambda_{t-1}, & \text{if } Y_{t-1} \leq 3, \\ 1.0 + 0.6Y_{t-1}, & \text{if } Y_{t-1} > 3. \end{cases} \quad (4.24)$$

In model (4.23), the autoregressive coefficient of  $\lambda_{t-1}$  of the first regime was set to zero. It should be noted that the estimation theory developed in Chapter 2 applies for this model when this autoregressive coefficient takes values between

zero and one. The results in Table 4.11 for threshold value  $r = 2$  and the parameters the last model  $d_1 = 1.2, a_1 = 0.45, d_2 = 0.6, a_2 = 0.2, b_2 = 0.5$  have been used also in Table 4.7 last model for comparison purposes. Similarly, for  $r = 2$  and the parameters  $d_1 = 0.9, a_1 = 0.2, b_2 = 0.3, d_2 = 0.6, a_2 = 0.5$  the results have been used in Table 4.8 last model. In addition, results for  $r = 3$  and the parameters  $d_1 = 1.2, b_1 = 0.4, d_2 = 0.6, a_2 = 0.3, b_2 = 0.6$  were used in Table 4.9, and for  $r = 3$  and the parameters  $d_1 = 1.5, a_2 = 0.3, b_1 = 0.4, d_2 = 1.0, b_2 = 0.6$  were used in Table 4.10. Table 4.11 shows two different cases in different models. In the first case, the coefficients of autoregression of the past observation  $Y_{t-1}$  disappeared, and in the other case, the autoregressive parameters of the past value  $\lambda_{t-1}$  disappeared. In Table 4.11, it can be concluded that the AIC performs well in small samples, with correct model detection percentage of 53.1 in the first model. The other criteria have one scenario in which they perform around chance level (i.e., 33.3%) or much lower. The HQIC and AIC become best criteria when the sample size increases. The performance of BIC improves as the sample size increases from 300 to 800 in the current framework. In contrast, in the second case, the table shows that the HQIC performs best when looking at the two different models with correct model detection percentage of 54.4 in the first model, and the performance of BIC and HQIC criteria gets best as the sample size increases. Moreover, the performance of all four information criteria increases as the sample size increases. In both cases, the BIC2 performs worse in small and moderate samples.

Table 4.8: Model selection for a linear GARCH (1,1) model and two different PTGARCH (1,1;2,1) models for sample sizes.

Threshold	Regime 1	Regime 2	Size	Model	<i>AIC</i>	<i>BIC</i>	<i>BIC2</i>	<i>HQIC</i>
$r=2$	$d_1=0.9$ $a_1=0.2$ $b_1=0.3$		300	a	<b>83.5</b>	<b>99.3</b>	<b>100</b>	<b>96.6</b>
				b	8.6	0.3	0.0	1.3
				c	7.9	0.4	0.0	2.1
	500		a	<b>82.9</b>	<b>99.7</b>	<b>100</b>	<b>97.1</b>	
			b	9.0	0.3	0.0	0.7	
			c	8.1	0.0	0.0	2.2	
	800		a	<b>86.7</b>	<b>99.9</b>	<b>100</b>	<b>97.9</b>	
			b	8.0	0.1	0.0	0.9	
			c	5.3	0.0	0.0	1.2	
$r=2$	$d_1=0.9$ $a_1=0.2$ $b_1=0.3$	$d_2=0.6$ $a_2=0.5$ $b_2=0.4$	300	a	9.6	69.3	99.6	33.8
				b	<b>90.1</b>	<b>29.4</b>	<b>0.3</b>	<b>65.2</b>
				c	0.3	1.3	0.1	1.0
	500	a	0.1	40.8	95.3	8.7		
		b	<b>98.5</b>	<b>59.1</b>	<b>4.7</b>	<b>91.3</b>		
		c	0.0	0.1	0.0	0.0		
	800	a	0.0	11.2	78.3	0.7		
		b	<b>100</b>	<b>88.8</b>	<b>21.7</b>	<b>99.3</b>		
		c	0.0	0.0	0.0	0.0		
$r=2$	$d_1=0.9$ $a_1=0.2$ $b_1=0.3$	$d_2=0.6$ $a_2=0.5$	300	a	12.7	63.5	96.1	34.1
				b	37.8	4.4	0.0	17.6
				c	<b>49.5</b>	<b>32.1</b>	<b>3.9</b>	<b>48.3</b>
	500	a	3.1	37.8	86.0	13.4		
		b	35.7	5.5	0.0	19.2		
		c	<b>61.2</b>	<b>56.7</b>	<b>14.0</b>	<b>67.4</b>		
	800	a	0.5	11.3	62.8	2.3		
		b	38.3	10.3	0.8	21.5		
		c	<b>61.2</b>	<b>78.4</b>	<b>36.4</b>	<b>76.2</b>		

NOTE: Simulation results for a linear GARCH (1,1) model and two different PTGARCH (1,1;2,1) models. The correct model detection percentages are printed in bold. Results are based on 1000 simulations.

Table 4.9: Model selection for a linear GARCH(1,1) model and two different PT-GARCH(1,1;2,1) model for different sample sizes.

Threshold	Regime 1	Regime 2	Size	Model	<i>AIC</i>	<i>BIC</i>	<i>BIC2</i>	<i>HQIC</i>
$r=2$	$d_1=0.3$		300	<b>a</b>	<b>86.8</b>	<b>99.9</b>	<b>100</b>	<b>98.1</b>
				b	12.7	0.1	0.0	1.8
				c	0.5	0.0	0.0	0.1
	$a_1=0.5$		500	<b>a</b>	<b>88.9</b>	<b>100</b>	<b>100</b>	<b>98.8</b>
				b	11.1	0.0	0.0	1.2
				c	0.0	0.0	0.0	0.0
	$b_1=0.4$		800	<b>a</b>	<b>87.5</b>	<b>100</b>	<b>100</b>	<b>99.0</b>
				b	12.5	0.0	0.0	1.0
				c	0.0	0.0	0.0	0.0
$r=3$	$d_1=0.9$	$d_2=0.5$	300	a	23.3	82.2	99.3	53.1
				<b>b</b>	<b>70.7</b>	<b>11.0</b>	<b>0.2</b>	<b>38.9</b>
				c	6.0	6.8	0.5	8.0
	$a_1=0.3$	$a_2=0.3$	500	a	6.8	65.0	97.6	27.6
				<b>b</b>	<b>90.9</b>	<b>28.6</b>	<b>0.8</b>	<b>68.1</b>
				c	2.3	6.4	1.6	4.1
	$d_1=0.3$	$b_2=0.6$	800	a	0.7	34.3	92.8	7.4
				<b>b</b>	<b>99.2</b>	<b>64.1</b>	<b>5.1</b>	<b>91.8</b>
				c	0.1	1.6	2.1	0.8
$r=3$	$d_1=1.2$	$d_2=0.6$	300	a	1.5	28.5	78.6	7.8
				b	51.9	23.4	2.2	37.8
				<b>c</b>	<b>46.6</b>	<b>48.1</b>	<b>19.2</b>	<b>54.4</b>
	$b_1=0.4$	$a_2=0.3$	500	a	0.1	4.0	39.7	0.3
				b	50.4	33.1	11.2	40.5
				<b>c</b>	<b>49.3</b>	<b>62.9</b>	<b>49.1</b>	<b>59.2</b>
		$b_2=0.6$	800	a	0.0	3.9	41.8	0.3
				b	50.4	32.7	11.2	39.8
				<b>c</b>	<b>49.4</b>	<b>63.4</b>	<b>47.0</b>	<b>59.9</b>

NOTE: Simulation results for a linear GARCH (1,1) model and two different PTGARCH (1,1;2,1) models. The correct model detection percentages are printed in bold. Results are based on 1000 simulations.

Table 4.10: Model selection for a linear GARCH (1,1) model and two different PTGARCH (1,1;2,1) models for different sample sizes.

Threshold	Regime 1	Regime 2	Size	Model	<i>AIC</i>	<i>BIC</i>	<i>BIC2</i>	<i>HQIC</i>	
$r=3$	$d_1=1.0$ $a_1=0.4$ $b_1=0.3$		300	a	<b>81.7</b>	<b>99.5</b>	<b>100</b>	<b>95.8</b>	
				b	8.7	0.0	0.0	1.0	
				c	9.6	0.5	0.0	3.2	
				500	a	<b>84.8</b>	<b>99.7</b>	<b>100</b>	<b>97.1</b>
					b	9.7	0.0	0.0	1.0
					c	5.5	0.3	0.0	1.9
				800	a	<b>88.4</b>	<b>100</b>	<b>100</b>	<b>98.7</b>
					b	9.3	0.0	0.0	0.6
					c	2.3	0.0	0.0	0.7
$r=2$	$d_1=0.8$ $a_1=0.7$ $b_1=0.2$	$d_2=0.5$ $a_2=0.5$ $b_2=0.2$	300	a	6.7	56.7	95.5	23.1	
				b	<b>59.2</b>	<b>18.2</b>	<b>0.6</b>	<b>40.2</b>	
				c	34.1	25.1	3.9	36.7	
				500	a	1.5	31.0	85.5	6.7
					b	<b>68.6</b>	<b>25.3</b>	<b>2.4</b>	<b>52.0</b>
					c	29.9	43.7	12.1	41.3
				800	a	0.2	7.9	59.2	1.1
					b	<b>81.9</b>	<b>42.9</b>	<b>5.4</b>	<b>68.8</b>
					c	17.9	49.2	35.4	30.1
$r=3$	$d_1=1.5$ $a_1=0.3$ $b_1=0.4$	$d_2=1.0$ $b_2=0.6$	300	a	15.8	68.0	96.2	36.1	
				b	28.0	2.5	0.1	10.9	
				c	<b>56.2</b>	<b>29.5</b>	<b>3.7</b>	<b>53.0</b>	
				500	a	2.1	40.6	87.2	12.0
					b	27.5	2.5	0.2	13.0
					c	<b>70.4</b>	<b>56.9</b>	<b>12.6</b>	<b>75.0</b>
				800	a	0.4	12.5	67.1	2.9
					b	30.1	3.3	0.1	11.2
					c	<b>69.5</b>	<b>84.2</b>	<b>32.8</b>	<b>85.9</b>

NOTE: Simulation results for a linear GARCH (1,1) model and two different PTGARCH (1,1;2,1) models. The correct model detection percentages are printed in bold. Results are based on 1000 simulations.

Table 4.11: Model selection percentage for different nonlinear PTG-ARCH(1,1;2,1) models for different sample sizes.

Threshold	Regime 1	Regime 2	Size	$S$	$AIC$	$BIC$	$BIC2$	$HQIC$	
$r=2$	$d_1=1.2$ $a_1=0.45$	$d_2=0.6$ $a_2=0.2$ $b_2=0.5$	300	a	20.8	77.4	98.8	46.7	
				b	26.1	2.8	0.0	12.2	
				<b>c</b>	<b>53.1</b>	<b>19.8</b>	<b>1.2</b>	<b>41.1</b>	
	500				a	7.8	55.1	95.3	24.3
					b	30.4	5.6	0.1	15.2
					<b>c</b>	<b>61.8</b>	<b>39.3</b>	<b>4.6</b>	<b>60.5</b>
					a	1.6	31.6	82.3	8.1
					b	34.8	9.2	0.4	20.3
					<b>c</b>	<b>63.6</b>	<b>59.2</b>	<b>17.3</b>	<b>71.6</b>
	$r=2$	$d_1=0.9$ $a_1=0.2$ $b_1=0.3$	$d_2=0.6$ $a_2=0.5$	300	a	12.7	63.5	96.1	34.1
					b	37.8	4.4	0.0	17.6
					<b>c</b>	<b>49.5</b>	<b>32.1</b>	<b>3.9</b>	<b>48.3</b>
500					a	3.1	37.8	86.0	13.4
					b	35.7	5.5	0.0	19.2
					<b>c</b>	<b>61.2</b>	<b>56.7</b>	<b>14.0</b>	<b>67.4</b>
					a	0.5	11.3	62.8	2.3
					b	38.3	10.3	0.8	21.5
					<b>c</b>	<b>61.2</b>	<b>78.4</b>	<b>36.4</b>	<b>76.2</b>
$r=3$		$d_1=1.2$ $b_1=0.4$	$d_2=0.6$ $a_2=0.3$ $b_2=0.6$	300	a	1.5	28.5	78.6	7.8
					b	51.9	23.4	2.2	37.8
					<b>c</b>	<b>46.6</b>	<b>48.1</b>	<b>19.2</b>	<b>54.4</b>
	500				a	0.1	4.0	39.7	0.3
					b	50.4	33.1	11.2	40.5
					<b>c</b>	<b>49.3</b>	<b>62.9</b>	<b>49.1</b>	<b>59.2</b>
					a	0.0	3.9	41.8	0.3
					b	50.4	32.7	11.2	39.8
					<b>c</b>	<b>49.4</b>	<b>63.4</b>	<b>47.0</b>	<b>59.9</b>
	$r=3$	$d_1=1.5$ $a_1=0.3$ $b_1=0.4$	$d_2=1.0$ $b_2=0.6$	300	a	15.8	68.0	96.2	36.1
					b	28.0	2.5	0.1	10.9
					<b>c</b>	<b>56.2</b>	<b>29.5</b>	<b>3.7</b>	<b>53.0</b>
500					a	2.1	40.6	87.2	12.0
					b	27.5	2.5	0.2	13.0
					<b>c</b>	<b>70.4</b>	<b>56.9</b>	<b>12.6</b>	<b>75.0</b>
					a	0.4	12.5	67.1	2.9
					b	30.1	3.3	0.1	11.2
					<b>c</b>	<b>69.5</b>	<b>84.2</b>	<b>32.8</b>	<b>85.9</b>

NOTE: Simulation results for two-regime PTGARCH (1,1;2,1) model with different orders ( $p \neq q$ ), when fitting a linear PGARCH (1,1) model ( $s = 1$ ), a PTGARCH(1,1;2,1) model ( $s = 2$ ) with ( $p = q = 1$ ) and a PTGARCH (1,1;2,1) model ( $s = 3$ ) with different orders ( $p \neq q$ ). The correct model detection percentages are printed in bold. Results are based on 1000 simulations.

#### 4.4.4 Experiment study 4: Nonlinear PTGARCH (1,1:3,1) with different coefficients

The main purpose of this experiment is distinguishing between two nonlinear PTGARCH (1,1;3,1) models using the information criteria. The first model was simulated with  $d_1 = 3.0, a_1 = 0.4, b_1 = 0.2; d_2 = 1.5, a_2 = 0.2, b_2 = 0.3, ; d_3 = 0.7, a_3 = 0.3, b_3 = 0.4$ :

$$\lambda_t = \begin{cases} 3.0 + 0.2Y_{t-1} + 0.4\lambda_{t-1}, & \text{if } Y_{t-1} \leq 3, \\ 1.5 + 0.3Y_{t-1} + 0.2\lambda_{t-1}, & \text{if } 3 < Y_{t-1} \leq 7, \\ 0.7 + 0.4Y_{t-1} + 0.3\lambda_{t-1}, & \text{if } Y_{t-1} > 7, \end{cases} \quad (4.25)$$

and the second model was simulated with  $d_1 = 2.5, a_1 = 0.39, b_1 = 0.6; d_2 = 0.6, a_2 = 0.3, b_2 = 0.45; d_3 = 1.0, a_3 = 0.2, b_3 = 0.35$ :

$$\lambda_t = \begin{cases} 2.5 + 0.6Y_{t-1} + 0.39\lambda_{t-1}, & \text{if } Y_{t-1} \leq 3, \\ 0.6 + 0.45Y_{t-1} + 0.3\lambda_{t-1}, & \text{if } 3 < Y_{t-1} \leq 7, \\ 1.0 + 0.35Y_{t-1} + 0.2\lambda_{t-1}, & \text{if } Y_{t-1} > 7. \end{cases} \quad (4.26)$$

The threshold values of both models are given ( $r_1 = 3, r_2 = 7$ ). In Table 4.12 the outcomes are given for the data generated with two different nonlinear PTGARCH (1,1:3,1) processes, when comparing a nonlinear PTGARCH (1,1:2,1) model to a nonlinear PTGARCH (1,1:3,1) model. The threshold value of the nonlinear PTGARCH (1,1:2,1) model is given ( $r = 4$ ) and then the remaining parameters are estimated. From the outcomes in Table 4.12, it can be concluded that AIC performs best when focusing on the small size ( $N=200, 300$ ), with correct model detection percentages of 44.2% for the first model and 68.6% for the second model. The other criteria perform at a much lower chance level 50.0% or lower. In contrast, from the outcomes for large sample size ( $N=800$ ), it is obvious that the AIC, the BIC and the HQIC have a tendency to outperform and their

performance gets best when the sample sizes increases.

In Table 4.13, the results are given for three-regime PTGARCH (1,1:3,1) when fitting a linear PGARCH (1,1) model ( $s=1$ ), a PTGARCH (1,1:2,1) model ( $s=2$ ) and a PTGARCH (1,1:3,1) model with correct model detection percentages. It can be seen that the AIC and HQIC perform well above chance level (i.e., 33.3%) with percentages between 42.8 and 97.5 for moderate to large sample size. The measures increase as sample size increases.

Table 4.12: The percentage of correct model detection for a nonlinear PTGARCH(1,1:3,1) model when compared to a nonlinear PTGARCH(1,1:2,1) model for different threshold values, coefficients and sample sizes.

Threshold	Regime 1	Regime 2	Regime 3	Size	AIC	BIC	BIC2	HQIC
$r=4$	$d_1=3$	$d_2=1.5$	$d_3=0.7$	200	44.2	8.2	0.1	22.3
$r_1=3$	$a_1=0.4$	$a_2=0.2$	$a_3=0.3$	300	65.2	20.3	0.4	45.9
$r_2=7$	$b_1=0.2$	$b_2=0.3$	$b_3=0.4$	500	88.5	47.0	4.1	75.8
				800	97.2	79.6	24.2	93.6
$r=4$	$d_1=2.5$	$d_2=0.6$	$d_3=1.0$	300	68.6	25.3	2.3	52.4
$r_1=3$	$a_1=0.39$	$a_2=0.3$	$a_3=0.2$	500	89.9	55.8	9.7	78.1
$r_2=7$	$b_1=0.6$	$b_2=0.45$	$b_3=0.35$	800	98.7	90.7	45.4	97.5
				800	99.8	98.8	88.7	99.7

NOTE: Percentage of correctly choice three-regime PTGARCH models with  $p = q = 1$  based on 1000 simulations.



Table 4.13: Model selection percentages for a nonlinear PTGARCH(1,1;3,1) model for different sample sizes.

Threshold	Regime 1	Regime 2	Regime 3	Size	$s$	$AIC$	$BIC$	$BIC2$	$HQIC$
$r=4$	$d_1=3.0$	$d_2=1.5$	$d_3=0.7$		1	2.4	32.2	86.1	11.4
$r_1=3$	$a_1=0.4$	$a_2=0.2$	$a_3=0.3$	200	2	54.4	61.4	13.9	66.3
$r_2=7$	$b_1=0.2$	$b_2=0.3$	$b_3=0.4$		<b>3</b>	<b>43.2</b>	<b>6.4</b>	<b>0.0</b>	<b>22.3</b>
					1	0.3	12.7	72.1	1.8
				300	2	35.5	70.4	27.7	55.4
					<b>3</b>	<b>64.2</b>	<b>16.5</b>	<b>0.2</b>	<b>42.8</b>
					1	0.0	0.7	26.4	0.1
				500	2	11.5	51.2	69.2	24.2
					<b>3</b>	<b>88.5</b>	<b>48.1</b>	<b>4.4</b>	<b>75.7</b>
					1	0.0	0.0	2.5	0.0
				800	2	2.5	17.0	71.1	5.9
					<b>3</b>	<b>97.5</b>	<b>83.0</b>	<b>26.4</b>	<b>94.1</b>
$r=4$	$d_1=2.5$	$d_2=0.6$	$d_3=1.0$		1	8.9	67.4	98.2	27.8
$r_1=3$	$a_1=0.39$	$a_2=0.3$	$a_3=0.2$	200	2	45.1	24.8	1.8	44.8
$r_2=7$	$b_1=0.6$	$b_2=0.45$	$b_3=0.35$		<b>3</b>	<b>46.0</b>	<b>7.8</b>	<b>0.0</b>	<b>24.4</b>
					1	1.4	49.3	96.8	13.8
				300	2	34.0	28.5	2.6	37.9
					<b>3</b>	<b>64.6</b>	<b>22.2</b>	<b>0.6</b>	<b>48.3</b>
					1	0.0	9.4	84.5	0.7
				500	2	4.2	14.5	8.3	6.6
					<b>3</b>	<b>95.8</b>	<b>76.1</b>	<b>7.2</b>	<b>92.7</b>
					1	0.0	1.2	40.2	0.1
				800	2	4.1	13.2	18.2	6.4
					<b>3</b>	<b>95.9</b>	<b>85.6</b>	<b>41.6</b>	<b>93.5</b>

NOTE: Percentages of correctly selected three-regime PTGARCH (1,1;3,1) models with  $p = q=1$ . The simulation results for three-regime PTGARCH process when fitting a linear PGARCH (1,1) model ( $s=1$ ), a nonlinear PTGARCH (1,1;2,1) model ( $s=2$ ) and nonlinear PTGARCH (1,1;3,1) model ( $s=3$ ). The correct model detection percentages are printed in bold. Results are based on 1000 simulations.

## 4.5 A real data example

As an example of application, we study the real data (as described in section (3.1)), to illustrate our methodology. Figure 4.2 shows the time series of the data and the sample autocorrelation function of the transaction series. The plot reports that there is dependence between the transactions. Table 4.14 shows the descriptive statistics for the number of transactions. In this data example, we have determined the maximum time delay ( $d$ ), which is equal to 6.

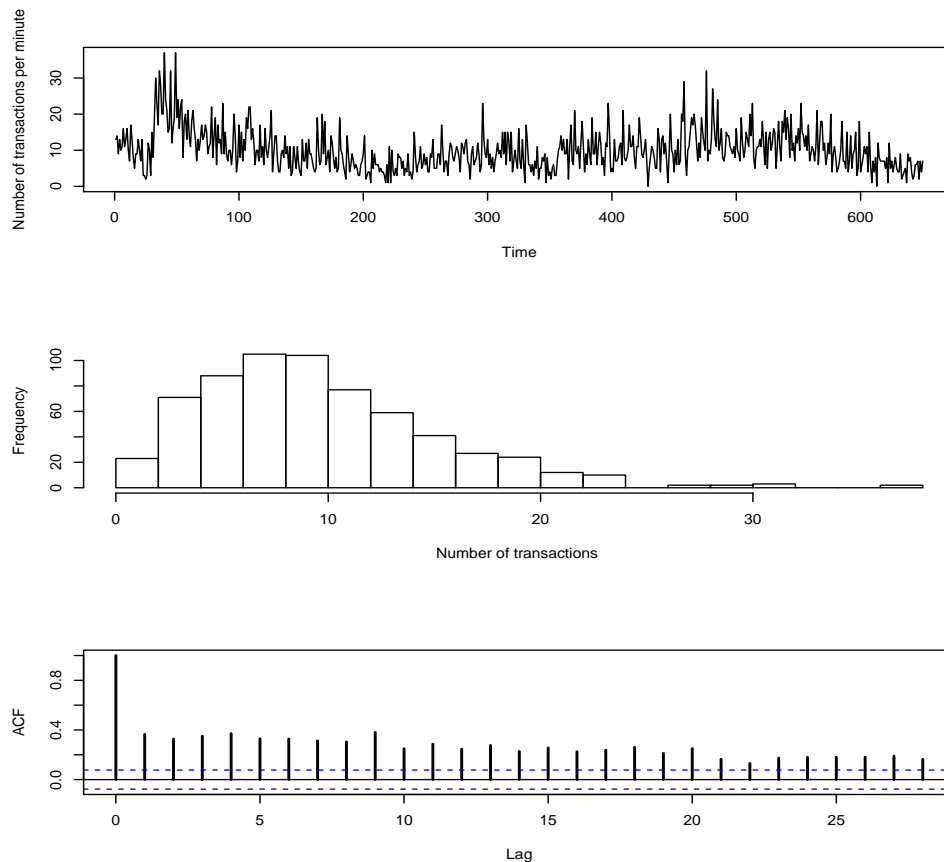


Figure 4.2: From top to bottom: Time plot of the number of transactions per minute for the stock Ericsson B in the period July 2-22, 2002, histogram of the observation numbers and the sample autocorrelation function.

In this example, we try to determine the number of regimes for TGARCH modelling. To do this, we compute the maximum likelihood estimates of the

Table 4.14: Descriptive statistics for transaction data.

$N$	Lower 25%	Median	Mean	75%	Upper	Var
650	0.00	6.00	9.00	10.08	13.00	37.00
						31.59247

parameters  $d_i, a_i, b_i, i = 1, 2, 3$  and calculate the normalized information criteria which are defined by

$$NIC(r, d) = \frac{IC(r, d)}{N - N_d},$$

where  $N_d = \max(d, L)$ , where  $d = 1, 2, \dots, D$ , and  $L$  is the maximum order for each of the  $s$  piecewise linear GARCH models, to determine the threshold value and time delay.

Table 4.15 reports the estimated parameters for the TGARCH model when the model has two regimes (*i.e.*  $m = 2$ ). Table 4.15 shows the possible threshold values ( $r$ ) between 5 and 13, and time delay ( $d$ ).

Table 4.15: Normalized information criteria of the PTGARCH (1,1;2, $d$ ) model fitted to the transactions' number for  $1 \leq d \leq 6$ .

Suggested time delay ( $d$ )	Possible threshold values ( $r$ )	Information criteria				The estimated parameters					
		$AIC$	$BIC$	$BIC2$	$HQIC$	$d_1$	$a_1$	$b_1$	$d_2$	$a_2$	$b_2$
2	5	-26.4600	-26.4195	-26.3610	-26.4443	1.764e-07	0.9757	1.991e-02	0.4581	0.7855	0.1727
2	6	-26.4689	-26.4284	-26.3699	-26.4532	2.482e-05	0.9696	1.048e-02	0.6673	0.7765	0.1682
1	7	-26.4727	-26.4323	-26.3739	-26.4570	3.879e-01	0.8949	1.951e-07	0.2915	0.7764	0.1892
2	8	-26.4585	-26.4180	-26.3596	-26.4428	8.175e-05	0.9043	9.286e-02	0.7948	0.7528	0.1789
2	9	-26.4870	-26.4465	-26.3880	-26.4713	1.712e-01	0.9122	4.899e-02	1.5881	0.6909	0.1884
2	10	-26.4903	-26.4498	-26.3913	-26.4746	3.767e-02	0.9224	6.833e-02	1.3712	0.6645	0.2272
2	11	-26.4939	-26.4535	-26.3950	-26.4783	4.152e-02	0.9156	7.607e-02	1.4155	0.6540	0.2398
3	12	-26.5005	-26.4600	-26.4014	-26.4848	1.264e-01	0.9079	6.960e-02	2.3965	0.5826	0.2456
2	13	-26.4524	-26.4120	-26.3535	-26.4368	1.906e-02	0.8692	1.369e-01	0.7452	0.6682	0.2580

NOTE: The columns from three to six report the normalized information criteria. The other columns report the estimated parameters for the TGARCH model fitted to the transactions' number.

The normalized information criteria are smallest when  $r = 12$  and  $d = 3$ .

Therefore, the suggested TGARCH model is:

$$\hat{\lambda}_t = \begin{cases} 0.1264 + 0.0696Y_{t-1} + 0.9079\hat{\lambda}_{t-1}, & \text{if } Y_{t-3} \leq 12, \\ 2.3965 + 0.2456Y_{t-1} + 0.5826\hat{\lambda}_{t-1}, & \text{if } Y_{t-3} > 12. \end{cases}$$

Tables 4.16 and 4.17 report autoregressive coefficients, the possible threshold value and time delay for the TGARCH model when the coefficients of autoregression of the past values ( $\lambda_{t-1}$ ) disappeared (*i.e.*  $a_i = 0, i = 1, 2$ ) respectively. Table 4.16 shows the possible threshold values ( $r$ ) between 5 and 13, and time delay ( $d$ ).

Table 4.16: Normalized information criteria of the PTGARCH (1,1;2, $d$ ) model fitted to the transactions' number for  $1 \leq d \leq 6$ .

Suggested time delay ( $d$ )	Possible threshold values ( $r$ )	Information criteria				The estimated parameters				
		$AIC$	$BIC$	$BIC2$	$HQIC$	$d_1$	$b_1$	$d_2$	$a_2$	$b_2$
3	5	-26.3509	-26.3171	-26.2683	-26.3378	5.6678	0.2458	1.0745	0.7019	0.2039
5	6	-26.3437	-26.3098	-26.2608	-26.3305	5.5252	0.2501	1.6093	0.6822	0.1961
4	7	-26.2528	-26.2189	-26.1701	-26.2397	5.7455	0.2932	2.0206	0.5987	0.2381
3	8	-26.2315	-26.1977	-26.1489	-26.2184	5.1944	0.3594	2.6186	0.5807	0.2160
4	9	-26.2294	-26.1955	-26.1467	-26.2162	5.5087	0.3298	3.2982	0.4844	0.2778
2	10	-26.2127	-26.1790	-26.1302	-26.1996	6.4593	0.2472	1.1533	0.6443	0.2670
3	11	-26.2166	-26.1828	-26.1340	-26.2035	5.6866	0.3391	3.8495	0.4980	0.2384
4	12	-26.1985	-26.1646	-26.1158	-26.1854	5.7283	0.3510	3.2515	0.5697	0.2388
3	13	-26.1901	-26.1563	-26.1075	-26.1770	5.7304	0.3540	4.3740	0.4583	0.2923

NOTE: The columns from three to six report the normalized information criteria. The other columns report the estimated parameters for the TGARCH model fitted to the transactions' number.

In Table 4.16 the information criteria are smallest when  $r = 5$  and  $d = 3$ . Therefore, the possible TGARCH model can be expressed as:

$$\hat{\lambda}_t = \begin{cases} 5.6678 + 0.2458Y_{t-1}, & \text{if } Y_{t-3} \leq 5, \\ 1.0745 + 0.2039Y_{t-1} + 0.7019\hat{\lambda}_{t-1}, & \text{if } Y_{t-3} > 5. \end{cases}$$

In contrast, the smallest values of the different information criteria in Table 4.17 were reached when  $r = 12, d = 3$ . Therefore, the estimated TGARCH model is:

Table 4.17: Normalized information criteria of the PTGARCH (1,1;2, $d$ ) model fitted to the transactions' number for  $1 \leq d \leq 6$ .

Suggested time delay( $d$ )	Possible threshold values ( $r$ )	Information criteria				The estimated parameters				
		$AIC$	$BIC$	$BIC2$	$HQIC$	$d_1$	$a_1$	$b_1$	$d_2$	$b_2$
4	5	-26.1129	-26.0790	-26.0302	-26.0998	2.8835	0.3799	0.15270	5.7257	0.4741
5	6	-26.1396	-26.1057	-26.0568	-26.1265	1.3511	0.5835	0.12150	5.9724	0.4750
5	7	-26.1974	-26.1635	-26.1146	-26.1843	1.1859	0.6285	0.15430	6.4413	0.4218
5	8	-26.2528	-26.2189	-26.1699	-26.2396	0.3450	0.7604	0.14560	6.3821	0.4553
3	9	-26.3046	-26.2708	-26.2220	-26.2915	0.8092	0.7347	0.11840	7.8115	0.3633
2	10	-26.3016	-26.2679	-26.2192	-26.2885	0.2575	0.8929	0.05698	6.3640	0.5209
3	11	-26.3398	-26.3061	-26.2573	-26.3268	0.2756	0.8412	0.10960	7.5620	0.3698
3	12	-26.3991	-26.3653	-26.3166	-26.3860	0.1455	0.8633	0.11140	7.1071	0.4597
3	13	-26.3337	-26.2999	-26.2511	-26.3206	0.5543	0.7706	0.15740	6.7477	0.5416

NOTE: The columns from three to six report the normalized information criteria. The other columns report the estimated parameters for the TGARCH model fitted to the transactions' number.

$$\hat{\lambda}_t = \begin{cases} 0.1455 + 0.11140Y_{t-1} + 0.8633\hat{\lambda}_{t-1}, & \text{if } Y_{t-3} \leq 12, \\ 7.1071 + 0.4597Y_{t-1}, & \text{if } Y_{t-3} > 12. \end{cases}$$

Tables 4.18 and 4.19 show the parameters to be estimated ( $d_i, a_i, b_i, r$  and  $d, i = 1, 2$ ) for the TGARCH model when the autoregressive parameters of the past observation ( $Y_{t-1}$ ) disappeared (*i.e.*  $b_i = 0, i = 1, 2$ ) respectively. In Table 4.18 the values of information criteria support that the best choice of the threshold value and time delay are 9 and 2 respectively. Therefore, the suggested TGARCH model for the example data can be expressed as:

$$\hat{\lambda}_t = \begin{cases} 0.27540 + 0.9363\hat{\lambda}_{t-1}, & \text{if } Y_{t-2} \leq 9, \\ 2.1554 + 0.1906Y_{t-1} + 0.6489\hat{\lambda}_{t-1}, & \text{if } Y_{t-2} > 9. \end{cases}$$

In contrast, according to the information criteria in Table 4.19, the suggested threshold value and the time delay are 13 and 6 respectively. Therefore, the possible TGARCH model is:

Table 4.18: Normalized information criteria of the PTGARCH(1,1;2, $d$ ) model fitted to the transactions' number for  $1 \leq d \leq 6$ .

Suggested time delay( $d$ )	Possible threshold values ( $r$ )	Information criteria				The estimated parameters				
		$AIC$	$BIC$	$BIC2$	$HQIC$	$d_1$	$a_1$	$d_2$	$a_2$	$b_2$
1	5	-26.4597	-26.4260	-26.3774	-26.4467	0.47620	0.8723	0.2461	0.7850	0.1847
2	6	-26.4707	-26.4370	-26.3883	-26.4576	0.02218	0.9734	0.7869	0.7692	0.1652
1	7	-26.4743	-26.4407	-26.3920	-26.4613	0.44670	0.8843	0.5642	0.7814	0.1639
1	8	-26.4768	-26.4431	-26.3945	-26.4637	0.56100	0.8776	0.8072	0.7763	0.1511
2	9	-26.4816	-26.4478	-26.3991	-26.4685	0.27540	0.9363	2.1554	0.6489	0.1906
2	10	-26.4735	-26.4398	-26.3911	-26.4604	0.18850	0.9539	2.3395	0.5886	0.2374
2	11	-26.4659	-26.4322	-26.3835	-26.4528	0.21180	0.9524	2.4534	0.5893	0.2425
3	12	-26.4645	-26.4307	-26.3819	-26.4514	0.18360	0.9599	3.5053	0.5049	0.2461
3	13	-26.3094	-26.2757	-26.2269	-26.2964	0.57210	0.9125	3.7557	0.4700	0.3005

NOTE: The columns from three to six report the normalized information criteria. The other columns report the estimated parameters for the TGARCH model fitted to the transactions' number.

$$\hat{\lambda}_t = \begin{cases} 1.0130 + 0.2302Y_{t-1} + 0.6415\hat{\lambda}_{t-1}, & \text{if } Y_{t-6} \leq 13, \\ 1.7387 + 0.9067\hat{\lambda}_{t-1}, & \text{if } Y_{t-6} > 13. \end{cases}$$

Table 4.20 reports the estimated autoregressive coefficients, the corresponding threshold values and time delay for the TGARCH model when the model has three regimes ( $s = 3$ ). The table shows that the values of normalized criteria are smallest when  $r_1 = 7, r_2 = 12$  and  $d = 3$ .

Therefore, the estimated TGARCH model is:

$$\hat{\lambda}_t = \begin{cases} 0.1384 + 0.02834Y_{t-1} + 0.9340\hat{\lambda}_{t-1}, & \text{if } Y_{t-3} \leq 7, \\ 0.3768 + 0.07388Y_{t-1} + 0.8881\hat{\lambda}_{t-1}, & \text{if } 7 < Y_{t-3} \leq 12, \\ 2.4144 + 0.2380Y_{t-1} + 0.5901\hat{\lambda}_{t-1}, & \text{if } Y_{t-3} > 12. \end{cases}$$

Table 4.21 presents the six estimated TGARCH models in Tables 4.15, 4.16, 4.17, 4.18, 4.19 and 4.20, and one linear model. The table reports the mean square of

Table 4.19: Normalized information criteria of the PTGARCH(1,1;2, $d$ ) model fitted to the transactions' number for  $1 \leq d \leq 6$ .

Suggested time delay ( $d$ )	Possible threshold values ( $r$ )	Information criteria				The estimated parameters				
		$AIC$	$BIC$	$BIC2$	$HQIC$	$d_1$	$a_1$	$b_1$	$d_2$	$a_2$
5	5	-26.1311	-26.0972	-26.0483	-26.1180	1.1035	0.4490	3.409e-01	1.4928	0.8818
2	6	-26.1784	-26.1446	-26.0959	-26.1653	1.0656	0.7374	1.020e-02	2.1445	0.8392
1	7	-26.2797	-26.2460	-26.1973	-26.2666	0.5063	0.8520	1.532e-08	2.1542	0.8440
6	8	-26.2498	-26.2158	-26.1668	-26.2366	0.8720	0.6316	2.042e-01	2.6114	0.8066
1	9	-26.3835	-26.3498	-26.3011	-26.3704	0.7435	0.8366	1.635e-02	2.8784	0.8085
1	10	-26.3684	-26.3347	-26.2860	-26.3553	0.4482	0.8222	9.909e-02	3.0455	0.8052
2	11	-26.3375	-26.3038	-26.2551	-26.3244	0.6567	0.8686	1.103e-02	3.5807	0.7727
6	12	-26.3695	-26.3356	-26.2866	-26.3564	0.6791	0.7077	1.911e-01	2.9345	0.8144
6	13	-26.3934	-26.3594	-26.3104	-26.3802	1.0130	0.6415	2.302e-01	1.7387	0.9067

NOTE: The columns from three to six report the normalized information criteria. The other columns report the estimated parameters for the TGARCH model fitted to the transactions' number.

the Pearson residuals (MSPR) for each model. According to the results in Table 4.21 the data follows a nonlinear model. In particular, the values of the average of LM test for the model which have two regimes are very large compared to the value of the average of LM test for the linear model, and also the mean square error of the Pearson residuals are slightly smallest than the corresponding mean square error of the Pearson residuals obtained from the linear model fit. In addition, all experiment studies have shown that the BIC2 has a strong performance for the data generated from linear model in different sample sizes with probability one. However, the values of BIC2 are close to -26.40 when  $s = 2$  and  $s = 2, b_1 = 0$  compared with -26.36 for the linear model ( $s = 1$ ). In addition, the normalized AIC value for the linear model was clearly larger than those for the nonlinear models ( $s = 2, s = 2, b_1 = 0$  and  $s = 3$ ) (Al-Awadhi and Jolliffe 1998). An important issue in Table 4.21 is how to determine the appropriate number of regimes ( $s$ ). In Table 4.21, there are two models which have the autoregressive coefficients of the past value ( $\lambda_{t-1}$ ) disappeared. When  $a_1 = 0$ , the normalized AIC value is equal to -26.3509, in comparison to -26.3991 when  $a_2 = 0$ . Therefore,

Table 4.20: Normalized information criteria of the PTGARCH (1,1;2,d) model fitted to the transactions' number for  $1 \leq d \leq 6$ .

Suggested time delay (d)	Possible threshold values		Information criteria				The estimated parameters							
	$r_1$	$r_2$	AIC	BIC	HQIC	$d_1$	$a_1$	$b_1$	$d_2$	$a_2$	$b_2$	$d_3$	$a_3$	$b_3$
2	5	14	-26.4484	-26.3877	-26.4249	4.826e-08	0.9844	1.491e-02	0.3802	0.7939	0.16940	2.0841	0.5803	0.2681
2	6	13	-26.4539	-26.3931	-26.4303	5.897e-02	0.9615	8.677e-08	0.6422	0.8303	0.11970	2.4677	0.5294	0.2790
3	7	12	-26.4962	-26.4354	-26.4727	1.384e-01	0.9340	2.834e-02	0.3768	0.8881	0.07388	2.4144	0.5901	0.2380
2	8	11	-26.4867	-26.4260	-26.4632	8.902e-02	0.9106	6.744e-02	0.3734	0.9122	0.05370	1.4757	0.6615	0.2300

NOTE: The columns from four to seven report the normalized information criteria. The other columns report the estimated parameters for the TGARCH model fitted to the number of transactions.



we select the model which has the smallest of the two values of AIC. A similar thing happens when the parameters of autoregression of the past observation  $Y_{t-1}$  disappear. The smallest value of AIC is -26.4816 when  $b_1 = 0$ . The Table 4.21 can be reduced to the four estimated models as in Table 4.22.

Table 4.21: Linear and nonlinear autoregressive models with different number of thresholds for the number of transactions per minute for the Ericsson B stock on July 2, 2002 for  $1 \leq d \leq 6$ .

Suggested time delay ( $d$ )	Possible threshold values		Model	Average LM	Mean square error of the Pearson residuals		Information criteria				The estimated parameters							
	$r = r_1$	$r_2$			$s$	AIC	BIC	BIC2	HQIC	$d_1$	$a_1$	$b_1$	$d_2$	$a_2$	$b_2$	$d_3$	$a_3$	$b_3$
			1	0.442	2.3659	-26.41778	-26.3976	-26.36846	-26.40997	0.3076	0.8002	0.1693						
3	12		$s=2$	54.38	2.2495	-26.5005	-26.4600	-26.4014	-26.4848	0.1264	0.9079	0.06960	2.3965	0.5826	0.2456			
3	5		$s=2, a_1 = 0$	92.12	2.4288	-26.3509	-26.3171	-26.2683	-26.3378	5.6678		0.2458	1.0745	0.7019	0.2039			
3	12		$s=2, a_2 = 0$	54.15	2.3407	-26.3991	-26.3653	-26.3166	-26.3860	0.1455	0.8633	0.11140	7.1071		0.4597			
2	9		$s=2, b_1 = 0$	70.29	2.2563	-26.4816	-26.4478	-26.3991	-26.4685	0.27540	0.9363		2.1554	0.6489	0.1906			
6	13		$s=2, b_2 = 0$	91.36	2.3818	-26.3934	-26.3594	-26.3104	-26.3802	1.0130	0.6415	0.2302	1.7387	0.9067				
3	7	12	$s=3$		2.2591	-26.4962	-26.4354	-26.3476	-26.4727	0.1384	0.9340	0.02834	0.3768	0.8881	0.07388	2.4144	0.5901	0.2380

NOTE: The fourth column reports the number of regimes. The columns from five to seven report the normalized information criteria. The other columns report the estimated parameters for the TGARCH model fitted to the number of transactions.

Table 4.22 shows four different TGARCH models. The simulation studies show that AIC, BIC and HQIC perform quite well when one of the autoregressive parameters disappears. Therefore, the values of AIC, BIC and HQIC are close to -26.482, -26.45 and -26.469 respectively, for the model  $s = 2, b_1 = 0$ . When  $s = 2, a_2 = 0$ , the values of the three criteria are -26.40, -26.37 and -26.39 respectively. Hence, the model  $s = 2, b_1 = 0$  may be better.

Table 4.23 reports three different TGARCH models. The simulation results report that the AIC and HQIC are the most appropriate measures to use for nonlinear models  $s = 2$  and  $s = 2, b_1 = 0$ . However, the performance of AIC and HQIC criteria is very high for the nonlinear model ( $s = 2$ ) and good for nonlinear model ( $s = 2, b_1 = 0$ ). The NAIC and NBIC values for the nonlinear model ( $s = 2$ ) are smaller than those for nonlinear model ( $s = 2, b_1 = 0$ ). When ( $s = 2$ ), the coefficient of autoregression of the past observation ( $Y_{t-1}$ ) in the first regime is 0.0696. We can apply an approximate likelihood ratio test (LR) to test  $H_0 : b_1 = 0$ . In this case, the test statistic is 11.85 and the corresponding  $p$ -value of the test close to 0.001. Hence, there is clearly significant benefit at the 5% level of significance in using the bigger model ( $s = 2$ ). In light of the above, the nonlinear model ( $s = 2$ ) might be better than the nonlinear model  $s = 2, b_1 = 0$ .

The simulation results to select the number of regimes in different sample sizes show that the AIC and HQIC should be preferred for the nonlinear model ( $s = 2$  and  $s = 3$ ). The value of AIC in both models is close to -26.50 and the values of HQIC criteria close to -26.48 and -26.47, for nonlinear models  $s = 2$  and  $s = 3$  respectively. When  $s = 2$ , the value of HQIC is slightly smaller than that of  $s = 3$ . In addition, the value of MSEPR of nonlinear model ( $s = 2$ ) is slightly smaller than the corresponding MSEPR obtained from the nonlinear ( $s = 3$ ) model fit. Therefore, there is no point to select the nonlinear model ( $s = 3$ ). Hence, the estimated model of the number of transactions per minute for July 2, 2002 could

Table 4.22: Normalized information criteria of the PTGARCH (1,1;2,d) model fitted to the transactions' number for  $1 \leq d \leq 6$ .

Suggested time delay ( $d$ )	Possible threshold values		Model	Information criteria					The estimated parameters							
	$r = r_1$	$r_2$		AIC	BIC	BIC2	HQIC	$d_1$	$a_1$	$b_1$	$d_2$	$a_2$	$b_2$	$d_3$	$a_3$	$b_3$
3	12		S	-26.5005	-26.4600	-26.4014	-26.4848	0.1264	0.9079	0.06960	2.3965	0.5826	0.2456			
3	12		$s=2$	-26.3991	-26.3653	-26.3166	-26.3860	0.1455	0.8633	0.11140	7.1071	0.4597				
2	9		$s=2, a_2 = 0$	-26.4816	-26.4478	-26.3991	-26.4685	0.27540	0.9363		2.1554	0.6489	0.1906			
3	7	12	$s=2, b_1 = 0$	-26.4962	-26.4354	-26.3476	-26.4727	0.1384	0.9340	0.02834	0.3768	0.8881	0.07388	2.4144	0.5901	0.2380
			$s=3$													

NOTE: The fourth column reports the number of regimes. The columns from five to seven report the normalized information criteria. The other columns report the estimated parameters for the TGARCH model fitted to the number of transactions.

Table 4.23: Normalized information criteria of the PTGARCH (1,1;2,d) model fitted to the transactions' number for  $1 \leq d \leq 6$ .

Suggested time delay ( $d$ )	Possible threshold values		Model $s$	Information criteria				The estimated parameters								
	$r = r_1$	$r_2$		$AIC$	$BIC$	$BIC2$	$HQIC$	$d1$	$a1$	$b1$	$d2$	$a2$	$b2$	$d3$	$a3$	$b3$
3	12		$s=2$	-26.5005	-26.4600	-26.4014	-26.4848	0.1264	0.9079	0.06960	2.3965	0.5826	0.2456			
2	9		$s=2, b_1 = 0$	-26.4816	-26.4478	-26.3991	-26.4685	0.27540	0.9363		2.1554	0.6489	0.1906			
3	7	12	$s=3$	-26.4962	-26.4354	-26.3476	-26.4727	0.1384	0.9340	0.02834	0.3768	0.8881	0.07388	2.4144	0.5901	0.2380

NOTE: The fourth column reports the number of regimes. The columns from five to seven report the normalized information criteria. The other columns report the estimated parameters for the TGARCH model fitted to the number of transactions.

be nonlinear model ( $s = 2$ ). The estimated model can be expressed as

$$\hat{\lambda}_t = \begin{cases} 0.1264 + 0.0696Y_{t-1} + 0.9079\hat{\lambda}_{t-1}, & \text{if } Y_{t-3} \leq 12, \\ 2.3965 + 0.2456Y_{t-1} + 0.5826\hat{\lambda}_{t-1}, & \text{if } Y_{t-3} > 12. \end{cases}$$

Table 4.24 summarizes the results of the estimated parameters for nonlinear model ( $s = 2$ ). Table 4.24 shows that the relatively large standard error of the para-

Table 4.24: The estimated parameters for model (3.1).

Model	Threshold value	time delay	First regime, $n_1 = 467$			Second regime, $n_2 = 182$		
Parameters	$r$	$d$	$d_1$	$a_1$	$b_1$	$d_2$	$a_2$	$b_2$
Estimated value	12	3	0.1264	0.9079	0.0696	2.3965	0.5826	0.2456
Standard error			0.1135	0.0236	0.0184	0.4850	0.0507	0.0315

meters in the second regime reflect that there are few observations in that regime comparing with the first regime. The plot 4.3 shows the time series of the number of transactions in the period July 2-22, 2002. The first regime (blue line) has values from 0 to 12, while the second regime (red line) takes values above 12. There are 467 and 182 observations in the first and second regime respectively. The plot 4.4 shows the sequence of residuals and their cumulative periodogram plot, and the autocorrelation function of the Pearson residuals for the transaction data. The cumulative periodogram plot obviously indicates that the Pearson residuals obtained after the application of the nonlinear model ( $s = 2$ ) do not deviate from a white noise sequence. The plot of the autocorrelation function reveals no significant serial correlation in the sequence of residuals.

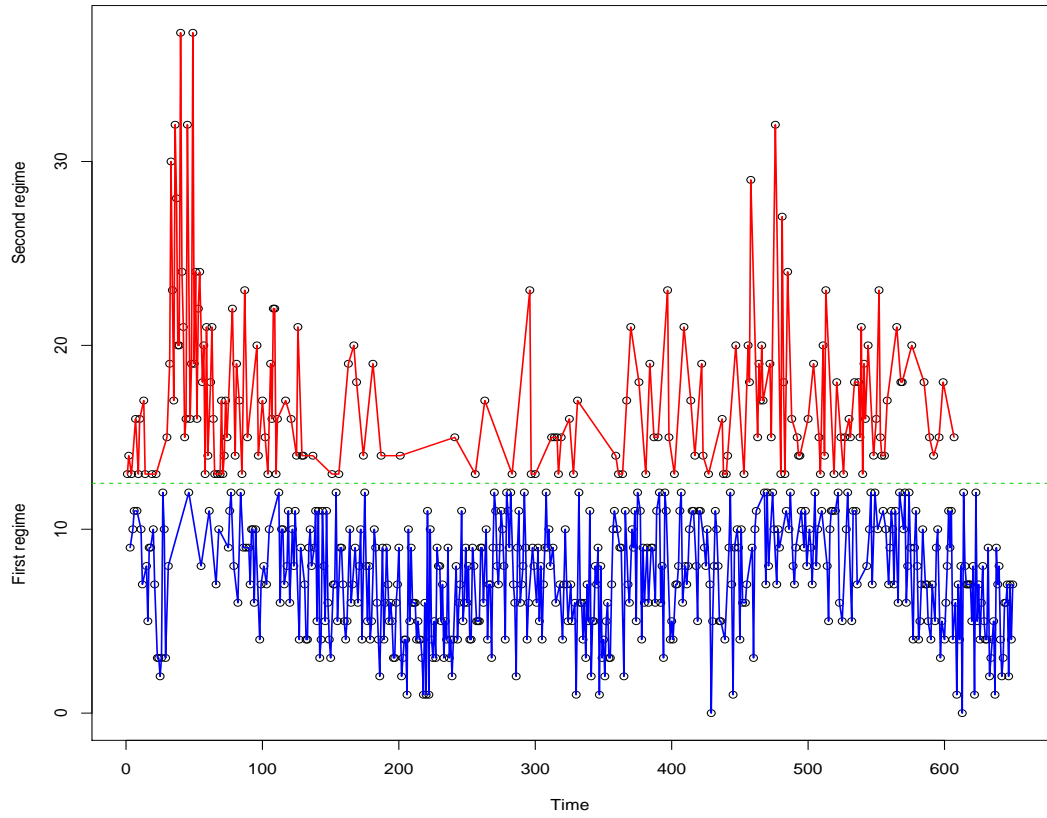


Figure 4.3: Time plot of two-regime of the number of transactions per minute for the stock Ericsson B in the period July 2-22, 2002. The green dotted line is the threshold value, the blue and the red lines correspond to the observations in the first and the second regime respectively.

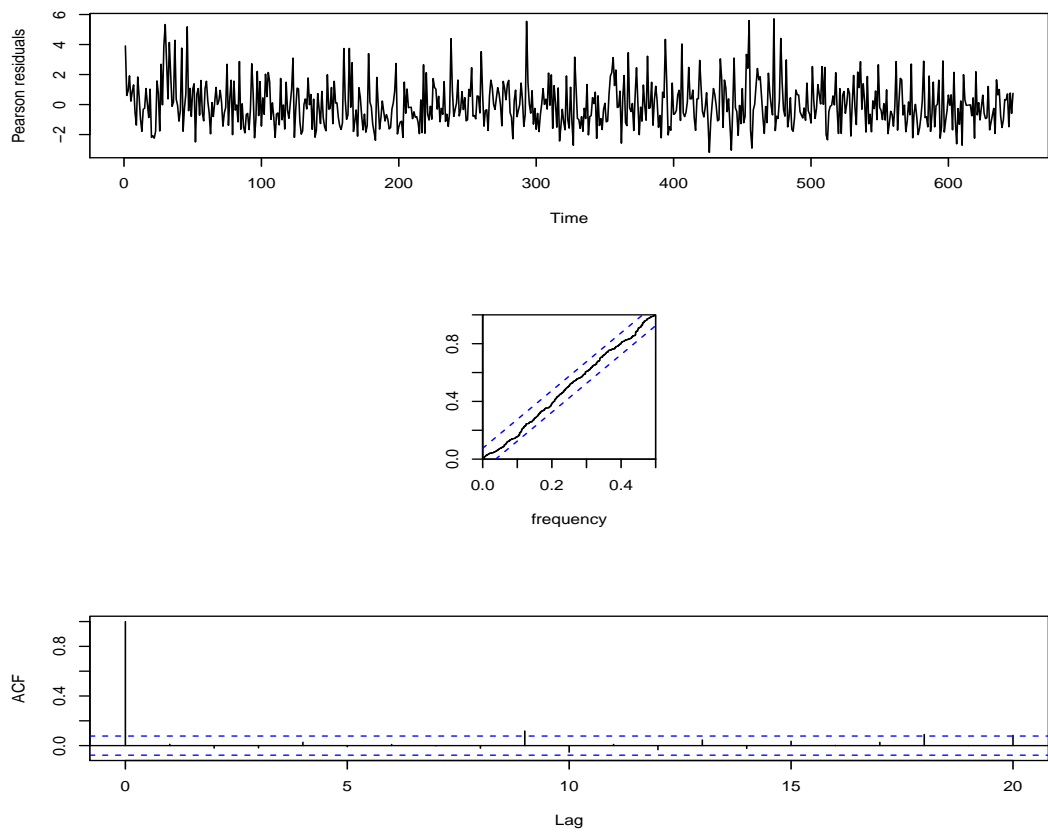


Figure 4.4: Pearson residuals (top) and their cumulative periodogram plot (middle), and the sample autocorrelation function (bottom) for the nonlinear model fitted to the number of transactions.



## 4.6 Summary

A PTGARCH model is a type of time series model that may be of use in various fields such as social and scientific research that focuses on processes. Many of these processes are characterized by some type of switches between two or more sub-models, where switching is based on the threshold variable ( $Y_{t-d}$ ).

The main purpose in PTGARCH modelling is how to determine the number of regimes. There are many ways to determine the number of regimes, one of them is by using various information criteria. However, there is no criterion which can be considered as the most suitable for this goal.

In this project, the performance of four information criteria was compared in the situation of defining the number of regimes. This study is an extension of existing studies by Gonzalo and Pitarakis (2002) who focused on moderate to large sample sizes, Strikholm and Teräsvirta (2006) who considered small to moderate sample sizes, and Hamaker (2009) who used also small to moderate sample sizes.

The threshold nonlinearity tests (LM test and LR test) are investigated to define the type of model, either one regime (linear) or more than one regime (nonlinear). If the model has one regime, it can be concluded that the BIC and BIC2 perform well with correct model detection percentages of close to 100%. In contrast, for nonlinear model with order ( $p = q = 1$ ), the AIC or HQIC is suggested, whereas when the autoregressive coefficients of the past observation ( $Y_{t-1}$ ) have disappeared, ( $b_i = 0, i = 1, 2$ ) (with different orders), it can be seen that in small sample sizes the AIC is the most suitable criterion, and HQIC is the most appropriate choice for large sample sizes. In addition, when the autoregressive parameters of the past values ( $\lambda_{t-1}$ ) disappeared, ( $a_i = 0, i = 1, 1$ ), the BIC and HQIC could be preferred. To sum up, the Table 4.25 shows the criteria that are appropriate measures for selecting the number of regimes in different sample sizes. Moreover, a real data example is given. So this chapter provides some practical

methods for determining the number of regimes of a PTGARCH model.

Table 4.25: The performance of diverse information criteria for selecting the number of regimes in different sample sizes.

Experiment	Sample Sizes	AIC		BIC		BIC2		HQIC	
		Linear ( $s=1$ )	Nonlinear ( $s=2$ )	Linear ( $s=1$ )	Nonlinear ( $s=2$ )	Linear ( $s=1$ )	Nonlinear ( $s=2$ )	Linear ( $s=1$ )	Nonlinear ( $s=2$ )
2	300		*	*		*		*	
	500		*	*		*		*	*
	800		*	*		*		*	*
	$N$	$b_1$ or $b_2 = 0$	$a_1$ or $a_2 = 0$	$b_1$ or $b_2 = 0$	$a_1$ or $a_2 = 0$	$b_1$ or $b_2 = 0$	$a_1$ or $a_2 = 0$	$b_1$ or $b_2 = 0$	$a_1$ or $a_2 = 0$
3	300	*							*
	500	*	*		*			*	*
	800	*			*			*	*
	$N$	$s=3$		$s=3$		$s=3$		$s=3$	
4	200,300		*						
	500		*						*
	800		*		*				*

NOTE: \* denotes the criteria which should be preferred, and  $s$  denotes the number of regimes.

# Chapter 5

## Simultaneous Determination of the Time Delay and the Threshold Value

### 5.1 Introduction

In the previous chapters, the parameter estimation and the determination of the threshold value  $r$  of the PTGARCH model, as well as the performance of various information criteria for the determination of the number of regimes were investigated in the cases of different sample sizes. The main goal of the present chapter is to complete the general framework of the PTGARCH model. We discuss how to choose a suitable delay parameter  $d$  as well as how to determine the appropriate threshold value and the delay parameter together when the number of regimes is given. To this end, we carry out some simulations and real data analysis.

## 5.2 Determination of the delay parameter $d$

Many studies have investigated the issue of determining the delay parameter  $d$  in a TAR model using information criteria such as AIC, BIC and test statistics such as the  $F$ -statistic in different sample sizes. Given the parameters of a TAR model, the value of threshold and the order of the AR process in each regime, Tong and Lim (1980) used the AIC value to determine the delay parameter  $d$ . Tsay (1989) suggested a different procedure to identify the delay parameter  $d$  which depends on the order of the AR model, by using the  $F$ -statistic before locating threshold values. Aide et al (1988), and Al-Awadhi and Jolliffe (1998) used the normalized value of AIC. Both studies have been used to describe and predict different phenomena such as Karst spring flow and climatic data respectively. Wong and Li (1998) and also De Gooijer (2001) concentrated on determining the delay parameter  $d$  and the order of the AR model in each regime by using the small sample corrected AIC, which was developed for TAR models.

In this part, we run some simulation studies to identify the delay parameter  $d$  for the PTGARCH models. We consider a two-regime ( $s = 2$ ) model defined by (3.1). To determine the delay parameter  $d$ , we use normalized information criteria (NIC) such as NIAC, NIBC and NIHQC, which are related to the information criteria. For example, Al-Awadhi and Jolliffe (1998) defined a NIAC as follows

$$\begin{aligned} NIAC(d) &= \frac{AIC(d, r)}{N - d}, \\ NIAC(\hat{d}) &= \min\{NIAC(d)\}, \end{aligned} \quad (5.1)$$

where  $N - d$  is the effective sample size and  $\hat{d}$  is the estimation of the delay parameter.

The main purpose here is to determine the delay parameter  $d$  and all the parameters of the model defined by (3.1) are given. For illustration, we carry out many simulations for the model defined by (3.1). The PTGARCH (1,1;2, $d$ )

models that were used to generate the data are

$$\lambda_t = \begin{cases} 0.4 + 0.3Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-d} \leq 2, \\ 0.5 + 0.4Y_{t-1} + 0.3\lambda_{t-1}, & \text{if } Y_{t-d} > 2, \end{cases} \quad (5.2)$$

and

$$\lambda_t = \begin{cases} 0.8 + 0.3Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-d} \leq 3, \\ 0.5 + 0.3Y_{t-1} + 0.5\lambda_{t-1}, & \text{if } Y_{t-d} > 3, \end{cases} \quad (5.3)$$

in Table 5.1 the results are obtained by averaging out of the smallest AIC, BIC and HQIC values for all simulations with different delay parameters (*i.e.*  $d = 1, 2, 3, 4$ ). Table 5.1 illustrates the relative rates of correct choices of the delay parameter  $d$  and the average of smallest information criteria (AIC, BIC and HQIC) values for the model described by equations (5.2) and (5.3). Clearly, the results show that the relative rates of correct choices improve as the sample size increases from 1000 to 2000. Table 5.1 reports the percentages of correctly selected delay parameters  $d$ , which improve when the sample size increases. The BIC values decrease when sample sizes increases.

Table 5.2 contains the outcomes for the data that were generated with the PTGARCH  $(1, 1; 2, d, r)$  model. To identify how the difference in the delay parameter affects the model selection results, we fix a specific threshold value. That is, the models have the same threshold value ( $r = 2$ ), while the delay parameters differ ( $1 \leq d \leq 4$ ). The models are given by (5.2). Table 5.2 shows the percentages of correct choices for each possible value of the delay parameter ( $1 \leq d \leq 4$ ). The percentages of suggested delay parameter improve significantly when the sample size increases. The percentage of false selection of the delay parameter  $d$  decreases when the sample size increases. In contrast, the percentage of the rest of the possible values of the delay parameter decreases, when the sample size increases. The average of the smallest information criterion reduces when the

delay parameters were 1 and 2 for all sample size cases. From the results in Table 5.2, it can be seen that the model with threshold value ( $r = 2$ ) and the delay parameters ( $d = 2$ ) is the best, because it has the smallest value in all information criteria.

Table 5.1: The percentage of correct choice of threshold delay for different PTGARCH(1,1:2, $d$ ) models and the average of AIC, BIC and HQIC values for different threshold values, coefficients and sample sizes.

$r$ and $d$	Regime 1	Regime 2	Size	Percentage of correct choices	AIC	BIC	HQIC	
$r = 2$	$d_1=0.4$	$d_2=0.5$	1000	79.1	0.9777	1.0071	0.9773	
	$d = 1$	$a_1=0.5$	$a_2=0.3$	1500	91.6	0.9750	0.9963	0.9750
		$b_1=0.3$	$b_2=0.4$	2000	95.8	0.9775	0.9943	0.9776
$d = 2$			1000	69.7	1.2955	1.3249	1.2951	
			1500	80.5	1.3028	1.3240	1.3027	
			2000	87.1	1.2985	1.3153	1.2985	
$d = 3$			1000	78.6	1.2630	1.2925	1.2626	
			1500	85.1	1.2578	1.2791	1.2578	
			2000	92.2	1.2544	1.2712	1.2545	
$d = 4$			1000	93.1	1.2283	1.2578	1.2279	
			1500	96.5	1.2332	1.2545	1.2332	
			2000	97.0	1.2320	1.2488	1.2321	
$r = 3$	$d_1=0.8$	$d_2=0.5$	1000	68.4	-1.5147	-1.4852	-1.5151	
	$d = 1$	$a_1=0.5$	$a_2=0.5$	1500	83.2	-1.5176	-1.4963	-1.5176
		$b_1=0.3$	$b_2=0.3$	2000	89.3	-1.5259	-1.5091	-1.5258
$d = 2$			1000	69.9	-1.5393	-1.5099	-1.5397	
			1500	81.4	-1.5590	-1.5378	-1.5591	
			2000	91.6	-1.5491	-1.5323	-1.5490	
$d = 3$			1000	70.8	-1.5723	-1.5428	-1.5727	
			1500	84.1	-1.5692	-1.5479	-1.5692	
			2000	91.3	-1.5787	-1.5619	-1.5786	
$d = 4$			1000	82.5	-1.5951	-1.5656	-1.5955	
			1500	91.0	-1.5946	-1.5733	-1.5946	
			2000	96.0	-1.6152	-1.5984	-1.6151	

NOTE: The results are based on 1000 simulations.

Table 5.2: Model selection percentages for different PTGARCH (1,1;2, $d$ ) models and different delay parameters  $d$  for different sample sizes.

Regime 1	Regime 2	Sample size	Possible values $d$	$AIC$ ( $\hat{r}, \hat{d}$ )	BIC ( $\hat{r}, \hat{d}$ )	$HQIC$ ( $\hat{r}, \hat{d}$ )	Percentage of correct choices
$d_1=0.5$	$d_2=0.4$	1000	<b>1</b>	<b>1.4189</b>	<b>1.4483</b>	<b>1.4185</b>	<b>65.4</b>
$a_1=0.4$	$a_2=0.5$		2	1.4275	1.4569	1.4271	2.4
$b_1=0.3$	$b_2=0.3$		3	1.4245	1.4540	1.4241	8.3
			4	1.4223	1.4517	1.4219	23.9
	$r=2$ $d=1$	1500	<b>1</b>	<b>1.4055</b>	<b>1.4267</b>	<b>1.4054</b>	<b>83.3</b>
			2	1.4148	1.4360	1.4147	0.9
			3	1.4126	1.4339	1.4126	4.0
			4	1.4106	1.4319	1.4106	11.8
		2000	<b>1</b>	<b>1.4034</b>	<b>1.4202</b>	<b>1.4035</b>	<b>90.7</b>
			2	1.4128	1.4296	1.4129	1.0
			3	1.4109	1.4277	1.4110	2.2
			4	1.4094	1.4262	1.4095	6.1
$d_1=0.5$	$d_2=0.4$	1000	1	1.1332	1.1626	1.1328	0.9
$a_1=0.4$	$a_2=0.5$		<b>2</b>	<b>1.1185</b>	<b>1.1480</b>	<b>1.1181</b>	<b>69.9</b>
$b_1=0.3$	$b_2=0.3$		3	1.1233	1.1527	1.1229	11.2
			4	1.1227	1.1521	1.1223	18.0
	$r=2$ $d=2$	1500	1	1.1256	1.1468	1.1255	0.3
			<b>2</b>	<b>1.1119</b>	<b>1.1331</b>	<b>1.1118</b>	<b>77.3</b>
			3	1.1169	1.1381	1.1168	9.5
			4	1.1170	1.1382	1.1169	12.9
		2000	1	1.1191	1.1359	1.1192	0.0
			<b>2</b>	<b>1.1056</b>	<b>1.1224</b>	<b>1.1057</b>	<b>86.5</b>
			3	1.11116	1.1279	1.11125	6.4
			4	1.11119	1.1279	1.11128	7.1
$d_1=0.5$	$d_2=0.4$	1000	1	1.1775	1.2069	1.1771	0.4
$a_1=0.4$	$a_2=0.5$		2	1.1706	1.2000	1.1702	3.7
$b_1=0.3$	$b_2=0.3$		<b>3</b>	<b>1.1624</b>	<b>1.1918</b>	<b>1.1620</b>	<b>75.1</b>
			4	1.1665	1.1959	1.1661	20.8
	$r=2$ $d=3$	1500	1	1.1624	1.1836	1.1623	0.5
			2	1.1558	1.1770	1.1558	2.3
			<b>3</b>	<b>1.1483</b>	<b>1.1696</b>	<b>1.1483</b>	<b>87.5</b>
			4	1.1533	1.1746	1.1533	9.7
		2000	1	1.1661	1.1829	1.1662	0.2
			2	1.1602	1.1770	1.1603	1.4
			<b>3</b>	<b>1.1532</b>	<b>1.1700</b>	<b>1.1532</b>	<b>89.8</b>
			4	1.1580	1.1748	1.1581	8.6
$d_1=0.5$	$d_2=0.4$	1000	1	1.3147	1.3441	1.3143	0.2
$a_1=0.4$	$a_2=0.5$		2	1.3123	1.3418	1.3119	2.5
$b_1=0.3$	$b_2=0.3$		3	1.3094	1.3388	1.3090	9.9
			<b>4</b>	<b>1.3061</b>	<b>1.3356</b>	<b>1.3057</b>	<b>87.4</b>
	$r=2$ $d=4$	1500	1	1.3243	1.3455	1.3242	0.5
			2	1.3228	1.3441	1.3228	2.4
			3	1.3205	1.3418	1.3205	8.9
			<b>4</b>	<b>1.3179</b>	<b>1.3391</b>	<b>1.3179</b>	<b>88.2</b>
		2000	1	1.3125	1.3293	1.3126	0.7
			2	1.3112	1.3280	1.3113	2.1
			3	1.3093	1.3261	1.3093	8.6
			<b>4</b>	<b>1.3069</b>	<b>1.3237</b>	<b>1.3069</b>	<b>88.6</b>

NOTE: The fourth column shows the suggested threshold delay  $d$ . Columns from fifth to seventh report the average of AIC, BIC and HQIC respectively. The last column shows the percentage of correct choices. The corrected delay parameter  $d$  detection percentages are printed in bold. Results are based on 1000 simulations.

## 5.3 Simultaneous determination of the delay parameter $d$ and the value of threshold $r$

In the previous section, we have treated the issue of determining the delay parameter  $d$ . We mentioned how to estimate the value of threshold  $r$  for a PTGARCH model in the Chapter 3. In the current section, we will focus on the determination of the value of threshold  $r$  and the delay parameter  $d$  at the same time. Many authors have adopted the Tong and Lim (1980) method to determine the threshold value. Tsay (1989) use the Tong and Lim method and scatter plot of the standardized predictive residuals against the threshold variable  $Y_{t-d}$  (in arranged autoregression). Here we adopt the following method:

1. Arrange the samples of the threshold variable  $Y_{t-d}$  in increasing order.
2. Divide the sample of the threshold variable into intervals which typically are defined to be from the  $u \times 100$ th percentile to the  $v \times 100$ th percentile of the data, e.g. 20th percentile to the 80th percentile of the data.
3. Choice of  $u$  and  $v$  which should guarantee that there are enough data falling into each of the sub-models, because the extreme threshold candidate value  $(u, v)$  means there are not enough data to estimate the parameters. Therefore, each regime contains at least 10% of the total observation number (Strikholm and Teräsvirta (2006)).
4. Calculate the normalized information criterion (NIC) which is defined by (5.1) for each threshold candidate value at the probable delay parameters.
5. Compute the minimum values of the normalized information criteria for each pair of threshold candidate value and probable delay parameter, and the smallest value of the criteria gives the required estimated threshold value and the delay parameter.



We give a few simulation results to illustrate the determination of the value of threshold  $r$  when the maximum of the proposed values of the delay parameter is 4, (*i.e.*  $d = 1, 2, 3, 4$ ) and the number of regimes are fixed ( $s = 2$ ). A simulated time series generated from model defined by (5.3) with different parameters for different lengths has very high frequency. Therefore, applying the 30% rule to the current simulation indicates that at least 300, 450 and 600 observations need to fall in each regime for sample size 1000, 1500 and 2000. If the threshold value estimate is close to the minimum or the maximum observation, may be not reliable because of the small sample size in one of the regimes.

In Table 5.3, the data have been generated by different PTGARCH  $(1,1;2,d)$  models. The table illustrates the relative rates of correct choices of the delay parameter  $d$  and the value of threshold  $r$  for the model described by equation (5.3). Clearly, the results show that the relative rates of correct choices for both  $d$  and  $r$  improve as the sample size increases from 1000 to 2000. Table 5.3 also reports the percentages of correctly selected delay parameters  $d$  and of the threshold value  $r$  becomes larger as the sample size increases, whereas the percentages of false choices of the delay parameters  $d$  and of the threshold value  $r$  become smaller as the sample size increases. All results in Table 5.3 are based on the equation (5.1).

Table 5.3: The percentage of correct choices of threshold delay  $d$  and the value of threshold  $r$  (together) for different nonlinear PTGARCH(1,1:2, $d$ ) models and different sample sizes.

Regime 1	Regime 2	Sizes $n$	proposed values $d$ $d$	Percentage of possible values $r$		
				$r = 2$	$r = 3$	$r = 4$
$d_1=0.8$	$d_2=0.5$	1000	<b>1</b>	3.1	<b>65.6</b>	5.3
$a_1=0.5$	$a_2=0.5$		2	1.1	9.2	2.4
$b_1=0.3$	$b_2=0.3$		3	0.8	2.7	2.4
$r = 3$	$d=1$		4	0.9	2.9	3.4
		1500	<b>1</b>	2.5	<b>82.8</b>	2.3
			2	0.3	5.0	1.2
			3	0.2	1.7	1.5
			4	0.0	0.6	1.8
		2000	<b>1</b>	1.5	<b>90.5</b>	1.5
			2	0.2	3.6	1.0
			3	0.0	0.5	0.3
			4	0.0	0.4	0.4
$d_1=0.8$	$d_2=0.5$	1000	1	0.6	3.0	0.6
$a_1=0.5$	$a_2=0.5$		<b>2</b>	3.4	<b>68.0</b>	4.4
$b_1=0.3$	$b_2=0.3$		3	1.6	8.2	2.2
$r=3$	$d=2$		4	0.4	4.9	2.0
		1500	1	0.2	2.4	0.1
			<b>2</b>	1.3	<b>82.4</b>	3.4
			3	0.0	6.6	1.5
			4	0.2	0.9	0.9
		2000	1	0.2	1.3	0.4
			<b>2</b>	0.7	<b>89.2</b>	1.8
			3	0.0	4.0	0.6
			4	0.1	1.2	0.5
$d_1=0.8$	$d_2=0.5$	1000	1	0.4	0.7	0.3
$a_1=0.5$	$a_2=0.5$		2	0.3	3.1	0.7
$b_1=0.3$	$b_2=0.3$		<b>3</b>	2.4	<b>73.3</b>	4.7
$r=3$	$d=3$		4	2.2	9.8	1.8
		1500	1	0.0	0.1	0.1
			2	0.3	3.0	0.2
			<b>3</b>	1.8	<b>84.6</b>	3.1
			4	0.2	5.8	0.8
		2000	1	0.0	0.4	0.0
			2	0.0	2.0	0.1
			<b>3</b>	0.8	<b>91.2</b>	1.7
			4	0.3	2.7	0.8
$d_1=0.8$	$d_2=0.5$	1000	1	0.2	0.0	0.0
$a_1=0.5$	$a_2=0.5$		2	0.1	0.2	0.6
$b_1=0.3$	$b_2=0.3$		3	1.1	4.0	0.7
$r=3$	$d=4$		<b>4</b>	4.0	<b>83.7</b>	5.1
		1500	1	0.1	0.0	0.1
			2	0.1	0.1	0.0
			3	0.5	2.6	0.2
			<b>4</b>	1.5	<b>91.4</b>	3.4
		2000	1	0.0	0.0	0.0
			<b>2</b>	0.1	0.3	0.0
			3	0.0	1.1	0.1
			<b>4</b>	1.0	<b>96.4</b>	1.0

NOTE: Experiment results for two-regime PTGARCH model with  $p = q = 1$ . The fourth column reports the proposed values of  $d$ . The other three columns show percentage of possible values of  $r$ . The percentages of correct threshold delay  $d$  and threshold value  $r$  are printed in bold. Results are based on 1000 simulations.

## 5.4 A real data example

We illustrate the application of the PTGARCH  $(1,1,2,d)$  model to data of daily admissions for asthma to a single hospital (at Campbelltown) in the Sydney metropolitan area from 1st January 1990 to 31st December 1993. There are 1461 observations available. The sample mean of the series is 1.9391 and the variance is equal to 2.7025. The data has previously been analyzed by Davis et al. (1999 and 2000) using a GLARMA and generalized linear model analysis respectively. Figure 5.1 illustrates the data, histogram of asthma counts and the respective autocorrelation function. The plot shows that there is some type of periodicity in these data and for relatively large lag values the autocorrelation function is still significant.

We present the estimation methods with the asthma count, using the normalized Akaike's information criterion (NAIC), which is related to AIC. Wilks (1985) and Katz and Skaggs (1981) have discussed the AIC in the framework of AR models. Table 5.4 shows the normalized AIC values of TAR model fitted to counts of daily admissions for asthma of a single hospital, using the NAIC method with the interval of threshold value approximately between the 25th percentile and the 75th percentile of the data. Therefore, the threshold values are between 1 and 4, while the maximum time delay is equal to 7, and the number of regimes are given. Table 5.5 shows estimated time delay and their normalized information criteria for each possible threshold value. The table illustrates that the normalized information criteria values are smallest when  $d = 3$  and  $r = 2$ , so the estimation of threshold value and of the delay parameter are 2 and 3 respectively. Table 5.6 reports the estimated parameters for the model described by equation (3.1). It shows that the relatively small standard deviations of the parameters in the first regime reflect that there are a lot of observations in that regime. After fitting the full model with the threshold value and time delay to the original data, we conclude with the PTGARCH  $(1,1,2)$  model for the counts of daily admissions

for asthma to a single hospital:

$$\hat{\lambda}_t = \begin{cases} 0.0651 + 0.0251Y_{t-1} + 0.9233\hat{\lambda}_{t-1}, & \text{if } Y_{t-3} \leq 2, \\ 0.2943 + 0.1560Y_{t-1} + 0.7476\hat{\lambda}_{t-1}, & \text{if } Y_{t-3} > 2. \end{cases}$$

The *aveLM* test value of the model is 15.5795, and the corresponding critical values of significance levels 10%, 5% and 1% of the test are 6.251, 7.815, 11.345 respectively, so there is clearly significant benefit at the 10%, 5% and 1% level of significance in using the nonlinear model. In addition, for the asthma data and for the linear model application, the MSPR is equal to 1.1465. The AIC (BIC) of the fit is equal to 1.1334 (1.1442, respectively), while the MSPR and the values of AIC and BIC for the nonlinear model application are 1.1287, 1.1179 and 1.1396 respectively which are slightly smaller than the corresponding MSPR, AIC and BIC are obtained from the linear model fit. Figure 5.2 illustrates the time series of asthma counts in a single hospital. The time plot shows that the first regime has more observations than the second regime. The plot 5.3 shows the sequence of residuals and their cumulative periodogram plot, as well as their autocorrelation function of the asthma data. The plot of the autocorrelation function reveals no significant serial correlation in the sequence of residuals.

## 5.5 Summary

The main target in this project is to estimate the parameter, number of regimes, threshold value and the delay parameters in the PTGARCH model. In this chapter, many simulations have been carried out to identify the delay parameter, and the delay parameter and threshold value simultaneously, using normalized information criteria. Moreover, an example of an application of the model to a real time series is given. Applying the method in this chapter, one can build a complete structure of the PTGARCH model.

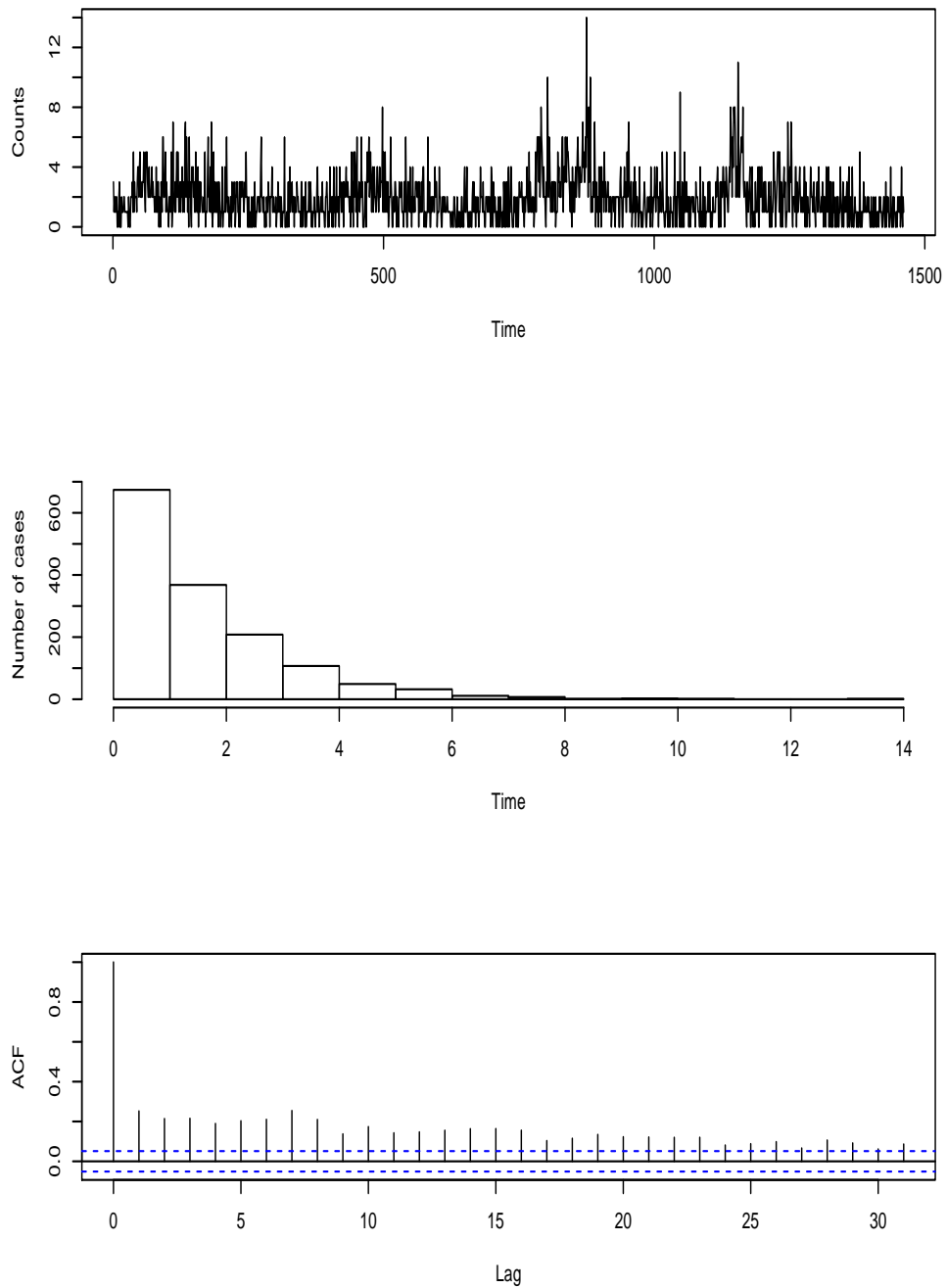


Figure 5.1: From top to the bottom: time-series plot, histogram and sample auto-correlation function of the asthma counts.

Table 5.4: Normalized information criteria for the PTGARCH (1,1,2,1) model fitted to counts of daily admissions for asthma of a single hospital.

Suggested threshold value ( $r$ )	Possible time delay ( $d$ )	Normalized information criteria			
		$AIC$	$BIC$	$BIC2$	$HQIC$
1	1	1.1369	1.1586	1.1886	1.1450
	2	1.1352	1.1569	1.1869	1.1433
	3	1.1286	1.1503	1.1803	1.1367
	4	1.1346	1.1563	1.1863	1.1427
	5	1.1309	1.1527	1.1827	1.1390
	6	1.1300	1.1518	1.1819	1.1382
	7	1.1264	1.1483	1.1783	1.1346
2	1	1.1365	1.1582	1.1881	1.1446
	2	1.1333	1.1551	1.1850	1.1414
	3	1.1202	1.1420	1.1720	1.1283
	4	1.1350	1.1567	1.1867	1.1431
	5	1.1316	1.1534	1.1834	1.1397
	6	1.1275	1.1493	1.1794	1.1357
	7	1.1286	1.1505	1.18052	1.1368
3	1	1.1362	1.1580	1.1879	1.1443
	2	1.1385	1.1602	1.1902	1.1466
	3	1.1263	1.1481	1.1781	1.1341
	4	1.1388	1.1606	1.1906	1.1469
	5	1.1340	1.1558	1.1858	1.1421
	6	1.1319	1.1537	1.1838	1.1401
	7	1.1297	1.1515	1.1815	1.1378
4	1	1.1374	1.1591	1.1890	1.1455
	2	1.1384	1.1601	1.1901	1.1465
	3	1.1270	1.1488	1.1788	1.1351
	4	1.1387	1.1605	1.1905	1.1469
	5	1.1682	1.1900	1.2200	1.1764
	6	1.1322	1.1540	1.1840	1.1403
	7	1.1298	1.1516	1.1817	1.1379

Table 5.5: Normalized information criteria for the PGARCH (1,1,2,1) model fitted to counts of daily admissions for asthma to a single hospital for the estimated the threshold values and time delays.

Estimated threshold value ( $r$ )	Estimated time delay ( $d$ )	Normalized information criteria			
		$AIC$	$BIC$	$BIC2$	$HQIC$
1	7	1.1264	1.1483	1.1783	1.1346
2	3	1.1202	1.1420	1.1720	1.1283
3	3	1.1263	1.1481	1.1781	1.1345
4	3	1.1270	1.1488	1.1788	1.1351

Table 5.6: The estimated parameters for model (3.1) and their standard errors.

Model	First regime, $n_1 = 1042$			Second regime, $n_2 = 419$				
Parameters	$r$	$d$	$d_1$	$a_1$	$b_1$	$d_2$	$a_2$	$b_2$
Estimated value	2	3	0.06509	0.9233	0.02508	0.2943	0.7476	0.1560
Standard error			0.0312	0.0243	0.0176	0.0872	0.0446	0.0257

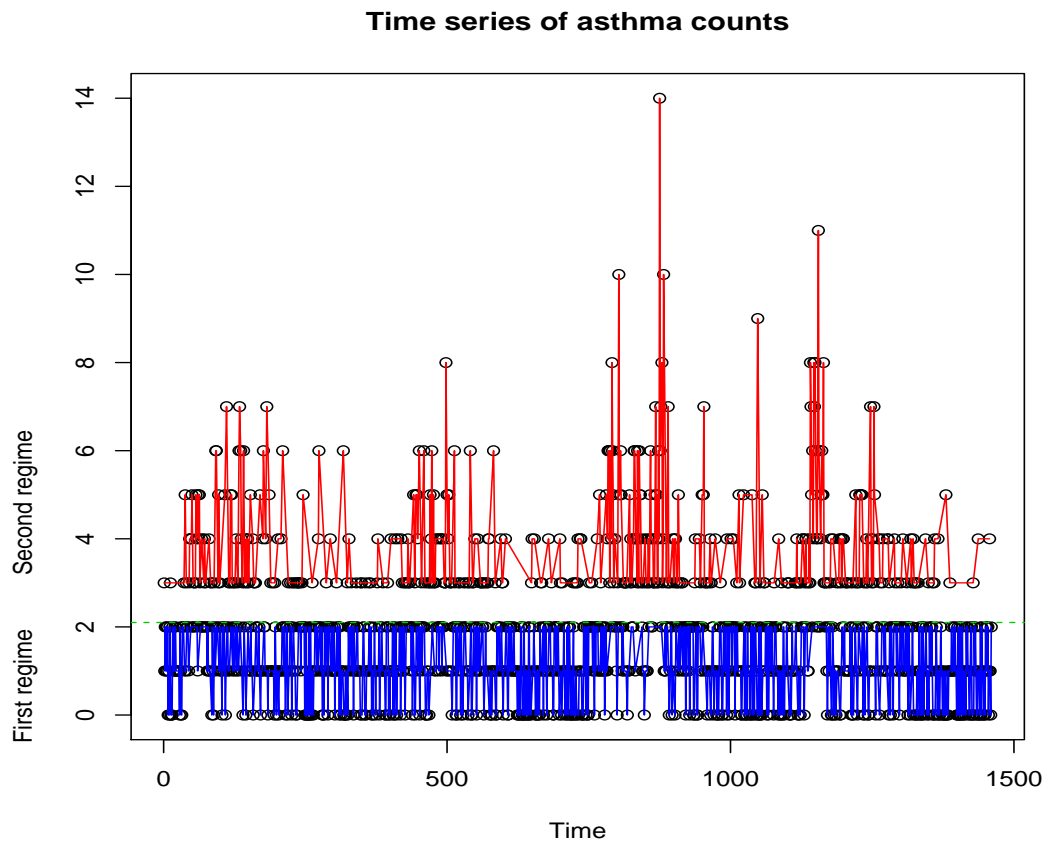


Figure 5.2: Time plot of two-regime of the counts of daily admissions for asthma to a single hospital. The green dotted line is the threshold value. The blue and the red lines correspond to the observations in the first and the second regime respectively.

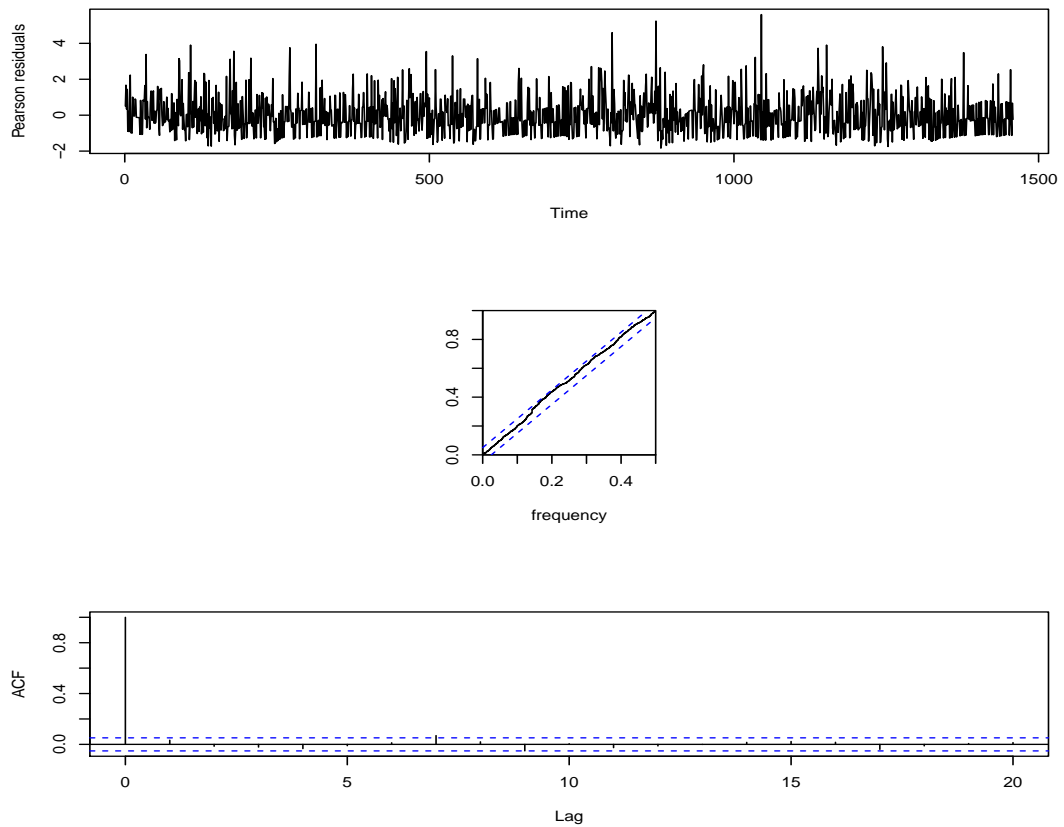


Figure 5.3: Pearson residuals (top), their cumulative periodogram plot (middle), and the sample autocorrelation function of Pearson residuals (bottom) for the nonlinear model fitted to counts of daily admissions for asthma to a single hospital.



# Chapter 6

## Poisson Markov Switching Model

### 6.1 Introduction

In the past few decades, there has been growing interest in nonlinear time series models, see Hamilton (1989), Tong and Series (1990) and Granger and Terasvirta (1993) for such examples. The Markov switching model of Hamilton (1989) is one of the most popular nonlinear models in time series. It consists of many regimes which can describe the time series pattern in various states. The switching between states is according to an unobservable (hidden) state, and each state follows a linear model. In this chapter, we discuss Poisson Markov switching models for count time series. Because the conditional mean coincides with the conditional variance for a Poisson process, we call our model a Markov switching Poisson generalized autoregressive conditional heteroscedastic (MS-PGARCH) model. We study the probability structure (ergodicity and stationarity) of the MS-PGARCH model first, and then discuss its maximum likelihood estimation. Finally, in order to give numerical evidence, we run a few simulation studies and fit a simple MS-PGARCH model to some real data sets.

## 6.2 The Markov Switching Model of Conditional Variance (Mean)

### 6.2.1 Definition of the model

Suppose that  $\{Y_t\}$  is a nonnegative integer-valued time series. Consider the model defined by

$$Y_t | \mathcal{F}_{t-1}^{Y, \lambda, S} \sim \text{Poisson}(\lambda_t), \quad \lambda_t = c(S_t) + a(S_t)\lambda_{t-1} + b(S_t)Y_{t-1} \quad (6.1)$$

where  $\mathcal{F}_{t-1}$  denotes the information set available at time  $t-1$ , which is a  $\sigma$ -field generated by  $\{(Y_{t-1}, S_{t-1}), (Y_{t-2}, S_{t-2}), \dots, (Y_0, S_0), (\lambda_0, S_0)\}$ .  $S_t$  is a stationary Markov chain with finite state space  $S = \{1, 2, \dots, m\}$  and an irreducible and aperiodic  $m \times m$  transition matrix,  $P$ , with typical element  $p_{ij} = Pr(S_t = j | S_{t-1} = i)$ , that is,

$$P = [p_{ij}] = [Pr(S_t = j | S_{t-1} = i)], i, j = 1, 2, \dots, m.$$

The stationary distribution  $\{S_t\}$  is denoted by  $\pi = (\pi_1, \pi_2, \pi_3, \dots, \pi_m)^\tau$ . The states of  $S_t$  represent the different regimes of the model. The coefficients  $(c_i, a_i, \text{ and } b_i, i = 1, 2, \dots, m)$  are assumed to be strictly positive and  $P_{ij} > 0$ . We call the model defined by (6.1) a Markov-switching Poisson generalized autoregressive heteroscedastic (MS-PGARCH) model, denoted by MS-PGARCH  $(m; 1, 1)$ , where  $m$  is called the number of regimes of the model.

The model (6.1) can be rewritten in a form of Poisson process by assuming that the time series of counts  $\{Y_t\}$  is equal to the number of events  $N_t(\lambda_t)$  of  $N_t(\cdot)$  in

the time interval  $[0, \lambda_t]$  as follows

$$\begin{cases} Y_t = N_t(\lambda_t) \\ \lambda_t = c(S_t) + a(S_t)\lambda_{t-1} + b(S_t)Y_{t-1} \end{cases} \quad (6.2)$$

for  $t \geq 1$  and both  $Y_0$  and  $\lambda_0$  are fixed.  $N_t(\lambda_t)$  denotes a Poisson process with intensity  $\lambda_t$ , and  $\lambda_t$  is defined by (6.1).

### 6.2.2 Ergodicity of a perturbed model

We study the stationarity and ergodicity of  $\{Y_t\}$  by proving these properties for  $\{(\lambda_t, S_t)\}$ , given a suitable initial distribution for  $(\lambda_0, S_0)$ . However, establishing  $\psi$ -irreducibility and finding a small set is quite complicated because of the fact that  $\lambda_t$  has discrete-valued random innovations. Given the value of  $\lambda_0$ , the set of possible values for  $\lambda_1$  is countable. In fact, the set of states that are reachable from a fixed starting state is also countable, and distinct initial values can have distinct sets of reachable locations. To avoid this issue, we consider a model with  $\varepsilon$ -perturbation defined by

$$Y_t^n = N_t(\lambda_t^n), \quad \lambda_t^n = c(S_t) + a(S_t)\lambda_{t-1}^n + b(S_t)Y_{t-1}^n + \varepsilon_{t,n}, \quad (6.3)$$

where

$$\varepsilon_{t,n} = c_n U_t, \quad c_n > 0, \quad c_n \rightarrow 0, \quad \text{as } n \rightarrow \infty,$$

and  $\{U_t\}$  is a sequence of iid uniform random variables on  $(0, 1)$  and such that  $U_t$  is independent of  $\mathcal{F}_{t-1}^{Y, \lambda, S}$ .

We have the following results on the ergodicity of model (6.3). There it is first proved that the unobserved process  $\{(\lambda_t^n, S_t)\}$  is geometrically ergodic.

Before stating main results, we need some notations. Let

$$M_u := [p_{ji}(a(i) + b(i))^u], \quad i, j = 1, 2, \dots, m; \quad u \geq 0.$$

Define a norm  $\|D\| = \sum_{ij} d_{ij}$  for a nonnegative vector or matrix  $D = [d_{ij}]$ , and  $\rho(D)$  for the spectral radius of matrix  $D$ . Let  $A := \{i : 0 < a(i) < 1, 1 \leq i \leq m\}$  and  $\lambda^* := \min_{i \in A} \{c(i)/(1 - a(i))\}$ , and  $\phi$  and  $v$  denote the Lebesgue and the counting measure of a set, respectively. If  $A \neq \emptyset$ , without loss of generality, write  $\lambda^* := c(1)/(1 - a(1))$ . Moreover, for convenience, write

$$A_t := [1(S_t = i, S_{t-1} = j)a(i)], \quad B_t := [1(S_t = i, S_{t-1} = j)b(i)],$$

$$C_{t,kl} := [1(S_t = i, S_{t-1} = j)a^k(i)b^l(i)], \quad M_{kl} := [p_{ji}a^k(i)b^l(i)], i, j = 1, \dots, m,$$

where  $1(\cdot)$  is the indicator function. Again, let  $\mathbf{1}_{S_t} := (1(S_t = 1), \dots, 1(S_t = m))^\tau$ ,  $|J|$  denote the length of the interval  $J$ , and  $Z_t^n := (\lambda_t^n, S_t)$ .

We are now ready to state the ergodicity of the process  $\{Y_t^n\}$  defined by (6.3). But, for the proof of ergodicity, we need the following lemma.

**Lemma 6.1** Let  $\{Z_t^n\}$  be a Markov chain defined by (6.3). If  $A \neq \emptyset$ , then every point in  $D = \{(\lambda, s) \in \mathbb{R}_+ \times \mathbb{R}_+ : (\lambda, s) \in \bigcup_{i=1}^m D_i\}$  is reachable, where  $D_i = \{(\lambda, s) \in \mathbb{R}_+ \times \mathbb{R}_+ : \lambda \geq c(i) + a(i)\lambda^*, s \in T_i\}$  with  $T_i = \{j \in S : c(j) + a(j)\lambda^* \leq c(i) + a(i)\lambda^*\}$ ,  $i = 1, 2, \dots, m$ .

**Proof.** Notice that if  $\lambda_{t-1}^n = \lambda \geq \lambda^*$ , then

$$\begin{aligned} \lambda_t^n &= c(S_t) + a(S_t)\lambda + b(S_t)N_{t-1}(\lambda) + \varepsilon_{t,n} \\ &\geq c(S_t) + a(S_t)\lambda = (c(S_t) + a(S_t)\lambda)(1(a(S_t) \geq 1) + 1(a(S_t) < 1)) \\ &= (c(S_t) + a(S_t)\lambda)1(a(S_t) \geq 1) + (c(S_t) + a(S_t)\lambda)1(a(S_t) < 1) \\ &\geq \lambda^*1(a(S_t) \geq 1) + (c(S_t) + a(S_t)\lambda^*)1(a(S_t) < 1) \\ &= \lambda^*1(a(S_t) \geq 1) + (1 - a(S_t)) \left( \frac{c(S_t)}{1 - a(S_t)} + \frac{a(S_t)}{1 - a(S_t)}\lambda^* \right) 1(a(S_t) < 1) \\ &\geq \lambda^*1(a(S_t) \geq 1) + (1 - a(S_t)) \left( \lambda^* + \frac{a(S_t)}{1 - a(S_t)}\lambda^* \right) 1(a(S_t) < 1) = \lambda^*. \end{aligned}$$

Hence, if  $\lambda_0^n = \lambda \geq \lambda^*$  is true, we know that  $\lambda_t^n \geq \lambda^*$  for all  $t \geq 1$ . Now consider a

point  $(d, s_i) \in D_i \subset D$ . We will show that there exists a  $j \geq 1$  and an interval  $J_d$  such that  $d \in J_d$ ,  $|J_d| \leq \varepsilon$  for sufficiently small  $\varepsilon > 0$ , and  $P^j(z_0, G) > 0$ , where  $z_0 = (\lambda, s_0) \in D$  and  $G = \{(\lambda, s) \in D : \lambda \in J_d, s = s_i\}$ .

First assume that  $d = c(i) + a(i)\lambda^*$  and  $s_i = i$ . In this case, define  $j$  to be the smallest positive integer such that  $a(1)^j a(i)(\lambda - \lambda^*) < \varepsilon$ , since  $a(1) < 1$ . Consider a path such that  $\lambda_0^n = \lambda$ ,  $S_0 = s_0$ ,  $Y_0^n = Y_1^n = \dots = Y_j^n = 0$ ,  $S_1 = S_2 = \dots = S_j = 1$ , and  $S_{j+1} = i$ . This implies that

$$\lambda_{j+1}^n = c(i) + a(i)\lambda^* + a(1)^j a(i)(\lambda - \lambda^*) + a(i) \sum_{i=0}^{j-1} a(1)^i \varepsilon_{j-i,n} + \varepsilon_{j+1,n}. \quad (6.4)$$

From (6.4) together with

$$\begin{aligned} P(S_1 = S_2 = \dots = S_j = 1, S_{j+1} = i | Z_0^n = z_0) &> 0, \\ P(Y_0^n = Y_1^n = \dots = Y_j^n = 0 | Z_0^n = z_0, W_{ij}) &> 0, \quad \text{and} \\ P(a(i) \sum_{i=0}^{j-1} a(1)^i \varepsilon_{j-i,n} + \varepsilon_{j+1,n} < \varepsilon - a(1)^j a(i)(\lambda - \lambda^*)) &> 0, \end{aligned}$$

we conclude that for  $G = \{(\lambda, s) \in D : \lambda \in J_d, s = i\}$  with  $J_d = [d, d + \varepsilon)$ ,

$$\begin{aligned} P^{j+1}(z_0, G) &\geq P(Y_0^n = Y_1^n = \dots = Y_j^n = 0, S_1 = S_2 = \dots = S_j = 1, S_{j+1} = i, \\ &\quad a(i) \sum_{i=0}^{j-1} a(1)^i \varepsilon_{j-i,n} + \varepsilon_{j+1,n} < \varepsilon - a(1)^j a(i)(\lambda - \lambda^*) | Z_0^n = z_0) \\ &= P(Y_0^n = \dots = Y_j^n = 0 | Z_0^n = z_0, W_{ij}) \\ &\quad \times P(S_1 = \dots = S_j = 1, S_{j+1} = i | Z_0^n = z_0) \\ &\quad \times P(a(i) \sum_{i=0}^{j-1} a(1)^i \varepsilon_{j-i,n} + \varepsilon_{j+1,n} < \varepsilon - a(1)^j a(i)(\lambda - \lambda^*)) > 0, \end{aligned}$$

where  $W_{ij} = \{\omega : S_1 = \dots = S_j = 1, S_{j+1} = i, a(i) \sum_{i=0}^{j-1} a(1)^i \varepsilon_{j-i,n} + \varepsilon_{j+1,n} < \varepsilon - a(1)^j a(i)(\lambda - \lambda^*)\}$ .

Next assume that  $d = c(i) + a(i)\lambda^*$  and  $s_i = k \neq i$ . Obviously,  $c(k) + a(k)\lambda^* \leq$

$c(i) + a(i)\lambda^*$ . Hence  $\delta := c(i) + a(i)\lambda^* - c(k) - a(k)\lambda^* \geq 0$ . If  $\delta = 0$ , replace  $i$  with  $k$  in (6.4), the assertion is also true. If  $\delta > 0$ , define  $j$  to be the smallest positive integer such that  $a(1)^{j-1}a(k)b(1) < \varepsilon$  and  $a(1)^ja(k)(\lambda - \lambda^*) < \delta/2$ , since  $a(1) < 1$ . Consider a path such that  $\lambda_0^n = \lambda$ ,  $S_0 = s_0$ ,  $Y_0^n = N$ ,  $Y_1^n = \dots = Y_j^n = 0$ ,  $S_1 = S_2 = \dots = S_j = 1$ , and  $S_{j+1} = k$ . This shows that

$$\begin{aligned} \lambda_{j+1}^n &= c(k) + a(k)(\lambda^* + a(1)^j(\lambda - \lambda^*)) + a(1)^{j-1}b(1)N + \sum_{i=0}^{j-1} a(1)^i \varepsilon_{j-i,n} \\ &\quad + \varepsilon_{j+1,n} \\ &= \lambda_{k,j+1}^n(N) + a(k) \sum_{i=0}^{j-1} a(1)^i \varepsilon_{j-i,n} + \varepsilon_{j+1,n}, \end{aligned}$$

where  $\lambda_{k,j+1}^n(N) := c(k) + a(k)(\lambda^* + a(1)^j(\lambda - \lambda^*)) + a(1)^{j-1}b(1)N$ . If there exists an  $N$  such that  $\lambda_{k,j+1}^n(N) = d$ , in an analogous way as above  $s = i$  case, one can show that  $P^{j+1}(z_0, G) > 0$ , where  $G = \{(\lambda, s) \in D : \lambda \in J_d, s = k\}$  with  $J_d = [d, d + \varepsilon)$ . Otherwise, let  $N := N_{j+1}$  be the least integer such that  $\lambda_{k,j+1}^n(N - 1) < d < \lambda_{k,j+1}^n(N)$  ( $\lambda_{k,j+1}^n(0) < d$ ). Taking  $J_d = [\lambda_{k,j+1}^n(N - 1), \lambda_{k,j+1}^n(N))$ , we easily obtain that  $P^{j+1}(z_0, G) > 0$  and  $|J_d| = a(1)^{j-1}a(k)b(1) < \varepsilon$ , where  $G = \{(\lambda, s) : \lambda \in J_d, s = k\}$ .

Now, we consider the case that  $d > c(i) + a(i)\lambda^*$ . If  $s_i = i$ , define  $j$  to be the smallest positive integer such that  $a(1)^{j-1}b(1)a(i) < \varepsilon$  and  $a(1)^ja(i)(\lambda - \lambda^*) < (d - c(i) - a(i)\lambda^*)/2$ , and consider a path such that  $Z_0^n = z_0$ ,  $Y_0^n = N$ ,  $Y_1^n = Y_2^n = \dots = Y_j^n = 0$ ,  $S_1 = S_2 = \dots = S_j = 1$ , and  $S_{j+1} = i$ . We can also obtain that there exists a  $j \geq 1$  and  $|J_d| \leq \varepsilon$  such that  $P^{j+1}(z_0, G) > 0$ , where  $G = \{(\lambda, s) \in D : \lambda \in J_d, s = i\}$  with  $J_d = [d, d + \varepsilon)$  if there exists an  $N$  such that  $\lambda_{i,j+1}^n(N) = d$ , otherwise  $J_d = [\lambda_{i,j+1}^n(N - 1), \lambda_{i,j+1}^n(N))$  with  $N := N_{j+1}$  being the least integer such that  $\lambda_{i,j+1}^n(N - 1) < d < \lambda_{i,j+1}^n(N)$  ( $\lambda_{i,j+1}^n(0) < d$ ). If  $s = k \neq i$  and  $c(k) + a(k)\lambda^* < c(i) + a(i)\lambda^*$ , define  $j$  to be the smallest positive integer such that  $a(1)^{j-1}b(1)a(k) < \varepsilon$  and  $a(1)^ja(k)(\lambda - \lambda^*) < (d - c(k) - a(k)\lambda^*)/2$ , and consider a path such that  $Z_0^n = z_0$ ,  $Y_0^n = N$ ,  $Y_1^n = Y_2^n = \dots = Y_j^n =$

0,  $S_1 = S_2 = \dots = S_j = 1$ , and  $S_{j+1} = k$ . If there exists an  $N$  such that  $\lambda_{k,j+1}^n(N) = d$ , in an analogous way as above, one can show that  $P^j(z_0, G) > 0$ , where  $G = \{(\lambda, s) \in D : \lambda \in [d, d + \varepsilon), s = k\}$ . Otherwise, let  $N := N_{j+1}$  be the least integer such that  $\lambda_{k,j+1}^n(N-1) < d < \lambda_{k,j+1}^n(N)$  ( $\lambda_{k,j+1}^n(0) < d$ ). Taking  $G = \{(\lambda, s) \in D : \lambda \in [\lambda_{k,j+1}^n(N-1), \lambda_{k,j+1}^n(N)), s = k\}$ , we easily obtain that  $P^j(z_0, G) > 0$  and  $|\lambda_{k,j+1}^n(N-1) - \lambda_{k,j+1}^n(N)| = a(1)^{j-1}a(k)b(1) < \varepsilon$ . Similarly, we may get that  $P^j(z_0, G) > 0$  if  $s = k \neq i$  and  $c(k) + a(k)\lambda^* = c(i) + a(i)\lambda^*$ , where  $G$  is defined as above in the same manner. Since  $\varepsilon$  is arbitrary,  $(d, s_i)$  can be approximated arbitrarily closely, and  $(d, s_i)$  is reachable if  $(d, s_i) \in D$ . This completes the proof of Lemma 6.1.

**Proposition 6.1.** Suppose that  $\rho(M_1) < 1$ . Then, given an appropriate initial distribution for  $(\lambda_0^n, S_0)$ , the process  $\{(\lambda_t^n, S_t)\}$  defined as in (6.3) is a stationary and geometrically ergodic Markov chain with finite first moment. Moreover, the process  $\{Y_t^n\}$  defined as in (6.3) is also a stationary and ergodic Markov chain with finite first moment.

**Proof.** We will first prove that the Markov chain  $\{Z_t^n\}$  is  $\psi$ -irreducible, aperiodic, and positive Harris recurrent. These properties will imply that  $\{Z_t^n\}$  has a unique stationary distribution  $\pi(\lambda, S)$ , and that if  $Z_0^n \sim \pi(\lambda, S)$  then  $\{Z_t^n\}$  is stationary and geometrically ergodic.

First note that  $A \neq \emptyset$  if  $\rho(M_1) < 1$ . The state space is equipped with  $\mathcal{F}_D$ , the Borel  $\sigma$ -algebra on  $\mathbb{R}_+ \times \mathbb{R}_+$  restricted to  $D$ . The measure  $\psi$  is the product measure  $\phi \otimes v$  on  $(D, \mathcal{F}_D)$ . Let  $G \in \mathcal{F}_D$  such that  $\psi(G) > 0$ . Then there exists a point  $(d, k) \in G \subset D_i$  such that  $\psi(D_\delta^+) > 0$ , where  $D_\delta^+ = G \cap B_\delta^+$  with  $B_\delta^+ = \{(\lambda, s) : \lambda \in [d, d + \delta/2), s = k\}$  for any  $\delta > 0$ . Write  $D_\delta^{d+} = \{(x, s) : x = y - d, y \in D_\delta^+, s = k\}$ . Since  $d \geq c(i) + a(i)\lambda^* \geq c(k) + a(k)\lambda^*$ , then  $(d - c(k))/a(k) \geq \lambda^*$ . Thus using the technique of proof of Lemma 6.1 for some  $j$ ,  $(\lambda_j^n, S_j)$  will be arbitrarily close to  $((d - c(k))/a(k), 1)$  by choosing  $j$  large enough

and  $S_j = 1$ . In particular,  $j$  can be chosen so that  $|c(k) + a(k)\lambda_j^n - d| < \varepsilon/2$ , where  $\varepsilon < \delta$ . Consider a path of the next step such that  $Y_j^n = 0$  and  $S_{j+1} = k$ , we have

$$\begin{aligned} P((\lambda', 1), G) &\geq P((\lambda', 1), D_\delta^+) \\ &\geq P(\varepsilon_{j+1,n} \in D_{\delta-\varepsilon}^{d+}(1))P(S_{j+1} = k|S_j = 1)P(Y_j^n = 0|\lambda_j^n = \lambda') \\ &= \frac{1}{c_n}\phi(D_{\delta-\varepsilon}^{d+}(1) \cap (0, c_n))P(S_{j+1} = k|S_j = 1)P(Y_j^n = 0|\lambda_j^n = \lambda') > 0, \end{aligned}$$

where  $D_{\delta-\varepsilon}^{d+}(1) = \{x : (x, s) \in D_{\delta-\varepsilon}^{d+}\}$ . Hence  $P^{j+1}(z_0, G) > 0$ , which implies  $\psi$ -irreducibility. It remains to prove the existence of a small set, aperiodicity and positive Harris recurrent.

Let  $C = \{(\lambda, s) \in D : \lambda \leq M^*\}$  for any  $M^* > \lambda^*$ , and define  $j$  to be the smallest positive integer such that  $a(1)^{j-1}(M^* - \lambda^*) < \varepsilon/2$  for sufficiently small  $\varepsilon$ . Then

$$\begin{aligned} \inf_{\lambda \in C} P(S_1 = S_2 = \dots = S_{j-1} = 1|Z_0^n = z_0) &> 0, \\ \inf_{\lambda \in C} P(Y_0^n = Y_1^n = \dots = Y_{j-2}^n = 0|Z_0^n = z_0, W_{j-1}) &> 0, \text{ and} \\ P(\varepsilon_{1,n} + \varepsilon_{2,n} + \dots + \varepsilon_{j-1,n} < \frac{\varepsilon}{2} - a(1)^{j-1}(M^* - \lambda^*)) &> 0, \end{aligned}$$

where  $W_{j-1} = \{\omega : S_1 = \dots = S_{j-1} = 1, \varepsilon_{1,n} + \varepsilon_{2,n} + \dots + \varepsilon_{j-1,n} < \varepsilon/2 - a(1)^{j-1}(M^* - \lambda^*)\}$ . Thus  $\inf_{z \in C} P^{j-1}(z, H) > 0$ , where  $H = \{(\lambda, s) \in D : \lambda \in [\lambda^*, \lambda^* + \varepsilon/2]\}$ . Taking  $\mu = \text{Unif}(\lambda^* + \varepsilon/2, \lambda^* + \varepsilon)$  and

$$\gamma = \inf_{z \in C} P^{j-1}(z, H) \inf_{z \in H} \{P(Y_{j-1}^n = 0|Z_{j-1}^n = z)P(S_j = 1|Z_{j-1}^n = z)\}/c_n > 0,$$



then, for all  $G \in \mathcal{B}(\mathbb{R}_+ \times \mathbb{R}_+)$ ,

$$\begin{aligned}
 P^j(z, G) &\geq P^j(z, G \cap H_\varepsilon) = \int_{\mathbb{R}_+ \times \mathbb{R}_+} P(y, G \cap H_\varepsilon) P^{j-1}(z, dy) \\
 &\geq \int_H P(y, G \cap H_\varepsilon) P^{j-1}(z, dy) \\
 &\geq \inf_{z \in H} \{P(Y_{j-1}^n = 0 | Z_{j-1}^n = z) P(S_j = 1 | Z_{j-1}^n = z)\} \\
 &\quad \times \int_H P(\varepsilon_{j,n} \in G_y \cap (\varepsilon/2 - a(1)(y - \lambda^*), \varepsilon - a(1)(y - \lambda^*))) P^{j-1}(z, dy) \\
 &\geq \gamma P(U_j \in G_{\lambda^*} \cap (\varepsilon/2, \varepsilon)) = \gamma \mu(G),
 \end{aligned}$$

which establishes  $C$  as a small set, where  $H_\varepsilon = \{(\lambda, s) \in D : \lambda \in (\lambda^* + \varepsilon/2, \lambda^* + \varepsilon)\}$  and  $G_u = \{x : x = z_1 - \lambda^* - a(1)(u - \lambda^*), (z_1, z_2) \in G\}$ .

We now show that  $\{Z_t^n\}$  is aperiodic. Consider the small set  $C = \{(\lambda, s) \in D : \lambda \leq M^*\}$  with  $M^* > \lambda^*$ . Note that  $\psi(C) > 0$ . If  $Z_{t-1}^n = (\lambda, s) \in C$ , consider a path such that  $Y_{t-1}^n = 0$ ,  $S_t = 1$ , and  $\varepsilon_{t,n} \leq (1 - a(1))(M^* - \lambda^*)$ , then

$$\lambda^* \leq \lambda_t^n = c(1) + a(1)\lambda + \varepsilon_{t,n} \leq c(1) + a(1)M^* + (1 - a(1))(M^* - \lambda^*) = M^*.$$

This implies

$$\begin{aligned}
 &P(Z_t^n \in C | Z_{t-1}^n = (\lambda, s) \in C) \\
 &\geq P(Y_{t-1}^n = 0, S_t = 1, \varepsilon_{t,n} \leq (1 - a(1))(M^* - \lambda^*) | Z_{t-1}^n = (\lambda, s) \in C) \\
 &= P(N_{t-1}(\lambda) = 0) P(S_t = 1 | S_{t-1} = s) P(\varepsilon_{t,n} \leq (1 - a(1))(M^* - \lambda^*)) > 0.
 \end{aligned}$$

Similarly,

$$\begin{aligned}
 &P(Z_{t+1}^n \in C | Z_{t-1}^n = (\lambda, s) \in C) \\
 &\geq P(Y_t^n = Y_{t-1}^n = 0, S_{t+1} = S_t = 1, \varepsilon_{t+1,n} \leq (1 - a(1))(M^* - \lambda^*), \\
 &\quad \varepsilon_{t,n} \leq (1 - a(1))(M^* - \lambda^*) | Z_{t-1}^n = (\lambda, s) \in C) > 0.
 \end{aligned}$$

It follows that  $\{Z_t^n\}$  is aperiodic by Chan (1990)( Prop. A1.1).

Finally, we prove that  $\{Z_t^n\}$  is positive Harris recurrent. It is well-known that  $M_1^k \rightarrow 0$  as  $k \rightarrow \infty$ , if  $\rho(M_1) < 1$ . Thus, for some positive constant  $\alpha < 1$ , there exists a positive integer  $N$  such that  $\|M_1^k\| \leq \alpha$  for all  $k \geq N$ . Consider  $V(x) = 1 + x_1$ , where  $x = (x_1, x_2)$  with  $x_1 > 0$ . Define the small set  $C = \{(\lambda, s) \in D : \lambda \leq M^*, s \in S\}$ , where  $M^* = K/(1 - 2\varepsilon - \alpha)$  with  $\varepsilon > 0$  such that  $1 - 2\varepsilon - \alpha > 0$ ,  $M^* = K/(1 - 2\varepsilon - \alpha) > \max\{1, \lambda^*\}$ , and constant  $K := (c_n/2 + c) \sum_{i=1}^N \|M^{i-1}\|$  with  $c = \max\{c(1), \dots, c(m)\}$ . First note that  $S_t$  and  $Y_{t-1}^n$  are conditionally independent under given  $(S_{t-1}, \lambda_{t-1}^n)$ , which implies that

$$\begin{aligned} E(\|B_t \mathbf{1}_{S_{t-1}} Y_{t-1}^n\| | \lambda_{t-1}^n, S_{t-1}) &= E(\|B_t \mathbf{1}_{S_{t-1}}\| | \lambda_{t-1}^n, S_{t-1}) E(Y_{t-1}^n | \lambda_{t-1}^n, S_{t-1}) \\ &= \lambda_{t-1}^n E(\|B_t \mathbf{1}_{S_{t-1}}\| | \lambda_{t-1}^n, S_{t-1}) = \lambda_{t-1}^n \|M_{01} \mathbf{1}_{S_{t-1}}\|. \end{aligned}$$

In addition,

$$E(\|A_t \mathbf{1}_{S_{t-1}} \lambda_{t-1}^n\| | \lambda_{t-1}^n, S_{t-1}) = \lambda_{t-1}^n \|M_{10} \mathbf{1}_{S_{t-1}}\|.$$

It follows that

$$E(\|A_t \mathbf{1}_{S_{t-1}} \lambda_{t-1}^n + B_t \mathbf{1}_{S_{t-1}} Y_{t-1}^n\| | \lambda_{t-1}^n, S_{t-1}) = \lambda_{t-1}^n \|M_1 \mathbf{1}_{S_{t-1}}\|.$$

Therefore, for any  $\lambda > M^*$ ,

$$\begin{aligned}
& E[V(Z_{Nt}^n) | Z_{N(t-1)}^n = z = (\lambda, s)] = E[1 + \lambda_{Nt}^n | Z_{N(t-1)}^n = z] \\
&= 1 + E[\|\mathbf{1}_{S_{Nt}} \lambda_{Nt}^n\| | Z_{N(t-1)}^n = z] \\
&= 1 + E[\|\mathbf{1}_{S_{Nt}} c(S_{Nt}) + A_{Nt} \mathbf{1}_{S_{Nt-1}} \lambda_{Nt-1}^n + B_{Nt} \mathbf{1}_{S_{Nt-1}} Y_{Nt-1}^n \\
&\quad + \mathbf{1}_{S_{Nt}} \varepsilon_{Nt,n}\| | Z_{N(t-1)}^n = z] \\
&= 1 + E\varepsilon_{Nt,n} + E[c(S_{Nt}) | Z_{N(t-1)}^n = z] \\
&\quad + E[E(\|A_{Nt} \mathbf{1}_{S_{Nt-1}} \lambda_{Nt-1}^n + B_{Nt} \mathbf{1}_{S_{Nt-1}} Y_{Nt-1}^n\| | \mathcal{F}_{N(t-1), Nt-1}^{\lambda^n, S}) | Z_{N(t-1)}^n = z)] \\
&= 1 + E\varepsilon_{Nt,n} + E[c(S_{Nt}) | Z_{N(t-1)}^n = z] + E[\|M_1 \mathbf{1}_{S_{Nt-1}} \lambda_{Nt-1}^n\| | Z_{N(t-1)}^n = z] \\
&= \dots \\
&= 1 + \sum_{i=1}^N \|M_1^{i-1} E[\mathbf{1}_{S_{Nt-i+1}} \varepsilon_{Nt-i+1,n} | Z_{N(t-1)}^n = z]\| \\
&\quad + \sum_{i=1}^N \|M_1^{i-1} E[\mathbf{1}_{S_{Nt-i+1}} c(S_{Nt-i+1}) | Z_{N(t-1)}^n = z]\| \\
&\quad + \|M_1^N E[\mathbf{1}_{S_{N(t-1)}} \lambda | Z_{N(t-1)}^n = z]\| \\
&\leq 1 + \frac{c_n}{2} \sum_{i=1}^N \|M_1^{i-1}\| + c \sum_{i=1}^N \|M_1^{i-1}\| + \|M_1^N\| \lambda = 1 + K + \|M_1^N\| \lambda \\
&\leq 1 + \lambda \left( \alpha + \frac{K}{\lambda} \right) \leq 1 + \lambda (\alpha + K) \leq 1 + \lambda (1 - 2\varepsilon) = 1 - \lambda\varepsilon + \lambda(1 - \varepsilon) \\
&\leq 1 - \varepsilon + \lambda(1 - \varepsilon) = (1 - \varepsilon)V(\lambda, s),
\end{aligned}$$

where  $\mathcal{F}_{u,t}^{\lambda^n, S}$  is the  $\sigma$ -algebra generated by  $\{Z_v^n, u \leq v \leq t\}$ . Furthermore, it is easy to see that, for  $\lambda \in C$ ,

$$E[V(Z_{Nt}^n) | Z_{N(t-1)}^n = z] \leq 1 + K + \|M_1^N\| \lambda \leq L$$

for some constant  $L$ . This implies that the chain  $\{Z_{Nt}^n\}$  has a unique station-

ary distribution  $\pi(\lambda, s)$ , and that if  $Z_0^n \sim \pi(\lambda, s)$  then  $\{Z_{Nt}^n\}$  is stationary and geometrically ergodic, and that  $E\lambda_{Nt}^n < \infty$ . Furthermore, by Lemma 3.1 in Tjøstheim (1990),  $\{Z_t^n\}$  is also stationary and geometrically ergodic. Finally, note that

$$Y_t^n = b^{-1}(S_{t+1})(\lambda_{t+1}^n - c(S_{t+1}) - a(S_{t+1})\lambda_t^n - \varepsilon_{t+1,n}) := g(Z_{t+1}^n, Z_t^n, \varepsilon_{t+1,n}),$$

which shows that  $Y_t^n$  depends on  $\{Z_t^n\}$  only through  $Z_t^n$  and  $Z_{t+1}^n$ . Since  $\{(Z_t^n, \varepsilon_{t,n})\}$  is stationary and ergodic, Theorem 36.4 of Billingsley (1995) shows that  $\{Y_t^n\}$  is stationary and ergodic. It is not difficult to see that  $EY_t^n < \infty$ . This completes the proof of proposition 6.1.

In order to prove proposition 6.2, we need the following Lemma.

**Lemma 6.2.** If  $\rho(M_{u_2}) < 1$ , then  $\rho(M_{u_1}) < 1$  for  $0 < u_1 < u_2$ .

**Proof.** First recall some properties of nonnegative matrices and nonnegative irreducible matrices (see Berman and Plemmons (1979) and Horn and Johnson (1991)). For a  $m \times m$  nonnegative matrix  $R = [r_{ij}]$ ,

$$\min_j \left\{ \sum_{i=1}^m r_{ij} \right\} \leq \rho(R) \leq \max_j \left\{ \sum_{i=1}^m r_{ij} \right\}. \quad (6.5)$$

In particular, for two nonnegative irreducible matrices  $Q = [q_{ij}]$  and  $R = [r_{ij}]$ , if  $q_{ij} \geq r_{ij}$ ,  $i, j = 1, 2, \dots, m$ , and  $Q \neq R$ , then  $\rho(Q) > \rho(R)$ . This immediately implies that if  $R$  is a nonnegative irreducible matrix and  $\min_j \{\sum_{i=1}^m r_{ij}\} < \max_j \{\sum_{i=1}^m r_{ij}\}$ , then, by (6.5),

$$\min_j \left\{ \sum_{i=1}^m r_{ij} \right\} < \rho(R) < \max_j \left\{ \sum_{i=1}^m r_{ij} \right\}. \quad (6.6)$$

Next notice that the transition matrix  $P$  is nonnegative irreducible and  $a(S_t) + b(S_t) > 0$ . Hence it is easy to prove that  $M_u$  is nonnegative irreducible for any  $u \geq 0$ . From (6.6) and  $\rho(M_{u_2}) < 1$ , we conclude that there exists some  $1 \leq i_1 \leq m$

such that  $a(i_1) + b(i_1) < 1$ . If  $\max_i \{a(i) + b(i)\} \leq 1$ , by (6.5) and (6.6), we know that Lemma 6.2 is true. Otherwise suppose that there exists some  $1 \leq i_2 \leq m$  such that  $a(i_2) + b(i_2) > 1$ . For convenience, now define a function

$$f(u) = \rho(M_u), u \geq 0.$$

We will show that  $\log f(u)$  is a convex function on  $[0, \infty)$ .

Since the transition matrix  $P$  is irreducible and aperiodic, and  $a(i) + b(i) > 0$ ,  $i = 1, 2, \dots, m$ , this implies that  $M_u$  is also an irreducible and aperiodic  $m \times m$  matrix. Therefore, by Theorem 8.5.1 in Horn and Johnson (1991), we may get that

$$\lim_{n \rightarrow \infty} f^{-n}(u) M_u^n = L_u,$$

where  $L_u$  is a constant positive matrix. It follows that

$$\lim_{n \rightarrow \infty} [\mathbf{1}^\tau M_u^n \pi]^{1/n} = f(u),$$

where  $\mathbf{1} = (1, 1, \dots, 1)^\tau$ . Again write  $M_u^{(n)}(\delta) = [((M_u^n)_{ij})^\delta]$ , where  $(M_u^n)_{ij}$  denotes the  $(i, j)$ -element of  $M_u^n$ . Then, we have

$$\begin{aligned} \log f(\gamma v_1 + (1 - \gamma)v_2) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log(\mathbf{1}^\tau M_{\gamma v_1 + (1 - \gamma)v_2}^n \pi) \\ &\leq \lim_{n \rightarrow \infty} \frac{1}{n} \log[\mathbf{1}^\tau (M_{v_1}^{(n)}(\gamma) \circ M_{v_2}^{(n)}(1 - \gamma)) \pi] \\ &\leq \lim_{n \rightarrow \infty} \frac{1}{n} \log[(\mathbf{1}^\tau M_{v_1}^n \pi)^\gamma (\mathbf{1}^\tau M_{v_2}^n \pi)^{1 - \gamma}] \\ &= \gamma \lim_{n \rightarrow \infty} \frac{1}{n} \log(\mathbf{1}^\tau M_{v_1}^n \pi) + (1 - \gamma) \lim_{n \rightarrow \infty} \frac{1}{n} \log(\mathbf{1}^\tau M_{v_2}^n \pi) \\ &= \gamma \log f(v_1) + (1 - \gamma) \log f(v_2), \end{aligned}$$

where  $0 < \gamma < 1$  and  $v_1, v_2 \geq 0, v_1 \neq v_2$ , and  $\circ$  denotes the Hadamard product

of matrices, the two inequalities are applications of the Hölder's inequality. This shows that  $\log f(u)$  is a convex function. Hence, together with  $f(0) = 1$  and  $f(u_2) < 1$ ,

$$\begin{aligned}\rho(M_{u_1}) &= f(u_1) = \exp\{\log f(u_1)\} = \exp\{\log f(\gamma \times 0 + (1 - \gamma)u_2)\} \\ &\leq \exp\{\gamma \log f(0) + (1 - \gamma) \log f(u_2)\} < 1,\end{aligned}$$

where  $\gamma = 1 - u_1/u_2$ .

**Remark 6.1.** It is not difficult to verify that  $\rho(M_1) = a + b$ , if  $a(S_t) \equiv a$  and  $b(S_t) \equiv b$ .

For the existence of high moments for the process  $\{Y_t^n\}$ , we have the following proposition.

**Proposition 6.2.** Suppose that  $\rho(M_k) < 1$ , where  $k \geq 1$  is a positive integer. Given an appropriate initial distribution for  $(\lambda_0^n, S_0)$ . Then the process  $\{Y_t^n\}$  defined as in (6.3) is a stationary and ergodic Markov chain with finite moment of order  $k$ .

**Proof.** By Lemma 6.2 and Proposition 6.1, we can obtain the stationarity and ergodicity of the process  $\{Y_t^n\}$  defined as in (6.3). A similar argument of the proof of Proposition 6.1 can be used to show the process  $\{Y_t^n\}$  with finite moment of order  $k$ .

By  $\rho(M_k) < 1$ , we know that there exists a positive integer  $N$  such that  $\|M_k^l\| \leq \alpha_k$  with some  $\alpha_k < 1$  for all  $l \geq N$ . Consider  $V(x) = 1 + x_1^k$ , where  $x = (x_1, x_2)$  with  $x_1 > 0$ . Define the small set  $C = \{(\lambda, s) \in D : \lambda \leq M^*, s \in S\}$ , where  $M^* = \beta/(1 - 2\varepsilon - \alpha_k)$  with  $\varepsilon > 0$  such that  $1 - 2\varepsilon - \alpha_k > 0$ ,  $\beta/(1 - 2\varepsilon - \alpha_k) >$

$\max\{1, \lambda^*\}$ , and constant  $\beta = \sum_{j=0}^{k-1} d_j$  is defined as follows. Note that

$$\begin{aligned}
 & E[\|A_{Nt}\mathbf{1}_{S_{Nt-1}}\lambda_{Nt-1}^n\|^i \|B_{Nt}\mathbf{1}_{S_{Nt-1}}Y_{Nt-1}^n\|^{k-i} | \lambda_{Nt-1}^n, S_{Nt-1}] \\
 = & (\lambda_{Nt-1}^n)^i E[(\mathbf{1}^\tau A_{Nt}\mathbf{1}_{S_{Nt-1}})^i (\mathbf{1}^\tau B_{Nt}\mathbf{1}_{S_{Nt-1}})^{k-i} | \lambda_{Nt-1}^n, S_{Nt-1}] \\
 & \times E[(Y_{Nt-1}^n)^{k-i} | \lambda_{Nt-1}^n, S_{Nt-1}] \\
 = & (\lambda_{Nt-1}^n)^k E[\mathbf{1}^\tau C_{Nt,i,k-i}\mathbf{1}_{S_{Nt-1}} | \lambda_{Nt-1}^n, S_{Nt-1}] \\
 & + \sum_{l=1}^{k-1} (\lambda_{Nt-1}^n)^l f_l(a(S_{Nt-1}), b(S_{Nt-1})) \\
 = & \|M_{i,k-i}\mathbf{1}_{S_{Nt-1}}(\lambda_{Nt-1}^n)^k\| + \sum_{l=1}^{k-1} (\lambda_{Nt-1}^n)^l f_l(a(S_{Nt-1}), b(S_{Nt-1}))
 \end{aligned}$$

and

$$\sum_{i=0}^k C_k^i M_{i,k-i} = M_k.$$

Therefore, for any  $\lambda > M^*$ ,

$$\begin{aligned}
 & E[V(Z_{Nt}^n) | Z_{N(t-1)}^n = z = (\lambda, s)] = E[1 + (\lambda_{Nt}^n)^k | Z_{N(t-1)}^n = z] \\
 = & 1 + E[\|\mathbf{1}_{S_{Nt}}\lambda_{Nt}^n\|^k | Z_{N(t-1)}^n = z] \\
 = & 1 + E[\|\mathbf{1}_{S_{Nt}}c(S_{Nt}) + A_{Nt}\mathbf{1}_{S_{Nt-1}}\lambda_{Nt-1}^n + B_{Nt}\mathbf{1}_{S_{Nt-1}}Y_{Nt-1}^n \\
 & + \mathbf{1}_{S_{Nt}}\varepsilon_{Nt,n}\|^k | Z_{N(t-1)}^n = z] \\
 = & 1 + E[\|A_{Nt}\mathbf{1}_{S_{Nt-1}}\lambda_{Nt-1}^n + B_{Nt}\mathbf{1}_{S_{Nt-1}}Y_{Nt-1}^n\|^k | Z_{N(t-1)}^n = z] \\
 & + \sum_{j=1}^k C_k^j E[\|\mathbf{1}_{S_{Nt}}c(S_{Nt}) + \mathbf{1}_{S_{Nt}}\varepsilon_{Nt,n}\|^j \\
 & \times \|A_{Nt}\mathbf{1}_{S_{Nt-1}}\lambda_{Nt-1}^n + B_{Nt}\mathbf{1}_{S_{Nt-1}}Y_{Nt-1}^n\|^{k-j} | Z_{N(t-1)}^n = z]
 \end{aligned}$$

$$\begin{aligned}
 &= 1 + \sum_{i=0}^k C_k^i E[E(\|A_{Nt} \mathbf{1}_{S_{Nt-1}} \lambda_{Nt-1}^n\|^i \\
 &\quad \times \|B_{Nt} \mathbf{1}_{S_{Nt-1}} Y_{Nt-1}^n\|^{k-i} | \mathcal{F}_{N(t-1), Nt-1}^{\lambda^n, S} | Z_{N(t-1)}^n = z] \\
 &\quad + \sum_{j=1}^k C_k^j E[E(\|\mathbf{1}_{S_{Nt}} c(S_{Nt}) + \mathbf{1}_{S_{Nt}} \varepsilon_{Nt,n}\|^j \\
 &\quad \times \|A_{Nt} \mathbf{1}_{S_{Nt-1}} \lambda_{Nt-1}^n + B_{Nt} \mathbf{1}_{S_{Nt-1}} Y_{Nt-1}^n\|^{k-j} | \mathcal{F}_{N(t-1), Nt-1}^{\lambda^n, S} | Z_{N(t-1)}^n = z] \\
 &= 1 + \sum_{i=0}^k C_k^i E[\|M_{i, k-i} \mathbf{1}_{S_{Nt-1}} (\lambda_{Nt-1}^n)^k\| | Z_{N(t-1)}^n = z] \\
 &\quad + \sum_{l=1}^{k-1} E[(\lambda_{Nt-1}^n)^l f_l(a(S_{Nt-1}), b(S_{Nt-1})) | Z_{N(t-1)}^n = z] \\
 &\quad + \sum_{j=1}^k C_k^j E[E(\|\mathbf{1}_{S_{Nt}} c(S_{Nt}) + \mathbf{1}_{S_{Nt}} \varepsilon_{Nt,n}\|^j \\
 &\quad \times \|A_{Nt} \mathbf{1}_{S_{Nt-1}} \lambda_{Nt-1}^n + B_{Nt} \mathbf{1}_{S_{Nt-1}} Y_{Nt-1}^n\|^{k-j} | \mathcal{F}_{N(t-1), Nt-1}^{\lambda^n, S} | Z_{N(t-1)}^n = z] \\
 &= \dots \\
 &= 1 + \|M_k^N E[\mathbf{1}_{S_{N(t-1)}} \lambda^k | Z_{N(t-1)}^n = z]\| + \sum_{j=0}^{k-1} d_j \lambda^j \\
 &\leq 1 + \lambda^k \left( \alpha_k + \sum_{j=0}^{k-1} \frac{d_j}{\lambda^{k-j}} \right) \leq 1 + \lambda^k (\alpha_k + \beta) \leq 1 + \lambda^k (1 - 2\varepsilon) \\
 &= (1 - \varepsilon)V(\lambda, s).
 \end{aligned}$$

Furthermore, it is easy to see that, for  $\lambda \leq M^*$ ,

$$E[V(Z_{Nt}^n) | Z_{N(t-1)}^n = z] \leq L_1$$

for some constant  $L_1$ . This implies that there exists finite moment of order  $k$  for  $\{Z_{Nt}^n, t \geq 0\}$ . It follows that there exists finite moment of order  $k$  for  $\{Y_t^n, t \geq 0\}$ .

**Remark 6.2.** A similar argument as Remark 2.1, we easily know that  $\rho(M_k) = (a+b)^k$ , if  $a(S_t) \equiv a$  and  $b(S_t) \equiv b$ . Obviously,  $a+b < 1$  if and only if  $(a+b)^k < 1$ .



Hence, Proposition 6.1 and 6.2 reduce to Proposition 2.1 and 2.2 in Fokianos et al. (2009), if  $a(S_t) \equiv a$  and  $b(S_t) \equiv b$ .

The following proposition quantifies the difference between (6.2) and (6.3), as  $n \rightarrow \infty$ , and shows that the perturbed model can be made arbitrarily close to the unperturbed model.

**Proposition 6.3.** Suppose that  $\rho(M_2) < 1$  and  $\{(Y_t, \lambda_t)\}$  and  $\{(Y_t^n, \lambda_t^n)\}$  are defined by (6.2) and (6.3) respectively, and  $(\lambda_0^n, S_0) =_d (\lambda_0, S_0)$ , then the following statements hold:

$$(1). |E(\lambda_t^n - \lambda_t)| = |E(Y_t^n - Y_t)| \leq \delta_{1,n},$$

$$(2). E(\lambda_t^n - \lambda_t)^2 \leq \delta_{2,n},$$

$$(3). E(Y_t^n - Y_t)^2 \leq \delta_{3,n}$$

where  $\delta_{i,n} \rightarrow 0$  as  $n \rightarrow \infty$  for  $i = 1, 2, 3$ . Furthermore, almost surely for any  $\delta > 0$ , with  $n$  large enough

$$|\lambda_t^n - \lambda_t| \leq \delta \quad \text{and} \quad |Y_t^n - Y_t| \leq \delta.$$

**Proof.** Since  $\rho(M_2) < 1$ , by Lemma 6.2, we have that  $\rho(M_1) < 1$  and

$$\sum_{k=0}^{\infty} M_1^k = (I - M_1)^{-1}, \quad \sum_{k=0}^{\infty} M_2^k = (I - M_2)^{-1},$$

where  $I$  denotes the identity matrix. By the equations (6.2) and (6.3),

$$\lambda_t^n - \lambda_t = a(S_t)(\lambda_{t-1}^n - \lambda_{t-1}) + b(S_t)(Y_{t-1}^n - Y_{t-1}) + \varepsilon_{t,n}. \quad (6.7)$$

Therefore, by using (6.7),

$$\begin{aligned}
& E(\lambda_t^n - \lambda_t) \\
= & E[E\|A_t \mathbf{1}_{S_{t-1}}(\lambda_{t-1}^n - \lambda_{t-1}) + B_t \mathbf{1}_{S_{t-1}}(Y_{t-1}^n - Y_{t-1})\| | S_{t-1}, \lambda_{t-1}^n, \lambda_{t-1}] + E(\varepsilon_{t,n}) \\
= & E[\|M_1 \mathbf{1}_{S_{t-1}}(\lambda_{t-1}^n - \lambda_{t-1})\|] + E(\varepsilon_{t,n}) = \|M_1 E[\mathbf{1}_{S_{t-1}}(\lambda_{t-1}^n - \lambda_{t-1})]\| + E(\varepsilon_{t,n}) \\
= & \|M_1 E[\mathbf{1}_{S_{t-1}}\|A_{t-1} \mathbf{1}_{S_{t-2}}(\lambda_{t-2}^n - \lambda_{t-2}) + B_{t-1} \mathbf{1}_{S_{t-2}}(Y_{t-2}^n - Y_{t-2}) + \varepsilon_{t-1,n}]\| \\
& + E(\varepsilon_{t,n}) \\
= & \|M_1 E[A_{t-1} \mathbf{1}_{S_{t-2}}(\lambda_{t-2}^n - \lambda_{t-2}) + B_{t-1} \mathbf{1}_{S_{t-2}}(Y_{t-2}^n - Y_{t-2})]\| \\
& + E(\varepsilon_{t,n}) + \|M_1 E[\mathbf{1}_{S_{t-1}} \varepsilon_{t-1,n}]\| \\
= & \dots \\
= & E(\varepsilon_{t,n}) \|(I + M_1 + \dots + M_1^{t-1})\| + \|M_1^t E[\mathbf{1}_{S_0}(\lambda_0^n - \lambda_0)]\| \\
= & \frac{c_n}{2} \|(I + M_1 + \dots + M_1^{t-1})\| \leq \frac{c_n}{2} \|(I + M_1 + \dots + M_1^{t-1} + \dots)\| \\
= & \frac{c_n}{2} \|(I - M_1)^{-1}\| \leq \frac{c_n}{2} \|(I - M_1)^{-1}\| := \delta_{1,n} \rightarrow 0
\end{aligned}$$

as  $n \rightarrow \infty$ , which implies the first assertion.

Next consider the second statement. By using (6.7) again,

$$\begin{aligned}
E(\lambda_t^n - \lambda_t)^2 &= E[a^2(S_t)(\lambda_{t-1}^n - \lambda_{t-1})^2] + E[b^2(S_t)(Y_{t-1}^n - Y_{t-1})^2] \\
&\quad + 2E[a(S_t)b(S_t)(\lambda_{t-1}^n - \lambda_{t-1})(Y_{t-1}^n - Y_{t-1})] + E(\varepsilon_{t,n}^2) \\
&\quad + 2E[a(S_t)(\lambda_{t-1}^n - \lambda_{t-1})\varepsilon_{t,n}] + 2E[b(S_t)(Y_{t-1}^n - Y_{t-1})\varepsilon_{t,n}].
\end{aligned}$$

We first calculate the second term of the above expression, by using the properties of the Poisson process and the conditional expectation, assuming first, that  $\lambda_{t-1}^n \geq \lambda_{t-1}$  and  $Y_{t-1}^n - Y_{t-1} \sim \text{Poisson}(\lambda_{t-1}^n - \lambda_{t-1})$  independently of  $Y_{t-1} \sim \text{Poisson}(\lambda_{t-1})$  (i.e  $\lambda_1^n > \lambda_1$  say  $\lambda_1^n = \lambda_1 + \mu_1$ )

$$Y_1^n = Y_1 + Z_1 \quad \text{and } Z_1 \sim \text{Poisson}(\mu_1), \quad Y_1^n > Y_1$$

$$\lambda_2^n > \lambda_2 \quad \text{say } \lambda_2^n = \lambda_2 + \mu_2$$

$$Y_2^n = Y_2 + Z_2 \quad \text{and } Z_2 \sim \text{Poisson}(\mu_2) \cdots).$$

Now we have that

$$\begin{aligned} & E[b^2(S_t)(Y_{t-1}^n - Y_{t-1})^2] = E[E(b^2(S_t)(Y_{t-1}^n - Y_{t-1})^2 | S_{t-1}, \lambda_{t-1}^n, \lambda_{t-1})] \\ &= E[E(b^2(S_t) | S_{t-1}, \lambda_{t-1}^n, \lambda_{t-1}) E(Y_{t-1}^n - Y_{t-1})^2 | S_{t-1}, \lambda_{t-1}^n, \lambda_{t-1})] \\ &= E[\|M_{02} \mathbf{1}_{S_{t-1}}\| E(N_{t-1}(\lambda_{t-1}^n) - N_{t-1}(\lambda_{t-1}))^2] \\ &= E[\|M_{02} \mathbf{1}_{S_{t-1}}\| (\lambda_{t-1}^n - \lambda_{t-1})^2] + E[\|M_{02} \mathbf{1}_{S_{t-1}}\| (\lambda_{t-1}^n - \lambda_{t-1})] \\ &\leq \|M_{02}\| [E(\lambda_{t-1}^n - \lambda_{t-1})^2 + E(\lambda_{t-1}^n - \lambda_{t-1})]. \end{aligned}$$

Similarly,

$$E[a^2(S_t)(\lambda_{t-1}^n - \lambda_{t-1})^2] \leq \|M_{20}\| E(\lambda_{t-1}^n - \lambda_{t-1})^2$$

and

$$2E[a(S_t)b(S_t)(\lambda_{t-1}^n - \lambda_{t-1})(Y_{t-1}^n - Y_{t-1})] \leq 2\|M_{11}\| E(\lambda_{t-1}^n - \lambda_{t-1})^2.$$

Finally, with  $K$  a positive constant,

$$\begin{aligned} & E(\varepsilon_{t,n}^2) + 2E[a(S_t)(\lambda_{t-1}^n - \lambda_{t-1})\varepsilon_{t,n}] + 2E[b(S_t)(Y_{t-1}^n - Y_{t-1})\varepsilon_{t,n}] \\ &\leq c_n^2/3 + c_n^2\|M_{10}\| \|(I - M_1)^{-1}\| + c_n^2\|M_{01}\| \|(I - M_1)^{-1}\| \\ &= \left( \frac{1}{3} + \|M_1\| \|(I - M_1)^{-1}\| \right) c_n^2 := Kc_n^2. \end{aligned}$$

Thus, by simple recursion,

$$\begin{aligned} E(\lambda_t^n - \lambda_t)^2 &\leq \|M_2\| E(\lambda_{t-1}^n - \lambda_{t-1})^2 + \|M_{02}\| \delta_{1,n} + Kc_n^2 \\ &\leq \|(I - M_2)^{-1}\| (\|M_{02}\| \delta_{1,n} + Kc_n^2) := \delta_{2,n} \rightarrow 0 \end{aligned}$$

as  $n \rightarrow \infty$ , which implies the second assertion. The third assertion follows that

$$\begin{aligned} E(Y_{t-1}^n - Y_{t-1})^2 &= E(N_{t-1}(\lambda_{t-1}^n) - N_{t-1}(\lambda_{t-1}))^2 \\ &= E(\lambda_{t-1}^n - \lambda_{t-1})^2 + E(\lambda_{t-1}^n - \lambda_{t-1}) \leq \delta_{2,n} + \delta_{1,n} := \delta_{3,n}. \end{aligned}$$

As to last statement of the proposition this follows by the second and third assertions of the proposition.

To make the analysis simple, a special case of the model defined by (6.1), will be described in the next section.

### 6.2.3 A simple model

Assume that the unobservable Markov chain  $S_t$  has two states: 1 and 2. A simple Markov switching model for the process  $\lambda_t$  contains two GARCH specifications:

$$\begin{aligned} Y_t | \mathcal{F}_{t-1}^{Y, \lambda, S} &\sim \text{Poisson}(\lambda_t), \\ \lambda_t &= \begin{cases} c_1 + a_1 \lambda_{t-1} + b_1 Y_{t-1}, & \text{if } S_t = 1, \\ c_2 + a_2 \lambda_{t-1} + b_2 Y_{t-1}, & \text{if } S_t = 2, \end{cases} \end{aligned} \quad (6.8)$$

where  $t \geq 1$ , the initial values  $Y_0, \lambda_0$  are fixed, and  $c_i, a_i, b_i, i = 1, 2$ , are assumed to be positive.  $S_t$  denotes an unobservable stationary Markov chain with finite state space  $S = \{1, 2\}$ , and  $\mathcal{F}_{t-1}$  denotes the information set available at time  $t - 1$ , which is a  $\sigma$ -field generated by  $\{(Y_{t-1}, S_{t-1}), (Y_{t-2}, S_{t-2}), \dots, (Y_0, S_1), (\lambda_0, S_0)\}$ . A model specification for the process  $\lambda_t$  contains two regimes, and each of them follows a *GARCH*(1,1) model. The process  $\lambda_t$  switches between two regimes according to the value of the state variable  $S_t$ . The limiting unconditional mean (as  $t \rightarrow \infty$ ) of the process is  $\frac{c_1 P(s_t=1) + c_2 P(s_t=2)}{1 - [(a_1 + b_1) P(s_t=1) + (a_2 + b_2) P(s_t=2)]}$ . In this model, the process  $\lambda_t$  is governed by two regimes with special means, and the transition between regimes is determined by the value of the state variable  $S_t$ . We call the model defined by (6.8) a Markov Switching Poisson Generalized Autoregressive

Heteroscedastic model with 2 regimes, denoted by  $MS - PGARCH(2; 1, 1)$ . The model (6.8) can be rewritten in a form of Poisson process as follows

$$Y_t = N_t(\lambda_t), \quad (6.9)$$

where  $N_t(\lambda_t)$  denotes a Poisson process with intensity  $\lambda_t$ . That is  $Y_t$  is equal to the number of events  $N_t(\lambda_t)$  of  $N_t(\cdot)$  in the time interval  $[0, \lambda_t]$ , where  $\lambda_t$  is defined by (6.8). This process can be approximated by the following process.

$$Y_t^m = N_t(\lambda_t^m),$$

$$\lambda_t^m = \begin{cases} c_1 + a_1 \lambda_{t-1}^m + b_1 Y_{t-1}^m + \varepsilon_{1,tm}, & \text{if } S_t = 1, \\ c_2 + a_2 \lambda_{t-1}^m + b_2 Y_{t-1}^m + \varepsilon_{2,tm}, & \text{if } S_t = 2, \end{cases} \quad (6.10)$$

where  $\varepsilon_{t,m} = C_m U_t$ ,  $C_m > 0$ ,  $C_m \rightarrow \infty$  as  $m \rightarrow \infty$ , and  $\{U_t\}$  is a sequence of independent identical distribution (iid) uniform variables on  $(0, 1)$  and such that the  $\{U_t\}$  is dependent of  $\mathcal{F}_{t-1}^{Y,\lambda,S}$ .

Proposition 6.3. shows that the model (6.10) converges to the model (6.9) when  $m \rightarrow \infty$ . Assume  $S_t$  in model (6.9) follows a first-order Markov chain with the transition matrix  $P_2 = [p_{ij}] = Pr(s_t = j | s_{t-1} = i)$ ,  $i, j = 1, 2$ , i.e.

$$P_2 = \begin{bmatrix} Pr(s_t = 1 | s_{t-1} = 1) & Pr(s_t = 2 | s_{t-1} = 1) \\ Pr(s_t = 1 | s_{t-1} = 2) & Pr(s_t = 2 | s_{t-1} = 2) \end{bmatrix}$$

$$=: \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{bmatrix}$$

where  $p_{ij}$ ,  $(i, j = 1, 2)$  denote the transition probabilities of the state variable  $S_t = j$  given  $S_{t-1} = i$ . The transition probabilities satisfy  $\sum_{j=1}^2 p_{ij} = 1$ , and  $0 \leq p_{ij} \leq 1$ ,  $i = 1, 2$ . A small value of the transition probabilities  $p_{12}$  and  $p_{21}$  means that the process tends to stay longer in state 1 and 2 respectively. In general,

the expected duration of the model to stay in the state 1 is  $\sum_{k=1}^{\infty} k p_{11}^{k-1} (1 - p_{11}) = \frac{1}{1-p_{11}}$ , and the expected duration of the other state is  $\frac{1}{1-p_{22}}$  (see, Hamilton (1989)). The matrix above shows that the transition matrix  $P_2$  contains only two parameters  $(p_{11}, p_{22})$ . The stationary (steady-state) distribution of the Markov chain  $\{S_t\}$  is

$$\pi = \begin{bmatrix} Pr(S_t = 1) \\ Pr(S_t = 1) \end{bmatrix} = \begin{bmatrix} \frac{1-p_{22}}{2-p_{11}-p_{22}} \\ \frac{1-p_{11}}{2-p_{11}-p_{22}} \end{bmatrix}.$$

According to the definition of the model (6.9), the *MS-PGARCH*(2, 1, 1) model uses an unobservable Markov chain to govern the transition from one conditional mean (variance) to another. This is different from that of a *PTGARCH* model defined in Chapter 2, for which the switch is determined by a particular lagged variable  $(Y_{t-d})$ . We mentioned previously that the nonlinear time series models have some drawbacks. The model defined by (6.2) has also the drawback that the state variables  $(S_t)$  are independent of  $(Y_{t-1})$  (over time) and consequently may be difficult to apply in time series data.

### 6.3 Statistical Inference

In the last few decades, many authors have discussed the different methods to estimate the Markov switching model, for example, Hamilton (1989,1990 and 1994), Kim (1994), Kim and Nelson (2003), Francq and Zakoian (2001) and, Safaei (2012). Here, we concentrate on the *MS-PGARCH*(2,1,1) model of order (1,1). However, the procedure of high order *MS-PGARCH* model would be straightforward.

This is because the states are not directly observable. Therefore, it is quite difficult to estimate a *MSAR* model than other models. Many methods exist for the estimation of the unknown parameters. For example, the EM algorithm (Hamilton (1989)), and the Gibbs sampler (Albert and Chib (1993)). The first

method is more difficult to implement in the presence of AR lags (see Filardo (1994)), while the second approach requires heavy computation. Hence, in the next section, we estimate the parameters in the model defined by (6.9) using a MLE numerical method.

### 6.3.1 Maximum likelihood estimation of MS-PGARCH model

We will discuss Maximum likelihood estimation (MLE) for the MS-PGARCH model. The main goal of this method is to define the coefficients that maximize the probability of observations. We concentrate on a *MS – PGARCH*(2, 1, 1) model to make the analysis simple, and extension to a model with  $m$  regimes would be straightforward. The vector of parameters for model defined by (6.9) is  $\theta = (c_i, a_i, b_i, p_{11}, p_{22}, i = 1, 2)$ . The estimate of  $\theta$  can be found by maximizing the conditional likelihood function of the observations  $Y_1, Y_2, \dots, Y_n$  given the prediction probability  $Pr(s_t = i | \mathcal{F}_{t-1}^{Y, \lambda, S}, \theta)$  which are based on the information prior to time  $t$ . The conditional log likelihood function is given by

$$\ell_T(\theta) = \frac{1}{T} \sum_{t=1}^T \log Pr(Y_t = k | \mathcal{F}_{t-1}^{Y, \lambda, S}, \theta), \quad (6.11)$$

where  $Pr(Y_t = k | \mathcal{F}_{t-1}^{Y, \lambda, S}, \theta)$  is the distribution of  $Y_t$  conditional on  $\mathcal{F}_{t-1}^{Y, \lambda, S}$  given  $Pr(s_t = i | \mathcal{F}_{t-1}^{Y, \lambda, S}, \theta)$ , that is

$$\begin{aligned} Pr(Y_t = k | \mathcal{F}_{t-1}^{Y, \lambda, S}, \theta) &= Pr(s_t = 1 | \mathcal{F}_{t-1}, \theta) Pr(Y_t = k | S_t = 1, \mathcal{F}_{t-1}) \\ &\quad + Pr(s_t = 2 | \mathcal{F}_{t-1}, \theta) Pr(Y_t = k | S_t = 2, \mathcal{F}_{t-1}) \end{aligned} \quad (6.12)$$

where  $Pr(Y_t = k|S_t = i, \mathcal{F}_{t-1}^{Y,\lambda,S}, \theta)$  is the distribution of  $Y_t$  conditional on  $\mathcal{F}_{t-1}^{Y,\lambda,S}$  and  $S_t = i, (i = 1, 2)$ , that is

$$\begin{aligned} & Pr(Y_t = k|S_t = i, \mathcal{F}_{t-1}^{Y,\lambda,S}, \theta) \\ &= \frac{\lambda_t^k \exp(-\lambda_t)}{k!} \\ &= \frac{(c_i + a_i \lambda_{t-1} + b_i Y_{t-1})^k \exp[-(c_i + a_i \lambda_{t-1} + b_i Y_{t-1})]}{k!}. \end{aligned} \quad (6.13)$$

For  $i = 1, 2$ , the filtering probabilities of  $S_t$  which are based on the past and present information are

$$Pr(S_t = i|\mathcal{F}_t, \theta) = \frac{Pr(s_t = i|\mathcal{F}_{t-1}, \theta)Pr(Y_t = k|S_t = i, \mathcal{F}_{t-1}^{Y,\lambda,S}, \theta)}{Pr(Y_t = k|\mathcal{F}_{t-1}^{Y,\lambda,S}, \theta)}. \quad (6.14)$$

By the Bayes theorem, and the relationship between prediction and filtering probabilities, this leads to

$$Pr(S_{t+1} = i|\mathcal{F}_{t-1}, \theta) = p_{1i}Pr(S_t = 1|\mathcal{F}_{t-1}, \theta) + p_{2i}Pr(S_t = 2|\mathcal{F}_{t-1}, \theta), \quad (6.15)$$

where  $p_{1i} = Pr(S_{t+1} = i|S_t = 1)$  and  $p_{2i} = Pr(S_{t+1} = i|S_t = 2)$  are transition probabilities. Note that the equations (6.12)-(6.15) form a recursive system for  $t = h, \dots, T$ , with initial values  $Pr(S_h = i|\mathcal{F}_{h-1}, \theta)$ . We can calculate the filtering probabilities  $Pr(S_t = i|\mathcal{F}_t, \theta)$  as well as the distribution of  $Y_t$  conditional on  $\mathcal{F}_{t-1}^{Y,\lambda,S}$  for  $t = h, \dots, T$ . The equation (6.11) is a complex nonlinear function of  $\theta$ . The MLE  $\hat{\theta}_T$  can be calculated by using a numerical search algorithm. There are many programs to compute the MLE  $\hat{\theta}_T$  such as the GAUSS program and R program and both adopt the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. We use the second program to estimate the model parameters. Both prediction probabilities and filtering probabilities are computed by plugging  $\hat{\theta}_T$  into their formulae.

To calculate the smoothing probabilities  $Pr(S_t = i|\mathcal{F}_T, \theta)$ , which are based on



all the information in the sample, we use the method of Kim's (1994) smoothing algorithm as follows. Consider the joint probability that  $S_t = i$  and  $S_{t+1} = j$  based on the full-sample information:

$$\begin{aligned}
 & Pr(S_t = i, S_{t+1} = j | \mathcal{F}_T, \theta) \\
 = & Pr(S_{t+1} = j | \mathcal{F}_T, \theta) \times Pr(S_t = i | S_{t+1} = j, \mathcal{F}_T, \theta) \\
 \approx & Pr(S_{t+1} = j | \mathcal{F}_T, \theta) \times Pr(S_t = i | S_{t+1} = j, \mathcal{F}_t, \theta) \\
 = & \frac{Pr(S_{t+1} = j | \mathcal{F}_T, \theta) \times Pr(S_t = i, S_{t+1} = j | \mathcal{F}_t, \theta)}{Pr(S_{t+1} = j | \mathcal{F}_t, \theta)} \\
 = & \frac{Pr(S_{t+1} = j | \mathcal{F}_T, \theta) \times Pr(S_t = i | \mathcal{F}_t, \theta) \times Pr(S_{t+1} = j | S_t = i, \mathcal{F}_t, \theta)}{Pr(S_{t+1} = j | \mathcal{F}_t, \theta)}.
 \end{aligned}$$

In addition,

$$\begin{aligned}
 & Pr(S_t = i | \mathcal{F}_T, \theta) \tag{6.16} \\
 = & \sum_{j=1}^2 Pr(S_t = i, S_{t+1} = j | \mathcal{F}_T, \theta) \\
 = & \frac{Pr(S_{t+1} = 1 | \mathcal{F}_T, \theta) \times Pr(S_t = i | \mathcal{F}_t, \theta) \times Pr(S_{t+1} = 1 | S_t = i, \mathcal{F}_t, \theta)}{Pr(S_{t+1} = 1 | \mathcal{F}_t, \theta)} \\
 + & \frac{Pr(S_{t+1} = 2 | \mathcal{F}_T, \theta) \times Pr(S_t = i | \mathcal{F}_t, \theta) \times Pr(S_{t+1} = 2 | S_t = i, \mathcal{F}_t, \theta)}{Pr(S_{t+1} = 2 | \mathcal{F}_t, \theta)} \\
 \approx & Pr(S_t = i | \mathcal{F}_t, \theta) \\
 & \times \left( \frac{p_{i1} Pr(S_{t+1} = 1 | \mathcal{F}_T, \theta)}{Pr(S_{t+1} = 1 | \mathcal{F}_t, \theta)} + \frac{p_{i2} Pr(S_{t+1} = 2 | \mathcal{F}_T, \theta)}{Pr(S_{t+1} = 2 | \mathcal{F}_t, \theta)} \right). \tag{6.17}
 \end{aligned}$$

The above procedure can be iterated for  $t = T - 1, T - 2, T - 3, \dots, 2, 1$ , and the smoothing probabilities  $Pr(S_t = i, | \mathcal{F}_T, \theta)$  for  $t = T - 1, T - 2, T - 3, \dots, 2, 1$  are estimates of smoothed probabilities.  $Pr(S_T | \mathcal{F}_T, \theta)$  is estimated at the last iteration of the filtering probabilities. These probabilities are function of  $\theta$ , plugging the MLE  $\hat{\theta}_T$  into (6.12), (6.14) and (6.15) formulae yields the estimated smoothing probabilities.

### 6.3.2 Simulation

In this part, we carry out a few simulation studies on calculation of the MLEs of the model (6.9) and (6.10). The MLEs were computed by optimizing the log-likelihood function (6.11). All simulation cases are performed using the R function *Optim* of package *stats*, which adopts the BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm. Five different cases will be examined, using different gradient functions, different methods to obtain the initial values of the parameters, different numbers of simulation times and different relative convergence tolerances.

#### Case 1

In this case, the true parameters of the model will be used as the initial values. No gradient function is used, and the results are obtained over 1000 simulations with relative convergence tolerance being  $1e^{-4}$ . Larger values of convergence tolerance may produce inaccurate results since convergence tolerance applies to each step.

Table 6.1 contains the results for the data that was generated with

$$(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.3; 0.6, 0.3, 0.4, 0.7, 0.4)$$

for different sample sizes. It reports the maximum likelihood estimates (MLEs) of the model defined by (6.9) and (6.10) for the parameters  $(c_i, a_i, b_i, i = 1, 2)$  with their mean square error (MSE) (in parentheses) as well as absolute average error (AAE) for different sample sizes. The estimates of parameters and their mean square errors are obtained by averaging out the results from all simulations. We see the mean square error (MSE) as well as average absolute error (AAE) reduce significantly as the sample size increases from 500 to 2000. In Table 6.1, it can be seen that the sample sizes play an important role in accuracy of estimation (see the MSE and AAE). Both the MSE and the AAE were calculated using the simulation output. Table 6.2 shows standard errors which are found using the standard deviation (SD) of the simulated estimators of model (6.9) and (6.10).

Table 6.1: Maximum likelihood estimates with their mean square errors (in parentheses), and the average absolute errors of estimates for model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.3; 0.6, 0.3, 0.4, 0.7, 0.4)$ , for different sample sizes. Results obtained from 1000 simulations.

Sample size		Maximum Likelihood Estimates (MLE)							Average Absolute Error
$n$	$c_1$	$a_1$	$b_1$	$c_2$	$a_2$	$b_2$	$p_{11}$	$p_{22}$	AAE
500	0.8134 (0.0368)	0.5045 (0.0092)	0.2901 (0.0051)	0.6037 (0.0101)	0.3021 (0.0108)	0.3995 (0.0087)	0.7023 (0.0026)	0.3968 (0.0030)	0.0032
1000	0.7974 (0.0051)	0.5028 (0.0027)	0.2972 (0.0020)	0.5995 (0.0015)	0.3046 (0.0024)	0.3991 (0.0021)	0.7013 (0.0004)	0.3995 (0.0002)	0.0011
1500	0.8005 (0.0012)	0.5049 (0.0013)	0.2967 (0.0011)	0.5989 (0.0005)	0.2989 (0.0012)	0.3976 (0.0010)	0.7005 (0.0002)	0.3997 (0.0001)	0.0007
2000	0.8007 (0.00003)	0.5031 (0.0001)	0.2995 (0.0001)	0.5999 (0.00001)	0.2997 (0.0001)	0.3988 (0.00004)	0.7005 (0.00001)	0.3997 (0.00001)	0.0005

They are obtained from 1000 simulations using different sample sizes. As the sample size increases from small to large, the simulated standard error clearly decreases.

Figure 6.1 provides the boxplot of the AAE of MLEs for the model defined by (6.9) and (6.10) for different sample sizes. The plot 6.1 shows the AAE reduces gradually as the sample size increases from 500 to 2000. The larger the sample size is the smaller the number of outliers is. The distribution of the AAE values is positive (right) skewed independently of the sample sizes. Table 6.3 shows the

Table 6.2: The simulated standard error for model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.3; 0.6, 0.3, 0.4, 0.7, 0.4)$ , for different sample sizes. The results are based on 1000 simulations each.

Sample size		Simulated Standard Error							
$n$	$c_1$	$a_1$	$b_1$	$c_2$	$a_2$	$b_2$	$p_{11}$	$p_{22}$	
500	0.1920	0.0960	0.0711	0.1007	0.1038	0.0932	0.0513	0.0547	
1000	0.0717	0.0517	0.0444	0.0390	0.0490	0.0462	0.0190	0.0157	
1500	0.0351	0.0362	0.0337	0.0232	0.0348	0.0309	0.0126	0.0092	
2000	0.0174	0.0257	0.0267	0.0109	0.0272	0.0202	0.0088	0.0072	

distribution of the maximum likelihood estimates for model (6.9) and (6.10), descriptive statistic of the parameters as well as the root mean square estimations (RMSE) which is defined as  $RMSE(\hat{\theta}) = \sqrt{MSE(\hat{\theta})} = \sqrt{E(\hat{\theta} - \theta)^2}$  for different sample sizes. Table 6.3 shows that the value of RMSE for each parameter decreases gradually as the sample size increases.

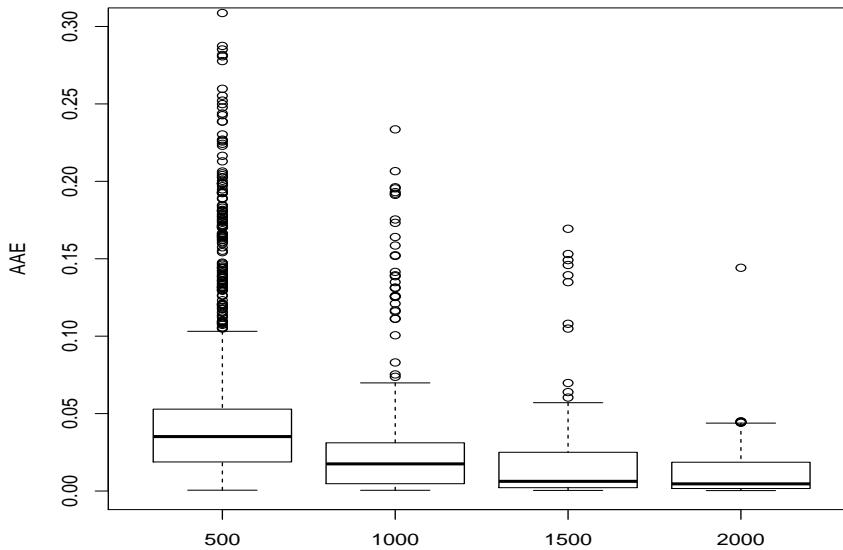


Figure 6.1: Boxplot of AAE for maximum likelihood estimator of model (6.9) and (6.10) for different sample sizes. Results are based on 1000 simulations.

Table 6.3: Distribution of the maximum likelihood estimates of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.3; 0.6, 0.3, 0.4, 0.7, 0.4)$  for different sample sizes. Results are based on 1000 simulations.

Sample size	Parameter	Value	Mean	RMSE	Summary				
					Lower	25%	Median	75%	Upper
500	$c_1$	0.8	0.8134	0.0831	-0.0297	0.7794	0.8006	0.8244	2.3230
	$a_1$	0.5	0.5045	0.0659	0.0302	0.4580	0.5016	0.5526	1.2930
	$b_1$	0.3	0.2901	0.0538	-0.1369	0.2445	0.2936	0.3312	0.5826
	$c_2$	0.6	0.6037	0.0478	0.0648	0.5874	0.6008	0.6199	1.1780
	$a_2$	0.3	0.3021	0.0666	-0.3262	0.2665	0.3025	0.3503	0.6630
	$b_2$	0.4	0.3995	0.0564	-0.1791	0.3740	0.4013	0.4377	0.8204
	$p_{11}$	0.7	0.7023	0.0270	0.1925	0.6871	0.7010	0.7178	1.0000
	$p_{22}$	0.4	0.3968	0.0253	0.0000	0.3869	0.3993	0.4106	0.9453
1000	$c_1$	0.8	0.7974	0.0245	0.1236	0.7930	0.8002	0.8073	1.3910
	$a_1$	0.5	0.5028	0.0321	0.1760	0.4845	0.5010	0.5184	0.8819
	$b_1$	0.3	0.2972	0.0318	0.1196	0.2757	0.2997	0.3189	0.4931
	$c_2$	0.6	0.5995	0.0151	0.2340	0.5954	0.6001	0.6053	0.9260
	$a_2$	0.3	0.3046	0.0303	0.0487	0.2902	0.3009	0.3182	0.5789
	$b_2$	0.4	0.3991	0.0245	0.0022	0.3903	0.4007	0.4122	0.6538
	$p_{11}$	0.7	0.7013	0.0112	0.5476	0.6952	0.7002	0.7064	0.8712
	$p_{22}$	0.4	0.3995	0.0086	0.2495	0.3956	0.3999	0.4031	0.5045
1500	$c_1$	0.8	0.8005	0.0130	0.3399	0.7964	0.8002	0.8038	1.1860
	$a_1$	0.5	0.5049	0.0225	0.3018	0.4925	0.5013	0.5144	0.7335
	$b_1$	0.3	0.2969	0.0227	0.1088	0.2839	0.2989	0.3075	0.4171
	$c_2$	0.6	0.5989	0.0090	0.2249	0.5970	0.6001	0.6027	0.8283
	$a_2$	0.3	0.2989	0.0203	0.0808	0.2930	0.3003	0.3080	0.5187
	$b_2$	0.4	0.3976	0.0167	0.0702	0.3921	0.3997	0.4047	0.6379
	$p_{11}$	0.7	0.7005	0.0074	0.5894	0.6978	0.7002	0.7033	0.8115
	$p_{22}$	0.4	0.3997	0.0054	0.3298	0.3979	0.3999	0.4015	0.5037
2000	$c_1$	0.8	0.8007	0.0069	0.7085	0.7981	0.8000	0.8024	1.1930
	$a_1$	0.5	0.5031	0.0158	0.3961	0.4962	0.5013	0.5099	0.5911
	$b_1$	0.3	0.2995	0.0167	0.1964	0.2938	0.3004	0.3057	0.4041
	$c_2$	0.6	0.5999	0.0053	0.5321	0.5986	0.6001	0.6017	0.7367
	$a_2$	0.3	0.2997	0.0147	0.0778	0.2967	0.3004	0.3057	0.6630
	$b_2$	0.4	0.3988	0.0110	0.2882	0.3962	0.4001	0.4039	0.5363
	$p_{11}$	0.7	0.7005	0.0051	0.6497	0.6990	0.7001	0.7023	0.7340
	$p_{22}$	0.4	0.3997	0.0039	0.3263	0.3987	0.3999	0.4008	0.4277

Case 2

In this case, the initial values of the intercept and the autoregressive parameters were obtained by the ARIMA (1,1) fit to the data, and the starting values for the transition probabilities ( $p_{11}$  and  $p_{22}$ ) were obtained by counting the number of states in regimes 1 and 2 and then dividing by the sample size for each simulation. 500 simulations are taken and the relative convergence tolerance is  $1e^{-4}$ . As in case 1, no gradient function will be used.

Table 6.4 reports MLEs with their mean square errors (in parentheses), and the average absolute error of estimates for two various models for different sample sizes. The table shows the estimates of the parameters obtained by averaging out the results from all simulations. The MSE was computed by using the simulations output. From Table 6.4, it can be seen that the MSE and AAE of MLSE for both different models decreases gradually when the sample size increases from 2000 to 10000.

Table 6.4: Maximum likelihood estimates with their mean square errors (in parentheses), and the average absolute error of estimates for two various models for different sample sizes. Results obtained from 500 simulations.

Regime One	Regime Two	Sample size $n$	Maximum Likelihood Estimators (MLE)								Average Absolute Error
			$c_1$	$a_1$	$b_1$	$c_2$	$a_2$	$b_2$	$p_{11}$	$p_{22}$	AAE
$c_1=0.8$ $a_1=0.5$	$c_2=0.6$ $a_2=0.4$	2000	1.0221 (0.1301)	0.4086 (0.0140)	0.4149 (0.0014)	1.3965 (0.7019)	0.2933 (0.0181)	0.3880 (0.0037)	0.6642 (0.0185)	0.3649 (0.0064)	0.1891
$b_1=0.4$ $p_{11}=0.6$	$b_2=0.4$ $p_{22}=0.4$	3000	1.0197 (0.1085)	0.4057 (0.0130)	0.4175 (0.0010)	1.3994 (0.6823)	0.2934 (0.0157)	0.3860 (0.0025)	0.6556 (0.0152)	0.3685 (0.0052)	0.1821
		5000	1.0595 (0.1150)	0.3969 (0.0139)	0.4190 (0.0008)	1.4050 (0.6761)	0.2947 (0.0133)	0.3907 (0.0015)	0.6386 (0.0099)	0.3800 (0.0028)	0.1814
		10000	1.0741 (0.1041)	0.3911 (0.0137)	0.4213 (0.0006)	1.4134 (0.6762)	0.2947 (0.0120)	0.3962 (0.0008)	0.6240 (0.0058)	0.3868 (0.0018)	0.1781
$c_1=0.6$ $a_1=0.54$	$c_2=0.7$ $a_2=0.45$	2000	1.2926 (0.5569)	0.3917 (0.0342)	0.4260 (0.0067)	0.8301 (0.0851)	0.4219 (0.0068)	0.3407 (0.0037)	0.3678 (0.0062)	0.6863 (0.0224)	0.1791
$b_1=0.45$ $p_{11}=0.4$	$b_2=0.3$ $p_{22}=0.6$	3000	1.2841 (0.5319)	0.3873 (0.0341)	0.4297 (0.0046)	0.8251 (0.0823)	0.4239 (0.0059)	0.3407 (0.0033)	0.3694 (0.0040)	0.6821 (0.0194)	0.1740
		5000	1.2874 (0.5126)	0.3788 (0.0335)	0.4274 (0.0032)	0.8486 (0.0831)	0.4186 (0.0053)	0.3455 (0.0033)	0.3699 (0.0033)	0.6708 (0.0163)	0.1734
		10000	1.2712 (0.4763)	0.3815 (0.0308)	0.4340 (0.0022)	0.8352 (0.0662)	0.4213 (0.0042)	0.3450 (0.0032)	0.3697 (0.0027)	0.6759 (0.0160)	0.1659

Table 6.5 illustrates the simulated standard errors for two different models for different sample sizes. The table shows that the standard errors of most

parameters decreases when the sample size increases for both models. Table 6.6

Table 6.5: The simulated standard errors for two different models for different sample sizes. The results are based on 500 simulations each.

Regime 1	Regime 2	Sample size	Simulated Standard Error							
$c_1=0.8$	$c_2=0.6$	2000	0.3608	0.1184	0.0372	0.8378	0.1347	0.0608	0.1362	0.0798
$a_1=0.5$	$a_2=0.4$	3000	0.3293	0.1139	0.0323	0.8260	0.1253	0.0502	0.1234	0.0719
$b_1=0.4$	$b_2=0.4$	5000	0.3392	0.1179	0.0281	0.8222	0.1155	0.0391	0.0993	0.0526
$p_{11}=0.6$	$p_{22}=0.4$	10000	0.3226	0.1180	0.0261	0.8223	0.1096	0.0284	0.0761	0.0424
$c_1=0.6$	$c_2=0.7$	2000	0.7463	0.1850	0.0830	0.2917	0.0822	0.0610	0.0788	0.1497
$a_1=0.54$	$a_2=0.45$	3000	0.7293	0.1848	0.0678	0.2868	0.0767	0.0575	0.0630	0.1391
$b_1=0.45$	$b_2=0.3$	5000	0.7159	0.1830	0.0566	0.2883	0.0730	0.0576	0.0583	0.1275
$p_{11}=0.6$	$p_{22}=0.4$	10000	0.6901	0.1755	0.0469	0.2572	0.0650	0.0572	0.0519	0.1266

shows the variance of AAE for two different models for different sample sizes.

The variance of AAE reduces when the sample sizes increases.

Table 6.6: The variance of average absolute error for two different models for different sample sizes. The results are based on 500 simulations each.

Regime 1	Regime 2	Sample Size	Variance of AAE
$c_1=0.8$	$c_2=0.6$	2000	0.002664
$a_1=0.5$	$a_2=0.4$	3000	0.001917
$b_1=0.4$	$b_2=0.4$	5000	0.001082
$p_{11}=0.6$	$p_{22}=0.4$	10000	0.000514
$c_1=0.6$	$c_2=0.7$	2000	0.002884
$a_1=0.54$	$a_2=0.45$	3000	0.002257
$b_1=0.45$	$b_2=0.3$	5000	0.001454
$p_{11}=0.4$	$p_{22}=0.6$	10000	0.000980

Table 6.7 presents the descriptive information about AAE for two different models for different sample sizes. The mean of AAE for both models decreases when the sample size increases.

Table 6.7: Summary of average absolute error of two various models for different sample sizes. Results are based on 500 simulations.

Regime		Sample Size	Summary					
One	Two	$n$	Lower	25%	Median	Mean	75%	Upper
$c_1=0.8$	$c_2=0.6$	2000	0.06038	0.15414	0.18752	0.18912	0.22221	0.35875
$a_1=0.5$	$a_2=0.4$	3000	0.05016	0.15020	0.18221	0.18213	0.21169	0.34767
$b_1=0.4$	$b_2=0.4$	5000	0.08418	0.15815	0.18271	0.18137	0.20361	0.28980
$p_{11}=0.6$	$p_{22}=0.4$	10000	0.11205	0.16270	0.17880	0.17810	0.19200	0.25070
$c_1=0.6$	$c_2=0.7$	2000	0.06133	0.14173	0.17435	0.17912	0.20735	0.35491
$a_1=0.54$	$a_2=0.45$	3000	0.04764	0.14055	0.17341	0.17403	0.20372	0.31549
$b_1=0.45$	$b_2=0.3$	5000	0.06719	0.14950	0.17314	0.17344	0.19820	0.27875
$p_{11}=0.4$	$p_{22}=0.6$	10000	0.06601	0.14396	0.16572	0.16591	0.18841	0.25307

Table 6.8 reports the maximum likelihood estimation of model (6.9) and (6.10) when

$(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.4, 0.6)$  and the descriptive statistics of these parameters. The RMSE values of the parameters are given. In Table 6.8, it can be seen that the RMSE of the parameters of the second regime decreases when the sample size increases from 2000 to 10000.

Table 6.9 reports the maximum likelihood estimation of model (6.9) and (6.10) when

$(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.4, 0.6)$  and the descriptive statistics of these estimates. The RMSE values of the parameters are given.

Figure 6.2 shows the boxplot of the values of estimated parameters for model when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.4, 0.6)$ , for two different sample sizes. The plot illustrates that the range of boxplot of the estimated parameters values reduce when the sample size increases from 2000 to 5000. We see that as the sample size increases, the values of outliers of the estimated parameters decrease. From Figure 6.2, it can be deduced that the MLE estimates become more efficient with small range of boxplot of parameters and outliers less



Table 6.8: Distribution of the maximum likelihood estimates of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 0.6, 0.4, 0.4, 0.6, 0.4)$ , for different sample sizes. Results are based on 500 simulations.

Sample size	Parameter	Value	Mean	RMSE	Summary				
					Lower	25%	Median	75%	Upper
2000	$c_1$	0.8	1.0220	03136	0.3186	0.8479	1.0650	1.2190	1.7210
	$a_1$	0.5	0.4086	0.1026	0.2379	0.3546	0.3982	0.4569	0.6557
	$b_1$	0.4	0.4149	0.0296	0.3206	0.3917	0.4159	0.4373	0.5490
	$c_2$	0.6	1.3970	0.7966	0.6665	1.2190	1.3960	1.5600	2.2300
	$a_2$	0.4	0.2933	0.1125	0.0002	0.2420	0.2948	0.3507	0.5515
	$b_2$	0.4	0.3880	0.0427	0.0692	0.3605	0.3897	0.4203	0.5943
	$p_{11}$	0.6	0.6642	0.0719	00.3098	0.5959	0.6067	0.6436	0.9939
	$p_{22}$	0.4	0.3649	0.0437	0.1241	0.3764	0.3951	0.4052	0.5934
3000	$c_1$	0.8	1.0197	0.2909	0.3400	0.8960	1.0640	1.1870	1.6190
	$a_1$	0.5	0.4057	0.1010	0.2568	0.3625	0.3932	0.4419	0.6205
	$b_1$	0.4	0.41752	0.0261	0.3232	0.3993	0.4183	0.4368	0.5036
	$c_2$	0.6	1.3990	0.7994	0.67550	1.2730	1.3940	1.5330	2.0950
	$a_2$	0.4	0.2934	0.1098	0.0227	0.2554	0.2968	0.3353	0.5216
	$b_2$	0.4	0.3860	0.0342	0.1699	0.3691	0.3925	0.4161	0.4932
	$p_{11}$	0.6	0.6556	0.0597	0.5757	0.5976	0.6051	0.6194	0.9978
	$p_{22}$	0.4	0.3685	0.0359	0.1262	0.3828	0.3955	0.4027	0.4245
5000	$c_1$	0.8	1.0590	0.3160	0.3879	0.9997	1.1020	1.1990	1.5700
	$a_1$	0.5	0.3969	0.1082	0.2889	0.3578	0.3829	0.4158	0.6352
	$b_1$	0.4	0.4190	0.0230	0.3485	0.4059	0.4187	0.4319	0.4929
	$c_2$	0.6	1.4050	0.8050	0.6018	1.2910	1.4160	1.5190	1.8210
	$a_2$	0.4	0.2947	0.1056	0.1082	0.2646	0.2942	0.3271	0.4270
	$b_2$	0.4	0.3907	0.0264	0.1879	0.3771	0.3962	0.4139	0.4817
	$p_{11}$	0.6	0.6386	0.0426	0.5837	0.5972	0.6035	0.6115	0.9644
	$p_{22}$	0.4	0.3800	0.0243	0.1872	0.3895	0.3969	0.4030	0.4165
10000	$c_1$	0.8	0.1.0740	0.3118	0.3847	1.0510	1.1110	1.1650	1.3700
	$a_1$	0.5	0.3911	0.1118	0.3062	0.3654	0.3829	0.3991	0.5673
	$b_1$	0.4	0.4213	0.0225	0.3761	0.4116	0.4207	0.4309	0.4603
	$c_2$	0.6	1.4130	0.8134	0.9781	1.3400	1.4190	1.4970	1.7070
	$a_2$	0.4	0.2947	0.1053	0.1590	0.2762	0.2941	0.3152	0.3750
	$b_2$	0.4	0.3962	0.0180	0.2684	0.3882	0.4007	0.4113	0.4929
	$p_{11}$	0.6	0.6240	0.0263	0.5881	0.5985	0.6017	0.6059	0.9713
	$p_{22}$	0.4	0.3868	0.0159	0.1351	0.3945	0.3987	0.4018	0.4121

when the sample size increases from 2000 to 5000.

Figure 6.3 shows the boxplot of AAE for maximum likelihood estimator of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 0.6, 0.4, 0.4,$

Table 6.9: Distribution of the maximum likelihood estimates of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.4, 0.6)$ , for different sample sizes. Results are based on 500 simulations.

Sample size	Parameter	Value	Mean			RMSE			Summary		
			Lower	25%	Median	75%	Upper				
2000	$c_1$	0.6	1.2930	0.6942	0.4667	1.1180	1.2930	1.4680	2.1700		
	$a_1$	0.54	0.3917	0.1643	0.07792	0.31810	0.37360	0.45690	0.7653		
	$b_1$	0.45	0.4260	0.0685	0.2485	0.3698	0.4131	0.4718	0.6921		
	$c_2$	0.7	0.8301	0.2450	0.1722	0.6136	0.8529	1.0270	1.4630		
	$a_2$	0.45	0.4219	0.0676	0.2191	0.3626	0.4219	0.4825	0.6323		
	$b_2$	0.3	0.3407	0.0510	0.2039	0.3132	0.3435	0.3741	0.4555		
	$p_{11}$	0.4	0.3678	0.0489	0.1700	0.3334	0.3916	0.4040	0.7608		
	$p_{22}$	0.6	0.6863	0.0935	0.4024	0.6003	0.6136	0.7966	0.9821		
3000	$c_1$	0.6	1.2840	0.6842	0.5894	1.1040	1.2840	1.4630	1.9460		
	$a_1$	0.54	0.3873	0.1626	0.1911	0.3118	0.3645	0.4606	0.7310		
	$b_1$	0.45	0.4297	0.0565	0.2670	0.3867	0.4165	0.4661	0.6696		
	$c_2$	0.7	0.8251	0.2510	0.2809	0.5876	0.8874	1.0310	1.4390		
	$a_2$	0.45	0.4239	0.0645	0.2577	0.3681	0.4187	0.4808	0.6051		
	$b_2$	0.3	0.3407	0.0490	0.2330	0.3153	0.3423	0.3693	0.4346		
	$p_{11}$	0.4	0.3694	0.0379	0.1910	0.3570	0.3918	0.4018	0.5284		
	$p_{22}$	0.6	0.6822	0.0866	0.3876	0.6008	0.6123	0.7875	0.9830		
5000	$c_1$	0.6	1.2870	0.6874	0.6883	1.1690	1.3070	1.4170	1.8510		
	$a_1$	0.54	0.3788	0.1656	0.2109	0.3139	0.3547	0.4406	0.6373		
	$b_1$	0.45	0.4274	0.0483	0.3248	0.3924	0.4145	0.4561	0.6344		
	$c_2$	0.7	0.8486	0.2641	0.1909	0.5915	0.9277	1.0310	1.3140		
	$a_2$	0.45	0.4186	0.0642	0.2769	0.3701	0.4039	0.4717	0.6014		
	$b_2$	0.3	0.3455	0.0506	0.2193	0.3240	0.3487	0.3714	0.4239		
	$p_{11}$	0.4	0.3699	0.0339	0.1941	0.3447	0.3950	0.4012	0.4638		
	$p_{22}$	0.6	0.6708	0.0734	0.5825	0.5996	0.6069	0.7620	0.9411		
10000	$c_1$	0.6	1.2710	0.6712	0.8006	1.1760	1.2840	1.3790	1.7040		
	$a_1$	0.54	0.3815	0.1589	0.2263	0.3214	0.3604	0.4481	0.5773		
	$b_1$	0.45	0.4340	0.0411	0.3421	0.4005	0.4212	0.4671	0.5740		
	$c_2$	0.7	0.8352	0.2380	0.3970	0.5835	0.9261	1.0190	1.1950		
	$a_2$	0.45	0.4213	0.0586	0.3040	0.3748	0.4020	0.4747	0.5509		
	$b_2$	0.3	0.3450	0.0501	0.2108	0.3209	0.3520	0.3713	0.4100		
	$p_{11}$	0.4	0.3697	0.0320	0.2430	0.3396	0.3939	0.3999	0.4117		
	$p_{22}$	0.6	0.6759	0.0774	0.5884	0.6009	0.6076	0.7837	0.9049		

0.6, 0.4), for different sample sizes. In Figure 6.3, it can be seen that when the sample size increases the boxplot decreases slowly.

Figure 6.4 illustrates the histogram of AAE for the maximum likelihood estim-

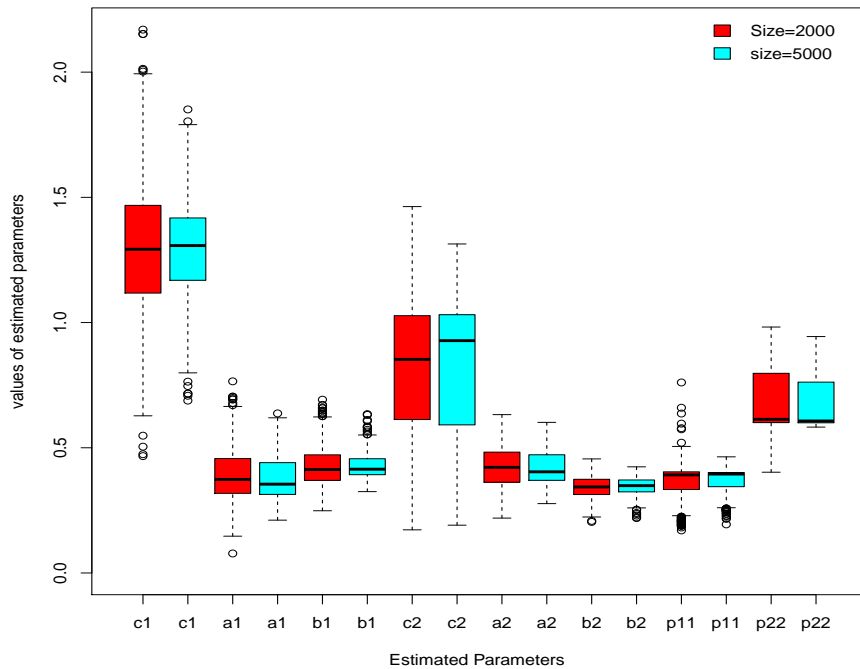


Figure 6.2: Boxplot of the values of estimated parameters of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.4, 0.6)$ , for different sample sizes. Results are based on 500 simulations. The initial values for the intercept and the slope parameters are obtained by the ARIMA(1,1) model.

ator of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 0.6, 0.4, 0.4, 0.6, 0.4)$ , for different sample sizes. The figure shows that the histogram of AAE values would be reasonably symmetric.

Figure 6.5 shows the boxplot of AAE for maximum likelihood estimator of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.4, 0.6)$ , for different sample sizes. In Figure 6.5, it can be seen that when the sample size increases the boxplot decreases slowly. Figure 6.6 illustrates the histogram of AAE for maximum likelihood estimator of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.4, 0.6)$ , for different sample sizes. The figure shows that the histogram of AAE values would be reasonably symmetric.

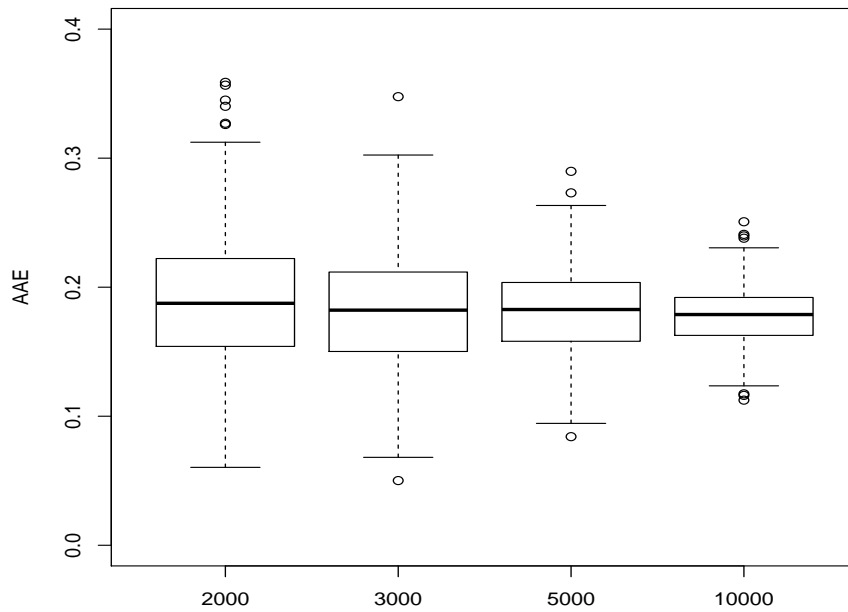


Figure 6.3: Boxplot of AAE for maximum likelihood estimator of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 0.6, 0.4, 0.4, 0.6, 0.4)$ , for different sample sizes. Results are based on 500 simulations. The initial values for the intercept and the slope parameters are obtained by the ARIMA (1,1) model.

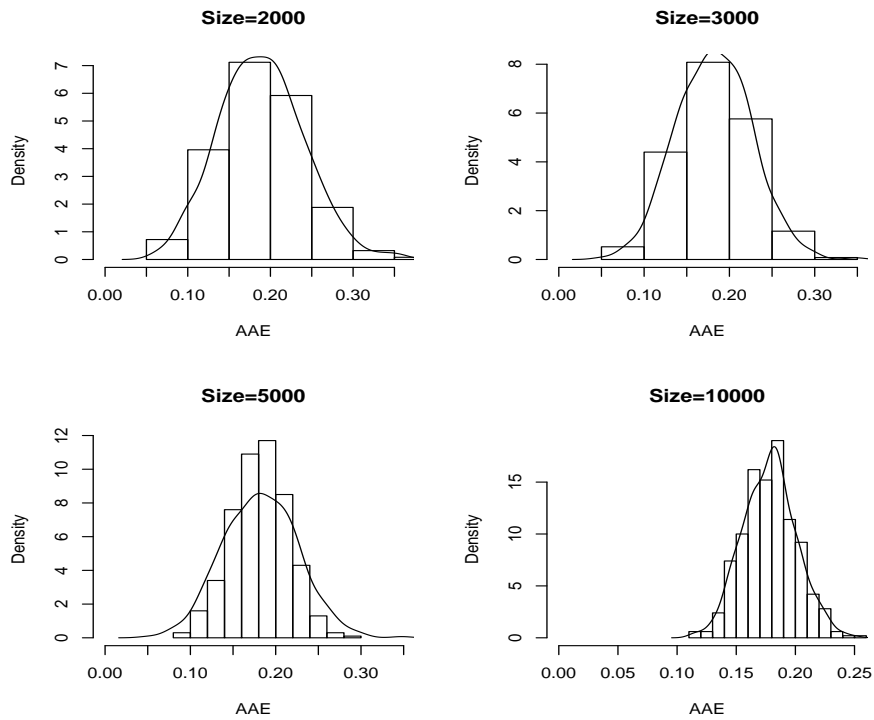


Figure 6.4: Histogram of AAE for maximum likelihood estimator of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 0.6, 0.4, 0.4, 0.6, 0.4)$ , for different sample sizes. Results are based on 500 simulations. The initial values for parameters are obtained by the ARIMA(1,1) model.

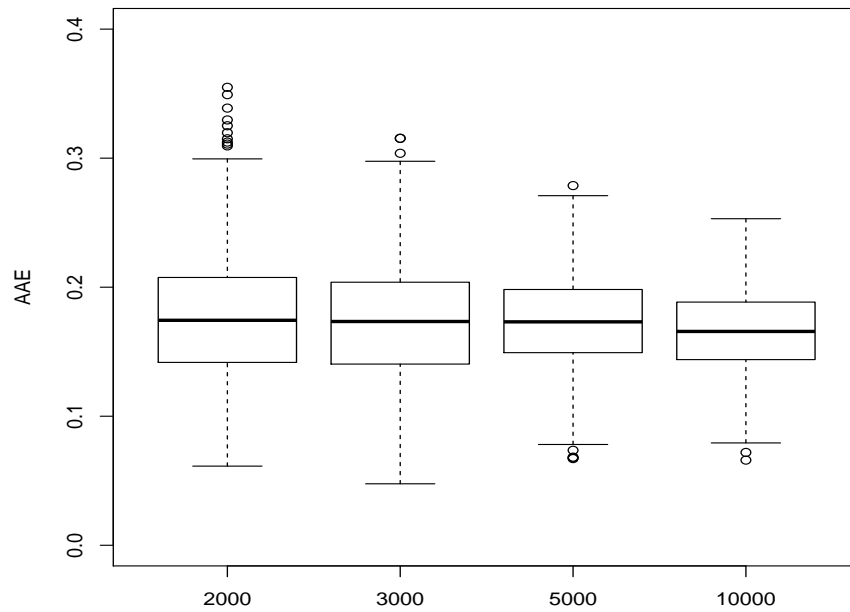


Figure 6.5: Boxplot of AAE for maximum likelihood estimator of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.4, 0.6)$ , for different sample sizes. Results are based on 500 simulations. The initial values for parameters are obtained by the ARIMA(1,1) model.

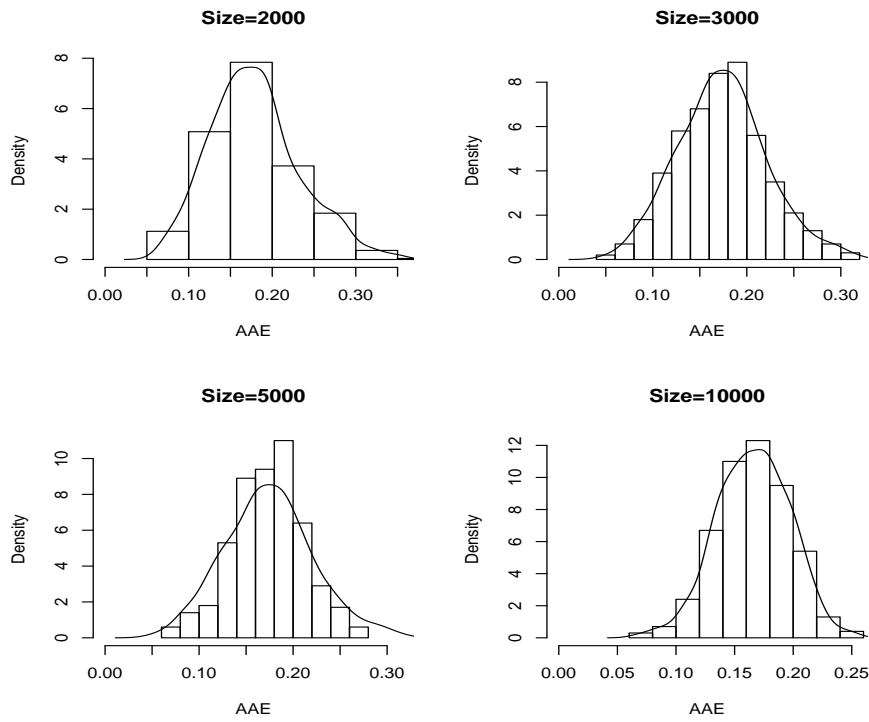


Figure 6.6: Histogram of AAE for maximum likelihood estimator of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.4, 0.6)$ , for different sample sizes. Results are based on 500 simulations. The initial values for parameters are obtained by the ARIMA (1,1) model.

**Case 3**

In the present case, the starting values of all parameters are similar to those in case 2. A numerical approximation of the gradient function will be used. The outcomes are based on 1000 simulations with relative convergence tolerance ( $e^{-8}$ ).

Table 6.10 shows MLE with their MSE (in parentheses) as well as AAE of the MLE for different sample sizes. The table shows that the values of MSE of the MLE for all parameters become smaller as the sample size increases from 2000 to 10000. In Table 6.10 the AAE values decrease gradually as the sample size increases from small to the large.

Table 6.10: Maximum likelihood estimates with their mean square errors (in parentheses), and the average absolute error of estimates for model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 0.6, 0.4, 0.4, 0.6, 0.4)$ , for different sample sizes. Results obtained from 1000 simulations.

Sample size	Maximum Likelihood Estimators (MLE)								Average Absolute Error
$n$	$c_1$	$a_1$	$b_1$	$c_2$	$a_2$	$b_2$	$p_{11}$	$p_{22}$	AAE
2000	0.6163 (0.1626)	0.4990 (0.0185)	0.4160 (0.0131)	1.0985 (0.4633)	0.3419 (0.0425)	0.3664 (0.0299)	0.6700 (0.0825)	0.4010 (0.1065)	0.2334
3000	0.5941 (0.1400)	0.5141 (0.0152)	0.4036 (0.0086)	1.1324 (0.4588)	0.3264 (0.0352)	0.3699 (0.0270)	0.6934 (0.0745)	0.4009 (0.0997)	0.2236
5000	0.6121 (0.1013)	0.5129 (0.0096)	0.4032 (0.0057)	1.0483 (0.3364)	0.3271 (0.0316)	0.3712 (0.0222)	0.6992 (0.0672)	0.3847 (0.0868)	0.1976
10000	0.6389 (0.0650)	0.5081 (0.0051)	0.4003 (0.0026)	1.0080 (0.2555)	0.3238 (0.0231)	0.3778 (0.0166)	0.7193 (0.0555)	0.3346 (0.0679)	0.1692

Table 6.11 reports simulated standard errors for model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 0.6, 0.4, 0.4, 0.6, 0.4)$  for different sample sizes. The simulated standard error for regressive coefficients and the transition probabilities reduces clearly as the sample size increases from 2000 to 10000.

Table 6.12 presents the variance of AAE for model (6.9) and (6.10) for different sample sizes.



Table 6.11: The simulated standard errors for model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 0.6, 0.4, 0.4, 0.6, 0.4)$ , for different sample sizes. The results are based on 1000 simulations each.

Sample size		Simulated Standard Error						
$n$	$c_1$	$a_1$	$b_1$	$c_2$	$a_2$	$b_2$	$p_{11}$	$p_{22}$
2000	0.4032	0.1360	0.1146	0.6807	0.2063	0.1728	0.2872	0.3263
3000	0.3742	0.1233	0.0926	0.6774	0.1876	0.1645	0.2729	0.3158
5000	0.3182	0.0981	0.0757	0.5800	0.1779	0.1489	0.2593	0.2946
10000	0.2550	0.0712	0.0509	0.5055	0.1519	0.1287	0.2355	0.2604

Table 6.12: The variance of average absolute error for model (6.9) and (6.10) for different sample sizes. The results are based on 1000 simulations each.

Regime 1	Regime 2	Sample Size	Variance of AAE
$c_1=0.8$	$c_2=0.6$	2000	0.00506
$a_1=0.5$	$a_2=0.4$	3000	0.00494
$b_1=0.4$	$b_2=0.4$	5000	0.00418
$p_{11}=0.6$	$p_{22}=0.4$	10000	0.00295

Table 6.13 presents the descriptive information of AAE for model (6.9) and (6.10) for different sample sizes. The values of different descriptive information of AAE for the model decrease when the sample size increases.

Table 6.13: Summary of average absolute error for model (6.10) for different sample sizes. Results are based on 1000 simulations.

Regime One	Regime Two	Sample Size $n$	Summary					
			Lower	25%	Median	Mean	75%	Upper
$c_1=0.8$	$c_2=0.6$	2000	0.05171	0.18670	0.23250	0.23340	0.28100	0.50960
$a_1=0.5$	$a_2=0.4$	3000	0.04243	0.17800	0.22160	0.22360	0.26900	0.44260
$b_1=0.4$	$b_2=0.4$	5000	0.03674	0.15290	0.19880	0.19760	0.24140	0.43690
$p_{11}=0.6$	$p_{22}=0.4$	10000	0.03094	0.13010	0.16790	0.16920	0.20540	0.40310

Table 6.14 reports the maximum likelihood estimation of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 0.6, 0.4, 0.4, 0.6, 0.4)$ , and the de-

scriptive statistics of these parameters. In addition, it shows that the RMSE values of the parameters reduce slowly as the sample size increases.

Table 6.14: Distribution of the maximum likelihood estimates of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 0.6, 0.4, 0.4, 0.6, 0.4)$ , for different sample sizes. Results are based on 1000 simulations.

Sample size	Parameter	Value	Mean	RMSE	Summary				
					Lower 25%	Median	75%	Upper	$n$
2000	$c_1$	0.8	0.6163	0.3126	0.0000	0.4019	0.6341	0.8157	2.6140
	$a_1$	0.5	0.4990	0.0941	0.0000	0.4326	0.4956	0.5576	1.1680
	$b_1$	0.4	0.4160	0.0760	0.0000	0.3632	0.4095	0.4599	1.2930
	$c_2$	0.6	1.0990	0.5675	0.0000	0.8214	1.0810	1.3870	2.9400
	$a_2$	0.4	0.3419	0.1567	0.0000	0.2263	0.3500	0.4517	1.7500
	$b_2$	0.4	0.3664	0.1304	0.0000	0.2745	0.3746	0.4683	0.9356
	$p_{11}$	0.6	0.6700	0.2463	0.0000	0.5027	0.7383	0.9053	1.0000
	$p_{22}$	0.4	0.4010	0.2833	0.0000	0.1052	0.3268	0.6954	0.9985
3000	$c_1$	0.8	0.5941	0.2914	0.0000	0.4193	0.6114	0.7577	2.2190
	$a_1$	0.5	0.5141	0.0817	0.0000	0.4508	0.4978	0.5654	1.2980
	$b_1$	0.4	0.4036	0.0620	0.0155	0.3643	0.4018	0.4427	0.9611
	$c_2$	0.6	1.1320	0.5778	0.0000	0.8752	1.1420	1.3950	2.5240
	$a_2$	0.4	0.3264	0.1459	0.0000	0.2173	0.3404	0.4373	0.9420
	$b_2$	0.4	0.3699	0.1229	0.0000	0.2872	0.3784	0.4634	0.9223
	$p_{11}$	0.6	0.6934	0.2369	0.0000	0.5329	0.7516	0.9120	1.0000
	$p_{22}$	0.4	0.4009	0.2699	0.0000	0.1385	0.3381	0.6613	0.9988
5000	$c_1$	0.8	0.6121	0.2441	0.0000	0.4879	0.6358	0.7604	1.5650
	$a_1$	0.5	0.5129	0.0655	0.1279	0.4595	0.5004	0.5494	1.0660
	$b_1$	0.4	0.4032	0.0506	0.0000	0.3676	0.4021	0.4379	0.9630
	$c_2$	0.6	1.0480	0.4979	0.0000	0.8309	1.0680	1.3000	2.8760
	$a_2$	0.4	0.3271	0.1377	0.0000	0.2237	0.3332	0.4235	0.8986
	$b_2$	0.4	0.3712	0.1112	0.0000	0.2944	0.3811	0.4627	0.7497
	$p_{11}$	0.6	0.6992	0.2234	0.0000	0.5617	0.7539	0.9002	1.0000
	$p_{22}$	0.4	0.3847	0.2503	0.0000	0.1479	0.3347	0.6271	0.9913
10000	$c_1$	0.8	0.6389	0.1987	0.0000	0.5429	0.6576	0.7459	1.5670
	$a_1$	0.5	0.5081	0.0460	0.2729	0.4701	0.4981	0.5353	1.0600
	$b_1$	0.4	0.4003	0.0347	0.0309	0.3775	0.4015	0.4254	0.6149
	$c_2$	0.6	1.0080	0.4408	0.0000	0.8275	1.0330	1.1960	1.8790
	$a_2$	0.4	0.3238	0.1187	0.0000	0.2462	0.3292	0.4307	0.7857
	$b_2$	0.4	0.3778	0.0948	0.0000	0.3100	0.3880	0.4504	0.8463
	$p_{11}$	0.6	0.7193	0.2022	0.0007	0.5871	0.7561	0.8813	0.9999
	$p_{22}$	0.4	0.3346	0.2179	0.0000	0.1342	0.3006	0.4929	0.9857

Figure 6.7 shows the values of AAE for maximum likelihood estimator of model

(6.9) and (6.10) for different sample sizes. The AAE values become smaller when the sample size increases. The plot shows the parameters have slight outliers.

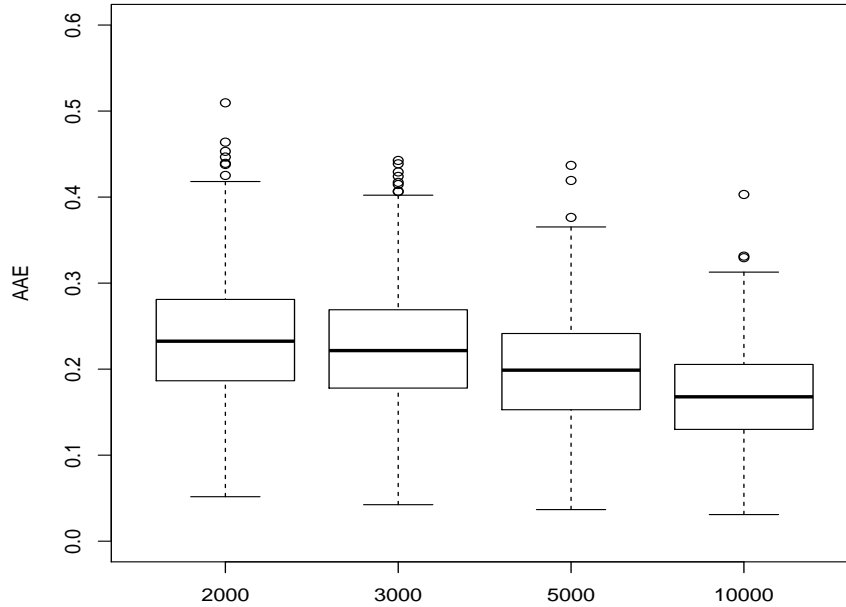


Figure 6.7: Boxplot of AAE for maximum likelihood estimator of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 0.6, 0.4, 0.4, 0.6, 0.4)$ , for different sample sizes. Results are based on 1000 simulations.

Figure 6.8 presents the boxplot of the estimated parameter values for model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 0.6, 0.4, 0.4, 0.6, 0.4)$ , for two different sample sizes. The values of the outliers of large sample sizes (10000) are smaller in comparison to the values of outliers of 3000. Figure 6.8 shows the tendency of the MLE estimates to become more efficient with smaller values of the outliers as the sample size increases from 3000 to 10000. Figure 6.9 illustrates the histogram of AAE for maximum likelihood estimator of model (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 0.6, 0.4, 0.4, 0.6, 0.4)$ , for different sample sizes. Figure 6.9 reports that the histogram of AAE could be symmetric.

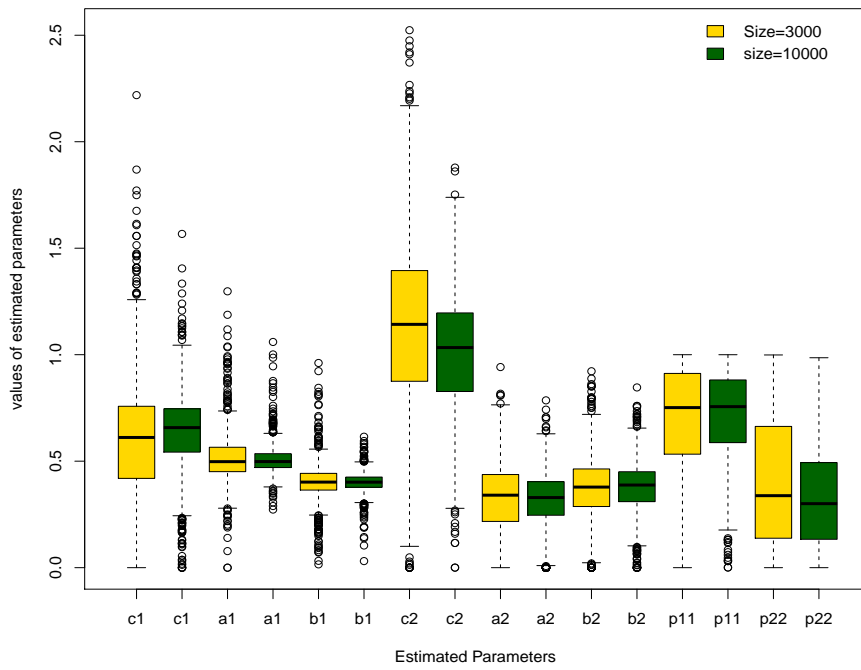


Figure 6.8: Boxplot of AAE for maximum likelihood estimator of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 0.6, 0.4, 0.4, 0.6, 0.4)$ , for different sample sizes. Results are based on 1000 simulations.

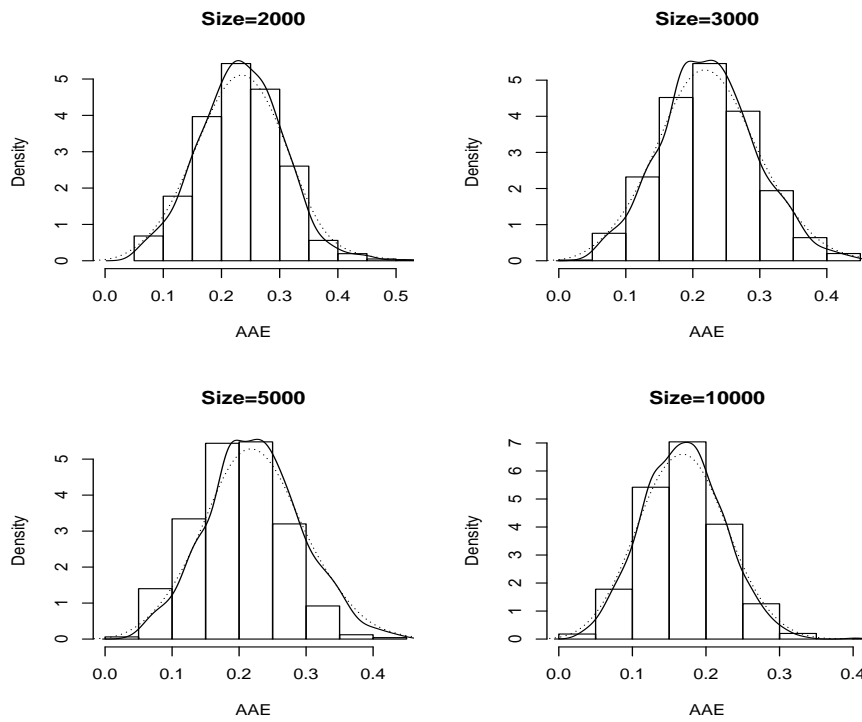


Figure 6.9: Histogram of AAE for maximum likelihood estimator of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 0.6, 0.4, 0.4, 0.6, 0.4)$ , for different sample sizes. Results are based on 1000 simulations.

**Case 4**

Here, the initial values of the parameters (autoregressive coefficients and intercept) are the same as in case 3. However, the initial values of the transition probabilities ( $p_{11}$  and  $p_{22}$ ) are fixed. That is,  $p_{11} = p_{22} = 0.5$ . The gradient function is obtained in a similar way to case 3, as well as, the number of simulations and tolerance level.

Table 6.15 reports the results of simulation for MS-GARCH (2;1,1) model with different parameters. The first model has the parameters  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 1.0, 0.4, 0.3, 0.7, 0.3)$ , whereas the parameters of the second model are  $(0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.7, 0.7)$ . In the first model, the transition probability is 0.3 from the first state ( $S_t = 1$ ) to the other, while it is 0.7 from the second state ( $S_t = 2$ ) to the first state. In contrast, the transition probability between states is 0.3 in the second model. Table 6.15 shows the MLEs with their MSE (in parentheses) as well as the AAE of the MLE for different sample sizes. The table shows that the MSE values of the MLE for all parameters become smaller as the sample size increases from 500 to 2000. In Table 6.15, the AAE values decrease gradually as the sample size increases in both models. Table 6.16 reports simulated standard error for two various models defined by (6.9) and (6.10) for different sample sizes. The simulated standard error for intercept, regressive coefficients and the transition probabilities reduces clearly as the sample size increases from 500 to 2000 in both models. Table 6.17 presents the variance of average absolute error for different models defined by (6.9) and (6.10) for different sample sizes. Clearly, it reduces gradually when the sample size increases in both models. Table 6.18 presents the descriptive information of AAE for model (6.9) and (6.10) for different sample sizes. The values of different descriptive information of AAE for the model decrease when the sample size increases. Table 6.18 reports the range of AAE for different models are 1.211 and 0.610 respectively when the sample size is 500, while they are 0.599 and 0.484 when the sample

Table 6.15: Maximum likelihood estimates with their mean square errors (in parentheses), and the average absolute error of estimates for two different MS-GARCH (2;1,1) models for different sample sizes. Results obtained from 1000 simulations.

Parameters		Sample	Maximum Likelihood Estimators (MLE)								Average Absolute Error
Regime 1	Regime 2	Sizes	$c_1$	$a_1$	$b_1$	$c_2$	$a_2$	$b_2$	$p_{11}$	$p_{22}$	AAE
$c_1=0.8$	$c_2=1.0$	500	0.8946	0.4586	0.4024	2.1639	0.3065	0.2911	0.7377	0.3504	0.3470
$a_1=0.5$	$a_2=0.4$		(0.3080)	(0.0322)	(0.0181)	(2.9043)	(0.0714)	(0.0350)	(0.0829)	(0.0947)	
$b_1=0.4$	$b_2=0.3$	1000	0.7571	0.4950	0.4003	1.8960	0.2887	0.2915	0.7405	0.3460	0.2885
$p_{11}=0.7$	$p_{22}=0.3$		(0.1836)	(0.0256)	(0.0149)	(1.5345)	(0.0565)	(0.0287)	(0.0719)	(0.1027)	
		1500	0.7187	0.4992	0.3988	1.8318	0.2975	0.2831	0.7527	0.3422	0.2686
			(0.1543)	(0.0206)	(0.0110)	(1.1610)	(0.0542)	(0.0278)	(0.0668)	(0.1054)	
		2000	0.6814	0.5044	0.3988	1.7864	0.2948	0.2841	0.7630	0.3211	0.2518
			(0.1279)	(0.0120)	(0.0082)	(1.0026)	(0.0470)	(0.0252)	(0.0599)	(0.0960)	
$c_1=0.6$	$c_2=0.7$	500	0.7082	0.5587	0.4282	0.7885	0.4684	0.2789	0.6280	0.6234	0.2454
$a_1=0.54$	$a_2=0.45$		(0.3883)	(0.0638)	(0.0413)	(0.3114)	(0.0419)	(0.0291)	(0.0860)	(0.0933)	
$b_1=0.45$	$b_2=0.3$	1000	0.6088	0.5881	0.4186	0.6748	0.4773	0.2665	0.6707	0.6574	0.1990
$p_{11}=0.7$	$p_{22}=0.7$		(0.2269)	(0.0435)	(0.0245)	(0.1893)	(0.0290)	(0.0208)	(0.0611)	(0.0700)	
		1500	0.5503	0.5867	0.4256	0.6778	0.4839	0.2692	0.6652	0.6677	0.1807
			(0.1761)	(0.0341)	(0.0227)	(0.1565)	(0.0253)	(0.0169)	(0.0567)	(0.0570)	
		2000	0.6814	0.5044	0.3988	1.7864	0.2948	0.2841	0.7630	0.3211	0.1652
			(0.1549)	(0.0224)	(0.0183)	(0.1354)	(0.0170)	(0.0112)	(0.0439)	(0.0460)	

NOTE: The estimates are obtained using starting values for the transition probabilities  $p_{11} = p_{22} = 0.5$  and the starting values for regressive coefficients were obtained by the ARIMA (1,1) fit to the data. Results are based on 1000 simulations.

size is 1500. In both models, the range of AAE decreases as the sample size increases. Table 6.19 reports the maximum likelihood estimation of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 1.0, 0.4, 0.3, 0.7, 0.3)$ , and the descriptive statistics of estimates. In addition, it shows that the RMSE values reduce slowly as the sample size increases. Figure 6.10 shows the values of AAE for maximum likelihood estimator of model (6.9) and (6.10) for different sample sizes. The AAE values become smaller when the sample size increases. The plot shows the estimates have slight outliers. Figure 6.11 presents the boxplot of the estimated parameter values for model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 1.0, 0.4, 0.3, 0.7, 0.3)$ , for two different sample sizes. The values of the outliers of large sample sizes (1000) are smaller in comparison to the values of the outliers of 500. Figure 6.11 shows the tendency of the MLE estimates to become more efficient with smaller values of the outliers as the sample size increases from 500 to 1000. Figure 6.12 illus-

Table 6.16: The simulated standard error for two different MS-GARCH (2;1,1) models for different sample sizes for different sample sizes. The results are based on 1000 simulations each.

Parameters		Sample	Simulated Standard Error							
Regime 1	Regime 2	Sizes	$c_1$	$a_1$	$b_1$	$c_2$	$a_2$	$b_2$	$p_{11}$	$p_{22}$
$c_1=0.8$	$c_2=1.0$	500	0.5550	0.1794	0.1347	1.7042	0.2672	0.1872	0.2879	0.3077
$a_1=0.5$	$a_2=0.4$	1000	0.4285	0.1600	0.1220	1.2387	0.2378	0.1694	0.2682	0.3205
$b_1=0.4$	$b_2=0.3$	1500	0.3929	0.1435	0.1051	1.0775	0.2328	0.1668	0.2585	0.3247
$p_{11}=0.7$	$p_{22}=0.3$	2000	0.3577	0.10957	0.0908	1.0013	0.2168	0.1589	0.2447	0.3099
$c_1=0.6$	$c_2=0.7$	500	0.6231	0.2526	0.2031	0.5581	0.2047	0.1707	0.2932	0.3054
$a_1=0.54$	$a_2=0.45$	1000	0.4763	0.2086	0.1567	0.4351	0.1703	0.1442	0.2471	0.2646
$b_1=0.45$	$b_2=0.3$	1500	0.4197	0.1848	0.1505	0.3956	0.1592	0.1298	0.2380	0.2387
$p_{11}=0.7$	$p_{22}=0.7$	2000	0.3577	0.16957	0.1321	0.3456	0.1368	0.1189	0.2147	0.2099

Table 6.17: The variance of average absolute error for two different MS-GARCH (2;1,1) models for different sample sizes for different sample sizes. The results are based on 1000 simulations each.

Regime 1	Regime 2	Sample Size	Variance of AAE
$c_1=0.8$	$c_2=1.0$	500	0.02421
$a_1=0.5$	$a_2=0.4$	1000	0.01087
$b_1=0.4$	$b_2=0.3$	1500	0.00701
$p_{11}=0.7$	$p_{22}=0.3$	2000	0.00670
$c_1=0.6$	$c_2=0.7$	500	0.00830
$a_1=0.54$	$a_2=0.45$	1000	0.00559
$b_1=0.45$	$b_2=0.3$	1500	0.0050
$p_{11}=0.7$	$p_{22}=0.7$	2000	0.00470

trates histogram of AAE for maximum likelihood estimator of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 1.0, 0.4, 0.3, 0.7, 0.3)$ , for different sample sizes. Figure 6.12 reports that the histogram of AAE could be symmetric.



Table 6.18: Summary of average absolute errors for two different MS-GARCH (2;1,1) models for different sample sizes for different sample sizes. Results are based on 1000 simulations.

Regime		Sample Sizes	Summary					
One	Two	$n$	Lower	25%	Median	Mean	75%	Upper
$c_1=0.8$	$c_2=1.0$	500	0.05784	0.24233	0.31300	0.34705	0.41400	1.26846
$a_1=0.5$	$a_2=0.4$	1000	0.06537	0.21919	0.27835	0.28850	0.34231	0.97494
$b_1=0.4$	$b_2=0.3$	1500	0.04115	0.21281	0.26278	0.26862	0.32127	0.64004
$p_{11}=0.7$	$p_{22}=0.3$	2000	0.03652	0.19366	0.24771	0.25179	0.30073	0.58622
$c_1=0.6$	$c_2=0.7$	500	0.04354	0.18127	0.23267	0.24541	0.29440	0.65320
$a_1=0.54$	$a_2=0.45$	1000	0.03277	0.14693	0.19212	0.19898	0.23872	0.52493
$b_1=0.45$	$b_2=0.3$	1500	0.03455	0.12819	0.17104	0.18069	0.22135	0.51901
$p_{11}=0.7$	$p_{22}=0.7$	2000	0.02452	0.11436	0.16541	0.17058	0.21457	0.42624

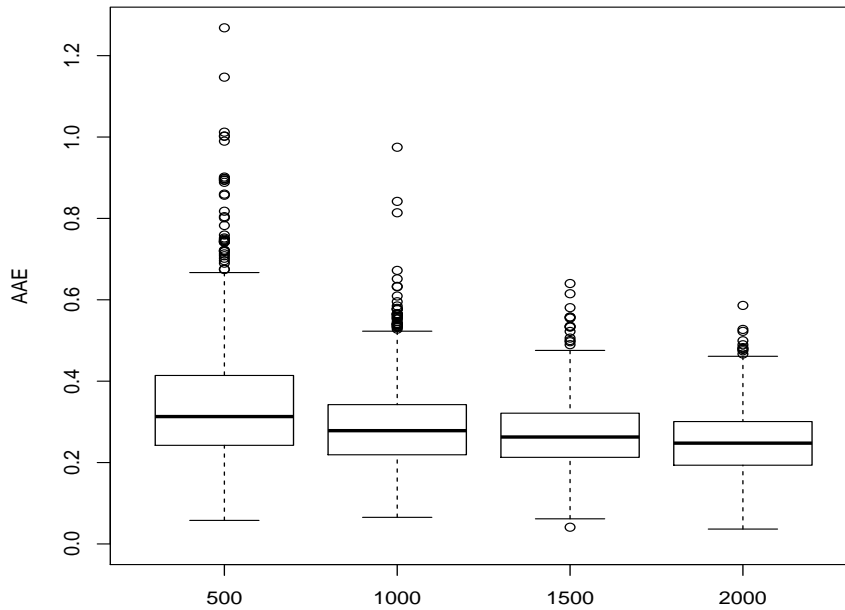


Figure 6.10: Boxplot of AAE for maximum likelihood estimator of model ((6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 1.0, 0.4, 0.3, 0.7, 0.3)$ , for different sample sizes. Results are based on 1000 simulations.

Table 6.19: Distribution of the maximum likelihood estimates of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 1.0, 0.4, 0.3, 0.7, 0.3)$ , for different sample sizes. Results are based on 1000 simulations.

Sample size $n$	Parameter	Value	Mean	RMSE	Summary				
					Lower 25%	Median	75%	Upper	
500	$c_1$	0.8	0.8946	0.4249	0.0000	0.5502	0.8716	1.2090	3.7190
	$a_1$	0.5	0.4586	0.1323	0.0000	0.3668	0.4567	0.5557	1.6090
	$b_1$	0.4	0.4024	0.0933	0.0000	0.3324	0.3874	0.4539	1.1150
	$c_2$	1.0	2.1640	1.2743	0.0000	1.3080	1.9500	2.8320	8.6520
	$a_2$	0.4	0.3065	0.2196	0.0000	0.0867	0.2882	0.4698	1.5230
	$b_2$	0.3	0.2911	0.1481	0.0000	0.1629	0.2981	0.4089	0.9703
	$p_{11}$	0.7	0.7377	0.2378	0.0000	0.5926	0.8385	0.9656	1.0000
	$p_{22}$	0.3	0.3504	0.2459	0.0000	0.0975	0.2872	0.5120	0.9975
1000	$c_1$	0.8	0.7571	0.3324	0.0000	0.4994	0.7617	1.0120	3.1850
	$a_1$	0.5	0.4950	0.1087	0.0000	0.4200	0.4879	0.5596	1.3620
	$b_1$	0.4	0.4003	0.0771	0.0000	0.3451	0.3891	0.4403	1.2590
	$c_2$	1.0	1.8960	0.9785	0.0000	1.3350	1.8370	2.3620	7.2980
	$a_2$	0.4	0.2887	0.1943	0.0000	0.1223	0.2885	0.4312	1.3560
	$b_2$	0.3	0.2915	0.1322	0.0000	0.1839	0.3015	0.3932	1.1200
	$p_{11}$	0.7	0.7405	0.2216	0.0000	0.6150	0.8370	0.9392	1.0000
	$p_{22}$	0.3	0.3460	0.2631	0.0000	0.0699	0.2605	0.5711	0.9971
1500	$c_1$	0.8	0.7187	0.2899	0.0000	0.4936	0.7382	0.9348	4.0960
	$a_1$	0.5	0.4992	0.0938	0.0000	0.4357	0.4932	0.5567	1.5220
	$b_1$	0.4	0.3988	0.0699	0.0000	0.3491	0.3869	0.4355	0.9283
	$c_2$	1.0	1.8320	0.8938	0.0000	1.3460	1.8140	2.2890	4.4480
	$a_2$	0.4	0.2975	0.1876	0.0000	0.1382	0.2976	0.4338	1.6450
	$b_2$	0.3	0.2831	0.1260	0.0000	0.1857	0.2912	0.3776	1.5050
	$p_{11}$	0.7	0.7527	0.2175	0.0000	0.6259	0.8492	0.9413	1.0000
	$p_{22}$	0.3	0.3422	0.2705	0.0000	0.0427	0.2477	0.3422	0.9982
2000	$c_1$	0.8	0.6814	0.2709	0.0000	0.4994	0.7057	0.8997	2.0640
	$a_1$	0.5	0.5044	0.0773	0.0000	0.4480	0.4997	0.5562	1.0740
	$b_1$	0.4	0.3988	0.0584	0.0000	0.3571	0.3887	0.4265	1.1550
	$c_2$	1.0	1.786	0.8407	0.0000	1.356	1.751	2.1800	3.8300
	$a_2$	0.4	0.2948	0.1770	0.0000	0.1537	0.2905	0.4313	0.9977
	$b_2$	0.3	0.2841	0.1243	0.0000	0.1841	0.2918	0.3890	0.9371
	$p_{11}$	0.7	0.7630	0.2037	0.0000	0.6667	0.8297	0.9389	1.0000
	$p_{22}$	0.3	0.3211	0.2622	0.0000	0.0191	0.2160	0.5491	0.9963

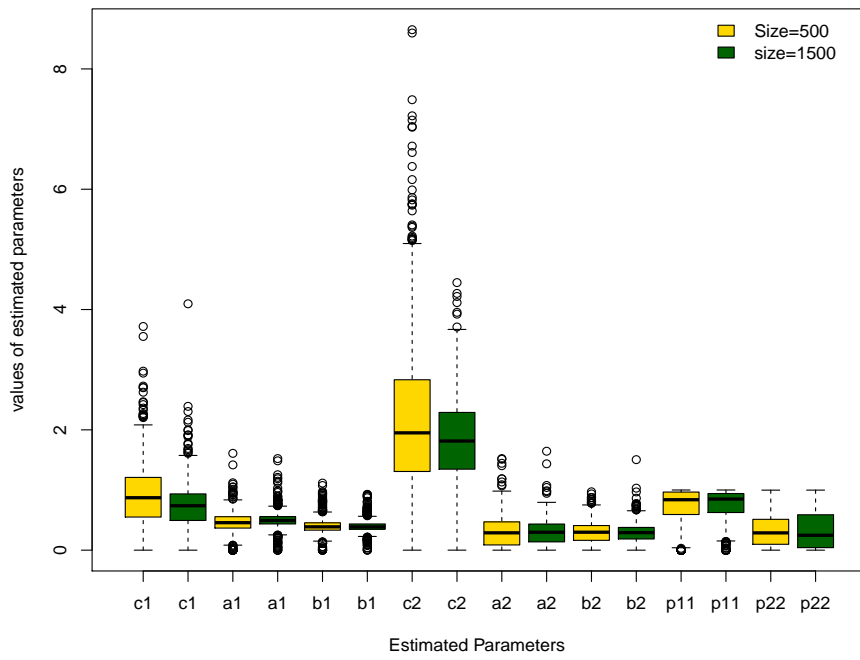


Figure 6.11: Boxplot of AAE for maximum likelihood estimator of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 1.0, 0.4, 0.3, 0.7, 0.3)$  for different sample sizes. Results are based on 1000 simulations.

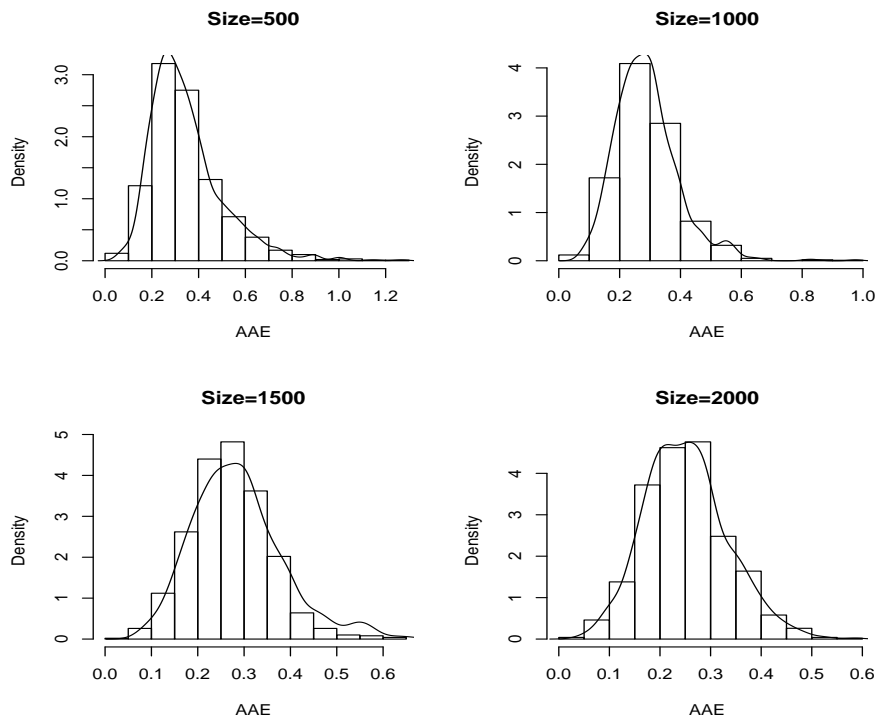


Figure 6.12: Histogram of AAE for maximum likelihood estimator of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.8, 0.5, 0.4; 1.0, 0.4, 0.3, 0.7, 0.3)$ , for different sample sizes. Results are based on 1000 simulations.

**Case 5**

In the last case, the estimates are obtained using the starting values  $p_{11} = p_{22} = 0.5$ . The starting values for autoregressive coefficients and intercept are obtained using a uniform distribution as follows:  $(c_i \sim \text{unif}(0.1, 0.9), a_i \sim \text{unif}(0.1, 0.9)$  and  $b_i \sim \text{unif}(0.1, 0.9), i = 1, 2)$ . The gradient function and the number of replications as well as the tolerance level are obtained in a similar way to cases 3 and 4.

Table 6.20 contains the results for the data that was generated in a similar way to the second model in the simulation case 4. That is,

$$\lambda_t = \begin{cases} 0.6 + 0.54\lambda_{t-1} + 0.45Y_{t-1}, & \text{if } S_t = 1, \\ 0.7 + 0.45\lambda_{t-1} + 0.3Y_{t-1}, & \text{if } S_t = 2. \end{cases} \quad (6.18)$$

It shows the MLEs with their MSE (in parentheses) for different sample sizes. The table reports that the values of AAE decrease clearly as the sample size increases from 500 to 2000.

Table 6.20: Maximum likelihood estimates with their mean square errors (in parentheses), and the average absolute errors of estimates for model (6.18) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.7, 0.7)$ , for different sample sizes. Results obtained from 1000 simulations.

Sample size	Maximum Likelihood Estimates (MLE)								Average Absolute Error
$n$	$c_1$	$a_1$	$b_1$	$c_2$	$a_2$	$b_2$	$p_{11}$	$p_{22}$	AAE
500	0.6476 (0.2890)	0.5249 (0.0544)	0.3738 (0.0498)	0.6719 (0.2781)	0.5403 (0.0643)	0.3673 (0.0519)	0.6149 (0.0919)	0.6057 (0.0914)	0.2442
1000	0.6167 (0.2116)	0.5549 (0.0384)	0.3574 (0.0443)	0.6088 (0.1971)	0.5389 (0.0506)	0.3612 (0.0374)	0.6481 (0.0680)	0.6439 (0.0701)	0.2125
1500	0.5866 (0.1494)	0.5543 (0.0301)	0.3617 (0.0388)	0.6133 (0.1641)	0.5458 (0.0389)	0.3571 (0.0314)	0.6543 (0.0631)	0.6624 (0.0613)	0.1928
2000	0.5639 (0.1107)	0.5480 (0.0246)	0.3553 (0.0353)	0.5673 (0.1342)	0.5512 (0.0341)	0.3573 (0.0262)	0.6590 (0.0530)	0.6695 (0.0513)	0.1759

Table 6.21 reports the simulated standard error of models defined by (6.18) for

different sample sizes. The simulated standard error for intercept, autoregressive coefficients and the transition probabilities reduces clearly as the sample size increases from 500 to 2000. Table 6.22 reports the maximum likelihood estimation

Table 6.21: The simulated standard error for model (6.18) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.7, 0.7)$ , for different sample sizes. The results are based on 1000 simulations each.

Sample size		Simulated Standard Error						
$n$	$c_1$	$a_1$	$b_1$	$c_2$	$a_2$	$b_2$	$p_{11}$	$p_{22}$
500	0.5376	0.2332	0.2232	0.5274	0.2535	0.2277	0.3031	0.3023
1000	0.4600	0.1959	0.2105	0.4439	0.2248	0.1934	0.2608	0.2647
1500	0.3866	0.1734	0.1969	0.4051	0.1972	0.1771	0.2512	0.2475
2000	0.3327	0.1570	0.1878	0.3664	0.1847	0.1617	0.2303	0.2265

of model (6.18) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.7, 0.7)$ , and the descriptive statistics of estimates. In addition, it shows that the RMSE values decrease slowly as the sample size increases.

Figure 6.13 shows the values of AAE for the maximum likelihood estimator of model (6.18) for different sample sizes. The AAE values become smaller when the sample size increases. The plot shows the estimates have slight outliers.

Figure 6.14 presents the boxplot of the estimated parameter values of model (6.10) for two different sample sizes. The values of the outliers of large sample sizes (2000) are smaller in comparison to the values of the outliers of sample size 1000. It shows the tendency of the MLE estimates to become more efficient with smaller values of the outliers as the sample size increases from 1000 to 2000. The histogram of AAE for maximum likelihood estimator of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.7, 0.7)$ , for different sample sizes is shown in Figure 6.15.

Table 6.22: Distribution of the maximum likelihood estimates of model (6.18) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.7, 0.7)$ , for different sample sizes. Results are based on 1000 simulations.

Sample size $n$	Parameter	Value	Mean	RMSE	Summary				
					Lower 25%	Median	75%	Upper	
500	$c_1$	0.6	0.6476	0.3859	0.0000	0.2910	0.5804	0.8834	5.5810
	$a_1$	0.54	0.5249	0.1699	0.0000	0.4032	0.5289	0.6435	1.7870
	$b_1$	0.45	0.3738	0.1758	0.0000	0.2549	0.3527	0.4720	1.4200
	$c_2$	0.7	0.6719	0.3924	0.0000	0.3129	0.6221	0.9032	3.8140
	$a_2$	0.45	0.5403	0.1873	0.0000	0.4140	0.5230	0.6653	1.5710
	$b_2$	0.3	0.3673	0.1642	0.0000	0.2476	0.3460	0.4747	1.6350
	$p_{11}$	0.7	0.6149	0.2399	0.0000	0.4367	0.6687	0.8644	1.0000
	$p_{22}$	0.7	0.6057	0.2386	0.0000	0.4113	0.6660	0.8487	1.0000
1000	$c_1$	0.6	0.6167	0.3390	0.0000	0.3062	0.5783	0.8259	3.3200
	$a_1$	0.54	0.5549	0.1405	0.0000	0.4548	0.5501	0.6485	1.4340
	$b_1$	0.45	0.3574	0.1636	0.0000	0.2568	0.3460	0.4421	1.4810
	$c_2$	0.7	0.6088	0.3468	0.0000	0.3125	0.5689	0.8248	2.7020
	$a_2$	0.45	0.5389	0.1617	0.0000	0.4458	0.5316	0.6303	1.8150
	$b_2$	0.3	0.3612	0.1412	0.0000	0.2556	0.3483	0.4487	1.3280
	$p_{11}$	0.7	0.6481	0.2038	0.0000	0.4904	0.7012	0.8569	1.0000
	$p_{22}$	0.7	0.6439	0.2036	0.0000	0.4977	0.7028	0.8402	1.0000
1500	$c_1$	0.6	0.5866	0.2944	0.0000	0.3408	0.5485	0.7856	2.5620
	$a_1$	0.54	0.5543	0.1217	0.0000	0.4624	0.5453	0.6385	1.3520
	$b_1$	0.45	0.3617	0.1521	0.0000	0.2718	0.3494	0.4303	1.4830
	$c_2$	0.7	0.6133	0.3167	0.0000	0.3611	0.5765	0.8100	2.7560
	$a_2$	0.45	0.5458	0.1479	0.0000	0.4612	0.5398	0.6253	1.5140
	$b_2$	0.3	0.3571	0.1241	0.0000	0.2713	0.3500	0.4289	1.5650
	$p_{11}$	0.7	0.6543	0.1936	0.0000	0.5187	0.7089	0.8437	1.0000
	$p_{22}$	0.7	0.6624	0.1922	0.0000	0.5224	0.7049	0.8603	1.0000
2000	$c_1$	0.6	0.5639	0.2575	0.0000	0.3492	0.5524	0.7491	2.0680
	$a_1$	0.54	0.5480	0.1064	0.0000	0.4679	0.5350	0.6100	1.5360
	$b_1$	0.45	0.3553	0.1488	0.0000	0.2672	0.3431	0.4272	1.5950
	$c_2$	0.7	0.5673	0.2908	0.0000	0.3445	0.5527	0.7616	2.6160
	$a_2$	0.45	0.5512	0.1402	0.0000	0.4752	0.5473	0.6235	1.4500
	$b_2$	0.3	0.3573	0.1173	0.0000	0.2747	0.3536	0.4248	1.5280
	$p_{11}$	0.7	0.6590	0.1762	0.0000	0.5324	0.7060	0.8302	1.0000
	$p_{22}$	0.7	0.6695	0.1698	0.0000	0.5573	0.7122	0.8288	1.0000

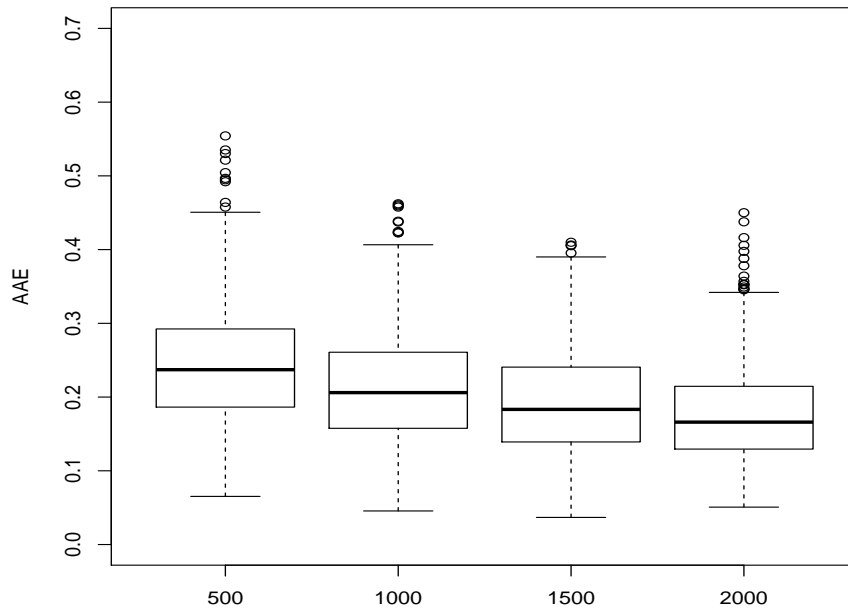


Figure 6.13: Boxplot of AAE for maximum likelihood estimator of model (6.18) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.7, 0.7)$ , for different sample sizes. Results are based on 1000 simulations.



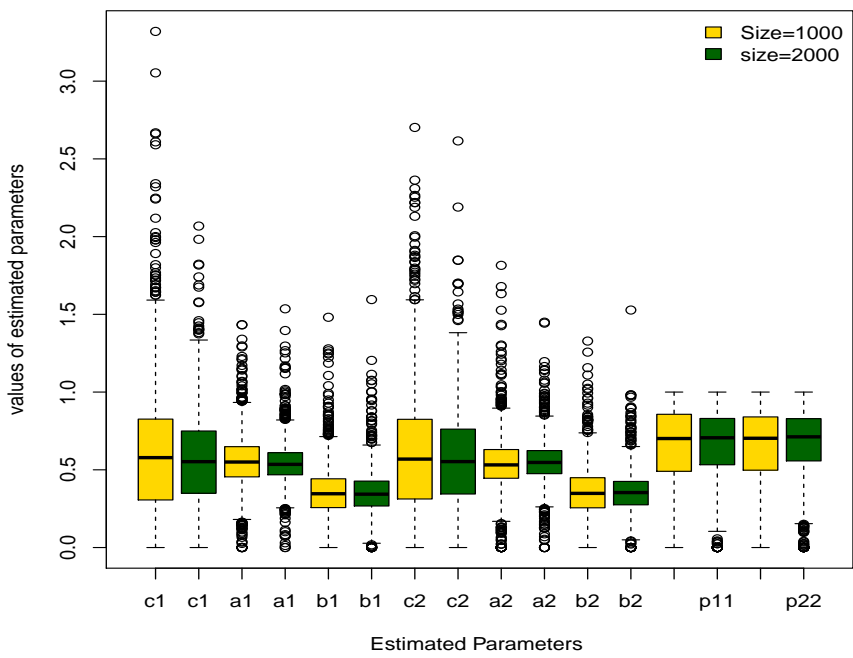


Figure 6.14: Boxplot of AAE for maximum likelihood estimator of model (6.9) and (6.10) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.7, 0.7)$  for different sample sizes. Results are based on 1000 simulations.

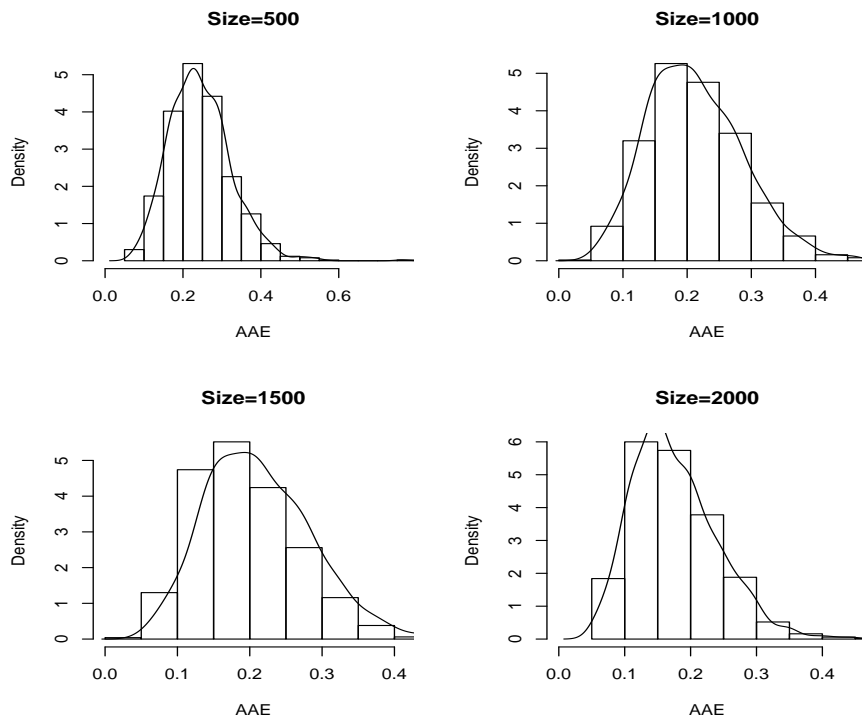


Figure 6.15: Histogram of AAE for maximum likelihood estimator of model (6.18) when  $(c_1, a_1, b_1; c_2, a_2, b_2, p_{11}, p_{22}) = (0.6, 0.54, 0.45; 0.7, 0.45, 0.3, 0.7, 0.7)$ , for different sample sizes. Results are based on 1000 simulations.

### 6.3.3 A real data example

There are several authors who have used a certain example data for different TAR and MSAR models. One example of such data is the growth rate of the U.S quarterly real gross national product from 1947 II to 1991 I. These data have been widely used in nonlinear analysis of economic time series. Tiao and Tsay (1994) and Potter (1995) employ TAR models, while Hamilton (1989) and McCulloch and Tsay (1994) used Markov switching models. Therefore, in this section we applied model (6.9) and (6.10) to illustrate our methodology of the real data (as described in section 3.1), which have been also applied to the TGARCH model in Chapter 4. The starting values of the model parameters in the real data case are the same to those used in case 5 of the simulation part of this chapter.

Table 6.23 shows the estimated parameters of models defined by (6.9). The estimates of the parameters are those corresponding to the minimum value of the log likelihood function over all simulations. The steady-state distribution of the Markov chain  $S_t$  is

$$\pi = \begin{bmatrix} 0.743 \\ 0.257 \end{bmatrix}$$

Table 6.23: The estimated parameters for model (6.9).

		First State				
Parameters	$c_1$	$a_1$	$b_1$	$p_{11}$	$p_{12}$	
Estimated value	0.053	0.870	0.099	0.716	0.284	
		Second State				
Parameters	$c_2$	$a_2$	$b_2$	$p_{21}$	$p_{22}$	
Estimated value	3.484	0.915	0.084	0.820	0.180	

NOTE: The estimates are obtained by using starting values from the uniform distribution. The results are based on 1000 simulations.

The results in Table 6.23 can be expressed as

$$\hat{\lambda}_t = \begin{cases} 0.053 + 0.099Y_{t-1} + 0.870\hat{\lambda}_{t-1}, & \text{if } S_t = 1, \\ 3.484 + 0.084Y_{t-1} + 0.915\hat{\lambda}_{t-1}, & \text{if } S_t = 2. \end{cases}$$

From Table 6.23, it can be seen that there are several interesting findings. First, the limiting unconditional mean (as  $t \rightarrow \infty$ ) of the number of transactions per minute of the original model for state 1 is  $\frac{c_1}{1-(a_1+b_1)} = 1.710$  and that of state 2 is  $\frac{c_2}{1-(a_2+b_2)} = 3,484$ . Hence, state 1 consists of minutes with a expansion period, while state 2 corresponds to minutes with contraction period. Second, the transition probabilities appear to be different for different states. The process tends to stay longer in state 1 than in state 2. Third, the expected duration for the process to stay in the state 1 is  $\frac{1}{1-p_{11}} = 3.521$ , while the expected duration of the other state is  $\frac{1}{1-p_{22}} = 1.220$ . That is, the expected duration for a contraction period and an expansion period are approximately 1.220 and 3.521 minutes. Figure 6.16 shows filtered and smoothed probabilities of the real data described in section (3.1). The lines depicting the filtered and smoothed probabilities of the real data clearly have similar shape (pattern). The filtered and smoothed probabilities in both states are observed to follow the same pattern, without practically any separation between the two lines.

## 6.4 Summary

Volatility clustering over time is also a stylized fact for discrete financial data. Therefore, the GARCH-type model is applied for studying this pattern of count time series data. The Markov switching Poisson GARCH model is used to describe the asymmetric phenomena in volatility of count data. We have studied the probability structure of the MS-PGARCH model. The maximum likelihood estimation of model (6.9) is investigated. Many simulations were carried out to

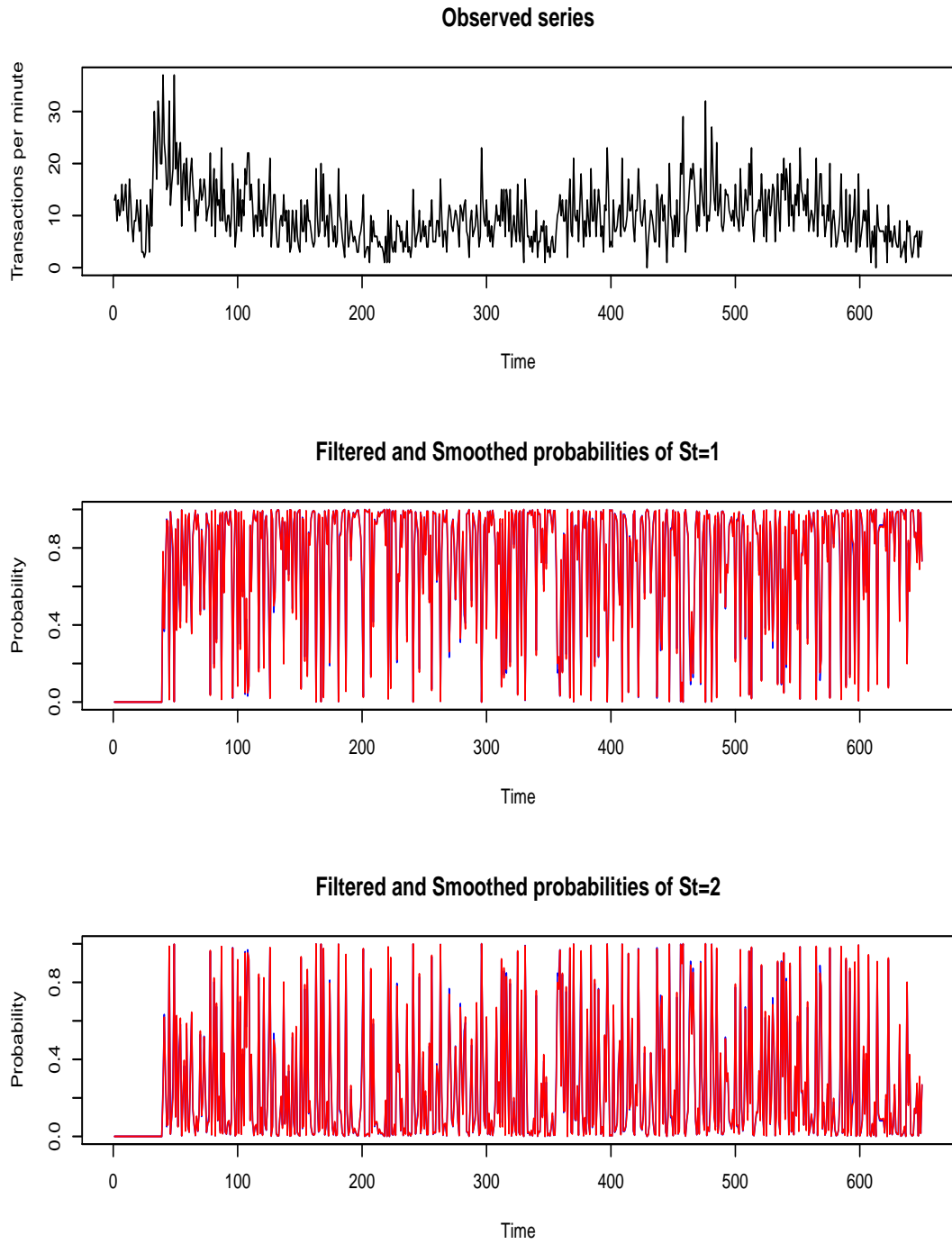


Figure 6.16: From top to bottom: Time plot of the number of transactions per minute for the stock Ericsson B in the period July 2-22,2002, filtered (blue line) and smoothed (red line) probabilities of the  $S_t = i, i = 1, 2$  respectively.

estimate the model parameters with different sample sizes and different initial values of the parameters. The starting values play an important role in accuracy of the estimated parameters. It is concluded that when the MLE can provide reasonable estimation of the MS-PGARCH model. An application of the MS-GARCH model to a real count time series data is presented.

# Chapter 7

## Conclusions and Further Work

### 7.1 Conclusions

The aims of this thesis include studying the ergodicity and statistical inference for a specific kind of threshold generalized autoregressive conditional heteroscedastic (PTGARCH) models for nonnegative integer-valued time series when the conditional distribution given historical information is the Poisson distribution, and also finding an answer to the question: "How to build a complete structure of PTGARCH models?" These models are characterized by frequent switches of two or more regimes, based on the threshold variable  $(Y_{t-d})$ . Taking into consideration the many numerical simulations and a real data set, we estimate the parameters  $d_i, a_i, b_i, s, r_i, i = 1, 2, \dots, s$  and  $d$  for the PTGARCH model.

In chapter 2, we established a simple sufficient condition for the geometric ergodicity of model (2.1), and we discussed the problem of statistical tools including the maximum likelihood estimators of parameters  $(d_i, a_i$  and  $b_i, i = 1, 2, \dots, s)$  and their asymptotic theory. We presented four different information criteria (AIC, BIC, BIC2 and HQIC) for determining the threshold value  $(r)$ , time delay  $(d)$  and the number of regimes  $(s)$ . We also discussed the use of two types of nonlinearity tests, Lagrange multiplier (LM) and Likelihood ratio (LR) test.

In chapter 3, we carried out many simulation results to estimate the parameters using MLE for two regimes of the PTGARCH model to make the analysis simple. We computed MLEs by optimizing the log-likelihood function (2.7), using a quasi-Newton method. We also reported a numerical simulation outcome for selecting the appropriate threshold value ( $r$ ) using information criteria. We tested the threshold nonlinearity using  $LM^*$  defined by (2.14) and (2.15). The simulation outcomes showed that our methodology is optimized for estimating the parameters of the PTGARCH model. In addition, the average absolute error (AAE) of the MLE decreases when the sample size increases, and the mean square error (MSE) of the MLE is lower than that of the conditional least square error (CLSE), for all sample sizes. We found that the Q-Q plot of the sampling distribution of the MLE of the parameters, supports the claim of asymptotic normality. Finally, we used a numerical example to illustrate our methodology.

In chapter 4, we investigated the performance of various information criteria for selecting the number of regimes in different sample sizes, when the threshold value ( $r$ ) and time delay ( $d$ ) are known. We used the threshold nonlinearity test to define the kind of model, either one regime ( $s=1$ ) or more than one regime ( $s > 1$ ). We found that the BIC and BIC2 have a strong performance for the simpler linear model ( $s=1$ ), and the AIC and HQIC perform very well for the complex PTGARCH models ( $s = 2, s = 3$ , and  $s = 2, b_i = 0, i = 1, 2$ ) in different sample sizes, while AIC, BIC and HQIC have good performance for nonlinear model ( $s = 2, a_i = 0, i = 1, 2$ ). Using various information criteria, we investigated the number of regimes in time series of the number of transactions per minute for the stock Ericsson B for the period July 2-3, 2002.

In chapter 5, we carried out many simulations to determine the delay parameter ( $d$ ), and to identify the time delay and the value of threshold ( $r$ ) simultaneously using the normalized information criteria. By applying this method, one can build a complete structure of a PTGARCH model. We applied the model (3.1)



to the asthma data when the number of regimes are given.

In the final chapter, we introduced another kind of the regime switching model. We focussed on a Markov-switching generalized autoregressive heteroscedastic (MS-PGARCH)-type Poisson time series model. We investigated the geometric ergodicity process by Markov theory and irreducibility. We treated the problem of maximum likelihood estimation of the parameters. We carried out many different simulations with regards to the optimization function arguments, for different sample sizes, to estimate the parameters of (MS-PGARCH) model. A real data example was used afterwards to confirm the results of the simulations.

## 7.2 Further work

In this field, there are many interesting open questions. For example,

- Is it possible to develop a theory for obtaining a simple sufficient condition for ergodicity or geometrical ergodicity for an unperturbed model?
- How to develop a theory to obtain a necessary and sufficient condition for stationarity for a given class of nonlinear count time series model?
- Is it possible to estimate the parameters of a model defined by (2.1) using other estimation methods such as Bayesian estimation and Gibbs sampler?
- Is there a measure (criterion) capable of determining the kind of model (linear or nonlinear), as well as of selecting the number of regimes, threshold value and time delay, e.g. MSE (Al-Awadhi and Jolliffe (1998))?

- Could all aspects of the research in this project be done in the general PTGARCH  $(p, q; s, r)$  model, defined as

$$\left\{ \begin{array}{l} Y_t | \mathcal{F}_{t-1}^{Y, \lambda} \sim \text{Poisson}(\lambda_t), \\ \lambda_t = \phi_{j0} + \sum_{i=1}^q \phi_{ji} Y_{t-i} + \sum_{m=1}^p \varphi_{jm} \lambda_{t-m}, \\ \text{when } r_{j-1} < Y_{t-d} \leq r_j, \quad j = 1, \dots, s \end{array} \right.$$

where  $\phi_{j0} > 0$ ,  $\phi_{ji} \geq 0, i = 1, 2, \dots, q$ ,  $\varphi_{jm} \geq 0, m = 1, 2, \dots, p$  and  $0 = r_1 < r_2 < \dots < r_{s-1} < r_s = \infty$ ?

# Bibliography

- Aide, D., Daoxian, Y., Wutian, C., and Bidaux, P. (1988). Threshold autoregressive model applied to prediction of karst spring flow. In : *Karst Hydrogeology and Karst Environment Protection, Proceedings of the 21<sup>st</sup> Congress of the International Association of Hydrogeologists, Guilin, China, 10-15 October 1988*, ed. Daoxian, Y., volume 2, pages 663–670. IAHS Press, Institute of Hydrology, Wallingford, Oxfordshire, UK.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In : *Information Theory, Proceedings of the 2<sup>nd</sup> International Symposium, 1973*, eds. Petrov, B.N. and Caski, F., pages 267–281. Akademinai Kiado, Budapest.
- Al-Awadhi, S. and Jolliffe, I. (1998). Time series modelling of surface pressure data. *International Journal of Climatology*, 18(4):443–455.
- Albert, J. H. and Chib, S. (1993). Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business & Economic Statistics*, 11(1):1–15.
- Andrews, D. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62(6):1383–1414.

- Basawa, I. (1991). Generalized score tests for composite hypotheses. In *Estimating Functions*, ed. Godambe, V.P., chapter 3, pages 121–131. Oxford, Clarendon Press.
- Berman, A. and Plemmons, R. J. (1979). *Nonnegative Matrices in the Mathematical Science*. Academic Press, New York.
- Billingsley, P. (1995). *Probability and measure*. 3<sup>rd</sup> Edition, Wiley, New York.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31(3):307–327.
- Brockwell, P. and Davis, R. (2009). *Time Series: Data Analysis and Theory*. 2<sup>nd</sup> Edition, New York, Springer.
- Brown, B. and Eagleson, G. (1971). Martingale convergence to infinitely divisible laws with finite variances. *Transactions of the American Mathematical Society*, 162:449–453.
- Buse, A. (1982). The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. *The American Statistician*, 36(3a):153–157.
- Caner, M. and Hansen, B. E. (2001). Threshold autoregression with a unit root. *Econometrica*, 69(6):1555–1596.
- Chan, K. (1990). Deterministic stability, stochastic stability, and ergodicity. In *Non-linear Time Series: A Dynamical Systems Approach*, ed. Tong, H., volume 6 of *Oxford Statistical Science Series*. Oxford, Clarendon Press.
- Chan, K. S. (1991). Percentage points of likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3):691–696.
- Chan, K. S. and Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, 7(3):179–190.

- Chan, K.-S. and Tsay, R. S. (1998). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika*, 85(2):413–426.
- Cryer, J. and Chan, K. (2008). *Time Series Analysis - with Applications in R*. 2<sup>nd</sup> Edition, Springer, New York, USA.
- De Gooijer, J. (2001). Cross-validation criteria for SETAR model selection. *Journal of Time Series Analysis*, 22(3):267–281.
- Enders, W. (2010). *Applied Economic Time Series*. 3<sup>rd</sup> Edition, Wiley, New York.
- Engel, C. and Hamilton, J. (1989). Long swings in the exchange rate: Are they in the data and do markets know it? *Journal of the American Economic Association*, 80(4):689–713.
- Engle, R. (1984). Wald, likelihood ratio, and lagrange multiplier tests in econometrics. *Handbook of Econometrics*, 2:775–826.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 50(4):987–1007.
- Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis*, 27(6):923–942.
- Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics*, 12(3):299–308.
- Fokianos, K. and Fried, R. (2010). Interventions in INGARCH processes. *Journal of Time Series Analysis*, 31(3):210–225.
- Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, 104(488):1430–1439.

- Francq, C. and Zakoian, J. (2001). Stationarity of multivariate Markov-switching ARMA models. *Journal of Econometrics*, 102(2):339–364.
- Gao, J., Tjostheim, D., and Yin, J. (2009). Estimation in threshold autoregressive models with nonstationarity. *Working Papers*.
- Ghysels, E. (1994). On the periodic structure of the business cycle. *Journal of Business & Economic Statistics*, 12(3):289–298.
- Gonzalo, J. and Pitarakis, J. (2002). Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics*, 110(2):319–352.
- Goodwin, T. (1993). Business-cycle analysis with a Markov-switching model. *Journal of Business & Economic Statistics*, 11(3):331–339.
- Granger, C. W. and Terasvirta, T. (1993). *Modelling non-linear economic relationships*. Oxford University Press, Oxford, UK.
- Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application*. Academic press, New York, USA.
- Hamaker, E. (2009). Using information criteria to determine the number of regimes in threshold autoregressive models. *Journal of Mathematical Psychology*, 53(6):518–529.
- Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 57(2):357–384.
- Hamilton, J. and Lin, G. (1998). Stock market volatility and the business cycle. *Journal of Applied Econometrics*, 11(5):573–593.
- Hamilton, J. D. and Lin, G. (1996). Stock market volatility and the business cycle. *Journal of Applied Econometrics*, 11(5):573–593.

- Hannan, E. and Quinn, B. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(2):190–195.
- Hansen, B. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica: Journal of the Econometric Society*, 64(2):413–430.
- Hansen, B. (2011). Threshold autoregression in economics. *Statistics and its Interface*, 4(2):123–128.
- Higgins, M. L. and Bera, A. K. (1992). A class of nonlinear ARCH models. *International Economic Review*, 33(1):137–158.
- Horn, R. and Johnson, C. (1991). *Topics in matrix analysis*. Cambridge University Press, Cambridge, New York.
- Huang, C. (1999). Phases and characteristics of Taiwan business cycles: A Markov switching analysis. *Taiwan Economic Review*, 27:185–214.
- Hwang, S. and Basawa, I. (2004). Stationarity and moment structure for Box-Cox transformed threshold GARCH (1, 1) processes. *Statistics & Probability Letters*, 68(3):209–220.
- Hwang, S. and Kim, T. Y. (2004). Power transformation and threshold modeling for ARCH innovations with applications to tests for ARCH structure. *Stochastic Processes and their Applications*, 110(2):295–314.
- Isaacson, D. (1979). A characterization of geometric ergodicity. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 49(3):267–273.
- Jensen, S. and Rahbek, A. (2004). Asymptotic inference for nonstationary GARCH. *Econometric Theory*, 20(6):1203–1226.

- Jensen, S. T. and Rahbek, A. (2007). Notes and problems on the law of the large numbers for (geometrically) ergodic Markov chains. *Econometric Theory*, 23:761–766.
- Katz, R. and Skaggs, R. (1981). On the use of autoregressive-moving average processes to model meteorological time series. *Monthly Weather Review*, 109(3):479–484.
- Kedem, B. and Fokianos, K. (2002). *Regression Models for Time Series Analysis*, volume 323. Wiley, New Jersey, New York.
- Kim, C. (1994). Dynamic linear models with Markov switching. *Journal of Econometrics*, 60(1):1–22.
- Kim, C. and Nelson, C. (2003). *State-space models with regime switching: classical and Gibbs-sampling approaches with applications*. MA: MIT Press, Cambridge.
- Kuan, C. (2002). Lecture on the Markov Switching Model. *Academia Sinica Peking*.
- Li, C. and Li, W. (1998). On a double-threshold autoregressive heteroscedastic time series model. *Journal of Applied Econometrics*, 11(3):253–274.
- McCulloch, R. E. and Tsay, R. S. (1994). Statistical analysis of economic time series via Markov switching models. *Journal of Time Series Analysis*, 15(5):523–539.
- McKenzie, E. (1985). Some simple models for discrete variate time series. *Journal of the American Water Resources Association*, 21(4):645–650.
- McKenzie, E. (2003). Discrete variate time series. In *Handbook of statistics*, eds. D.N. Shanbhag and C.R. Rao, volume 21, pages 573–606. Elsevier.



- Medeiros, M., Veiga, A., and Resende, M. (2002). A combinatorial approach to piecewise linear time series analysis. *Journal of Computational and Graphical Statistics*, 11(1):236–258.
- Meitz, M. and Saikkonen, P. (2008). Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. *Econometric Theory*, 24(05):1291–1320.
- Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability*. Springer, London, UK.
- Mike, E., So, P., Lam, K., and Li, W. (1998). A stochastic volatility model with markov switching. *Journal of Business & Economic Statistics*, 16(2):244–253.
- Pan, J. and Ling, S. (2000). On the tail behaviours of first-order multiple threshold AR models (preprint). *Unpublished*.
- Pan, J., Wang, H., and Tong, H. (2008). Estimation and tests for power-transformed and threshold GARCH models. *Journal of Econometrics*, 142(1):352–378.
- Peña, D. and Rodriguez, J. (2005). Detecting nonlinearity in time series by model selection criteria. *International Journal of Forecasting*, 21(4):731–748.
- Potter, S. M. (1995). A nonlinear approach to US GNP. *Journal of Applied Econometrics*, 10(2):109–125.
- Ramchand, L. and Susmel, R. (1998). Volatility and cross correlation across major stock markets. *Journal of Empirical Finance*, 5(4):397–416.
- Rao, C. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. In *Proceedings of the Cambridge Philosophical Society*, volume 44, pages 50–57. Cambridge Univ Press, Cambridge.

- Safaei, M. (2012). Maximum likelihood estimator for markov switching autoregressive conditional heteroscedasticity model. *Journal of Basic and Applied Scientific Research*, 2(6):5652–5655.
- Schaller, H. and Van Norden, S. (1997). Regime switching in stock market returns. *Applied Financial Economics*, 7(2):177–191.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.
- Silvey, S. (1959). The Lagrangian multiplier test. *The Annals of Mathematical Statistics*, 30(2):389–407.
- So, M. E. P., Lam, K., and Li, W. K. (1998). A stochastic volatility model with Markov switching. *Journal of Business & Economic Statistics*, 16(2):244–253.
- Strikholm, B. and Teräsvirta, T. (2005). Determining the number of regimes in a threshold autoregressive model using smooth transition autoregressions. SSE/EFI Working Paper Series in Economics and Finance 578, The Economic Research Institute, Stockholm School of Economics, Stockholm.
- Strikholm, B. and Teräsvirta, T. (2006). A sequential procedure for determining the number of regimes in a threshold autoregressive model. *The Econometrics Journal*, 9(3):472–491.
- Tiao, G. C. and Tsay, R. S. (1994). Some advances in non-linear and adaptive modelling in time-series. *Journal of Forecasting*, 13(2):109–131.
- Tjøstheim, D. (1990). Non-linear time series and Markov chains. *Advances in Applied Probability*, 22(3):587–611.
- Tong, H. (1978). On a threshold model. In Chen, C., editor, *Pattern Recognition and Signal Processing*, volume 29 of *NATO ASI Series E: Applied Sciences*, pages 575–586. Sijthoff & Noordhoff, Netherlands.

- Tong, H. and Lim, K. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(3):245–292.
- Tong, H. and Series, N. (1990). *Non-Linear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford, UK.
- Tsay, R. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, 84(405):231–240.
- Tsay, R. (2005). *Analysis of Financial Time Series*. Wiley-Interscience, New Jersey, USA.
- Tsay, R. S. (1987). Conditional heteroscedastic time series models. *Journal of the American Statistical Association*, 82(398):590–604.
- Tweedie, R. L. (1975). Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stochastic Processes and their Applications*, 3(4):385–403.
- Wang, C., Liu, H., Yao, J.-F., Davis, R. A., and Li, W. K. (2014). Self-excited threshold poisson autoregression. *Journal of the American Statistical Association*, (published online 2014).
- Weiss, C. and Testik, M. (2009). CUSUM monitoring of first-order integer-valued autoregressive processes of Poisson counts. *Journal of Quality Technology*, 41(4):389–400.
- Wilks, D. (1985). *Statistical Methods in the Atmospheric Sciences*. Academic Press, San Diego.
- Wong, C. and Li, W. (1998). A note on the corrected Akaike information criterion for threshold autoregressive models. *Journal of Time Series Analysis*, 19(1):113–124.

Wu, J. (2010). Threshold GARCH Model: Theory and Application. *Department of Economics, University of Western Ontario, London, Ontario.*

Zucchini, W. and MacDonald, I. (2009). *Hidden Markov models for Time Series: an Introduction Using R.* Monographs on statistics and applied probability. Chapman and Hall/CRC, Boca Raton, Florida.