UNIVERSITY OF CALIFORNIA SAN DIEGO

Prediction in Time Series Models and Model-free Inference with a Specialization in Financial Return Data

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Mathematics (with a specialization in Statistics)

by

Jie Chen

Committee in charge:

Professor Dimitris N. Politis, Chair
Professor Ery Arias-Castro
Professor Brendan K. Beare
Professor Jelena Bradic
Professor James D. Hamilton

2018

The dissertation of Jie Chen is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2018

DEDICATION

To

My Family

For their unconditional love

Pengfei Chen

For his love and support

My Friends, Amir Babaeian, Srinjoy Das and Nan Zou, etc

For their help

EPIGRAPH

To laugh often and much;

To win the respect of intelligent people and the affection of children;

To earn the appreciation of honest critics and endure the betrayal of false friends;

To appreciate beauty;

To find the best in others;

To leave the world a bit better, whether by a healthy child, a garden patch, or a redeemed social condition;

To know even one life has breathed easier because you have lived.

This is to have succeeded.

—Ralph Waldo Emerson

TABLE OF CONTENTS

Signature Page . . . iii

Dedication . . . iv

Epigraph . . . v

Table of Contents . . . vi

List of Figures . . . viii

List of Tables . . . ix

Acknowledgements . . . xii

Vita . . . xiii

Abstract of the Dissertation . . . xiv

Chapter 1 Time-varying NoVaS vs. GARCH: Point Prediction, Volatility Estimation and Prediction Intervals . . . 1
1.1 Introduction . . . 2
1.2 Point Prediction . . . 5
1.2.1 GARCH(1,1) and NoVaS Transformation for Stationary Data . . . 5
1.2.2 Local Stationarity . . . 15
1.2.3 Simulations and Results . . . 16
1.3 Estimation of Realized Volatility . . . 19
1.3.1 Data and Summary Statistics . . . 21
1.3.2 NoVaS and GARCH(1,1) Optimization and Estimating Specifications . . . 21
1.3.3 Results and Conclusions . . . 22
1.4 Bootstrap Prediction Intervals . . . 24
1.4.1 Description of Interval Prediction Algorithms . . . 24
1.4.2 Local Stationarity . . . 28
1.4.3 Finite-Sample Performance of Model-Free and Model-Based Prediction Intervals . . . 29
1.5 Concluding Remarks . . . 33
1.6 Acknowledgements . . . 34

Chapter 2 Optimal Multistep-ahead Prediction of Nonlinear Time Series Models and Model-free Inference . . . 42
2.1 Introduction . . . 42
2.2 Optimal Multistep-ahead Point Prediction . . . 43

2.2.1 For ARCH(p) and GARCH(1,1) models 44
2.2.2 For NoVaS 47
2.2.3 Generalization 48
2.3 Prediction Intervals 52
2.4 Simulation and Finite Sample performance 56
2.4.1 Settings 56
2.4.2 Results and Discussions 58
2.5 Conclusions 67
2.6 Acknowledgements 67

Chapter 3 Simplified Models for Autoregression under Stationary and Non-stationary Errors 75
3.1 Introduction 75
3.2 Simplified Models— AR(p) with $i.i.d$ Errors 77
3.2.1 Properties of parameter estimates 79
3.2.2 MSE-based comparison of model predictions 82
3.2.3 Example 1. $p_1 = 1$ 85
3.3 Simplified Models-AR(p) under Heteroskedasticity 87
3.3.1 Properties of parameter estimates 88
3.3.2 MSE-based comparison of model predictions 89
3.4 Strategy for Assessing Uncertainty about Which Model is Better . . . 91
3.4.1 Illustrative examples 92
3.4.2 Results and discussions 92
3.5 Conclusions 94
3.6 Acknowledgements 95

Appendix A Proofs 100
A.1 Proof of Lemma 3.2.2 in Chapter 3 100
A.2 Proof of Lemma 3.3.1 in Chapter 3 101
A.3 Proof of Lemma 3.2.1 in Chapter 3 103

Bibliography 104

LIST OF FIGURES

Figure 1.1: MAD of prediction of squared returns obtained by fitting TV-GARCH vs. TV-NoVaS; data from CP-GARCH model. 19
Figure 1.2: MAD of prediction of squared returns obtained by fitting TV-GARCH vs. TV-NoVaS; data from TV-GARCH model. 20
Figure 1.3: Daily realized volatility for IBM stock; the top one is daily realized volatility for data from 07-01-2010 to 07-31-2012 and the bottom is for the series from 01-02-2013 to 11-10-2016. 36

LIST OF TABLES

Table 1.1: MADs and RMSEs of estimation errors for the daily realized volatility of IBM stock from July 1, 2010 to July 31, 2012. 23
Table 1.2: MADs and RMSEs of estimation errors for the daily realized volatility of IBM stock from January 2, 2013 to November 10, 2016. 23
Table 1.3: A stationary process generated by GARCH(1,1) with $C = 10^{-5}$, $B = 0.73$, $A = 0.10$ 34
Table 1.4: A stationary process generated by GARCH(1,1) with $C = 10^{-5}$, $B = 0.93$, $A = 0.05$ 35
Table 1.5: Data generated by TV-GARCH models 37
Table 1.6: Data generated by CP-GARCH models 38
Table 1.7: Foreign exchange rate data 39
Table 1.8: S&P500 Stock index data 40
Table 1.9: Stock price series (IBM) data 41

Table 2.1: MADs of L_1 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .8895, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.N(0,1)$. 59
Table 2.2: MADs of L_2 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .8895, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.N(0,1)$. 60
Table 2.3: MSEs of L_1 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .8895, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.N(0,1)$. 60
Table 2.4: MSEs of L_2 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .8895, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.N(0,1)$. 60
Table 2.5: MADs of L_1 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.N(0,1)$. 60
Table 2.6: MADs of L_2 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.N(0,1)$. 61
Table 2.7: MSEs of L_1 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.N(0,1)$. 61
Table 2.8: MSEs of L_2 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.N(0,1)$. 61
Table 2.9: MADs of L_1 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.\ t$ distributed with degree of freedom 5. 61
Table 2.10: MADs of L_2 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.\ t$ distributed with degree of freedom 5. 62
Table 2.11: MSEs of L_1 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.\ t$ distributed with degree of freedom 5. 62
Table 2.12: MSEs of L_2 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.\ t$ distributed with degree of freedom 5. 62
Table 2.13: MADs of L_1 predictions for Data generated from GARCH(1,1) with slowing varying parameters(TV-GARCH) 63

Table 2.14: MADs of L_2 predictions for Data generated from GARCH(1,1) with slowing varying parameters(TV-GARCH) . . . 63
Table 2.15: MSEs of L_1 predictions for Data generated from GARCH(1,1) with slowing varying parameters(TV-GARCH) . . . 63
Table 2.16: MSEs of L_2 predictions for Data generated from GARCH(1,1) with slowing varying parameters(TV-GARCH) . . . 63
Table 2.17: MADs of L_1 predictions for Data generated from Two-state Markov Switching GARCH(1,1) (MS-GARCH) . . . 64
Table 2.18: MADs of L_2 predictions for Data generated from Two-state Markov Switching GARCH(1,1) (MS-GARCH) . . . 64
Table 2.19: MSEs of L_1 predictions for Data generated from Two-state Markov Switching GARCH(1,1) (MS-GARCH) . . . 64
Table 2.20: MSEs of L_2 predictions for Data generated from Two-state Markov Switching GARCH(1,1) (MS-GARCH) . . . 65
Table 2.21: MADs of L_1 predictions for Data generated from Smooth Transition GARCH(1,1)(ST-GARCH) . . . 65
Table 2.22: MADs of L_2 predictions for Data generated from Smooth Transition GARCH(1,1)(ST-GARCH) . . . 65
Table 2.23: MSEs of L_1 predictions for Data generated from Smooth Transition GARCH(1,1)(ST-GARCH) . . . 65
Table 2.24: MSEs of L_2 predictions for Data generated from Smooth Transition GARCH(1,1)(ST-GARCH) . . . 66
Table 2.25: MADs of L_1 predictions for Data generated from Stochastic Volatility Model(SV-GARCH) . . . 66
Table 2.26: MADs of L_2 predictions for Data generated from Stochastic Volatility Model(SV-GARCH) . . . 66
Table 2.27: MSEs of L_1 predictions for Data generated from Stochastic Volatility Model(SV-GARCH) . . . 66
Table 2.28: MSEs of L_2 predictions for Data generated from Stochastic Volatility Model(SV-GARCH) . . . 67
Table 2.29: Interval predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .8895, \theta = .10$ and $\varepsilon \sim i.i.d\ N(0,1)$. . . . 68
Table 2.30: Interval predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\varepsilon \sim i.i.d\ N(0,1)$. . . . 69
Table 2.31: Results of interval predictions for data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\varepsilon \sim i.i.d\ t_5$. . . . 70
Table 2.32: Results of interval predictions for data generated from TV-GARCH(1,1) . . . 71
Table 2.33: Results of interval predictions for data generated from MS-GARCH(1,1) . . . 72
Table 2.34: Results of interval predictions for data generated from ST-GARCH(1,1) . . . 73
Table 2.35: Results of interval predictions for data generated from SV-GARCH(1,1) . . . 74

Table 3.1: Performance of different sample size with $\theta_1 = 0.4$, $\sigma^2 = 1$, $\theta_2 = 0.5$ 93
Table 3.2: Performance of different sample size with $\theta_1 = 0.1$, $\sigma^2 = 1$, $\theta_2 = 0.5$ 93

Table 3.3: Performance of different sample size with $\theta_1 = 0.01$, $\sigma^2 = 1$, $\theta_2 = 0.5$ . . . 94
Table 3.4: Performance of different sample size with $\theta_1 = 0.001$, $\sigma^2 = 1$, $\theta_2 = 0.5$ . . . 94
Table 3.5: Performance of different σ^2 with $\theta_1 = 0.4$, $N = 100$, $\theta_2 = 0.5$ 95
Table 3.6: Performance of different σ^2 with $\theta_1 = 0.1$, $N = 100$, $\theta_2 = 0.5$ 95
Table 3.7: Performance of different σ^2 with $\theta_1 = 0.01$, $N = 100$, $\theta_2 = 0.5$ 96
Table 3.8: Performance of different σ^2 with $\theta_1 = 0.001$, $N = 100$, $\theta_2 = 0.5$ 96
Table 3.9: Performance of different N with $\theta_1 = 0.4$, $\theta_2 = 0.5$, $\delta = 0.2$, $m = 2$ 96
Table 3.10: Performance of different n with $\theta_1 = 0.1$, $\theta_2 = 0.5$, $\delta = 0.2$, $m = 2$ 97
Table 3.11: Performance of different N with $\theta_1 = 0.05$, $\theta_2 = 0.5$, $\delta = 0.2$, $m = 2$ 97
Table 3.12: Performance of different N with $\theta_1 = 0.01$, $\theta_2 = 0.5$, $\delta = 0.2$, $m = 2$ 97
Table 3.13: Performance of different N with $\theta_1 = 0.001$, $\theta_2 = 0.5$, $\delta = 0.2$, $m = 2$ . . . 98
Table 3.14: Performance of different N with $\theta_1 = 0.4$, $\theta_2 = 0.5$, $\delta = 5$, $m = 2$ 98
Table 3.15: Performance of different N with $\theta_1 = 0.1$, $\theta_2 = 0.5$, $\delta = 5$, $m = 2$ 98
Table 3.16: Performance of different N with $\theta_1 = 0.05$, $\theta_2 = 0.5$, $\delta = 5$, $m = 2$ 99
Table 3.17: Performance of different N with $\theta_1 = 0.01$, $\theta_2 = 0.5$, $\delta = 5$, $m = 2$ 99
Table 3.18: Performance of different N with $\theta_1 = 0.001$, $\theta_2 = 0.5$, $\delta = 5$, $m = 2$ 99

ACKNOWLEDGEMENTS

Special thanks to my advisor, Dimitris N.Politis. For the past 6 years including my master program and doctoral program, he has been so helpful and patient. I'm very grateful for his guidance, encouragement, and support.

I would like to thank all my committee members, Ery Arias-Castro, Brendan K. Beare, Jelena Bradic, James D. Hamilton for their time and comments.

Many thanks to Srinjoy Dos and Nan Zou, for helping me out so many times, and for so many useful and interesting discussions.

Thanks also go to all professors who have taught me and others who helped me during my graduate study.

Last but not least, many thanks to my family members and Pengfei Chen, for their support, understanding, and love.

Chapter 1, in full is currently being prepared for submission for publication of the material. Chen, Jie; Politis, Dimitris N. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in part is currently being prepared for submission for publication of the material. Chen, Jie; Politis, Dimitris N. The dissertation author was the primary investigator and author of this material.

Chapter 3, in full is currently being prepared for submission for publication of the material. Chen, Jie; Politis, Dimitris N. The dissertation author was the primary investigator and author of this material.

VITA

2012	Bachelor of Science in Biology, Peking University
2012	Bachelor of Arts in Economics, Peking University
2014	Master of Science in Statistics, University of California San Diego
2018	Doctor of Philosophy in Mathematics (with specialization in Statistics), University of California San Diego

ABSTRACT OF THE DISSERTATION

Prediction in Time Series Models and Model-free Inference with a Specialization in Financial Return Data

by

Jie Chen

Doctor of Philosophy in Mathematics (with a specialization in Statistics)

University of California San Diego, 2018

Professor Dimitris N. Politis, Chair

The main aim of this dissertation is to study the prediction of financial returns or squared financial returns. As is known, financial returns data have the distribution with fatter tails than the normal and often show significant correlation and the phenomenon of volatility clustering. To capture these features is the most challenging thing in modeling financial returns data. The most popular nonlinear time series models for financial returns data now are ARCH (Autoregressive Conditional Heteroscedasticity) and GARCH (Generalized Autoregressive Conditional Heteroscedasticity) models. Early this century, a Model-free prediction approach was also derived to understand this complex type of data. One application of the model-free approach, NoVaS

(Normalizing and Variance-Stabilizing) transformation has been proved to outperform ARCH and GARCH models under stationary financial data. In Chapter 1, we extend the realm of application of NoVaS to non-stationary data and compare the performance with GARCH in the one-step point prediction and prediction intervals of squared financial returns. In addition, we show the applicability of NoVaS transformation for estimating realized volatility. A new approach to the multi-step ahead prediction of squared financial returns is defined and analyzed in Chapter 2. Our work on linear time series model such as Autoregression are shown in the last two chapters. Chapter 3 describes in detail the situations that a simplified autoregressive models should be considered and the theoretical support was also given there for further study. In the last Chapter, we construct the prediction intervals of regression with model selections in the bootstrap world, which give better performance than the standard bootstrap methods. All the work here are mainly focusing on financial returns time series.

Chapter 1

Time-varying NoVaS vs. GARCH: Point Prediction, Volatility Estimation and Prediction Intervals

The NoVaS methodology was introduced by Politis (2003, 2007) for stationary data in prediction of squared financial returns. In this chapter, we extend the realm of applicability of NoVaS methodology to non-stationary data(local stationarity and/or structural breaks) for one-step ahead point prediction of squared returns. In addition, we show the applicability of NoVaS transformation for estimating realized volatility. Finally, we construct prediction intervals of squared returns for both stationary and non-stationary data. Our empirical results show that the NoVaS methodology performs better than the benchmark GARCH(1,1) models for point and interval prediction as well as in the estimation of realized volatility in situations where (global) stationarity of financial returns fails. Furthermore, it's shown that the point prediction based on NoVaS outperform those based on the benchmark GARCH(1,1) even when the true data generating model is GARCH(1,1). In change point problems, the NoVaS methodology adapts fast to the new regime that occurs after the unknown/undetected change point.

1.1 Introduction

In empirical financial research work, accurate volatility forecasting is always of great importance. Auto-Regressive Conditional Heteroscedasticity models (ARCH) and Generalized Auto-Regressive Conditional Heteroscedasticity models (GARCH) have gained prominence and are widely used in financial engineering since they were introduced by Engle (1982) and Bollerslev (1986). The simple GARCH(1,1) model is particularly most popular; see book of Francq and Zakoian (2011). Nevertheless, ARCH/GARCH models with respect to normal errors can account only partly for the degree of heavy tails empirically found in the distribution of returns. Consequently, researchers and practitioners have been resorting to ARCH/GARCH models with heavy-tailed errors; for example, the t-distribution with degrees of freedom empirically chosen to match the apparent degree of heavy tails in the residuals; see Shephard (1996) and the references therein. However, this situation is not satisfactory since the choice of a t-distribution seems quite arbitrary. Perhaps the real issue is that a simple and neat parametric model such as ARCH/GARCH could not be expected to perfectly capture the behavior of a complicated real-world phenomenon such as the evolution of financial returns that—almost by definition of market *efficiency*—ranks at the top in terms of difficulty of modeling/prediction. The Model-free prediction principle was first introduced in Politis (2003) to understand this complex type of data. The normalizing and variance-stabilizing transformation (NoVaS, for short) is a straightforward application of the Model-free principle in prediction of squared financial returns. The original development of the NoVaS approach was made in Politis (2003, 2007) having as its 'spring board' the popular ARCH model with normal innovations. Politis (2007) showed that NoVaS methods outperform the benchmark GARCH(1,1) model in prediction of squared returns under the assumption of stationarity.

In addition, the most crucial problem in GARCH models is that the GARCH models are not robust with respect to violation from the stationarity assumption; see Mikosch and Starica

(2004); Polzehl and Spokoiny (2006). Then the theory of time-varying ARCH/GARCH process was developed for non-stationary time series; see Dahlhaus et al. (2006). The work of Polzehl and Spokoiny (2006) indicates that time-varying GARCH(1,1) models demonstrates a relatively good predicting performance as far as the short term forecasting horizon is considered. To deal with time varying data, we can consider and propose a time-varying version of NoVaS transformation. Our interest in this chapter is to apply different time-varying NoVaS methods to predicting squared financial returns in situations where (global) stationarity for returns fails such as the cases of local stationarity and/or structural breaks and/or model uncertainty.

In this chapter, we focus on studying the predictive power of different NoVaS methods to non-stationary data in point prediction of squared returns as well as estimation of realized volatility. In addition, we use NoVaS methodology to construct prediction intervals of squared returns for both stationary and non-stationary time series. A very comprehensive simulation and real world data analysis are conducted to study about the relative forecasting performance of time-varying NoVaS methods compared with that of the benchmark time-varying GARCH(1,1) model. The evaluation of forecasting performance for both NoVaS transformation and the benchmark GARCH(1,1) models is addressed via the L_1-norm instead of the usual mean squared error(MSE), since the case was made that financial returns might not have finite 4th moment; see Politis (2007).

The literature on volatility modeling, predicting and the evaluation of volatility prediction is huge and also varies in topics. Here, we just selectively list some recent literature related to the volatility prediction problems: Mikosch and Starica (2004) for change in structure in volatility time series and GARCH modeling;Peng and Yao (2003) for robust least absolute deviations estimation of GARCH models; Poon and Granger (2003) for assessing the forecasting performance of various volatility models; Hansen et al. (2003) on selecting volatility models; Andersen et al. (2004, 2005) on analytic evaluation of volatility forecasts and the use of realized volatility in evaluating volatility forecasts; Hansen and Lunde (2006a) for using a semi-parametric,

transformation-based approach to forming predictive intervals; Ghysels et al. (2006) on the use and predictive power of absolute returns; Francq et al. (2005), Lux and Morales-Arias (2010) and Choi et al. (2010) on switching regime GARCH models, structural breaks and long memory in volatility; Hillebrand (2005) on GARCH models with structural breaks; Hansen and Lunde (2005, 2006b) for comparing forecasts of volatility models against the standard GARCH(1,1) model and for consistent ranking of volatility models and the use of an appropriate series as the 'true' volatility; Ghysels et al. (2006) for predicting volatility by mixing data at different frequencies and Ghysels and Sohn (2009) for the type of power variation that predicts well volatility in the context of mixed data frequencies. Andersen et al. (2007) for modeling realized volatility when jump components are included; Chen et al. (2008) examine volatility forecasting in the context of threshold models coupled with volatility measurement based on intraday range. The whole line of work of Andersen, Bollerslev, Diebold and their various co-authors on realized volatility and volatility forecasting is nicely summarized in their review article "Volatility and Correlation Forecasting", in the Handbook of Economic Forecasting, see Andersen et al. (2006). The book of Francq and Zakoian (2011) provides a comprehensive and systematic approach to understanding GARCH time series models and their applications whilst presenting the most advanced results concerning the theory and practical aspects of GARCH. Bandi et al. (2008) discuss the selection of optimal sampling frequency in realized volatility estimation and forecasting; Patton and Sheppard (2009) present results on optimal combinations of realized volatility estimators in the context of volatility forecasting while Patton and Sheppard (2015) discuss the signed jumps and the persistence of volatility. Priestley (1965, 1988) and Fryzlewicz et al. (2006, 2008) and Dahlhaus et al. (2006, 2007) all work in the context of local stationarity and a new class of ARCH processes with slowly varying parameters. Of course this list is by no means complete.

The chapter is organized as follows: Section 2 presents the work on the point prediction of squared financial returns; Section 3 addresses the comparison of the performance of GARCH(1,1) and NoVaS transformation for estimation of realized volatility; Section 4 presents the Model-Free

algorithms for interval prediction of squared returns and illustrate the numerical performance by means of some simulated examples and applications to real world data; the concluding remarks is provided in Section 5.

1.2 Point Prediction

In this section we compare the performance of NoVaS transformation and GARCH(1,1) in one-step ahead point prediction of squared returns under the assumption of non-stationarity.

1.2.1 GARCH(1,1) and NoVaS Transformation for Stationary Data

The Benchmark: GARCH(1,1)

Consider $\{Y_t, t \in \mathbf{Z}\}$ is a financial time series of returns. For now, we assume that $\{Y_t\}$ is (strictly) stationary with mean zero, which implies that the trends and other nonstationarities have been successfully removed.

A typical ARCH(p) model is described by an equation of the type:

$$Y_t = Z_t \sqrt{\alpha + \sum_{i=1}^{p} a_i Y_{t-i}^2} \tag{1.2.1}$$

where the series $\{Z_t\}$ is assumed to be $i.i.d.N(0,1)$ and p is a positive integer indicating the order of the model.

Let $\mathcal{F}_n$ be a short-hand for the observed information set, i.e., $\mathcal{F}_n = \{Y_t, 1 \leq t \leq n\}$. Note that under the above ARCH(p) model, the L_2 optimal predictor of Y_{n+1}^2 based on $\mathcal{F}_n$ is given by

$$E(Y_{n+1}^2 | \mathcal{F}_n) = \alpha + \sum_{i=1}^{p} a_i Y_{n+1-i}^2. \tag{1.2.2}$$

This conditional expectation $E(Y_{n+1}^2 | \mathcal{F}_n)$ is commonly referred to as the *volatility*, although the

same term is sometimes also used for its square root.

A standard GARCH(1,1) model is described by the equation:

$$Y_t = h_t Z_t \;\text{ with }\; h_t^2 = C + AY_{t-1}^2 + Bh_{t-1}^2 \tag{1.2.3}$$

where the series $\{Z_t\}$ are $i.i.d.\ (0,1)$, and the parameters A, B, C are assumed nonnegative. The quantity $h_t^2 = E(Y_t^2|\mathcal{F}_{t-1})$ is the *volatility* as defined in Eq. (1.2.2). Back solving in the right-hand-side of Eq. (1.2.3), it is easy to show that the GARCH(1,1) model is tantamount to the ARCH model of Eq. (1.2.1) with $p = \infty$ and the following identifications:

$$\alpha = \frac{C}{1-B}, \;\text{ and }\; a_i = AB^{i-1} \;\text{ for }\; i = 1, 2, \ldots \tag{1.2.4}$$

Under the objective of L_1-optimal prediction, the optimal predictor is the conditional median—*not* the conditional expectation. For an ARCH(p) process, the L_1 optimal predictor of Y_{n+1}^2 is given by

$$Median\left(Y_{n+1}^2|\mathcal{F}_n\right) = (\alpha + \sum_{i=1}^{p} a_i Y_{n+1-i}^2) Median(Z_{n+1}^2|\mathcal{F}_n). \tag{1.2.5}$$

Therefore, the aforementioned equivalence of GARCH(1,1) with an ARCH(∞) implies that Eq. (1.2.5) would also give the L_1 optimal GARCH(1,1) predictor of Y_{n+1}^2 by allowing $p = \infty$ with the ARCH coefficients $\alpha, a_1, a_2, \ldots$ following the structure given by Eq. (1.2.4).

GARCH(1,1) model is by far the most popular, and typically forms the benchmark for modeling financial returns. That is also the reason why we will compare the prediction ability of NoVaS methodology with that of GARCH(1,1).

NoVaS Methodology

Let us continue considering a zero mean and (strictly) stationary financial return time series $\{Y_t, t \in \mathbf{Z}\}$. The NoVaS methodology is trying to map the dataset $Y_1, \ldots, Y_n$ to a Gaussian series.

The starting point is the ARCH model of Eq. (1.2.1), under which the residual

$$\frac{Y_t}{\sqrt{\alpha + \sum_{i=1}^{p} a_i Y_{t-i}^2}} \tag{1.2.6}$$

is thought of as perfectly normalized and variance-stabilized as it is assumed to be $i.i.d.N(0,1)$, which is actually not true here. This ratio can be interpreted as an attempt to 'studentize' the return Y_t by dividing with a time-localized measure of the standard deviation of Y_t. However, there seems to be no reason to exclude the value of Y_t from an empirical, causal estimate of the standard deviation of Y_t; recall that a causal estimate is one involving present and past data only, i.e., the data $\{Y_s, s \leq t\}$.

Hence, Politis (2003) defined a new 'studentized' quantity as follows:

$$W_{t,\alpha} := \frac{Y_t}{\sqrt{\alpha s_{t-1}^2 + a_0 Y_t^2 + \sum_{i=1}^{p} a_i Y_{t-i}^2}} \quad \text{for} \quad t = p+1, p+2, \ldots, n. \tag{1.2.7}$$

In the above, s_{t-1}^2 is an estimator of $\sigma_Y^2 = Var(Y_1)$ based on the data up to (but not including[1]) time t; under the zero mean assumption for Y_1, the natural estimator is $s_{t-1}^2 = (t-1)^{-1} \sum_{k=1}^{t-1} Y_k^2$.

The definition in Eq. (1.2.7) describes the proposed normalizing and variance-stabilizing transformation under which the data series $\{Y_t\}$ is mapped to the new series $\{W_{t,\alpha}\}$. The order $p(\geq 0)$ and the vector of nonnegative parameters $(\alpha, a_0, \ldots, a_p)$ are chosen by the practitioner with the twin goals of normalization and variance stabilization.

[1]The reason for not including time t in the variance estimator is for purposes of notational clarity as well as the easy identifiability of the effect of the coefficient a_0 associated with Y_t^2 in the denominator of Eq. (1.2.7).

Also, the NoVaS transformation Eq. (1.2.7) can be re-arranged to yield:

$$Y_t = W_{t,\alpha} \sqrt{\alpha s_{t-1}^2 + a_0 Y_t^2 + \sum_{i=1}^{p} a_i Y_{t-i}^2}. \tag{1.2.8}$$

Formally, the only real difference between the NoVaS of Eq. (1.2.8) and the ARCH of Eq. (1.2.1) is the presence of the term Y_t^2 paired with the coefficient a_0. Replacing the term α in Eq. (1.2.1) by the term αs_{t-1}^2 in Eq. (1.2.8) is only natural since the former has—by necessity—units of variance; in other words, the term α in Eq. (1.2.1) is not scale invariant, whereas the term α in Eq. (1.2.8) is.

Given the assumed structure of the return series, the target of variance stabilization, which amounts to constructing a local estimator of scale for studentization purposes, requires:

$$\alpha \geq 0, \;\; a_i \geq 0 \quad \text{for all} \;\; i \geq 0, \quad \text{and} \quad \alpha + \sum_{i=0}^{p} a_i = 1. \tag{1.2.9}$$

Eq. (1.2.9) has the interesting implication that the $\{W_{t,\alpha}\}$ series can be assumed to have an (unconditional) variance that is (approximately) unity. Nevertheless, note that p and $\alpha, a_0, \ldots, a_p$ must be carefully chosen to achieve a degree of conditional homoscedasticity as well; to do this, one must necessarily take p small enough—as well as α small enough or even equal to zero—so that a local (as opposed to global) estimator of scale is obtained. Politis (2003) provided two structures for the a_i coefficients satisfying Eq. (1.2.9). One is to let $\alpha = 0$ and $a_i = 1/(p+1)$ for all $0 \leq i \leq p$; this specification is called the *simple* NoVaS transformation, and involves only one parameter, namely the order p, to be chosen by the practitioner. The other one is given by the *exponential* decay NoVaS where $\alpha = 0$ and $a_i = c' e^{-ci}$ for all $0 \leq i \leq p$. The exponential scheme involves choosing two parameters: p and $c > 0$ since c' is determined by Eq. (1.2.9).

Now note that

$$\frac{1}{W_{t,\alpha}^2} = \frac{\alpha s_{t-1}^2 + a_0 Y_t^2 + \sum_{i=1}^{p} a_i Y_{t-i}^2}{Y_t^2} \geq a_0$$

if all the parameters are nonnegative, therefore,

$$|W_{t,\alpha}| \leq 1/\sqrt{a_0} \tag{1.2.10}$$

So one must be careful to ensure that the $\{W_{t,\alpha}\}$ variables have a large enough range such that the boundedness is not seen as spoiling the normality. Thus, we also require

$$\frac{1}{\sqrt{a_0}} \geq C \quad \text{i.e.,} \quad a_0 \leq 1/C^2 \tag{1.2.11}$$

for some appropriate C of the practitioner's choice.Recalling that 99.7% of the mass of the $N(0,1)$ distribution is found in the range ± 3, the simple choice $C = 3$ can be suggested; this choice seems to work reasonably well—at least for the usual samples sizes.

Then we proceeds to choose p and $\alpha, a_0, a_1, \ldots, a_p$ (the parameters needed to identify) with the optimization goal of making the $\{W_{t,\alpha}\}$ transformed series as close to normal as possible (normalization). To qualify this goal, one can use minimize a (pseudo)distance measuring departure of the transformed data from normality. Recall that it is a matter of common practice to assume that the distribution of financial returns is *symmetric* (at least to a first approximation) and therefore, the skewness of financial returns is often ignored. In contrast, the kurtosis is typically very large, indicating a heavy tailed distribution. Hence, the kurtosis can serve as a simple (pseudo)distance measuring the departure of a (non-skewed) dataset from normality. Let $KURT_n(Y)$ denote the empirical kurtosis of data $\{Y_t, t = 1, \ldots, n\}$, i.e.,

$$KURT_n(Y) = \frac{n^{-1}\sum_{t=1}^{n}(Y_t - \bar{Y})^4}{(n^{-1}\sum_{t=1}^{n}(Y_t - \bar{Y})^2)^2}$$

where $\bar{Y} = n^{-1}\sum_{t=1}^{n} Y_t$ is the sample mean.

The following are the algorithms to select the optimal parameters for NoVaS; see Politis (2003). Note that the only free parameter in Simple NoVaS is the order p; therefore, the Simple

NoVaS transformation will be denoted by $W^S_{t,p}$. In the Exponential NoVaS, to specify all the a_is, one just needs to specify the two parameters p and $c > 0$, in view of Eq. (1.2.9). However, because of the exponential decay, the parameter p is now of secondary significance; thus, we concisely denote the exponential NoVaS transformation by $W^E_{t,c}$.

ALGORITHM 2.1. SIMPLE NOVAS

1. Let $\alpha = 0$ and $a_i = 1/(p+1)$ for all $0 \leq i \leq p$.

2. Pick p such that $|KURT_n(W^S_{t,p}) - 3|$ is minimized.

2′. Pick p such that $KURT_n(W^S_{t,p}) \simeq 3$.

Step 2 in the above is described as an optimization problem for mathematical concreteness. Nevertheless, it could be better understood as a moment matching, see Step 2′, where of course the value 3 for kurtosis corresponds to the Gaussian distribution.

ALGORITHM 2.2. EXPONENTIAL NOVAS

1. Let p take a very high starting value, e.g., $p \simeq n/4$ or $n/5$. Then, let $\alpha = 0$ and $a_i = c' e^{-ci}$ for all $0 \leq i \leq p$, where $c' = 1/\sum_{i=0}^{p} e^{-ci}$ by Eq. (1.2.9).

2. Pick $c > 0$ in such a way that $|KURT_n(W^E_{t,c}) - 3|$ is minimized.

Technically, the above search is for $c \in (0, \infty)$ which appears formidable; what makes this minimization problem well-behaved is that we know that high values of c can not plausibly be solutions. To see why, note that if c is large, then $a_i \approx 0$ for all $i > 0$ and $W^E_{t,c} = Y_t$ which has kurtosis much larger than 3. It is apparent that the search for the optimal c will be practically conducted over a discrete grid of c-values spanning an interval of the type $(0, s]$ for some s of the order of one. A practical way to narrow in on the optimal c value is to run two grid searches, one coarse followed by a fine one: (i) use a coarse grid search over the whole interval $(0, s]$, and denote $\tilde{c}_0$ the minimizer over the coarse grid search; and (ii) run a fine grid search over a

neighborhood of $\tilde{c}_0$. Let c_0 denote the resulting minimizer from the above algorithm. If needed, the following range–adjustment step may be added.

3. *If c_0 as found above is such that Eq. (1.2.11) is not satisfied, then decrease c stepwise (starting from c_0) over the discrete grid until Eq. (1.2.11) is satisfied.*

Finally, the value of p must be trimmed for efficiency of usage of the available sample; to do this we can simply discard the a_i coefficients that are close to zero, i.e., those that fall below a certain threshold/tolerance level ε which is the practitioner's choice. A threshold value of ε =0.01 is reasonable in connection with the a_i which—as should be stressed—are normalized to sum to one.

4. *Trim the value of p by a criterion of the type: if $a_i < \varepsilon$, then let $a_i = 0$. If i_0 is the smallest integer such that $a_i < \varepsilon$ for all $i \geq i_0$, then let $p = i_0$ and* re-normalize *the a_is so that their sum (for $i = 0, 1, \ldots, i_0$) equals one.*

Although many different multi-parameter NoVaS schemes can be devised, we now elaborate on the possibility of a nonzero value for the parameter α (say $\alpha_1, \alpha_2, \ldots, \alpha_K$, that span a subset of the interval [0,1]) in Eq. (1.2.7) in connection with the Simple and Exponential NoVaS. We thus define the Generalized Simple (GS) and Generalized Exponential (GE) NoVaS denoted by $W^{GS}_{t;p,\alpha}$ and $W^{GE}_{t;c,\alpha}$ indicating their respective two free parameters; both are based on Eq. (1.2.7).

ALGORITHM 2.3. GENERALIZED SIMPLE NOVAS

A. For $k = 1, \ldots, K$ perform the following steps.

   1. Let $\alpha = \alpha_k$ and $a_i = (1-\alpha_k)/(p+1)$ for all $0 \le i \le p$ so that eq. (1.2.9) is satisfied while all the coefficients $a_0, a_1, \ldots, a_p$ are the same.

   2. Denote by p_k the minimizer of $|KURT_n(W^{GS}_{t,p}) - 3|$ over values of $p = 1, 2, \ldots$.

   3. If p_k (and a_0) as found above are such that eq. (1.2.11) is not satisfied, then increase p_k accordingly, i.e., re-define $p_k = \lfloor 1 + C^2(1-\alpha_k) \rfloor$, and let $a_i = (1-\alpha_k)/(p_k+1)$ for all $0 \le i \le p_k$ by Eq. (1.2.9); here, $\lfloor x \rfloor$ denotes the integer part of x.

B. Finally, compare the transformations $\{W^{GS}_{t;p_k,\alpha_k},\ k = 1, \ldots, K\}$ in terms of their volatility prediction performance, and pick the model with optimal performance.

ALGORITHM 2.4. GENERALIZED EXPONENTIAL NOVAS

A. For $k = 1, \ldots, K$, perform the following steps.

   1. Let p take a very high starting value, e.g., let $p \simeq n/4$ or $n/5$. Then, let $\alpha = \alpha_k$ and $a_i = c' e^{-ci}$ for all $0 \leq i \leq p$, where $c' = (1-\alpha_k)/\sum_{i=0}^{p} e^{-ci}$ by Eq. (1.2.9).

   2. Pick c in such a way that $|KURT_n(W^{GE}_{t;c,\alpha_k}) - 3|$ is minimized, and denote by c_k the minimizing value.[2]

   3. Trim the value of p to some value p_k as before: if $a_i < \varepsilon$, then set $a_i = 0$. Thus, if $a_i < \varepsilon$, for all $i \geq i_k$, then let $p_k = i_k$, and re-normalize the a_is so that their sum (for $i = 0, 1, \ldots, p_k$) equals $1 - \alpha_k$ by Eq. (1.2.9).

B. Finally, compare the transformations $\{W^{GE}_{t;c_k,\alpha_k},\ k = 1, \ldots, K\}$ in terms of their volatility prediction performance, and pick the model with optimal performance.

Remark. *For all the empirical work in this chapter, we set* $p = n/4$ *and* $\varepsilon = 0.01$ *when using the Exponential NoVaS and GE NoVaS algorithms. For Generalized NoVaS, we choose* $K = 8$ *and* $\alpha_k = 0, 0.1, 0.2, \ldots, 0.7$.

Suppose that the NoVaS parameters, i.e., the order $p(\geq 0)$ and the parameters $\alpha, a_0, \ldots, a_p$ have already been chosen. Re-arrange the NoVaS Eq. (1.2.7) and then yield:

$$Y_t^2 = \frac{W_{t,\alpha}^2}{1 - a_0 W_{t,\alpha}^2} \left(\alpha s_{t-1}^2 + \sum_{i=1}^{p} a_i Y_{t-i}^2 \right) \quad \text{for } t = p+1, \ldots, n \tag{1.2.12}$$

[2] As before, if c_k is such that Eq. (1.2.11) is not satisfied, then decrease it stepwise over its discrete grid until Eq. (1.2.11) is satisfied.

and

$$Y_t = \frac{W_{t,\alpha}}{\sqrt{1-a_0 W_{t,\alpha}^2}} \sqrt{\alpha s_{t-1}^2 + \sum_{i=1}^{p} a_i Y_{t-i}^2} \text{ for } t = p+1, \ldots, n. \tag{1.2.13}$$

Let $g(\cdot)$ be some (measurable) function of interest; examples include $g_0(x) = x,\ g_1(x) = |x|$, and $g_2(x) = x^2$, the latter being the function of interest for volatility prediction. Based on the Model-free Prediction Principle, the one-step ahead prediction problem of NoVaS can be defined. From Eq.(1.2.13) it follows that the predictive (given $\mathcal{F}_n$) distribution of $g(Y_{n+1})$ is identical to the distribution of the random variable

$$g\left(A_n \frac{W}{\sqrt{1-a_0 W^2}}\right) \tag{1.2.14}$$

where $A_n = \sqrt{\alpha s_n^2 + \sum_{i=1}^{p} a_i Y_{n+1-i}^2}$ is treated as a constant given the past $\mathcal{F}_n$, and the random variable W has the same distribution as the conditional (on $\mathcal{F}_n$) distribution of the random variable $W_{n+1,a}$.

Therefore, the L_1 optimal prediction of $g(Y_{n+1})$ given $\mathcal{F}_n$ is given by the median of the conditional (given $\mathcal{F}_n$) distribution of $g(Y_{n+1})$, i.e.,

$$\widehat{g(Y_{n+1})} := Median\left(g\left(A_n \frac{W_{n+1,a}}{\sqrt{1-a_0 W_{n+1,a}^2}}\right) | \mathcal{F}_n\right) \tag{1.2.15}$$

Specializing to the case of our interest, i.e., volatility prediction and the function $g_2(x) = x^2$ yields the NoVaS predictor:

$$\widehat{Y_{n+1}^2} = \mu_2 A_n^2 \tag{1.2.16}$$

where

$$\mu_2 = Median\left(\frac{W_{n+1,a}^2}{1-a_0 W_{n+1,a}^2} | \mathcal{F}_n\right).$$

1.2.2 Local Stationarity

When we consider a real-valued time series $Y_1, \ldots, Y_n$ spanning a long time interval, e.g., annual rainfall measurements spanning over 100 years or 30 minutes financial returns spanning several years, it may be unrealistic to assume that the stochastic structure of time series $\{Y_t, t \in \mathbf{Z}\}$ has stayed invariant over such a long stretch of time; hence, we can not assume that $\{Y_t, t \in \mathbf{Z}\}$ is stationary. It is therefore plausible to assume a slowly-changing stochastic structure, i.e., a *locally stationary model*; see Priestley (1965, 1988), Dahlhaus et al. (1997) and Dahlhaus (2012). The type of processes that can be described with locally stationarity are those which locally at each time point are close to a stationary process but whose characteristics (covariances, parameters, etc.) are gradually changing in an unspecific way as time evolves.

Time-varying GARCH and Time-varying NoVaS

To capture such a non-stationary phenomenon, the theory of Time-varying ARCH (TV-ARCH, for short) processes was developed and their asymptotic properties of weighted quasi-likelihood estimators were studied; see Dahlhaus et al. (2006).

The analysis of a Time-varying ARCH/GARCH model can be based on the premise of local stationarity. For example, in order to predict $g(Y_{t+1})$ based on $\mathcal{F}_t$ via a **Time–varying GARCH(1,1)**(TV-GARCH for short) model, we can simply fit GARCH(1,1) model of Eq. (1.2.3) using the the 'windowed' data $Y_{t-b+1}, \ldots, Y_t$, i.e., the coefficients of GARCH(1,1) models are varying with time. Here, the window size b should be large enough so that accurate estimation of the GARCH parameters is possible based on the subseries $Y_{t-b+1}, \ldots, Y_t$ but small enough so that such a subseries can plausibly be considered stationary.

In a similar vein, we can predict $g(Y_{t+1})$ by fitting one of the NoVaS algorithms (Simple vs. Exponential, Generalized or not) just using the 'windowed' data $Y_{t-b+1}, \ldots, Y_t$. In so doing, we are constructing a **Time-varying NoVaS** (TV-NoVaS) transformation. In numerical work, Politis and Thomakos (2006, 2013) showed that NoVaS fitting can be done more efficiently than

GARCH fitting by (numerical) MLE. Thus, it is expected that TV-NoVaS may be able to capture a slowly changing stochastic structure in a more flexible manner; stated in different term, the window size b required for accurate NoVaS fitting should be smaller than the one required for accurate GARCH fitting.

Structural Breaks: Changing points

An alternative form of non-stationarity is due to the possible presence of structural breaks, i.e., change points, occurring at some isolated time points. Kokoszka and Leipus (2000), and Berkes et al. (2004) have studied the detection/estimation of change points in ARCH/GARCH modeling. Mikosch and Starica (2004) and Stărică and Granger (2005) show the interesting effects that an undetected change point may have on our interpretation and analysis of ARCH/GARCH modeling. Polzehl and Spokoiny (2006) shows that time-varying models can give a better predictive power for time series with change points. In this section, we can also consider the time-varying NoVaS methods under short windowed data except in a neighborhood of the change points. Hence, in the simulation that follows, we also include a structural break model in order to see the effect of an undetected change point on the performance of TV-NoVaS predictors for squared returns.

1.2.3 Simulations and Results

We investigate these above conjecture and compare the one-step ahead prediction performance of NoVaS with that of the the standard benchmark GARCH(1,1) model in the following simulation experiment.

Simulation Design

For the simulation, 500 datasets $\underline{Y}_n = (Y_1, \ldots, Y_n)'$ were constructed using either a TV-GARCH or a change point GARCH (CP-GARCH); these were defined using the standard GARCH

model of Eq. (1.2.3) as building block with $C = 10^{-5}$. The *i.i.d.* errors Z_t are commonly assumed to have a Student t_5 distribution; instead, we use the simple assignment $Z_t \sim i.i.d.\ N(0,1)$ in the simulation in order to facilitate the convergence of the numerical (Gaussian) MLE in fitting the TV-GARCH model.

CP-GARCH: For $t \leq n/2$, let $A = 0.10$ and $B = 0.73$; for $t > n/2$, let $A = 0.05$ and $B = 0.93$. These values are close to the ones used by Mikosch and Starica (2004).

TV-GARCH: The value of A decreases as a linear function of t, starting at $A = 0.10$ for $t = 1$, and ending at $A = 0.05$ for $t = n$. At the same time, the value of B increases as a linear function of t, starting at $B = 0.73$ for $t = 1$, and ending at $B = 0.93$ for $t = n$.

The difference between CP-GARCH model and TV-GARCH model, is an abrupt vs. smooth transition spanning the same values. Some more information on the simulation follows.

- The prediction method employed was the conditional median obtained either from a TV-GARCH model with normal errors fitted by windowed Gaussian MLE, or via TV-NoVaS (Simple or Exponential); in either case, two window sizes were tried out, namely $b = 125$ or 250.

- The sample size was $n = 1001$ corresponding to about 4 years of daily data; so the choices $b = 125$ and 250 correspond to 6 and 12 months respectively.

- Training period for all methods was 250, i.e., the experiment amounted to predicting Y_{t+1}^2 from the 'windowed' data $Y_{t-b+1}, \ldots, Y_t$ for $t = 250, 251, \ldots, 1000$.

- Updating (re-estimation) of all methods would ideally be for each $t = 250, 251, \ldots, 1000$. To save computing time, updating in the simulation was only performed for t being an integer multiple of 50. In fairness, the performance of predictions was recorded and compared *only* at the moment of updating the model, i.e., at time points $250, 300, 350, \ldots, 1000$.

For each point, we give the mean absolute deviation (MAD) of the prediction error at the update time point averaged over the 500 replications.

Results and Conclusions

Figure 1.1 shows the MAD of volatility prediction errors of TV-GARCH as compared to that of TV-NoVaS (Simple or Exponential) with data from model CP-GARCH for the 16 time points where the updating and prediction occurred, i.e., the time points $250, 300, 350, \ldots, 1000$. The left panel depicts the case $b = 125$ while the right panel depicts the case $b = 250$. Figure 1.2 is similar but using data generated by a TV-GARCH model instead.

Some conclusions are as follows:

- Time points $250, 300, 350, 400, 450$, and 500 in the left panel of Figure 1.1 corroborate the aforementioned fact that NoVaS (Simple or Exponential) beats GARCH for prediction of squared returns even if the data generating model is (stationary) GARCH as long as the sample size available for model-fitting is small—equal to 125 in this case. The corresponding points in the right panel of Figure 1.1 indicate that GARCH manages to do as well as (or better than)[3] NoVaS when the effective sample size is increased to 250.

- Figure 1.1 shows that the change point at $t = 500$ wreaks havoc in GARCH model fitting and the associated predictions; this adds another dimension to the observations of Mikosch and Starica (2004). By contrast, both NoVaS methods seem to adapt immediately to the new regime that occurs after the unknown/undetected change point.

- Figure 1.2 shows that TV-NoVaS (Simple or Exponential) beats TV-GARCH for prediction of squared returns even when the data generating model is TV-GARCH. Not only is the MAD of prediction of TV-NoVaS just a small fraction of that of TV-GARCH, but the

[3]Note that here GARCH is fitted by Gaussian MLE with only three free parameters; in the more realistic case of four parameter MLE using the t distribution—the fourth parameter being the degrees of freedom—GARCH underperforms compared to NoVaS even with a sample size of 350; see Politis and Thomakos (2006, 2013).

wild swings associated with the latter indicate the inherent instability of GARCH model-fitting; this instability is prominent even in this simplistic case where the errors have a true Gaussian distribution, and Gaussian MLE is used for estimating just the three GARCH parameters.

- As seen in both Figure 1.1 and Figure 1.2, the performance of Simple NoVaS is practically indistinguishable from that of Exponential NoVaS although upon closer look the latter appears to be marginally better.

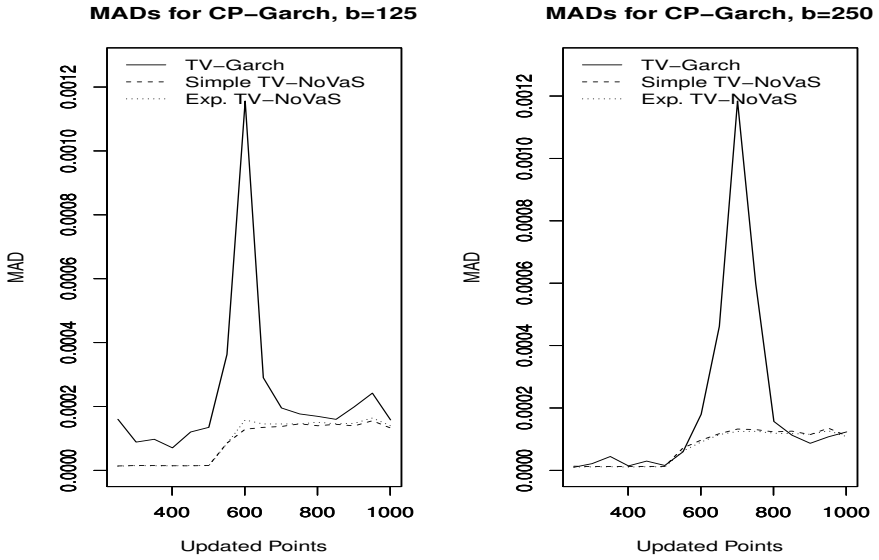

Figure 1.1: MAD of prediction of squared returns obtained by fitting TV-GARCH vs. TV-NoVaS; data from CP-GARCH model.

1.3 Estimation of Realized Volatility

In this section, we compare the performance of NoVaS methods and GARCH(1,1) in the estimation of daily realized volatility with real world time series. Based on the empirical work of Politis (2003, 2007), the NoVaS series $\{W_{t,\alpha}\}$ appears to be uncorrelated with several

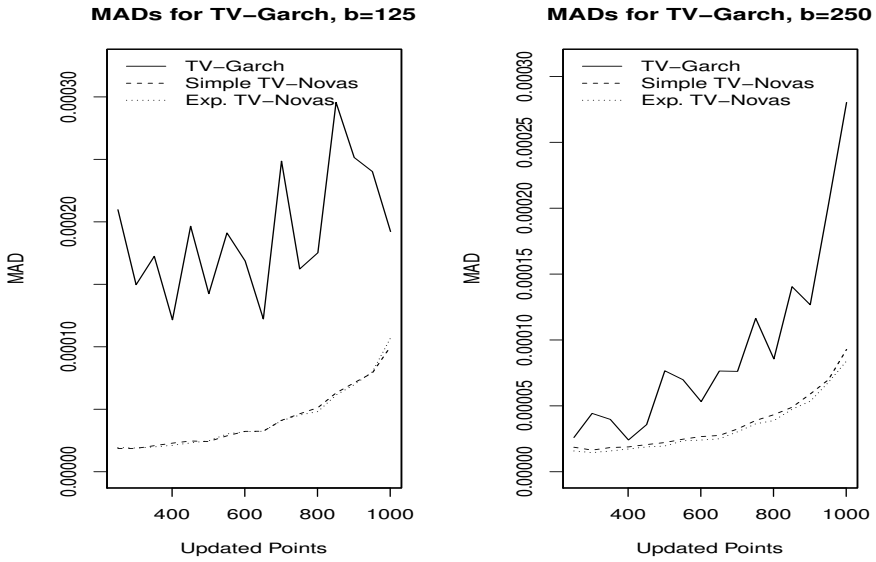

Figure 1.2: MAD of prediction of squared returns obtained by fitting TV-GARCH vs. TV-NoVaS; data from TV-GARCH model.

daily returns. Since $\{W_{t,\alpha}\}$ is (approximately) Gaussian, then we can infer that the series $\{W_{t,\alpha}\}$ is not only uncorrelated but also independent; therefore, it is straightforward to construct a Model-free estimate of the conditional expectation $E(Y_{n+1}^2|\mathcal{F}_n)$. In this case, Eq. (1.2.12) implies that $E(Y_{n+1}^2|\mathcal{F}_n) = A_n^2 E\left(\frac{W_{t,\alpha}^2}{1-a_0 W_{t,\alpha}^2}\right)$; a natural estimate of realized volatility thereof is

$$\frac{A_n^2}{n-p}\sum_{t=p+1}^{n}\left(\frac{W_{t,\alpha}^2}{1-a_0 W_{t,\alpha}^2}\right) \tag{1.3.1}$$

which has validity, e.g. consistency, under the sole assumption that Y_t has a finite second moment conditionally on $\mathcal{F}_n$ (and therefore unconditionally as well). To examine the performance of this estimate, we conduct the following empirical analysis for estimating the daily realized volatility by using the formula (1.3.1) and GARCH(1,1) model.

1.3.1 Data and Summary Statistics

We consider two real world time series. Both are intraday 30 minutes IBM stock returns and associated with the daily realized volatility. The sample period of the first dataset is from 07-01-2010 to 07-31-2012 for a total of N = 526 days. The sample period of the second dataset is from 01-02-2013 to 11-10-2016 for a total of N = 975 days. Weekends and holidays are excluded for both series. In Figure 1.3, we present graphs for the daily realized volatility of the two datasets. The associated daily realized volatility was constructed by summing all 30-minutes squared returns of one day for IBM stock. Specifically, if we denote by $r_{t,i}$, the i^{th} daily return for day t, then the daily realized volatility is defined as $h_t^2 \stackrel{\text{def}}{=} \sum_{i=1}^{n_t} r_{t,i}^2$, where n_t is the total number of 30-minutes intervals during day t of open market.

1.3.2 NoVaS and GARCH(1,1) Optimization and Estimating Specifications

To estimate the realized volatility for our two series, we continue using the TV-NoVaS transformation and TV-GARCH(1,1) models, since the fourth moments of the series may be infinite. All forecasts we make are *'honest'*, i.e., we use only observations prior to the time period to be forecasted. The parameters of NoVaS approach and GARCH(1,1) models are re-estimated as the window rolls over the entire evaluation sample. The window sizes we choose are m = 126 (six months) and m = 252 (one year). To compare the performance of the NoVaS approach, we estimate using not only standard GARCH(1,1) models with normal distribution but also GARCH(1,1) models, assuming a $t_{(v)}$ distribution with degree of freedom estimated from the data. Here we employ all NoVaS algorithms: Simple-NoVaS, Exp-NoVaS, GS NoVaS and GE NoVaS. For Exp-NoVaS and GE NoVaS algorithms, we set $p_{max} = n/4$ and the trimming threshold of 0.01, same as those in Section 1.2.

In the process of analysis, we always evaluate our estimation using the 'true' realized volatility measure given in the previous subsection and report the mean absolute deviation (MAD)

and root mean-squared error (RMSE) of the estimation errors $e_t \stackrel{\text{def}}{=} h_t^2 - \widehat{h_t}^2$, given by:

$$MAD(e) \stackrel{\text{def}}{=} \frac{1}{N-m} \sum_{t=m+1}^{N} |e_t - \bar{e}|, \text{ and } RMSE(e) \stackrel{\text{def}}{=} \sqrt{\frac{1}{N-m} \sum_{t=m+1}^{N} (e_t - \bar{e})^2}$$

where $\widehat{h_t}^2$ denotes the estimation of the daily realized volatility for any of the methods or models we use and $\bar{e} \stackrel{\text{def}}{=} \frac{1}{N-m} \sum_{t=m+1}^{N} e_t$.

1.3.3 Results and Conclusions

Our estimation results are summarized in Table 1.1 and Table 1.2. Table 1.1 is the results for the series from 07-01-2010 to 07-31-2012 and Table 1.2 is for the series from 01-02-2013 to 11-10-2016. In the second columns of each table are the MADs and RMSEs of TV-NoVaS and TV-GARCH(1,1) with window size 126. The third columns in both tables record the MADs and RMSEs with window size 252. Some general comments on the results are follows:

- For all methodologies, the MADs and RMSEs of the estimation errors of daily realized volatility are decreasing when the window size is small, i.e., 126. This may indicate that both data series are non-stationary as expected.

- The TV-GARCH(1,1) models with standard t innovations outperform the TV-GARCH(1,1) models with Normal innovations for both series. This should not be surprising since the distribution of financial returns has fatter tails than the normal.

- The performance of TV-Exp-NoVaS is always better than that of TV-Simple-NoVaS. Similarly, TV-GE-NoVaS also performs better than TV-GS-NoVaS. These mean that the method of exponential smoothing is more reasonable to obtain a *local* time series; see, Hamilton (1994).

- As seen in both Table 1.1 and Table 1.2, TV-Exp-NoVaS and TV-Simple-NoVaS are

producing bigger MADs and RMSEs of the estimation errors than Time-varying Generalized NoVaS, even sometimes bigger MADs than TV-GARCH(1,1) models. Interestingly, Time-varying Generalized NoVaS(both simple and exponential) transformation, especially TV-GE-NoVaS, perform best always with smaller MADs and RMSEs of estimation errors among all methods we use. Therefore, a non-zero α in Eq. (1.2.7) of NoVaS transformation is crucial for good performance when our goal is to estimate realized volatility, even though this is not true when predicting squared returns in Section 1.2.

Table 1.1: MADs and RMSEs of estimation errors for the daily realized volatility of IBM stock from July 1, 2010 to July 31, 2012.

Series 2010-2012	Window size = 126		Window size = 252	
Methods	MAD	RMSE	MAD	RMSE
GARCH(1,1) with normal error	1.418E-04	2.112E-04	1.732E-04	2.292E-04
GARCH(1,1) with t-error	1.397E-04	2.049E-04	1.706E-04	2.233E-04
Simple-NoVaS	1.450E-04	1.941E-04	1.793E-04	2.231E-04
Exp-NoVaS	1.356E-04	1.698E-04	1.589E-04	1.922E-04
GS NoVaS	1.199E-04	1.689E-04	1.439E-04	1.901E-04
GE NoVaS	1.115E-04	1.684E-04	1.343E-04	1.908E-04

Table 1.2: MADs and RMSEs of estimation errors for the daily realized volatility of IBM stock from January 2, 2013 to November 10, 2016.

Series 2013-2016	Window size = 126		Window size = 252	
Methods	MAD	RMSE	MAD	RMSE
GARCH(1,1) with normal error	1.362E-04	3.692E-04	1.446E-04	3.617E-04
GARCH(1,1) with t-error	1.343E-04	3.682E-04	1.390E-04	3.592E-04
Simple-NoVaS	1.655E-04	3.398E-04	1.664E-04	3.319E-04
Exp-NoVaS	1.546E-04	3.291E-04	1.531E-04	3.147E-04
GS NoVaS	1.175E-04	3.252E-04	1.191E-04	3.179E-04
GE NoVaS	1.114E-04	3.253E-04	1.138E-04	3.175E-04

1.4 Bootstrap Prediction Intervals

Beyond the one-step ahead point prediction of Y_{n+1}^2, we may try to construct a prediction interval that will contain Y_{n+1}^2 with (conditional) probability $1-\alpha$ asymptotically and compare its performance with that based on GARCH(1,1) models. As usual, the key idea of Model-Free Prediction Principle is to transform a given complex dataset into one that is $i.i.d$, and therefore easier to handle. For NoVaS transformation, we transform our dataset into $i.i.d$ normal distribution. For Model-Based Bootstrap, standard errors and confidence intervals are based on generating one-step ahead pseudo data by some estimated conditional distribution, e.g., the transition densities or the transition distribution functions; see Pan and Politis (2014, 2016). However, Model-Free Bootstrap re-samples the $i.i.d$ data and then transforms them back to obtain the desired one-step ahead prediction.

1.4.1 Description of Interval Prediction Algorithms

Recall that our best (in an L_1 sense) prediction of $g(Y_{n+1})$ given $\mathcal{F}_n$ was given, i.e.,

$$\widehat{g(Y_{n+1})} = Median\left(g\left(A_n\frac{W_{n+1,\alpha}}{\sqrt{1-a_0W_{n+1,\alpha}^2}}\right)|\mathcal{F}_n\right)$$

$$= Median\left(g\left(A_n\frac{W_{n+1,\alpha}}{\sqrt{1-a_0W_{n+1,\alpha}^2}}\right)\right);$$

where the second equality is due to the independence in the series $W_{t,\alpha}$. The above can give us a preliminary approximation to the predictive distribution of $g(Y_{n+1})$ given $\mathcal{F}_n$ in the form of the empirical distribution of the random variables $\{g\left(A_n\frac{W_{t,\alpha}}{\sqrt{1-a_0W_{t,\alpha}^2}}\right)$ for $t=p+1,\ldots,n\}$. However, as Politis (2015) remarked, this empirical distribution ignores the variability of estimated parameters in the construction of the NoVaS transformation; to incorporate this variability, Model-

free bootstrap is needed here as well. Note that the point predictor $g(\widehat{Y_{n+1}})$ is a function only[4] of $Y_n, \ldots, Y_{n-p+1}$, i.e., is a predictor of the type of a (nonlinear) AR model or Markov process of order p. Hence, to develop the relevant re-sampling algorithms, we can borrow some ideas from work of Pan and Politis (2014, 2016); in particular, we will adopt the *'forward'* bootstrap methodology, i.e., generate bootstrap series forward in time but also ensure that Y^*_{n+1} is constructed correctly; see also in Politis (2015).

The basic Model-free (MF) bootstrap algorithm for prediction intervals in the setting of financial returns goes as follows.

ALGORITHM 4.1 MF PREDICTION INTERVALS FOR $g(Y_{n+1})$

1. Use one of the NoVaS algorithms (Simple vs.Exponential, Generalized or not, etc.) to obtain the transformed data $\{W_{t,\alpha}$ for $t = p+1, \ldots, n\}$ that are assumed to be approximately *i.i.d*. Let p, α and a_i denote the fitted NoVaS parameters.

2. Calculate $g(\widehat{Y_{n+1}})$, the point predictor of $g(Y_{n+1})$, as the median of the set $\{g\left(A_n \frac{W_{t,\alpha}}{\sqrt{1-a_0 W^2_{t,\alpha}}}\right)$ for $t = p+1, \ldots, n\}$; recall that $A_n = \sqrt{\alpha s_n^2 + \sum_{i=1}^p a_i Y^2_{n+1-i}}$.

3. (a) Re-sample randomly (with replacement) the transformed variables $\{W_{t,\alpha}$ for $t = p+1, \ldots, n\}$ to create the pseudo-data $W^*_{p+1}, \cdots, W^*_{n-1}, W^*_n$ and W^*_{n+1}.

[4]In the case of Generalized NoVaS (Simple or Exponential), $g(\widehat{Y_{n+1}})$ is also a function of s_n^2 which, however, converges to EY_t^2 for large n; hence, it can be treated as constant for all practical purposes.

3. (b) Let $(Y_1^*,\ldots,Y_p^*)' = (Y_{1+I},\cdots,Y_{p+I})'$ where I is generated as a discrete random variable uniform on the values $0,1,\ldots,n-p$.

   (c) Generate the bootstrap pseudo-data Y_t^* for $t = p+1,\ldots,n$ using the following equations, i.e., let

$$Y_t^* = \frac{W_t^*}{\sqrt{1-a_0 W_t^{*2}}}\sqrt{\alpha s_{t-1}^{*2} + \sum_{i=1}^{p} a_i Y_{t-i}^{*2}} \quad \text{for } t = p+1,\ldots,n \tag{1.4.1}$$

   where $s_{t-1}^{*2} = (t-1)^{-1}\sum_{k=1}^{t-1} Y_k^{*2}$.

   (d) Based on the bootstrap data $Y_1^*,\ldots,Y_n^*$, re-estimate the NoVaS transformation yielding parameters p^*, α^*, $a_0^*, a_1^*,\ldots,a_p^*$.
   Let $A_n^* = \sqrt{\alpha^* s_n^2 + \sum_{i=1}^{p^*} a_i^* Y_{n+1-i}^2}$, and calculate the bootstrap predictor $g(\widehat{Y_{n+1}^*})$ as the median of the set

$$\{g\left(A_n^* \frac{W_{t,\alpha}}{\sqrt{1-a_0^* W_{t,\alpha}^2}}\right) \quad \text{for } t = p+1,\ldots,n\} \tag{1.4.2}$$

   using the convention[5]that when $1-a_0^* W_{t,\alpha}^2 \le 0$, we assign

$$\frac{1}{\sqrt{1-a_0^* W_{t,\alpha}^2}} = \infty.$$

3. (e) Calculate the bootstrap future value Y^*_{n+1} as

$$Y^*_{n+1} = \frac{W^*_{n+1}}{\sqrt{1 - a_0 W^{*2}_{n+1}}} \sqrt{\alpha s_n^2 + \sum_{i=1}^{p} a_i Y^2_{n-i+1}}. \tag{1.4.3}$$

   (f) Calculate the bootstrap root: $g(Y^*_{n+1}) - g(\widehat{Y^*_{n+1}})$.

4. Repeat step 3 above B times; the B bootstrap root replicates are collected in the form of an empirical distribution whose α-quantile is denoted $q(\alpha)$.

5. The $(1-\alpha)100\%$ equal-tailed prediction interval for $g(Y_{n+1})$ is given by

$$[g(\widehat{Y_{n+1}}) + q(\alpha/2),\ g(\widehat{Y_{n+1}}) + q(1-\alpha/2)].$$

[5]This is because the original NoVaS data satisfies $|W_{t,\alpha}| \leq 1/\sqrt{a_0}$ but a_0^* might turn out bigger (or smaller) than a_0. Alternatively, one can base Eq. (1.4.2) on the NoVaS transformed series $W^*_{t,\alpha}$ that corresponds to the bootstrap data $Y^*_1, \ldots, Y^*_n$, or on a Monte Carlo experiment using a $N(0,1)$ distribution truncated to $\pm 1/\sqrt{a_0^*}$.This is the case of LMF in (a′). All these options are practically indistinguishable as far as taking the median is concerned, and Eq. (1.4.2) is the most straightforward.

Note that the last p values from the *original* data, i.e., $Y_{n-p+1},\ldots,Y_n$, are used in both the creation of the bootstrap predictor in Eq.(1.4.2) and bootstrap future value in Eq.(1.4.3); this is in accordance with the 'forward' bootstrap methodology in work of Pan and Politis (2014) but also with the general Model-free Bootstrap described in Algorithm 2.4.1 in Politis (2015).

Another version of Algorithm 4.1 can also be devised in the spirit of the Limit Model-Free(LMF) Bootstrap of Politis (2015); it would amount to replacing Step 3 (a) by:
(a′) Generate $W^*_{p+1},\cdots,W^*_{n-1},W^*_n$ *and* W^*_{n+1} *as i.i.d. from a* $N(0,1)$ *distribution truncated to* $\pm 1/\sqrt{a_0}$.

Remark. *Algorithm 4.1 can also be applied to each windowed dataset for non-stationary time series, thus obtaining the time varying NoVaS prediction intervals.*

1.4.2 Local Stationarity

Under a model-free setup of a local stationary time series, Paparoditis and Politis (2002) proposed the Local Block Bootstrap in order to generate pseudo-series $Y^*_1,\ldots,Y^*_n$ whose probability structure mimics that of the observed data $Y_1,\ldots,Y_n$. The Local Block Bootstrap has been found useful for the construction of confidence intervals; see Dowla et al. (2003), Dowla et al. (2013). Politis (2015) put forth the algorithms of model-free and model-based prediction intervals for locally stationary time series. Like the work in Section 2, when the data is a locally stationary time series, we can conduct the local stationary models (TV-GARCH(1,1) or TV-NoVaS) on the windowed subseries. Similarly, we can also try to construct a prediction interval of NoVaS using the Algorithm 4.1 on each subseries to construct the prediction intervals for each data point we concern.

1.4.3 Finite-Sample Performance of Model-Free and Model-Based Prediction Intervals

Three Illustrative Datasets

In the work of prediction intervals, we also focus on another three real world datasets of daily returns taken from a foreign exchange rate, a stock price, and a stock index; a brief description of these datasets is as follows.

- **Example 1: Foreign exchange rate.** Daily returns from the Yen vs. Dollar exchange rate from January 1, 1988 to August 1, 2002; the data were downloaded from Datastream. The sample size is n = 3600 (weekends and holidays are excluded).
- **Example 2: Stock index.** Daily returns of the S&P500 stock index from October 1, 1983 to August 30, 1991; the data are available as part of the `GARCH` module in Splus. The sample size is n = 2000.
- **Example 3: Stock price.** Daily returns of the IBM stock price from February 1, 1984 to December 31, 1991; the data are again available as part of the `GARCH` module in Splus. The sample size is n = 2000.

More information of these three datasets can be found in Politis (2007, 2015)

Simulation

In the following simulation work, we will compare the performance in interval prediction of squared returns by using Simple-NoVaS, Exp-NoVaS, Limit Model-Free with Simple-NoVaS method (LMF Simple-NoVaS), Limit Model-Free Exp-NoVaS method (LMF Exp-NoVaS), Generalized Simple-NoVaS method (GS-NoVaS), Generalized Exp-NoVaS method (GE-NoVaS), Limit Model-Free with Generalized Simple-NoVaS method (LMF GS-NoVaS) and Limit Model-Free

with Generalized Exp-NoVaS method (LMF GE-NoVaS) method with that by using GARCH(1,1) models with normal errors. Each interval is constructed based on the windowed data series. So all the methods and models here are still time-varying models. For prediction intervals by using GARCH(1,1) models, we use the algorithms 9.2.1 of Model-Based prediction intervals for Y_{n+1} in the book of Politis (2015); see also Pan and Politis (2016).

As in Section 1.2 , we employ the same TV-GARCH(1,1) and CP-GARCH(1,1) models to generate the simulated datasets. For consistency, two stationary processes are generated by the following two standard GARCH(1,1) models:

- **Model 1.** $X_t = \sigma_t \varepsilon_t,\ \sigma_t^2 = .00001 + .93\sigma_{t-1}^2 + .05X_{t-1}^2,\ \{\varepsilon_t\} \sim i.i.d.N(0,1).$
- **Model 2.** $X_t = \sigma_t \varepsilon_t,\ \sigma_t^2 = .00001 + .73\sigma_{t-1}^2 + .10X_{t-1}^2,\ \{\varepsilon_t\} \sim i.i.d.N(0,1).$

Each dataset with size n = 1000. Also, three financial indexes datasets are conducted to compute the prediction intervals. For computational reasons, we chose $B = 500$ for bootstrap re-sampling. For simulated data, the window sizes for simulated data are b = 125 and b = 250, same as those in Section 2 of point predictions. For these three real world data, we use window size b = 250 and 500.

For each dataset, $n-b$ windowed datasets of size b are generated. For each windowed dataset, one of the bootstrap methods(GARCH(1,1) and different NoVaS algorithms) was used to create B bootstrap sample paths and B one-step ahead "future" values denoted by $X_{(b+1,j)}$ for $j = 1,2,\ldots,B$; The bootstrap prediction interval(L_i, U_i) was constructed for the "future" value $X_{(b+1)}$ of windowed dataset i, here $i = b+1, b+2, \ldots, n$. The corresponding empirical average coverage level(CVR) and the average length(LEN) of the constructed intervals and the standard error (St.err) associated with each length of the constructed intervals are calculated as

$$CVR = \frac{1}{n-b}\sum_{i=1}^{n-b} CVR_j$$

$$LEN = \frac{1}{n-b}\sum_{i=1}^{n-b} LEN_i \text{ and } St.err = \frac{1}{n-b}\sum_{i=1}^{n-b}(LEN_i - LEN)^2$$

where

$$CVR_i = \frac{1}{B}\sum_{j=1}^{B} \mathbf{1}_{[L_j,U_j]}X_{(b+1,i)} \text{ and } LEN_i = U_i - L_i.$$

Discussion of results

Our results are summarized in Tables 1.3 and 1.4for simulated stationary datasets, in Tables 1.5 and 1.6for simulated datasets generated by CP-GARCH(1,1) and TV-GARCH(1,1) models respectively and in Tables 1.7, 1.8 and 1.9 for three real world datasets. Each table has two subtables with different window sizes. The first two lines of each subtable are the results using the Simple NoVaS (Simple-NoVaS) and Exponential NoVaS (Exp-NoVaS) methods. Lines 3 and 4 of each subtable are the results using Limit Model-Free with Simple-NoVaS (LMF Simple-NoVaS) and Limit Model-Free Exp-NoVaS (LMF Exp-NoVaS) methods. Lines 5 and 6 of each subtable are the results of Generalized Simple-NoVaS (GS-NoVaS) and Generalized Exp-NoVaS (GE-NoVaS) methods. Lines 7 and 8 of each subtable are for Limit Model-Free with Generalized Simple-NoVaS method (LMF GS-NoVaS) and Limit Model-Free with Generalized Exp-NoVaS (LMF GE-NoVaS) method. The last line of each subtable gives the results by using GARCH(1,1) models with normal errors.

For simulated stationary time series data by GARCH(1,1) models, Tables 1.3 and 1.4 show that the performance of different NoVaS methods is better than that of GARCH(1,1), that is, NoVaS methods have average coverages very close to or equal to the nominal ones. Remarkably, the Generalized Simple and/or Exponential NoVaS methods beats Simple/Exp-NoVaS as well as LMF Simple/Exp-NoVaS. In addition, there are not significant differences between Limit Model-Free NoVaS and Generalized NoVaS methods when the data is stationary. It is also shown that the window size has little effect—the coverages are still close to nominal ones—for the performance of NoVaS methodology, while the reaction of GARCH(1,1) to the change of sample

size is relatively large. We can see that the average coverages are much closer to the nominal ones as the window size is increasing from 125 to 250.

Tables 1.6 and 1.7 give the results for data generated from time-varying GARCH(1,1) and CP-GARCH(1,1) models with normal errors respectively. It is shown that in each table, all NoVaS methods have the average coverages much closer to the nominal ones than GARCH(1,1) in both subtables for each table. When the window size is increased to 250, GARCH(1,1) models perform better than that with a smaller window size, i.e., 125. Meanwhile, the performance of NoVaS methods is not sensitive to changes in the sample size. This finding is in accordance with the results for point prediction of squared returns in Section 2. However, this effect of window size on GARCH(1,1) under data with nonstationarity is bigger than that for the case of stationarity if you compare the effect with that in Tables 1.3 and 1.4. It is also worthwhile to note that Limit Model-Free NoVaS and Generalized NoVaS methods can sometimes capture the nominal coverage exactly with smaller average lengths and standard deviations for the predicted intervals. Also, we can find that Limit Model-Free NoVaS methods perform better than Simple-NoVaS and Exponential NoVaS, while the performance of the Limit Model-Free Generalized NoVaS methods are not as good as the Generalized NoVaS under both simulated datasets. If we look at the results for three real world financial series in Tables Tables 1.7, 1.8 and 1.9, the NoVaS still outperforms the benchmark GARCH(1,1) models. Similarly with the case in Tables 1.5 and 1.6, GARCH(1,1) performs worse when the window size is increasing from 250 to 500, that is, the average coverage is smaller and the average length and standard errors of the predicted intervals are larger. It's also seen that the performance of Limit Model-Free Generalized NoVaS methods are indistinguishable from that of Generalized NoVaS although the latter appears to be marginally better upon closer comparison. All in all, when the data is non-stationary, the results from Tables Tables 1.6, 1.7, 1.8 and 1.9 are supporting the superior performance of NoVaS in interval prediction against the benchmark GARCH(1,1).

1.5 Concluding Remarks

In this chapter, we study the prediction ability of squared returns and realized volatility by using NoVaS based on the non-stationary time series. For nonstationarity, time-varying NoVaS and GARCH(1,1) models are selected to fit the windowed data for prediction. Also, we add the generalized NoVaS methodology and compared its relative performance in prediction with Simple/EXP-NoVaS and the benchmark GARCH(1,1) models.

In particular, we conducted an extensive simulation to study the one-step ahead point prediction of squared returns under non-stationary data using different time-varying NoVaS methods and time-varying GARCH(1,1) models. We compared their performance under different DGPs, including change-point GARCH(1,1) models and time-varying GARCH(1,1) models. It was that shown that the NoVaS methodology for point prediction of squared returns remains successful—with smaller mean absolute deviations of prediction errors—in situations where global stationarity fails such as the cases of structural breaks and local stationarity. We also work on the comparison of the performance of time-varying NoVaS and time-varying GARCH(1,1) models in estimation of realized volatility for real world data. The time-varying GE/GS NoVaS give smaller mean absolute deviation and root mean squared errors in the estimating realized volatility than time varying GARCH(1,1) models with normal or t distributions. In addition, we constructed the Model-Free algorithm of prediction intervals by NoVaS transformation. In the empirical work of interval prediction of squared returns under both real world datasets and simulated non-stationary series, we find that NoVaS gives a higher average coverage than GARCH(1,1) and the performance of NoVaS does not depend much on the window size. Remarkably, the Generalized Simple/Exponential NoVaS methods are the best ones for interval prediction of squared returns as well as estimation of volatility among all methods/models we use in this chapter. So a nonzero α in NoVaS transformation is crucial for good performance, i.e., Generalized NoVaS is a must. To sum up, the time-varying NoVaS methodology is robust against nonstationarity and

invariably outperforms the GARCH benchmark for prediction of squared returns.

1.6 Acknowledgements

Chapter 1, in full is currently being prepared for submission for publication of the material. Chen, Jie; Politis, Dimitris N. The dissertation author was the primary investigator and author of this paper.

Table 1.3: A stationary process generated by GARCH(1,1) with $C = 10^{-5}$, $B = 0.73$, $A = 0.10$

Window size 125	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.948	6.17E-04	5.84E-04	0.901	3.89E-04	2.25E-04
Exp-NoVaS	0.963	3.96E-03	1.25E-03	0.912	3.52E-04	2.01E-04
Limit Simple-NoVaS	0.951	4.00E-04	2.34E-04	0.899	2.73E-04	1.54E-04
Limit Exp-NoVaS	0.954	4.89E-04	2.05E-04	0.893	3.12E-04	1.22E-04
GS NoVaS	0.947	2.82E-04	9.98E-05	0.896	2.27E-04	1.08E-04
GE NoVaS	0.950	3.57E-04	2.59E-04	0.899	2.55E-04	1.54E-04
LMF GS-NoVaS	0.946	2.52E-04	1.14E-04	0.893	2.01E-04	8.70E-05
LMF GE-NoVaS	0.944	3.06E-04	8.44E-05	0.893	2.43E-04	6.47E-05
GARCH(1,1)	0.926	2.74E-04	1.25E-04	0.879	2.08E-04	9.40E-05
Window size 250	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.949	5.17E-04	3.20E-04	0.895	3.79E-04	2.16E-04
Exp-NoVaS	0.953	4.73E-04	2.37E-04	0.896	2.92E-04	1.12E-04
Limit Simple-NoVaS	0.931	4.10E-04	2.31E-04	0.886	2.89E-04	1.60E-04
Limit Exp-NoVaS	0.943	4.59E-04	1.70E-04	0.894	3.05E-04	1.10E-04
GS NoVaS	0.945	2.86E-04	1.10E-04	0.896	2.10E-04	1.21E-04
GE NoVaS	0.948	5.92E-04	5.63E-04	0.900	2.81E-04	1.57E-04
LMF GS-NoVaS	0.943	2.73E-04	1.48E-04	0.892	2.09E-04	1.10E-04
LMF GE-NoVaS	0.948	2.99E-04	8.76E-05	0.893	3.11E-04	1.75E-04
GARCH(1,1)	0.931	2.52E-04	9.64E-05	0.883	2.53E-04	7.28E-05

Table 1.4: A stationary process generated by GARCH(1,1) with $C = 10^{-5}$, $B = 0.93$, $A = 0.05$

Window size 125	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.950	4.05E-03	4.71E-03	0.901	2.42E-03	1.38E-03
Exp-NoVaS	0.960	7.41E-03	3.86E-02	0.907	2.60E-03	1.60E-03
Limit Simple-NoVaS	0.951	3.06E-03	1.88E-03	0.893	2.04E-03	1.18E-03
Limit Exp-NoVaS	0.949	3.88E-03	2.02E-03	0.886	2.49E-03	1.30E-03
GS-NoVaS	0.952	1.87E-03	6.74E-04	0.899	2.28E-03	1.78E-03
GE-NoVaS	0.947	2.67E-04	1.02E-04	0.898	2.78E-04	1.68E-04
LIMIT GS-NOVAS	0.947	3.13E-03	1.76E-03	0.893	2.48E-03	1.27E-03
LIMIT GE-NOVAS	0.949	3.30E-03	1.60E-03	0.894	2.81E-03	1.57E-03
GARCH(1,1)	0.914	2.37E-03	1.03E-03	0.866	2.09E-03	8.27E-03
Window size 250	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.958	4.06E-03	2.25E-03	0.906	2.81E-03	1.54E-03
Exp-NoVaS	0.960	3.14E-03	2.09E-03	0.903	2.06E-03	1.19E-03
Limit Simple-NoVaS	0.949	5.62E-03	3.81E-03	0.899	3.94E-03	2.66E-03
Limit Exp-NoVaS	0.938	3.64E-03	1.81E-03	0.888	2.42E-03	1.22E-03
GS-NoVaS	0.951	2.14E-03	9.14E-04	0.900	2.48E-03	3.71E-04
GE-NoVaS	0.949	2.71E-04	1.17E-04	0.900	2.19E-04	1.19E-04
LIMIT GS-NOVAS	0.946	2.67E-03	1.44E-03	0.894	2.09E-03	7.86E-04
LIMIT GE-NOVAS	0.949	3.33E-03	1.31E-03	0.899	2.71E-03	1.41E-03
GARCH(1,1)	0.919	2.16E-03	1.09E-03	0.873	1.87E-03	5.81E-03

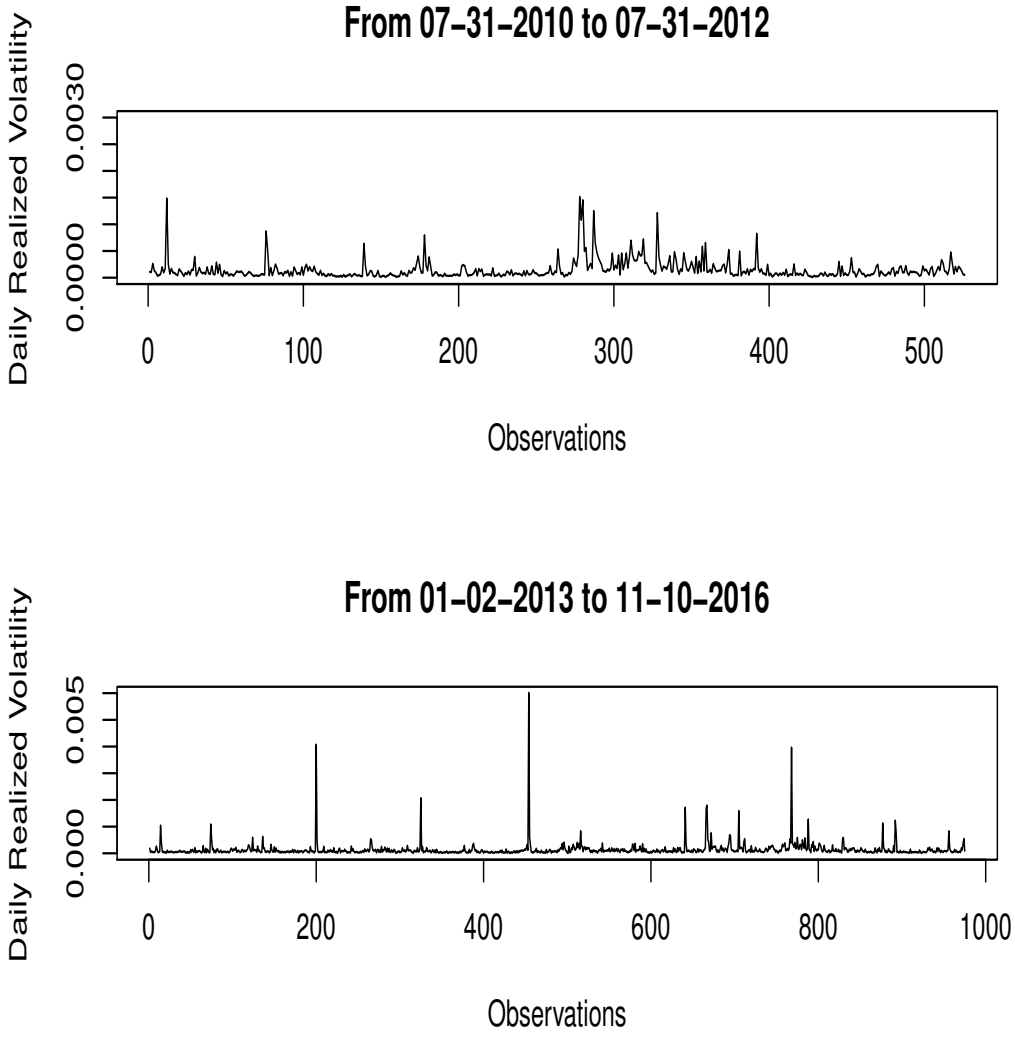

Figure 1.3: Daily realized volatility for IBM stock; the top one is daily realized volatility for data from 07-01-2010 to 07-31-2012 and the bottom is for the series from 01-02-2013 to 11-10-2016.

Table 1.5: Data generated by TV-GARCH models

Window size 125	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.954	1.24E-02	1.87E-03	0.900	2.05E-03	1.86E-03
Exp-NoVaS	0.965	2.63E-02	5.14E-03	0.918	8.87E-02	6.41E-03
LMF Simple-NoVaS	0.951	1.07E-03	7.81E-04	0.906	7.43E-04	5.29E-04
LMF Exp-NoVaS	0.945	1.09E-03	7.49E-04	0.890	7.12E-04	4.74E-04
GS-NoVaS	0.951	8.84E-04	5.78E-04	0.901	6.27E-04	3.99E-04
GE-NoVaS	0.950	9.25E-04	8.72E-04	0.892	6.66E-04	6.26E-04
LMF GS-NoVaS	0.944	8.10E-04	6.39E-04	0.897	6.83E-04	5.92E-04
LMF GE-NoVaS	0.947	9.25E-04	6.34E-04	0.887	5.15E-04	3.99E-04
GARCH(1,1)	0.885	5.03E-04	4.14E-04	0.842	4.07E-04	5.29E-04
Window size 250	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.939	2.61E-03	3.43E-03	0.884	1.22E-03	8.47E-03
Exp-NoVaS	0.968	9.69E-04	6.66E-04	0.911	6.85E-04	4.42E-04
LMF Simple-NoVaS	0.944	9.27E-04	7.90E-04	0.893	6.54E-04	4.96E-04
LMF Exp-NoVaS	0.924	1.12E-03	7.41E-04	0.881	7.51E-04	4.97E-04
GS-NoVaS	0.947	9.82E-04	6.39E-04	0.891	5.52E-04	2.97E-04
GE-NoVaS	0.948	9.25E-04	6.70E-04	0.888	5.25E-04	2.91E-04
LMF GS-NoVaS	0.940	8.71E-04	6.71E-04	0.891	1.02E-03	1.00E-03
LMF GE-NoVaS	0.940	8.11E-04	5.68E-04	0.895	6.36E-04	5.84E-04
GARCH(1,1)	0.909	4.19E-04	4.03E-04	0.854	3.28E-04	5.22E-04

Table 1.6: Data generated by CP-GARCH models

Window size 125	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.960	3.96E-03	4.21E-02	0.908	2.21E-03	2.36E-02
Exp-NoVaS	0.963	1.07E-02	1.92E-02	0.918	1.61E-03	1.80E-03
LMF Simple-NoVaS	0.958	1.73E-03	1.79E-03	0.897	1.18E-03	1.21E-03
LMF Exp-NoVaS	0.946	2.40E-03	2.58E-03	0.894	1.58E-03	1.70E-03
GS-NoVaS	0.950	1.10E-03	8.86E-04	0.901	8.10E-04	6.33E-04
GE-NoVaS	0.950	2.39E-03	2.35E-03	0.897	1.79E-03	1.79E-03
LMF GS-NoVaS	0.938	1.78E-03	1.72E-03	0.896	1.37E-03	1.36E-03
LMF GE-NoVaS	0.949	1.70E-03	2.03E-03	0.889	9.42E-04	9.35E-04
GARCH(1,1)	0.901	3.42E-03	1.52E-03	0.833	2.60E-03	5.74E-03
Window size 250	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.949	3.04E-03	5.25E-03	0.907	2.10E-03	3.19E-03
Exp-NoVaS	0.957	2.84E-03	3.01E-03	0.900	1.82E-03	1.78E-03
LMF Simple-NoVaS	0.952	1.80E-03	1.42E-03	0.892	1.25E-03	9.79E-04
LMF Exp-NoVaS	0.941	2.46E-03	1.85E-03	0.893	1.66E-03	1.23E-03
GS-NoVaS	0.949	3.86E-03	3.09E-03	0.906	4.84E-03	4.15E-03
GE-NoVaS	0.944	1.89E-03	1.24E-03	0.890	1.75E-03	1.36E-03
LMF GS-NoVaS	0.956	4.08E-03	4.86E-03	0.911	2.84E-03	3.03E-03
LMF GE-NoVaS	0.943	1.96E-03	1.73E-03	0.880	1.28E-03	1.00E-03
GARCH(1,1)	0.916	1.86E-03	1.47E-03	0.871	1.34E-03	7.87E-03

Table 1.7: Foreign exchange rate data

Window size 250	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.97	6.64E-04	6.25E-04	0.922	3.86E-04	3.32E-04
Exp-NoVaS	0.968	6.86E-04	5.63E-04	0.908	3.35E-04	2.33E-04
LMF Simple-NoVaS	0.958	4.71E-04	3.97E-04	0.898	3.02E-04	2.39E-04
LMF Exp-NoVaS	0.964	5.23E-04	3.47E-04	0.924	3.15E-04	2.14E-04
GS-NoVaS	0.950	4.45E-04	3.72E-04	0.906	3.09E-04	2.02E-04
GE-NoVaS	0.950	4.51E-04	2.78E-04	0.896	2.71E-04	1.72E-04
LMF GS-NoVaS	0.946	4.14E-04	2.85E-04	0.9	2.81E-04	1.54E-04
LMF GE-NoVaS	0.950	4.20E-04	1.82E-04	0.886	2.67E-04	9.95E-05
GARCH(1,1)	0.928	3.60E-04	1.73E-04	0.87	2.25E-04	1.37E-04
Window size 500	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.945	6.17E-04	8.48E-04	0.895	3.79E-04	5.07E-04
Exp-NoVaS	0.953	5.00E-04	5.27E-04	0.888	3.09E-04	3.16E-04
LMF Simple-NoVaS	0.918	4.86E-04	3.03E-04	0.862	6.09E-04	3.69E-04
LMF Exp-NoVaS	0.954	5.13E-04	5.18E-04	0.914	3.14E-04	3.10E-04
GS-NoVaS	0.949	3.05E-04	1.55E-04	0.906	2.08E-04	8.80E-05
GE-NoVaS	0.949	3.56E-04	1.90E-04	0.902	2.05E-04	9.05E-05
LMF GS-NoVaS	0.946	4.10E-04	1.98E-04	0.890	2.66E-04	1.21E-04
LMF GE-NoVaS	0.950	3.89E-04	1.20E-04	0.912	2.56E-04	6.66E-05
GARCH(1,1)	0.928	1.60E-04	1.73E-04	0.870	2.25E-04	1.37E-04

Table 1.8: S&P500 Stock index data

Window size 250	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.946	6.04E-04	8.10E-04	0.896	3.78E-04	4.11E-04
Exp-NoVaS	0.940	5.51E-04	4.37E-04	0.880	3.42E-04	2.35E-04
LMF Simple-NoVaS	0.934	5.02E-04	5.69E-04	0.880	3.24E-04	3.20E-04
LMF Exp-NoVaS	0.956	5.32E-04	4.19E-04	0.894	3.31E-04	2.56E-04
GS-NoVaS	0.948	4.12E-04	3.66E-04	0.896	2.99E-04	2.54E-04
GE-NoVaS	0.950	4.20E-04	2.76E-04	0.898	2.95E-04	1.70E-04
LMF GS-NoVaS	0.946	4.38E-04	3.47E-04	0.892	2.86E-04	2.14E-04
LMF GE-NoVaS	0.948	4.25E-04	2.93E-04	0.900	2.87E-04	1.99E-04
GARCH(1,1)	0.938	3.25E-04	2.27E-04	0.862	1.92E-04	1.67E-04
Window size 500	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.961	1.59E-03	5.35E-03	0.918	1.06E-03	3.84E-03
Exp-NoVaS	0.948	1.16E-03	8.49E-04	0.878	7.77E-04	5.68E-04
LMF Simple-NoVaS	0.929	1.49E-03	5.13E-03	0.872	9.68E-04	3.44E-03
LMF Exp-NoVaS	0.952	1.61E-03	4.67E-03	0.897	9.92E-04	2.81E-03
GS-NoVaS	0.949	5.26E-04	2.71E-04	0.894	3.43E-04	1.91E-04
GE-NoVaS	0.949	3.56E-04	1.90E-04	0.892	3.36E-04	1.51E-04
LMF GS-NoVaS	0.944	6.78E-04	4.87E-04	0.894	4.45E-04	3.18E-04
LMF GE-NoVaS	0.936	5.86E-04	3.01E-04	0.884	4.58E-04	2.75E-04
GARCH(1,1)	0.927	4.58E-04	3.04E-03	0.863	3.60E-04	1.82E-03

Table 1.9: Stock price series (IBM) data

Window size 250	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.964	1.42E-03	7.51E-04	0.918	7.67E-04	4.71E-04
Exp-NoVaS	0.954	1.31E-03	1.28E-03	0.894	7.75E-04	4.21E-04
LMF Simple-NoVaS	0.958	1.02E-03	5.73E-04	0.918	7.83E-04	4.42E-04
LMF Exp-NoVaS	0.956	1.08E-03	6.64E-04	0.914	7.48E-04	3.81E-04
GS-NoVaS	0.954	9.16E-04	4.45E-04	0.894	7.09E-04	4.37E-04
GE-NoVaS	0.946	8.50E-04	3.48E-04	0.896	6.99E-04	4.28E-04
LMF GS-NoVaS	0.950	9.17E-04	4.81E-04	0.890	6.88E-04	3.89E-04
LMF GE-NoVaS	0.952	1.01E-03	5.85E-04	0.898	6.67E-04	3.43E-04
GARCH(1,1)	0.940	8.99E-04	4.01E-04	0.878	5.60E-04	1.75E-04
Window size 500	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.957	1.26E-03	7.64E-03	0.910	8.47E-04	7.74E-04
Exp-NoVaS	0.949	1.16E-03	8.24E-03	0.879	7.74E-04	5.51E-04
LMF Simple-NoVaS	0.955	1.25E-03	1.16E-03	0.906	8.39E-04	7.67E-04
LMF Exp-NoVaS	0.957	1.33E-03	9.84E-04	0.879	8.67E-04	5.51E-04
GS-NoVaS	0.948	1.14E-03	1.91E-04	0.902	1.13E-04	2.61E-04
GE-NoVaS	0.949	2.13E-03	3.75E-04	0.898	1.71E-04	3.54E-04
LMF GS-NoVaS	0.948	1.68E-03	4.55E-04	0.904	1.33E-04	2.97E-04
LMF GE-NoVaS	0.950	1.93E-03	3.07E-03	0.896	1.30E-04	2.34E-03
GARCH(1,1)	0.934	5.22E-04	8.06E-04	0.878	1.56E-04	5.56E-04

Chapter 2

Optimal Multistep-ahead Prediction of Nonlinear Time Series Models and Model-free Inference

2.1 Introduction

Multistep-ahead prediction in a time series is to predict a sequence of future values using only the values observed in the past. It includes predicting the time series for crop yield, stock prices, traffic volume, and electrical power consumption. In this section, we focus on multistep-ahead prediction of squared financial returns, which is preformed in volatility as a conditional mean of the squared returns. A typical approach to solve this problem, known as multi-stage prediction, is to construct a single model from the past observed time series data and then apply the model step by step to predict its future values. It also uses the predicted value of the current time step to determine its value in the next time step. A lot of empirical evidences have been given that multi-stage prediction is susceptible to the error accumulation problem, i.e., errors committed in the past will be propagated into future predictions.

This chapter considers an alternative approach to multistep-ahead prediction of the squared returns for the nonlinear time series models, such as ARCH/GARCH models. Our method depends only on their past observations for each prediction step and try to eliminate the errors accumulation issue. Here we focus on the two most popular nonlinear time series models for financial returns series — ARCH/GARCH models— and also the normalizing and variance-stabilizing transformation (NoVaS, for short) method.

The chapter here is organized as follows: Section 2 presents the work on the optimal multistep-ahead point predictions for ARCH and GARCH processes and NoVaS transformation; Section 3 addresses the optimal multistep-ahead prediction methods for prediction intervals; Section 4 illustrate the numerical performance by means of some simulated examples; the concluding remarks is provided in Section 5.

2.2 Optimal Multistep-ahead Point Prediction

First consider a zero mean and (strictly) stationary financial returns time series $\{X_t, 1 \leq t \leq n\}$. Our goal is to predict the future squared returns X^2_{n+h} for any $h \geq 2$.

Let $\mathcal{F}_n$ be a short-hand for the observed information set, i.e., $\mathcal{F}_n = \{X_t, 1 \leq t \leq n\}$. In the L_2 sense, the optimal predictor of X^2_{n+h} based on $\mathcal{F}_n$ is the conditional mean and given by

$$\widehat{X^2_{n+h}} = E(X^2_{n+h}|\mathcal{F}_n). \tag{2.2.1}$$

Similarly, the optimal L_1 predictor is the conditional median as

$$\widehat{X^2_{n+h}} = Median(X^2_{n+h}|\mathcal{F}_n). \tag{2.2.2}$$

In the following parts of this section, we study the multistep-ahead prediction in the nonlinear financial models ARCH/GARCH(1,1) and NoVaS, which is an application of model-free approach

in financial returns data.

2.2.1 For ARCH(p) and GARCH(1,1) models

Suppose the data follow an ARCH(p) process which is defined as

$$X_t = \sigma_t \varepsilon_t, \ \textit{and} \ \sigma_t^2 = \alpha + a_1 X_{t-1}^2 + ... + a_p X_{t-p}^2 \tag{2.2.3}$$

where $\alpha \geq 0$, $a_j \geq 0$ for all $j = 1, \ldots, p$, and $\{\varepsilon_t\} \sim i.i.d.N(0,1)$.

First, let's look at the simplest case $h = 2$. Based on the model (2.2.3), we can express X_{n+1} and X_{n+2} in the following way:

$$X_{n+1} = \varepsilon_{n+1}\sqrt{\sigma_{n+1}^2}, \ \textit{and} \ \sigma_{n+1}^2 = \alpha + a_1 X_n^2 + ... + a_p X_{n-p+1}^2,$$

$$X_{n+2} = \varepsilon_{n+2}\sqrt{\sigma_{n+2}^2}, \ \textit{and} \ \sigma_{n+2}^2 = \alpha + a_1 X_{n+1}^2 + ... + a_p X_{n-p+2}^2.$$

Obviously, X_{n+1} can be easily written as a function of the past observations $X_n, \ldots, X_{n+1-p}$ and the unknown future error ε_{n+1}. Furthermore, we can also rewrite X_{n+2} to be a function of $X_n, \ldots, X_{n+1-p}$ and the unknown future errors ε_{n+1}, and ε_{n+2}. The notations are as followings:

$$\begin{aligned} X_{n+1} =& \varepsilon_{n+1}\sqrt{\alpha + a_1 X_n + a_2 X_{n-1}^2 + ... + a_p X_{n-p+1}^2} \\ =& f_1(X_1, ..., X_n; \varepsilon_{n+1}) \end{aligned} \tag{2.2.4}$$

$$\begin{aligned} X_{n+2} =& \varepsilon_{n+2}\sqrt{\alpha + a_1 X_{n+1}^2 + a_2 X_n^2 + ... + a_p X_{n-p+2}^2} \\ =& \varepsilon_{n+2}\sqrt{\alpha + a_1 \varepsilon_{n+1}^2(\alpha + a_1 X_n + a_2 X_{n-1}^2 + ... + a_p X_{n-p+1}^2) + a_2 X_n^2 + ... + a_p X_{n-p+2}^2} \\ =& f_2(X_1, ..., X_n; \varepsilon_{n+1}, \varepsilon_{n+2}) \end{aligned} \tag{2.2.5}$$

Recursively, we can express X_{n+h} for any $h \geq 1$ as a function of past observations $\{X_1, ..., X_n\}$ and the unknown future innovations $\{\varepsilon_{n+1}, \ldots, \varepsilon_{n+h}\}$ like

$$X_{n+h} = f_h(X_1, ..., X_n; \varepsilon_{n+1}, \ldots, \varepsilon_{n+h}). \tag{2.2.6}$$

Since $\{X_1, ..., X_n\}$ are given and known, we can write (2.2.6) simply as

$$X_{n+h} = f_h(\varepsilon_{n+1}, \ldots, \varepsilon_{n+h}), \text{ for any } h \geq 1. \tag{2.2.7}$$

So the squared financial returns can be rewritten as f_h^2 for any future values. Based on the assumption that ε_t is $i.i.d\ N(0,1)$, the conditional distribution function $F_{f_h^2}$ of the future squared returns $f_h^2(\cdot)$ can be derived. Hence, the optimal predictor (conditional median for L_1 or conditional mean for L_2) of x_{n+h}^2 is easy to be calculated by the $F_{f_h^2}$.

Take $h = 1$ and $h = 2$ as an example. By (2.2.4), (2.2.5), the L_2 optimal predictors of X_{n+1}^2 and X_{n+2}^2 are

$$\begin{aligned} \widehat{X_{n+1}^2} =& E\{\varepsilon_{n+1}^2(\alpha + a_1 X_n + a_2 X_{n-1}^2 + ... + a_p X_{n-p+1}^2)|\mathcal{F}_n\} \\ =& \alpha + a_1 X_n^2 + a_2 X_{n-1}^2 + ... + a_p X_{n-p+1}^2, \end{aligned} \tag{2.2.8}$$

$$\begin{aligned} \widehat{X_{n+2}^2} =& E[\varepsilon_{n+2}^2(\alpha + a_1\sigma_{n+1}^2\varepsilon_{n+1}^2 + a_2 X_n^2 + \ ... \ + a_p X_{n-p+2}^2)|X_1, ..., X_n] \\ =& \alpha + a_1\sigma_{n+1}^2 + a_2 X_n^2 + \ ... \ + a_p X_{n-p+2}^2 \end{aligned} \tag{2.2.9}$$

since $E(\varepsilon_{n+1}^2|\mathcal{F}_n) = 1$ and $E(\varepsilon_{n+2}^2|\mathcal{F}_n) = 1$ by the assumption. First, we can note that $\widehat{X_{n+1}^2} = f_1^2(\varepsilon_{n+1}^2 = 1)$ and $\widehat{X_{n+2}^2} = f_2^2(\varepsilon_{n+1}^2 = 1, \varepsilon_{n+2}^2 = 1)$. Actually, we can easily verify that for any $h \geq 1$,

$$\widehat{X^2_{n+h}} = f^2_h(\varepsilon^2_{n+1} = 1, \ldots, \varepsilon^2_{n+h} = 1), \text{ for any } h \geq 1 \text{ and in the } L_2 \text{ sense} \tag{2.2.10}$$

Also, because all ε_t's are independent with each other as well as the past values of time points less than n, the predictions in (2.2.10) are equivalent to the method of multi-stage prediction, which apply the predicted values of the current time step to determine its value in the next time step. However, for the L_1 case, because the median function is not a linear operator, this conclusion is not true. We have derived an approximate analytic form of the conditional distribution function of X^2_{n+2} as following:

$$F_{f^2_2}(x|\{X_n, \ldots, X_1\}) = \frac{\gamma(\frac{1}{2}, \frac{x}{2A})}{B\sqrt{\pi}}, \quad x > 0$$

where

$$A = \alpha + a_2 X^2_n + a_3 X^2_{n-1} + \ldots + a_p X^2_{n-p+2},$$

$$B = a_1 \sigma^2_{n+1} = a_1(\alpha + a_1 X^2_n + \ldots + a_p X^2_{n-p+1}),$$

$$\gamma(s,x) = \int_0^x t^{s-1} e^{-t} dt, \quad s > 0$$

Solve $F_{f^2_h}(x|\{X_n, \ldots, X_1\}) = \frac{1}{2}$, we get the

$$\widehat{X^2_{n+2}} \approx \frac{\pi}{8} AB^2 + \pi^4 B^4.$$

For the GARCH(1,1) process with *i.i.d* errors, we can use the similar way to predict the futures. Based on the exact formula of a GARCH(1,1) model, we can also write the future value $X_{n+h}(h \geq 1)$ as a function f_h of the past observations and the unknown *i.i.d* errors $\varepsilon_{n+1}, \ldots, \varepsilon_{n+h}$.

2.2.2 For NoVaS

Given a sequence of observations $\{X_1, ..., X_n\}$, we will fit the data by a special application of model-free methodology—NoVaS, which was first introduced by Politis (2003, 2007) for stationary data in prediction of squared financial returns. Let us continue considering a zero mean and (strictly) stationary financial return time series $\{X_t, t \leq n\}$. The NoVaS methodology is trying to map the dataset $X_1, \ldots, X_n$ to a *i.i.d* Gaussian series $\{W_t, t \leq n\}$, where

$$W_t := \frac{X_t}{\sqrt{\alpha s_{t-1}^2 + a_0 X_t^2 + \sum_{i=1}^{p} a_i X_{t-i}^2}} \quad \text{for} \quad t = p+1, p+2, \ldots, n. \tag{2.2.11}$$

More details of the NoVaS transformation can be found in Chapter 1. Suppose that the NoVaS parameters, i.e., the order $p(\geq 0)$ and the parameters $\alpha, a_0, \ldots, a_p$ have already been chosen by Simple NoVaS or Exponential NoVaS or Generalized NoVaS or not; see more details of different methods to choose the parameters in NoVaS in Politis (2007). Re-arrange the NoVaS Eq. (2.2.11) and then yield:

$$X_t = \frac{W_t}{\sqrt{1 - a_0 W_t^2}} \sqrt{\left(\alpha s_{t-1}^2 + \sum_{i=1}^{p} a_i X_{t-i}^2 \right)} \quad \text{for} \quad t = p+1, \ldots, n \tag{2.2.12}$$

Given $\{X_1, \ldots, X_n\}$ and using NoVaS transformation on them, similarly we can write X_{n+h} for $h \geq 1$ as some function of $\{X_1, \ldots, X_n\}$ and $\{W_t, t = 1, ..., h\}$.

$$X_{n+h} = f_h(X_1, \ldots, X_n; W_{n+1}, \ldots, W_{n+h}) \tag{2.2.13}$$

Since $\{X_1, \ldots, X_n\}$ are given, we can simplify (2.2.13) as

$$X_{n+h} = f_h(W_{n+1}, \ldots, W_{n+h}) \tag{2.2.14}$$

Let $\mathcal{F}_n$ be a short-hand for the observed information set, i.e., $\mathcal{F}_n = \{X_t, 1 \leq t \leq n\}$. In the L_2 sense, the optimal prediction of X^2_{n+h} based on $\mathcal{F}_n$ is given by

$$\widehat{X^2_{n+h}} = E(X^2_{n+h}|\mathcal{F}_n) = E\{f^2_h(W_{n+1},\ldots,W_{n+h})|\mathcal{F}_n)\} \tag{2.2.15}$$

Since W_t are $i.i.d$, we can get the similar results with those of the ARCH/GARCH models.

$$\widehat{X^2_{n+h}} = f^2_h(W^2_{n+1} = 1,\ldots,W^2_{n+h} = 1)) \tag{2.2.16}$$

Actually for any $h \geq 1$, we can use the similar idea with that in ARCH/GARCH cases to conduct multistep-ahead prediction in NoVaS by approximating the conditional mean or median from their conditional distribution functions.

2.2.3 Generalization

We can generalize the above prediction method to an interesting class of prediction functions $g(\cdot)$, namely the power family where $g(x) = x^k$ for some fixed k, and the power-absolute value family where $g(x) = |x|^k$. In the above case, we work on the prediction of X^2_{n+h}, that is, $g(x) = x^2$. Similarly, we can derive the best L_2 or L_1 predictor of $g(X_{n+h})$ given $\mathcal{F}_n$.

For $h = 2$ case above, we can easily get an analytic formula of the conditional distribution of X^2_{n+h}. However, when h is large or the data generating process is complex, even with the same independence and identical distribution assumptions of errors, it is still not easy or sometimes impossible to derive the analytic formula of the conditional distribution of $g(X_{n+h})$. To solve this problem, we can resort some numerical methods to approximate the conditional distribution and then compute the conditional mean and/or median.

Since the errors are independent and with the distribution function F_ε (F is normal distribution in the above setting), we can easily estimate the conditional median by generating

many ε_{n+i} $i.i.d$ from F_ε or the empirical distribution function $\hat{F}_\varepsilon$. The detailed algorithms are given as follows.

Algorithms to compute the conditional mean/median of ARCH/GARCH(1,1)

Suppose all parameters in the model are known, and $\{\varepsilon_t\} \sim i.i.d.$ with distribution F_ε. Under the independence of $\{\varepsilon_t\}$ for all $t \geq 1$, we can generate many $\varepsilon^*_{n+1}, \ldots, \varepsilon^*_{n+h} \sim i.i.d.$ from F_ε by Monte Carlo and compute the values of a sequence of $g(X^*_{n+h})$.

ALGORITHM 2.2.1. h-STEP AHEAD PREDICTION WITH PARAMETERS KNOWN

Step 1. Compute X_{n+h} based on (2.2.1) if it is an ARCH(p) process or (2.2.3) if it is a GARCH(1,1) process as a function of $\{X_1, \ldots, X_n\}$ and $\{\varepsilon_{n+1}, \ldots, \varepsilon_{n+h}\}$.

Step 2. Using Monte Carlo B times, each time generate $\{\varepsilon^*_{n+1}, \ldots, \varepsilon^*_{n+h}\}$ from F and plug in the function obtained in step.1 to compute the series of $\{X^{(1)}_{n+h}, \ldots, X^{(B)}_{n+h}\}$. Then compute the series of $\{g(X^{(1)}_{n+h}), \ldots, g(X^{(B)}_{n+h})\}$.

Step 3. Calculate the predictor $\widehat{g(X_{n+h})}$ of $g(X_{n+h})$ by taking the median (if L_1 measure) or mean (if L_2 measure) of the series $\{g(X^{(1)}_{n+h}), \ldots, g(X^{(B)}_{n+h})\}$.

If the parameters and F_ε in the model are unknown, firstly we need to select the lag order p if ARCH(p) and then estimate the parameters. F_ε can be approximated by $\hat{F}_\varepsilon$.

ALGORITHM 2.2.2. h-STEP AHEAD PREDICTION WITH PARAMETERS UNKNOWN

Step 1. Fit the data with an ARCH(p) or GARCH(1,1) model and get the estimators $\{\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_p\}$ if fitting ARCH(p) or $\{\hat{\alpha}, \hat{a}_1, \hat{b}_1\}$ if fitting GARCH(1,1) and also record the residuals $\{\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n\}$ with the distribution function $\hat{F}_\varepsilon$.

Step 2. Using the estimators of coefficients in Step 1, compute X_{n+h} as a function of $\{X_1, \ldots, X_n\}$ and $\{\varepsilon_{n+1}, \ldots, \varepsilon_{n+h}\}$.

Step 3. Using Monte Carlo B times, each time generate $\{\varepsilon^*_{n+1}, \ldots, \varepsilon^*_{n+h}\}$ from $\hat{F}_\varepsilon$ and then compute the series of $\{X^{(1)}_{n+h}, \ldots., X^{(B)}_{n+h}\}$. Then compute the series of $\{g(X^{(1)}_{n+h}), \ldots., g(X^{(B)}_{n+h})\}$.

Step 4. Calculate the predictor $g(\widehat{X_{n+h}})$ of $g(X_{n+h})$ by taking median (if L_1 measure) or mean (if L_2 measure) of the series $\{g(X^{(1)}_{n+h}), \ldots, g(X^{(B)}_{n+h})\}$.

Remark. *Bose and Mukherjee (2009) proposed a weighted linear estimator(WLE) to estimate the ARCH parameters. This method does not involve nonlinear optimization and gives a closed form expression, so it is computationally easier to obtain the estimator compared to QMLE. We use the WLE to get the estimators* $\{\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_p\}$ *of* $\{a_0, a_1, \ldots, a_p\}$ *in Algorithm 2.2.2.*

Remark. *If the innovations* $\{\varepsilon_t\}$ *are i.i.d* $N(0, \sigma^2)$*, we can directly generate* $\varepsilon^*_{n+1}, \ldots, \varepsilon^*_{n+h}$ *each time from* $N(0, \hat{\sigma}^2)$ *in step 3 of Algorithm 2.2.1.* $\hat{\sigma}^2$ *is given by*

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}(\hat{\varepsilon}_t - \bar{\varepsilon})^2 \text{ and } \bar{\varepsilon} = \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t.$$

In order to estimate the conditional mean or conditional median in NoVaS transformation, we should first use one of the NoVaS methods (Simple vs. Exponential, Generalized or not,

etc.) to get the coefficients $\alpha, a_0, a_1, \ldots, a_p$. Based on the independence and standard normal distribution of W_t, we can use Monte Carlo to generate different W^*_{n+k} for $k = 1, \ldots, h$. and then approximate the distribution of $f_h(W_{n+1}, \ldots, W_{n+h})$. The following algorithm is similar with Algorithm 2.2.2.

ALGORITHM 2.2.3. h-STEP AHEAD PREDICTION FOR NOVAS

Step 1. Use one of the NoVaS methods (Simple vs. Exponential, Generalized or not, etc.) to obtain the transformed data $\{W_t \text{ for } t = p+1, \ldots, n\}$ and the coefficients α, p and $a_0, a_1, \ldots, a_p$.

Step 2. Compute the analytic form of X_{n+h} as a function of $\{X_1, \ldots, X_n\}$ and $\{W_{n+1}, \ldots, W_{n+h}\}$ and $\{a_0, a_1, \ldots, a_p\}$ based on the NoVaS transformation method we use in Step 1.

Step 3. Use Monte Carlo B times, each time generate $\{W^*_{n+1}, \ldots, W^*_{n+h}\}$ from $N(0,1)$ distribution truncated to $\pm\frac{1}{\sqrt{a_0}}$. Then plug in the generated series $\{W^*_{n+1}, \ldots, W^*_{n+h}\}$ into the form obtained in the above Step 2 to compute $X^{(j)}_{n+h}$, where j = 1, . . . , B. Then to compute $g(X^{(j)}_{n+h})$, where j = 1, . . . , B.

Step 4. Calculate the predictor $\widehat{g(X_{n+h})}$ of $g(X_{n+h})$ by taking the median (if L_1 measure) or mean (if L_1 measure) of the series $\{g(X^{(1)}_{n+h}), \ldots, g(X^{(B)}_{n+h})\}$.

Remark. *From* (2.2.11)*, we note that*

$$\frac{1}{W_t^2} = \frac{\alpha s_{t-1}^2 + a_0 X_t^2 + \sum_{i=1}^p a_i X_{t-i}^2}{X_t^2} \geq a_0$$

if all the parameters are nonnegative, therefore,

$$|W_t| \leq 1/\sqrt{a_0} \tag{2.2.17}$$

So one must be careful to ensure that the $\{W_t\}$ variables have a large enough range such that the boundedness is not seen as spoiling the normality. Thus, we also require

$$\frac{1}{\sqrt{a_0}} \geq C \quad i.e., \quad a_0 \leq 1/C^2 \tag{2.2.18}$$

for some appropriate C of the practitioner's choice.Recalling that 99.7% of the mass of the $N(0,1)$ distribution is found in the range ± 3, the simple choice $C = 3$ can be suggested; this choice seems to work reasonably well—at least for the usual samples sizes. Therefore, in step 3 of Algorithm 2.2.3, we use a truncated norm distribution.

2.3 Prediction Intervals

For multi-step ahead interval predictions, we use the above methods as well as bootstrap to construct α-level confidence intervals of $g(X_{n+h})$. Given the financial returns series $\{X_1, \ldots, X_n\}$, now we construct L_1 and L_2 h-step ahead prediction intervals of $g(X_{n+h})$. The basic Model-free (MF, for short) bootstrap algorithm for one-step ahead prediction intervals in the setting of financial returns can be found in Chapter 1.

ALGORITHM 2.3.1 BOOTSTRAP PREDICTION INTERVALS FOR $g(X_{n+h})$ USING GARCH(1,1)

1. Use Algorithm 2.2.1 or Algorithm 2.2.2 to compute predictors $g(\widehat{X_{n+h}})$, the point predictor of $g(X_{n+h})$ and the residuals $\{\hat{\varepsilon}_t, t = 1,\ldots,n\}$.

3. (a) Re-sample (with replacement) the residuals $\{\varepsilon_t \text{ for } t = 1,\ldots,n\}$ to create the pseudo-data $\varepsilon^*_{p+1},\cdots,\varepsilon^*_{n-1},\varepsilon^*_n$ and $\varepsilon^*_{n+1},\cdots,W^*_{n+h}$. Then Generate the bootstrap pseudo-data X^*_t for $t = p+1,\ldots,n$ by iteration based on the fitted models in Step 1.

   (b) Calculate the bootstrap future value X^*_{n+h} by iteration

   (c) Based on the bootstrap data $X^*_1,\ldots,X^*_n$, re-estimate the parameters. With the data $X_{n-p+1},\ldots,X_n$ and the re-estimated parameters, use Algorithm 2.2.1 to calculate the bootstrap predictor $g(\widehat{X^*_{n+h}})$.

   (d) Calculate the bootstrap root: $g(X^*_{n+h}) - g(\widehat{X^*_{n+h}})$.

4. Repeat step 3 above B times; the B bootstrap root replicates are collected in the form of an empirical distribution whose α-quantile is denoted $q(\alpha)$.

5. The $(1-\alpha)100\%$ equal-tailed prediction interval for $g(X_{n+h})$ is given by

$$[g(\widehat{X_{n+h}}) + q(\alpha/2),\ g(\widehat{X_{n+h}}) + q(1-\alpha/2)].$$

ALGORITHM 2.3.2. MF BOOTSTRAP PREDICTION INTERVALS FOR $g(X_{n+h})$

1. Use one of the NoVaS algorithms (Simple vs.Exponential, Generalized or not, etc.) to obtain the transformed data $\{W_t \text{ for } t = p+1,\ldots,n\}$ that are assumed to be approximately *i.i.d*. Let p, α and a_i denote the fitted NoVaS parameters.

2. Use Algorithm 2.2.3 to calculate $g(\widehat{X_{n+h}})$, the point predictor of $g(X_{n+h})$.

3. (a) Re-sample randomly (with replacement) the transformed variables $\{W_t \text{ for } t = p+1,\ldots,n\}$ to create the pseudo-data $W^*_{p+1},\cdots,W^*_{n-1},W^*_n$ and $W^*_{n+1},\cdots,W^*_{n+h}$.

   (b) Let $(X^*_1,\ldots,X^*_p)' = (X_{1+I},\cdots,X_{p+I})'$ where I is generated as a discrete random variable uniform on the values $0,1,\ldots,n-p$.

   (c) Generate the bootstrap pseudo-data X^*_t for $t = p+1,\ldots,n$ using the following equations, i.e., let

$$X^*_t = \frac{W^*_t}{\sqrt{1-a_0 W^{*2}_t}}\sqrt{\alpha s^{*2}_{t-1} + \sum_{i=1}^{p} a_i X^{*2}_{t-i}} \text{ for } t = p+1,\ldots,n \tag{2.3.1}$$

   where $s^{*2}_{t-1} = (t-1)^{-1}\sum_{k=1}^{t-1} X^{*2}_k$.

3. (d) Calculate the bootstrap future value X^*_{n+h} by iteration as

$$X^*_{n+1} = \frac{W^*_{n+1}}{\sqrt{1-a_0 W^{*2}_{n+1}}}\sqrt{\alpha s_n^2 + \sum_{i=1}^{p} a_i X^2_{n-i+1}}$$

where $s_n^2 = n^{-1}\sum_{i=1}^{n} X_i^2$.

if $h < p$, for $j = 2,\ldots,h$

$$X^*_{n+j} = \frac{W^*_{n+j}}{\sqrt{1-a_0 W^{*2}_{n+j}}}\sqrt{\alpha s^2_{n+1-j} + \sum_{k=1}^{j-1} a_k X^{*2}_{n-k+j} + \sum_{i=j}^{p} a_i X^2_{n-i+j}} \tag{2.3.2}$$

if $h > p$, for $j = 2,\ldots,h$

$$X^*_{n+j} = \frac{W^*_{n+j}}{\sqrt{1-a_0 W^{*2}_{n+j}}}\sqrt{\alpha s^2_{n+1-j} + \sum_{i=1}^{p} a_i X^{*2}_{n-i+j}} \tag{2.3.3}$$

where $s^2_{n+1-j} = (n+j-1)^{-1}(\sum_{i=1}^{n} X_i^2 + \sum_{k=1}^{j-1} X^{*2}_{n+k})$.

(e) Based on the bootstrap data $X^*_1,\ldots,X^*_n$, re-estimate the NoVaS transformation yielding parameters p^*, α^*, $a^*_0, a^*_1, \ldots, a^*_p$. With the data $X_{n-p+1},\ldots,X_n$ and the parameters p^*, α^*, $a^*_0, a^*_1, \ldots, a^*_p$, use Algorithm 2.2.3 to calculate the bootstrap predictor $g(\widehat{X^*_{n+h}})$

(f) Calculate the bootstrap root: $g(X^*_{n+h}) - g(\widehat{X^*_{n+h}})$.

4. Repeat step 3 above B times; the B bootstrap root replicates are collected in the form of an empirical distribution whose α-quantile is denoted $q(\alpha)$.

5. The $(1-\alpha)100\%$ equal-tailed prediction interval for $g(X_{n+h})$ is given by

$$[g(\widehat{X_{n+h}}) + q(\alpha/2),\ g(\widehat{X_{n+h}}) + q(1-\alpha/2)].$$

2.4 Simulation and Finite Sample performance

In this section, we conduct some simulation to examine the finite sample performance of our algorithms.

2.4.1 Settings

In the simulation, 200 data-sets $X_N = (X_1, \ldots, X_N)'$ are generated separately by the following 7 different GARCH(1,1) models.

Model 1. Standard GARCH with Gaussian errors and finite fourth moment:

$X_t = \sigma_t \varepsilon_t,\ \sigma_t^2 = .00001 + .73\sigma_{t-1}^2 + .10X_{t-1}^2,\ \{\varepsilon_t\} \sim i.i.d.N(0,1).$

Model 2. Standard GARCH with Gaussian errors and infinite fourth moment:

$X_t = \sigma_t \varepsilon_t,\ \sigma_t^2 = .00001 + 0.8895\sigma_{t-1}^2 + .10X_{t-1}^2,\ \{\varepsilon_t\} \sim i.i.d.N(0,1).$

Model 3. Standard GARCH with student-t errors:

$X_t = \sigma_t \varepsilon_t,\ \sigma_t^2 = .00001 + .73\sigma_{t-1}^2 + .10X_{t-1}^2,\ \{\varepsilon_t\} \sim i.i.d.\ t$ distributed with degree of freedom 5.

Model 4. GARCH with slowing varying parameters(TV-GARCH):

The value of β decreases as a linear function of t, starting at $\beta_1 = 0.10$ for $t = 1$, and ending at $\beta = 0.05$ for $t = n$. At the same time, the value of α increases as a linear function of t, starting at $\alpha = 0.73$ for $t = 1$, and ending at $\alpha = 0.93$ for $t = n$. $\omega = 0.00001$ and $\{\varepsilon_t\} \sim i.i.d.N(0,1)$.

Model 5. Two-state Markov Switching GARCH(1,1) (MS-GARCH):

$$X_t = \sigma_t \varepsilon_t, \sigma_t^2 = \sum_{s=1}^{2} 1\{P(S_t = s)\}[\omega_s + \alpha_t \sigma_{t-1}^2 + \beta_s X_{t-1}^2$$

In the first regime, we set $\alpha_1 = 0.9$, $\beta_1 = 0.07, \omega_1 = 2.4e-5$. In the second regime, we set $\alpha_2 = 0.7$, $\beta_2 = 0.22, \omega_2 = 1.2e-4$. The transition probabilities for the first regime are $p_{11} = 0.9$ and $p_{12} = 0.1$ while for the second regime we use $p_{21} = 0.3$ and $p_{22} = 0.7$. $\{\varepsilon_t\} \sim i.i.d.N(0,1)$.

Model 6. Smooth Transition GARCH(ST-GARCH):

$$X_t = [a - b(t/T)]\sigma_t\varepsilon_t,\ \sigma_t^2 = \omega + \alpha\sigma_{t-1}^2 + \beta X_{t-1}^2$$

where $\{\varepsilon_t\} \sim i.i.d.N(0,1)$. $\omega = 1.2e-5, \alpha = 0.9, \beta = 0.07, a = \alpha + \beta = 0.97$, and $b = \beta/\alpha \approx 0.078$.

Model 7. Stochastic Volatility Model(SV-GARCH):

$$X_t|h_t \sim N(0, exp(h_t)),$$

$$h_t|h_{t-1} \sim N(\mu + \phi(h_{t-1} - \mu), \eta^2),\ h_0 \sim N(\mu, \eta^2/(1-\phi^2)),$$

where $\mu = -10$, $\phi = 0.95$, $\eta = 0.2$.

Each dataset is of size $n = 100$. We try to do up to 5 step ahead point predictions and interval predictions for each dataset. In the evaluating process, repeat Monte Carlo 5000 times. Five models or transformations are used to fit the data in both point predictions and interval predictions as follows: fitting a GARCH(1,1) model, Simple-NoVaS, Exponential NoVaS(Exp-NoVaS, for short), General Simple Novas(GS-NoVaS, for short) and General Exponential NoVaS(GE-NoVaS, for short).

In point predictions, the mean absolute deviations(MADs) and mean squared errors(MSEs) for 5 steps ahead point predictions in both the L_1 and L_2 senses(the absolute value or the square of the prediction error at the updated time point averaged over the 200 replications)

are recorded. Also, the bootstrap prediction interval(L_i, U_i) with a nominal coverage %95 was constructed for the future values X_{n+h} with $h = 1,\ldots,5$. The bootstrap replication B = 300. The corresponding empirical average coverage level(CVR) and the average length(LEN) of the constructed intervals and the standard error (St.err) associated with each length of the constructed intervals are calculated as

$$CVR = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{[L_i,U_i]} X_{(n+h,i)}$$

$$LEN = \frac{1}{N}\sum_{i=1}^{N} LEN_i \text{ and } St.err = \frac{1}{N}\sum_{i=1}^{N} (LEN_i - LEN)^2$$

where $LEN_i = U_i - L_i$..

2.4.2 Results and Discussions

The simulation results for point predictions are shown in Table 2.1 - 2.28. The following conclusions can be obtained from the results:

- When comparing the MADs between the L_1 and L_2 predictions by fitting the same models, we can find that MADs of L_1 predictions are always smaller than that of L_2 predictions. Also, we can find that MSEs of L_1 predictions are always bigger than that of L_2 predictions, when comparing the MSEs between the L_1 and L_2 predictions with same models' settings. These are both theoretically and empirically reasonable since in the L_1 sense, we try to minimize the mean absolute deviations. While for L_2, the loss function to be minimized is the mean squared errors.

- Furthermore, for each model fitting results, there are not obvious errors accumulation problems in the multistep-ahead prediction for both L_1 and L_2 measures. In this project, we just did up to five steps ahead predictions. Maybe we can conduct much more than five steps to examine if the errors accumulation issue becomes worse.

- NoVaS methods consistently perform better than GARCH(1,1) for all data generating processes. When the prediction step is higher, the difference of their performance are getting smaller.

For prediction intervals, the simulated results are summarized in Table 2.29-2.35. We can find the conclusions are similar with that of the point predictions. However, NoVaS methods give more accurate coverage than GARCH(1,1) in the L_1 sense prediction for all these data generating processes. In the L_2 sense, when the data are generated from a standard GARCH(1,1) with normal errors, GARCH(1,1) also gives good coverage as NoVaS. When we use other models to generate data, for example, GARCH(1,1) with t distributed errors and MS-GARCH(1,1) and ST-GARCH(1,1) and others, GARCH(1,1) performs very poorly while NoVaS methods are still as good. These results show the drawbacks of GARCH(1,1) again in the Chapter 1. NoVaS methods are more robust than GARCH(1,1) when the data is nonstationary or with structure breaks.

Table 2.1: MADs of L_1 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .8895, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.N(0,1)$.

Prediction step	1	2	3	4	5
Fitting a GARCH	8.23E-05	7.35E-05	6.93E-05	8.86E-05	1.49E-04
Simple-NoVaS	6.99E-05	7.88E-05	8.29E-05	1.06E-04	1.63E-04
Exp-NoVaS	7.21E-05	8.28E-05	8.72E-05	1.14E-04	1.70E-04
GS-NoVaS	6.30E-05	7.31E-05	8.07E-05	8.71E-05	9.94E-05
GE-NoVaS	7.02E-05	8.44E-05	8.76E-05	1.16E-04	1.71E-04

Table 2.2: MADs of L_2 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .8895, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.N(0,1)$.

Prediction step	1	2	3	4	5
Fitting a GARCH	1.48E-04	2.07E-04	2.67E-04	3.20E-04	4.60E-04
Simple-NoVaS	5.22E-05	6.44E-05	6.61E-05	8.82E-05	1.49E-04
Exp-NoVaS	5.13E-05	6.41E-05	6.58E-05	8.84E-05	1.50E-04
GS-NoVaS	4.71E-05	5.97E-05	6.06E-05	8.48E-05	1.46E-04
GE-NoVaS	4.84E-05	6.26E-05	6.38E-05	8.73E-05	1.48E-04

Table 2.3: MSEs of L_1 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .8895, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.N(0,1)$.

Prediction step	1	2	3	4	5
Fitting a GARCH	2.48E-08	4.42E-08	4.36E-08	1.94E-07	1.12E-06
Simple-NoVaS	1.23E-08	4.24E-08	4.17E-08	1.94E-07	1.12E-06
Exp-NoVaS	1.18E-08	4.21E-08	4.14E-08	1.94E-07	1.12E-06
GS-NoVaS	1.08E-08	4.09E-08	3.94E-08	1.92E-07	1.13E-06
GE-NoVaS	1.10E-08	4.14E-08	4.06E-08	1.92E-07	1.12E-06

Table 2.4: MSEs of L_2 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .8895, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.N(0,1)$.

Prediction step	1	2	3	4	5
Fitting a GARCH	1.03E-07	2.11E-07	3.92E-07	6.26E-07	1.86E-06
Simple-NoVaS	1.36E-08	4.15E-08	4.10E-08	1.93E-07	1.11E-06
Exp-NoVaS	1.31E-08	4.13E-08	4.07E-08	1.93E-07	1.12E-06
GS-NoVaS	1.03E-08	3.88E-08	3.73E-08	1.91E-07	1.13E-06
GE-NoVaS	1.13E-08	3.98E-08	3.93E-08	1.88E-07	1.11E-06

Table 2.5: MADs of L_1 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.N(0,1)$.

Prediction step	1	2	3	4	5
Fitting a GARCH	4.84E-05	4.69E-05	5.18E-05	5.45E-05	5.99E-05
Simple-NoVaS	4.94E-05	4.73E-05	5.26E-05	5.59E-05	6.00E-05
Exp-NoVaS	4.87E-05	4.69E-05	5.24E-05	5.55E-05	6.00E-05
GS-NoVaS	4.83E-05	4.67E-05	5.17E-05	5.44E-05	5.94E-05
GE-NoVaS	4.84E-05	4.69E-05	5.25E-05	5.43E-05	6.00E-05

Table 2.6: MADs of L_2 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.\, N(0,1)$.

Prediction step	1	2	3	4	5
Fitting a GARCH	5.68E-05	5.45E-05	5.54E-05	6.25E-05	6.26E-05
Simple-NoVaS	6.20E-05	6.05E-05	6.06E-05	6.96E-05	6.67E-05
Exp-NoVaS	6.20E-05	6.09E-05	6.15E-05	7.05E-05	6.87E-05
GS-NoVaS	5.98E-05	5.81E-05	5.84E-05	6.55E-05	6.37E-05
GE-NoVaS	5.50E-05	5.30E-05	5.75E-05	6.22E-05	6.23E-05

Table 2.7: MSEs of L_1 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.\, N(0,1)$.

Prediction step	1	2	3	4	5
Fitting a GARCH	7.94E-09	8.48E-09	7.80E-09	8.67E-09	1.03E-08
Simple-NoVaS	8.00E-09	8.35E-09	7.85E-09	8.88E-09	1.03E-08
Exp-NoVaS	7.95E-09	8.33E-09	7.83E-09	8.84E-09	1.03E-08
GS-NoVaS	7.78E-09	8.35E-09	7.63E-09	8.59E-09	1.02E-08
GE-NoVaS	8.02E-09	8.64E-09	7.99E-09	8.92E-09	1.02E-08

Table 2.8: MSEs of L_2 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.\, N(0,1)$.

Prediction step	1	2	3	4	5
Fitting a GARCH	7.28E-09	7.47E-09	6.50E-09	7.64E-09	8.65E-09
Simple-NoVaS	7.79E-09	7.50E-09	6.84E-09	8.67E-09	8.85E-09
Exp-NoVaS	7.48E-09	7.34E-09	6.64E-09	8.39E-09	8.70E-09
GS-NoVaS	7.12E-09	7.40E-09	6.41E-09	7.90E-09	8.43E-09
GE-NoVaS	6.99E-09	7.44E-09	6.43E-09	7.68E-09	8.53E-09

Table 2.9: MADs of L_1 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.\ t$ distributed with degree of freedom 5.

Prediction step	1	2	3	4	5
Fitting a GARCH	1.46E-04	1.26E-04	1.29E-04	1.84E-04	1.77E-04
Simple-NoVaS	1.44E-04	1.24E-04	1.27E-04	1.82E-04	1.76E-04
Exp-NoVaS	1.43E-04	1.22E-04	1.26E-04	1.80E-04	1.75E-04
GS-NoVaS	1.43E-04	1.23E-04	1.27E-04	1.80E-04	1.76E-04
GE-NoVaS	1.43E-04	1.23E-04	1.29E-04	1.81E-04	1.77E-04

Table 2.10: MADs of L_2 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.\ t$ distributed with degree of freedom 5.

Prediction step	1	2	3	4	5
Fitting a GARCH	1.72E-04	1.53E-04	1.63E-04	2.17E-04	2.16E-04
Simple-NoVaS	1.58E-04	1.35E-04	1.44E-04	1.97E-04	1.92E-04
Exp-NoVaS	1.57E-04	1.35E-04	1.46E-04	1.98E-04	1.93E-04
GS-NoVaS	1.56E-04	1.35E-04	1.45E-04	1.99E-04	1.92E-04
GE-NoVaS	1.59E-04	1.34E-04	1.48E-04	2.01E-04	1.94E-04

Table 2.11: MSEs of L_1 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.\ t$ distributed with degree of freedom 5.

Prediction step	1	2	3	4	5
Fitting a GARCH	1.06E-07	7.11E-08	8.97E-08	2.95E-07	3.81E-07
Simple-NoVaS	1.03E-07	6.86E-08	9.00E-08	2.93E-07	3.83E-07
Exp-NoVaS	1.03E-07	6.83E-08	9.00E-08	2.93E-07	3.82E-07
GS-NoVaS	1.03E-07	6.90E-08	9.05E-08	2.94E-07	3.84E-07
GE-NoVaS	1.05E-07	6.88E-08	9.23E-08	2.96E-07	3.85E-07

Table 2.12: MSEs of L_2 predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\{\varepsilon_t\} \sim i.i.d.\ t$ distributed with degree of freedom 5.

Prediction step	1	2	3	4	5
Fitting a GARCH	1.05E-07	6.94E-08	8.91E-08	2.84E-07	3.79E-07
Simple-NoVaS	9.24E-08	6.03E-08	8.32E-08	2.77E-07	3.66E-07
Exp-NoVaS	9.16E-08	5.96E-08	8.28E-08	2.76E-07	3.64E-07
GS-NoVaS	9.17E-08	6.07E-08	8.35E-08	2.77E-07	3.67E-07
GE-NoVaS	9.48E-08	6.12E-08	8.53E-08	2.80E-07	3.67E-07

Table 2.13: MADs of L_1 predictions for Data generated from GARCH(1,1) with slowing varying parameters(TV-GARCH)

Prediction step	1	2	3	4	5
Fitting a GARCH	1.94E-04	2.17E-04	2.01E-04	1.76E-04	2.09E-04
Simple-NoVaS	1.91E-04	2.12E-04	2.03E-04	1.72E-04	2.07E-04
Exp-NoVaS	1.91E-04	2.12E-04	2.02E-04	1.72E-04	2.06E-04
GS-NoVaS	1.91E-04	2.13E-04	2.02E-04	1.73E-04	2.06E-04
GE-NoVaS	1.97E-04	2.17E-04	2.02E-04	1.79E-04	2.13E-04

Table 2.14: MADs of L_2 predictions for Data generated from GARCH(1,1) with slowing varying parameters(TV-GARCH)

Prediction step	1	2	3	4	5
Fitting a GARCH	1.90E-04	2.13E-04	2.03E-04	1.73E-04	2.02E-04
Simple-NoVaS	1.94E-04	2.11E-04	2.07E-04	1.74E-04	2.10E-04
Exp-NoVaS	1.94E-04	2.11E-04	2.05E-04	1.74E-04	2.09E-04
GS-NoVaS	1.91E-04	2.10E-04	1.99E-04	1.71E-04	2.03E-04
GE-NoVaS	1.89E-04	2.09E-04	1.99E-04	1.72E-04	2.07E-04

Table 2.15: MSEs of L_1 predictions for Data generated from GARCH(1,1) with slowing varying parameters(TV-GARCH)

Prediction step	1	2	3	4	5
Fitting a GARCH	1.21E-07	1.70E-07	1.28E-07	1.06E-07	1.58E-07
Simple-NoVaS	1.15E-07	1.61E-07	1.23E-07	9.86E-08	1.53E-07
Exp-NoVaS	1.16E-07	1.61E-07	1.23E-07	9.95E-08	1.52E-07
GS-NoVaS	1.16E-07	1.62E-07	1.24E-07	9.99E-08	1.53E-07
GE-NoVaS	1.25E-07	1.70E-07	1.31E-07	1.08E-07	1.61E-07

Table 2.16: MSEs of L_2 predictions for Data generated from GARCH(1,1) with slowing varying parameters(TV-GARCH)

Prediction step	1	2	3	4	5
Fitting a GARCH	1.07E-07	1.55E-07	1.12E-07	9.44E-08	1.41E-07
Simple-NoVaS	9.60E-08	1.41E-07	1.06E-07	8.36E-08	1.32E-07
Exp-NoVaS	9.60E-08	1.40E-07	1.05E-07	8.39E-08	1.30E-07
GS-NoVaS	9.70E-08	1.41E-07	1.07E-07	8.51E-08	1.32E-07
GE-NoVaS	1.04E-07	1.47E-07	1.12E-07	9.02E-08	1.38E-07

Table 2.17: MADs of L_1 predictions for Data generated from Two-state Markov Switching GARCH(1,1) (MS-GARCH)

Prediction step	1	2	3	4	5
Fitting a GARCH	6.76E-04	7.67E-04	8.44E-04	7.80E-04	7.14E-04
Simple-NoVaS	7.00E-04	7.74E-04	8.85E-04	7.90E-04	7.26E-04
Exp-NoVaS	7.02E-04	7.75E-04	8.87E-04	7.92E-04	7.26E-04
GS-NoVaS	6.97E-04	7.70E-04	8.80E-04	7.91E-04	7.23E-04
GE-NoVaS	7.06E-04	7.75E-04	8.85E-04	7.98E-04	7.27E-04

Table 2.18: MADs of L_2 predictions for Data generated from Two-state Markov Switching GARCH(1,1) (MS-GARCH)

Prediction step	1	2	3	4	5
Fitting a GARCH	7.80E-04	9.01E-04	9.21E-04	9.17E-04	8.80E-04
Simple-NoVaS	6.77E-04	7.50E-04	8.59E-04	7.73E-04	7.10E-04
Exp-NoVaS	6.78E-04	7.50E-04	8.61E-04	7.75E-04	7.09E-04
GS-NoVaS	6.76E-04	7.50E-04	8.59E-04	7.76E-04	7.08E-04
GE-NoVaS	6.81E-04	7.50E-04	8.62E-04	7.79E-04	7.09E-04

Table 2.19: MSEs of L_1 predictions for Data generated from Two-state Markov Switching GARCH(1,1) (MS-GARCH)

Prediction step	1	2	3	4	5
Fitting a GARCH	1.27E-06	2.30E-06	2.35E-06	2.78E-06	1.49E-06
Simple-NoVaS	1.45E-06	2.46E-06	2.73E-06	3.09E-06	1.76E-06
Exp-NoVaS	1.45E-06	2.46E-06	2.73E-06	3.10E-06	1.76E-06
GS-NoVaS	1.43E-06	2.45E-06	2.71E-06	3.10E-06	1.74E-06
GE-NoVaS	1.46E-06	2.47E-06	2.74E-06	3.12E-06	1.76E-06

Table 2.20: MSEs of L_2 predictions for Data generated from Two-state Markov Switching GARCH(1,1) (MS-GARCH)

Prediction step	1	2	3	4	5
Fitting a GARCH	1.29E-06	2.28E-06	2.05E-06	2.55E-06	1.51E-06
Simple-NoVaS	1.34E-06	2.33E-06	2.58E-06	2.97E-06	1.64E-06
Exp-NoVaS	1.34E-06	2.33E-06	2.59E-06	2.98E-06	1.64E-06
GS-NoVaS	1.33E-06	2.32E-06	2.57E-06	2.97E-06	1.62E-06
GE-NoVaS	1.35E-06	2.33E-06	2.59E-06	2.99E-06	1.64E-06

Table 2.21: MADs of L_1 predictions for Data generated from Smooth Transition GARCH(1,1)(ST-GARCH)

Prediction step	1	2	3	4	5
Fitting a GARCH	1.83E-04	1.78E-04	2.01E-04	2.02E-04	2.22E-04
Simple-NoVaS	1.83E-04	1.79E-04	2.03E-04	2.02E-04	2.24E-04
Exp-NoVaS	1.82E-04	1.78E-04	2.03E-04	2.02E-04	2.24E-04
GS-NoVaS	1.81E-04	1.78E-04	2.02E-04	1.98E-04	2.21E-04
GE-NoVaS	1.82E-04	1.79E-04	2.05E-04	1.99E-04	2.24E-04

Table 2.22: MADs of L_2 predictions for Data generated from Smooth Transition GARCH(1,1)(ST-GARCH)

Prediction step	1	2	3	4	5
Fitting a GARCH	2.23E-04	2.14E-04	2.18E-04	2.39E-04	2.42E-04
Simple-NoVaS	1.91E-04	1.88E-04	2.02E-04	2.09E-04	2.19E-04
Exp-NoVaS	1.89E-04	1.86E-04	2.03E-04	2.09E-04	2.19E-04
GS-NoVaS	1.90E-04	1.89E-04	2.03E-04	2.09E-04	2.17E-04
GE-NoVaS	1.86E-04	1.83E-04	2.06E-04	2.03E-04	2.20E-04

Table 2.23: MSEs of L_1 predictions for Data generated from Smooth Transition GARCH(1,1)(ST-GARCH)

Prediction step	1	2	3	4	5
Fitting a GARCH	1.12E-07	1.23E-07	1.14E-07	1.13E-07	1.43E-07
Simple-NoVaS	1.16E-07	1.24E-07	1.19E-07	1.21E-07	1.48E-07
Exp-NoVaS	1.16E-07	1.24E-07	1.20E-07	1.21E-07	1.49E-07
GS-NoVaS	1.12E-07	1.22E-07	1.16E-07	1.16E-07	1.45E-07
GE-NoVaS	1.18E-07	1.27E-07	1.22E-07	1.21E-07	1.50E-07

Table 2.24: MSEs of L_2 predictions for Data generated from Smooth Transition GARCH(1,1)(ST-GARCH)

Prediction step	1	2	3	4	5
Fitting a GARCH	1.05E-07	1.13E-07	9.61E-08	1.04E-07	1.25E-07
Simple-NoVaS	1.01E-07	1.09E-07	9.96E-08	1.04E-07	1.23E-07
Exp-NoVaS	9.93E-08	1.08E-07	1.00E-07	1.03E-07	1.23E-07
GS-NoVaS	9.84E-08	1.08E-07	9.85E-08	1.00E-07	1.21E-07
GE-NoVaS	1.00E-07	1.11E-07	1.02E-07	1.02E-07	1.24E-07

Table 2.25: MADs of L_1 predictions for Data generated from Stochastic Volatility Model(SV-GARCH)

Prediction step	1	2	3	4	5
Fitting a GARCH	5.35E-05	5.77E-05	4.72E-05	4.28E-05	3.86E-05
Simple-NoVaS	5.34E-05	5.69E-05	4.79E-05	4.45E-05	3.78E-05
Exp-NoVaS	5.35E-05	5.66E-05	4.74E-05	4.35E-05	3.75E-05
GS-NoVaS	5.23E-05	5.68E-05	4.74E-05	4.23E-05	3.79E-05
GE-NoVaS	5.23E-05	5.71E-05	4.79E-05	4.31E-05	3.85E-05

Table 2.26: MADs of L_2 predictions for Data generated from Stochastic Volatility Model(SV-GARCH)

Prediction step	1	2	3	4	5
Fitting a GARCH	5.77E-05	6.23E-05	5.56E-05	5.29E-05	5.12E-05
Simple-NoVaS	6.02E-05	6.17E-05	5.85E-05	5.99E-05	5.41E-05
Exp-NoVaS	6.10E-05	6.19E-05	5.99E-05	6.04E-05	5.68E-05
GS-NoVaS	5.74E-05	6.29E-05	6.10E-05	5.85E-05	5.69E-05
GE-NoVaS	5.66E-05	5.96E-05	5.45E-05	5.28E-05	4.77E-05

Table 2.27: MSEs of L_1 predictions for Data generated from Stochastic Volatility Model(SV-GARCH)

Prediction step	1	2	3	4	5
Fitting a GARCH	1.04E-08	1.69E-08	1.08E-08	7.03E-09	5.06E-09
Simple-NoVaS	1.02E-08	1.65E-08	1.05E-08	7.06E-09	4.82E-09
Exp-NoVaS	1.04E-08	1.65E-08	1.06E-08	6.99E-09	4.78E-09
GS-NoVaS	1.00E-08	1.66E-08	1.04E-08	6.92E-09	4.80E-09
GE-NoVaS	1.03E-08	1.70E-08	1.08E-08	6.94E-09	5.04E-09

Table 2.28: MSEs of L_2 predictions for Data generated from Stochastic Volatility Model(SV-GARCH)

Prediction step	1	2	3	4	5
Fitting a GARCH	8.99E-09	1.51E-08	1.02E-08	6.76E-09	5.29E-09
Simple-NoVaS	8.70E-09	1.42E-08	9.99E-09	7.62E-09	5.49E-09
Exp-NoVaS	8.82E-09	1.39E-08	1.00E-08	7.41E-09	5.51E-09
GS-NoVaS	8.10E-09	1.42E-08	1.02E-08	6.92E-09	5.32E-09
GE-NoVaS	8.37E-09	1.46E-08	1.00E-08	6.62E-09	5.27E-09

2.5 Conclusions

In this chapter, we derive a new way of multistep-ahead predictions for ARCH/GARCH and NoVaS methods only based on the basic assumptions of models or transformations. This method has some good properties based on our theoretical methodology and simulated results. To sum up,

- This method works well for nonlinear time series processes.
- No obvious error accumulation issue.
- Free of model and only require the errors are independent.
- Easy to conduct and computationally friendly.
- Combining with the results in Chapter 1, NoVaS performs better than GARCH models in both one-step ahead prediction and multistep-ahead prediction.

2.6 Acknowledgements

Chapter 2, in full is currently being prepared for submission for publication of the material. Chen, Jie; Politis, Dimitris N. The dissertation author was the primary investigator and author of this paper.

Table 2.29: Interval predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .8895, \theta = .10$ and $\varepsilon \sim i.i.d\ N(0,1)$.

L2				L1			
GARCH(1,1)				GARCH(1,1)			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.714	1.28E-02	1.77E-02	1	0.744	1.01E-02	1.55E-02
2	0.746	1.28E-02	1.79E-02	2	0.734	1.13E-02	1.83E-02
3	0.746	1.37E-02	1.87E-02	3	0.768	1.19E-02	1.95E-02
4	0.766	1.18E-02	1.59E-02	4	0.734	1.20E-02	1.99E-02
5	0.786	1.28E-02	1.75E-02	5	0.744	1.28E-02	2.04E-02
EXP-NoVaS				EXP-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.958	2.00E-02	1.35E-02	1	0.936	1.99E-02	1.39E-02
2	0.952	2.07E-02	1.44E-02	2	0.936	2.19E-02	1.52E-02
3	0.952	1.98E-02	1.27E-02	3	0.944	1.94E-02	1.27E-02
4	0.946	2.17E-02	1.43E-02	4	0.944	2.08E-02	1.28E-02
5	0.950	2.19E-02	1.32E-02	5	0.936	2.19E-02	1.51E-02
Simple-NoVaS				Simple-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.946	1.98E-02	1.35E-02	1	0.934	1.84E-02	1.32E-02
2	0.942	1.78E-02	1.41E-02	2	0.946	1.91E-02	1.49E-02
3	0.946	1.92E-02	1.35E-02	3	0.936	2.04E-02	1.47E-02
4	0.946	2.07E-02	1.34E-02	4	0.964	1.93E-02	1.40E-02
5	0.956	2.21E-02	1.43E-02	5	0.954	2.07E-02	1.45E-02
GS-NoVaS				GS-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.948	2.05E-02	1.39E-02	1	0.95	1.61E-02	1.21E-02
2	0.942	1.78E-02	1.41E-02	2	0.94	2.07E-02	1.47E-02
3	0.952	1.93E-02	1.43E-02	3	0.936	1.62E-02	1.25E-02
4	0.948	2.14E-02	1.30E-02	4	0.936	1.70E-02	1.42E-02
5	0.954	2.26E-02	1.40E-02	5	0.94	1.78E-02	1.32E-02
GE-NoVaS				GE-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.948	1.73E-02	1.14E-02	1	0.958	1.98E-02	1.13E-02
2	0.948	1.66E-02	1.05E-02	2	0.942	1.68E-02	1.16E-02
3	0.952	1.98E-02	1.27E-02	3	0.944	1.94E-02	1.27E-02
4	0.946	2.17E-02	1.43E-02	4	0.944	2.08E-02	1.28E-02
5	0.950	2.19E-02	1.32E-02	5	0.942	2.32E-02	1.13E-02

Table 2.30: Interval predictions for Data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\varepsilon \sim i.i.d\ N(0,1)$.

L2				L1			
GARCH(1,1)				GARCH(1,1)			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.948	4.39E-03	2.11E-03	1	0.92	3.76E-03	3.18E-03
2	0.940	4.53E-03	2.23E-03	2	0.936	5.49E-03	5.21E-03
3	0.950	4.47E-03	2.74E-03	3	0.938	5.99E-03	5.52E-03
4	0.952	4.02E-03	1.89E-03	4	0.922	7.16E-03	6.31E-03
5	0.934	3.77E-03	2.74E-03	5	0.92	4.57E-03	4.21E-03
EXP-NoVaS				EXP-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.954	4.46E-03	2.69E-03	1	0.954	4.74E-03	2.29E-03
2	0.972	4.40E-03	2.55E-03	2	0.95	4.62E-03	2.19E-03
3	0.938	4.17E-03	2.59E-03	3	0.95	4.64E-03	2.12E-03
4	0.958	4.62E-03	2.59E-03	4	0.948	4.58E-03	2.12E-03
5	0.950	4.58E-03	2.49E-03	5	0.942	4.45E-03	1.88E-03
Simple-NoVaS				Simple-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.960	4.54E-03	2.50E-03	1	0.946	4.28E-03	2.45E-03
2	0.958	4.27E-03	2.97E-03	2	0.95	4.26E-03	2.33E-03
3	0.968	4.63E-03	2.87E-03	3	0.952	4.21E-03	2.62E-03
4	0.960	4.73E-03	2.85E-03	4	0.954	4.25E-03	2.55E-03
5	0.948	4.15E-03	2.93E-03	5	0.948	4.19E-03	2.32E-03
GS-NoVaS				GS-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.949	4.37E-03	2.53E-03	1	0.946	4.26E-03	2.37E-03
2	0.948	4.63E-02	2.78E-03	2	0.95	4.26E-03	2.33E-03
3	0.938	4.17E-03	2.59E-03	3	0.95	4.22E-03	2.42E-03
4	0.945	3.76E-03	2.71E-03	4	0.948	4.20E-03	1.91E-03
5	0.950	4.58E-03	2.49E-03	5	0.948	4.19E-03	2.32E-03
GE-NoVaS				GE-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.946	4.92E-03	2.57E-03	1	0.96	5.37E-03	2.00E-03
2	0.946	4.68E-03	2.38E-03	2	0.948	5.13E-03	3.34E-03
3	0.958	4.39E-03	2.35E-03	3	0.952	4.03E-03	2.05E-03
4	0.954	4.30E-03	2.03E-03	4	0.95	4.80E-03	2.03E-03
5	0.948	4.15E-03	2.93E-03	5	0.944	4.42E-03	2.78E-03

Table 2.31: Results of interval predictions for data generated from GARCH(1,1) with $\omega = .00001, \alpha = .73, \theta = .10$ and $\varepsilon \sim i.i.d\ t_5$.

L2				L1			
GARCH(1,1)				GARCH(1,1)			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.924	4.17E-03	3.18E-03	1	0.936	2.35E-03	8.60E-03
2	0.931	3.75E-03	2.57E-03	2	0.928	2.15E-03	7.50E-03
3	0.922	4.47E-03	2.24E-03	3	0.92	2.37E-03	8.52E-03
4	0.925	4.02E-03	2.63E-03	4	0.938	2.92E-03	6.95E-03
5	0.922	4.56E-03	2.79E-03	5	0.92	2.79E-03	8.20E-03
EXP-NoVaS				EXP-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.945	4.46E-03	2.66E-03	1	0.95	3.38E-03	2.95E-03
2	0.942	4.25E-03	2.72E-03	2	0.958	3.80E-03	2.72E-03
3	0.943	4.54E-03	2.77E-03	3	0.946	3.75E-03	2.40E-03
4	0.949	4.94E-03	2.72E-03	4	0.952	3.76E-03	2.43E-03
5	0.954	4.72E-03	3.08E-03	5	0.946	3.40E-03	2.84E-03
Simple-NoVaS				Simple-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.950	4.82E-03	2.35E-03	1	0.953	4.01E-03	3.07E-03
2	0.963	4.86E-03	3.26E-03	2	0.942	3.58E-03	2.73E-03
3	0.966	4.85E-03	2.82E-03	3	0.952	3.28E-03	2.48E-03
4	0.954	5.04E-03	3.05E-03	4	0.952	3.51E-03	2.54E-03
5	0.944	4.36E-03	2.51E-03	5	0.944	4.20E-03	3.03E-03
GS-NoVaS				GS-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.950	4.82E-03	2.35E-03	1	0.948	3.64E-03	4.49E-03
2	0.954	4.47E-03	2.93E-03	2	0.956	3.22E-03	5.92E-03
3	0.952	4.69E-03	3.03E-03	3	0.946	3.31E-03	4.08E-03
4	0.950	4.57E-03	2.99E-03	4	0.948	3.23E-03	4.52E-03
5	0.954	4.50E-03	3.06E-03	5	0.95	3.62E-03	4.54E-03
GE-NoVaS				GE-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.950	4.90E-03	2.57E-03	1	0.954	3.65E-03	2.78E-03
2	0.946	4.36E-03	2.93E-03	2	0.954	3.86E-03	2.84E-03
3	0.948	4.48E-03	2.82E-03	3	0.946	3.64E-03	2.81E-03
4	0.952	4.58E-03	2.78E-03	4	0.95	3.53E-03	2.89E-03
5	0.952	4.53E-03	2.94E-03	5	0.966	6.38E-03	3.92E-03

Table 2.32: Results of interval predictions for data generated from TV-GARCH(1,1)

L2				L1			
GARCH(1,1)				GARCH(1,1)			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.850	1.57E-03	1.07E-02	1	0.726	1.12E-03	4.97E-03
2	0.848	1.20E-03	1.60E-03	2	0.716	1.32E-03	5.88E-03
3	0.858	1.84E-03	2.77E-03	3	0.718	8.87E-03	3.90E-03
4	0.844	2.32E-03	2.02E-03	4	0.716	9.78E-03	4.39E-03
5	0.856	2.29E-03	1.68E-03	5	0.72	1.24E-03	5.77E-03
EXP-NoVaS				EXP-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.948	2.40E-03	2.20E-03	1	0.952	2.76E-03	2.45E-03
2	0.950	2.23E-03	2.17E-03	2	0.952	2.74E-03	2.50E-03
3	0.952	2.93E-03	2.15E-03	3	0.95	2.69E-03	2.62E-03
4	0.954	3.02E-03	2.18E-03	4	0.942	2.78E-03	2.65E-03
5	0.950	2.86E-03	2.12E-03	5	0.948	2.82E-03	2.61E-03
Simple-NoVaS				Simple-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.950	2.32E-03	2.37E-03	1	0.942	2.20E-03	2.30E-03
2	0.956	2.20E-03	2.36E-03	2	0.956	2.82E-03	2.45E-03
3	0.960	2.88E-03	2.15E-03	3	0.948	2.60E-03	2.50E-03
4	0.952	2.50E-03	2.27E-03	4	0.946	2.79E-03	2.36E-03
5	0.954	2.80E-03	2.01E-03	5	0.946	2.47E-03	2.62E-03
GS-NoVaS				GS-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.950	2.32E-03	2.37E-03	1	0.942	2.20E-03	2.30E-03
2	0.950	2.52E-03	2.57E-03	2	0.948	2.57E-03	2.33E-03
3	0.952	2.59E-03	2.53E-03	3	0.952	2.52E-03	2.18E-03
4	0.950	2.37E-03	2.05E-03	4	0.946	2.97E-03	2.22E-03
5	0.950	2.62E-03	2.12E-03	5	0.95	2.64E-03	2.21E-03
GE-NoVaS				GE-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.948	2.40E-03	2.20E-03	1	0.948	2.77E-03	2.66E-03
2	0.950	2.23E-03	2.17E-03	2	0.952	2.71E-03	2.83E-03
3	0.942	2.47E-03	2.25E-03	3	0.95	2.53E-03	2.50E-03
4	0.948	2.44E-03	2.17E-03	4	0.95	2.75E-03	2.49E-03
5	0.949	2.29E-03	2.12E-03	5	0.954	2.64E-03	2.47E-03

Table 2.33: Results of interval predictions for data generated from MS-GARCH(1,1)

L2				L1			
GARCH(1,1)				GARCH(1,1)			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.868	3.32E-02	1.78E-02	1	0.856	3.27E-02	1.25E-02
2	0.872	3.42E-02	1.50E-02	2	0.89	3.04E-02	1.10E-02
3	0.868	3.58E-02	1.62E-02	3	0.882	3.10E-02	1.07E-02
4	0.858	3.67E-02	1.86E-02	4	0.886	3.09E-02	1.12E-02
5	0.87	3.60E-02	2.10E-02	5	0.908	7.66E-03	1.28E-02
EXP-NoVaS				EXP-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.946	4.14E-02	2.18E-02	1	0.952	3.97E-02	2.67E-02
2	0.948	4.02E-02	2.26E-02	2	0.944	4.22E-02	2.80E-02
3	0.96	4.78E-02	2.19E-02	3	0.958	3.99E-02	2.74E-02
4	0.958	4.16E-02	2.06E-02	4	0.938	3.86E-02	2.80E-02
5	0.956	4.27E-02	2.07E-02	5	0.944	4.20E-02	2.97E-02
Simple-NoVaS				Simple-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.954	4.21E-02	2.87E-02	1	0.958	3.05E-02	2.03E-02
2	0.948	3.98E-02	2.81E-02	2	0.936	3.08E-02	2.26E-02
3	0.94	4.47E-02	2.91E-02	3	0.936	3.45E-02	2.13E-02
4	0.948	4.26E-02	2.80E-02	4	0.94	3.42E-02	2.28E-02
5	0.946	4.32E-02	2.93E-02	5	0.938	3.52E-02	2.16E-02
GS-NoVaS				GS-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.954	4.21E-02	2.87E-02	1	0.948	3.21E-02	2.16E-02
2	0.948	3.98E-02	2.81E-02	2	0.946	3.26E-02	2.18E-02
3	0.942	4.75E-02	2.84E-02	3	0.948	3.46E-02	2.28E-02
4	0.948	4.26E-02	2.80E-02	4	0.952	3.17E-02	2.21E-02
5	0.946	4.32E-02	2.93E-02	5	0.946	3.22E-02	2.02E-02
GE-NoVaS				GE-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.954	4.30E-02	2.08E-02	1	0.948	3.40E-02	2.19E-02
2	0.952	3.93E-02	2.09E-02	2	0.942	3.24E-02	2.09E-02
3	0.948	4.36E-02	2.03E-02	3	0.946	3.63E-02	2.15E-02
4	0.948	4.20E-02	2.05E-02	4	0.95	3.08E-02	2.39E-02
5	0.95	4.29E-02	2.07E-02	5	0.944	3.55E-02	2.90E-02

Table 2.34: Results of interval predictions for data generated from ST-GARCH(1,1)

L2				L1			
GARCH(1,1)				GARCH(1,1)			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.89	2.80E-03	2.14E-03	1	0.894	2.53E-03	3.80E-03
2	0.888	2.19E-03	2.79E-03	2	0.902	2.88E-03	4.28E-03
3	0.904	2.15E-03	2.87E-03	3	0.884	2.48E-03	3.20E-03
4	0.908	2.07E-03	2.04E-03	4	0.9012	2.73E-03	4.21E-03
5	0.896	2.00E-03	2.09E-03	5	0.89	2.02E-03	4.92E-03
EXP-NoVaS				EXP-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.956	3.35E-03	2.57E-03	1	0.958	3.70E-03	2.69E-03
2	0.96	3.40E-03	2.58E-03	2	0.944	3.60E-03	2.54E-03
3	0.946	3.60E-03	2.73E-03	3	0.966	3.64E-03	2.65E-03
4	0.942	3.42E-03	2.51E-03	4	0.956	3.52E-03	2.46E-03
5	0.944	3.56E-03	2.70E-03	5	0.962	3.66E-03	2.53E-03
Simple-NoVaS				Simple-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.952	3.44E-03	2.32E-03	1	0.944	3.44E-03	2.75E-03
2	0.956	3.49E-03	2.24E-03	2	0.938	3.65E-03	2.75E-03
3	0.958	3.46E-03	2.04E-03	3	0.934	3.67E-03	2.81E-03
4	0.954	3.40E-03	2.10E-03	4	0.946	3.60E-03	2.72E-03
5	0.946	3.73E-03	2.19E-03	5	0.938	3.51E-03	2.51E-03
GS-NoVaS				GS-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.952	3.49E-03	2.18E-03	1	0.946	3.73E-03	2.16E-03
2	0.956	3.30E-03	2.29E-03	2	0.956	3.86E-03	2.31E-03
3	0.948	3.51E-03	2.29E-03	3	0.954	3.84E-03	2.27E-03
4	0.948	3.54E-03	2.36E-03	4	0.95	3.82E-03	2.15E-03
5	0.944	3.56E-03	2.70E-03	5	0.946	3.76E-03	2.22E-03
GE-NoVaS				GE-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.946	3.37E-03	2.10E-03	1	0.954	3.64E-03	2.60E-03
2	0.944	3.51E-03	2.76E-03	2	0.952	3.60E-03	2.55E-03
3	0.946	3.60E-03	2.73E-03	3	0.956	3.61E-03	2.55E-03
4	0.948	3.44E-03	2.49E-03	4	0.95	3.76E-03	2.68E-03
5	0.948	3.69E-03	2.61E-03	5	0.944	3.46E-03	2.31E-03

Table 2.35: Results of interval predictions for data generated from SV-GARCH(1,1)

L2				L1			
GARCH(1,1)				GARCH(1,1)			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.878	2.24E-02	2.99E-02	1	0.884	1.22E-02	1.93E-02
2	0.862	2.11E-02	2.51E-02	2	0.882	1.28E-02	2.10E-02
3	0.896	2.49E-02	2.78E-02	3	0.872	1.17E-02	1.60E-02
4	0.87	2.14E-02	2.18E-02	4	0.878	1.26E-02	1.46E-02
5	0.892	2.32E-02	2.33E-02	5	0.876	1.33E-02	1.63E-02
EXP-NoVaS				EXP-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.954	2.71E-02	2.59E-02	1	0.944	1.84E-02	2.33E-02
2	0.952	2.80E-02	2.67E-02	2	0.93	2.01E-02	2.46E-02
3	0.956	2.79E-02	2.66E-02	3	0.954	1.96E-02	2.45E-02
4	0.95	2.84E-02	2.65E-02	4	0.942	2.19E-02	2.62E-02
5	0.968	3.07E-02	2.84E-02	5	0.93	1.88E-02	2.43E-02
Simple-NoVaS				Simple-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.952	2.68E-02	2.79E-02	1	0.942	2.08E-02	2.40E-02
2	0.95	2.77E-02	2.92E-02	2	0.932	2.46E-02	2.53E-02
3	0.946	2.78E-02	2.74E-02	3	0.95	2.63E-02	2.84E-02
4	0.958	3.05E-02	2.90E-02	4	0.926	2.25E-02	2.65E-02
5	0.954	2.66E-02	2.73E-02	5	0.934	2.07E-02	2.41E-02
GS-NoVaS				GS-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.952	2.68E-02	2.79E-02	1	0.95	2.17E-02	2.42E-02
2	0.95	2.77E-02	2.92E-02	2	0.96	2.05E-02	2.40E-02
3	0.952	2.84E-02	2.96E-02	3	0.95	2.18E-02	2.44E-02
4	0.944	2.28E-02	2.34E-02	4	0.95	2.22E-02	2.48E-02
5	0.946	2.34E-02	2.18E-02	5	0.942	2.28E-02	2.55E-02
GE-NoVaS				GE-NoVaS			
STEPS	CVR	LEN	ST.ERR	STEPS	CVR	LEN	ST.ERR
1	0.954	2.71E-02	2.59E-02	1	0.952	2.02E-02	2.35E-02
2	0.948	2.52E-02	2.29E-02	2	0.948	2.18E-02	2.47E-02
3	0.956	2.79E-02	2.66E-02	3	0.948	1.95E-02	2.29E-02
4	0.95	2.84E-02	2.65E-02	4	0.942	2.29E-02	2.58E-02
5	0.946	2.74E-02	2.46E-02	5	0.946	2.10E-02	2.34E-02

Chapter 3

Simplified Models for Autoregression under Stationary and Non-stationary Errors

Simplified models have many appealing properties and sometimes give better parameter estimates and model predictions, in sense of mean-squared-error, than extended models, especially when the limited data are available. In this chapter, we summarize extensive quantitative and qualitative results in the literature concerned with using simplified models on autoregression processes(AR, for short) with stationary and non-stationary errors. Also, we develop a practical strategy to help modellers decide whether a simplified model should be used in AR processes based on the data observed.

3.1 Introduction

In general, we assume that a process can be truly described by

$$y_i = f(x_i, \theta) + \varepsilon_i, i = 1, 2, ..., n \tag{3.1.1}$$

where x_i is a p dimensional vector of explanatory variables for the ith observation, θ is an m dimensional vector of model parameters, and ε_i is the random part of the model.

Typically, the practitioners fit the following model (3.1.2) either because they are lack of knowledge of the true model (3.1.1) or because they think this simpler model may give better results.

$$y_i = g(z_i, \beta) + e_i, i = 1, 2, ..., n \tag{3.1.2}$$

where z_i is a vector of explanatory variables, β is vector of model parameters and $g(z_i, \beta)$ is the function that the modeller believes (or hopes) relates (z_i, β) to y_i. The term e_i encompasses the stochastic component and any deterministic part that is not captured by the model. The function form of $g(z_i, \beta)$ may be specified from a fundamental understanding of the process or a desire to find a purely empirical representation; see Wu et al. (2007). In either case, the parameters are often estimated as the solution to the least-squares problem

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - g(z_i, \beta))^2 \tag{3.1.3}$$

If the case is the parameters non-linear in the model, a non-linear optimization algorithm is considered to determine $\hat{\beta}$. When the parameters are linearly entering the model, ordinary least squares(OLS) is commonly used (Montgomery and Runger (2007)).

There are actually many instances when one deliberately chooses a structural form that does not match the true process. Interesting questions are here:

1. Can simplified models give better parameter estimates model predictions than the correctly structured extended model in the auto-regression process?

2. Is there any strategy that can be used to determine whether simplified models are better?

In the work of Wu et al. (2007), the linear regression case was considered. It is shown that the simplified model gives better parameter estimates and model predictions on average than

the true or full model if the critical conditions are satisfied. In this chapter, we extend the use of simplified model to autoregression case and try to answer these two questions above.

This chapter is organized as follows: in section 2, the homoscedastic AR processes are considered; The use of simplified model for AR processes with heteroskedasticity is studied in section 3; In section 4, we give the strategy for assessing uncertainty about which situation simplified model works better and some simulation are done for evaluating the performance; Conclusions and remarks are given in section 5.

3.2 Simplified Models— AR(p) with $i.i.d$ Errors

Assume that the true process is described by AR(p)

$$y_t = \theta_1 y_{t-1} + ... + \theta_p y_{t-p} + \varepsilon_t \tag{3.2.1}$$

where $\theta_p \neq 0$ and the lag order p is finite and known. Suppose we observe a sample containing $T+p$ observations, denoted by $\{y_{-p+1}, y_{-p+2}, ..., y_T\}$ and $p \ll T$. $\Theta = (\theta_1, \theta_2, ..., \theta_p)'$ is the parameter vector of interest. We will refer this correctly structured model as the full model(FM, for short). Assumption 1 is made as follows:

(i) All roots of the polynomial $1 - \theta_1 z - \theta_2 z^2 - ... - \theta_p z^p = 0$ lie outside the unit circle.

(ii) $\{\varepsilon_t\}$ is $i.i.d$ with mean 0 and variance σ^2.

The practitioner believes that $\theta_1, ..., \theta_{p_1}$ are close to 0, for example $\theta_i = o(1)$ for any $i = 1, \ldots, p_1$, for some positive constant $0 < p_1 < p$, and assuming $p = p_1 + p_2$.Then the practitioner decides to use a model of the form

$$y_t = \theta_{p_2} y_{t-p_2} + ... + \theta_p y_{t-p} + \eta_t \tag{3.2.2}$$

where $\eta_t = \theta_1 y_{t-1} + ... + \theta_{p_1} + \varepsilon_t$ is the stochastic component combined with any model mismatch. We will refer (3.2.2) as the "simplified" model(SM, for short).

Under Assumption 1(i), the autoregressive coefficients are assumed to satisfy the usual stability conditions which would ensure that y_t is stationary or asymptotically covariance-stationary, depending on initial conditions. Also under Assumption 1(i) and Theorem 3.1.3 in Brockwell and Davis (1986), y_t has Wold representation

$$y_t = \sum_{i=-\infty}^{\infty} \alpha_i \varepsilon_{t-i}, \tag{3.2.3}$$

where the coefficients $\{\alpha_i\}$ satisfy the recursion

$$\alpha_i - \theta_1 \alpha_{i-1} - \cdots - \theta_p \alpha_{i-p} = 0 \text{ for } i > 0,$$

$$\alpha_0 = 1, \alpha_i = 0 \text{ for } i < 0,$$

and

$$\sum_{i=0}^{\infty} |\alpha_i| < \infty. \tag{3.2.4}$$

With the causality assumption, , the Wold representation (3.2.3) is written as

$$y_t = \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i}, \tag{3.2.5}$$

which is an MA(∞) process defined by (3.2.5). By Theorem 3.2.1 in Brockwell and Davis (1986), the autocovariance function has the extremely simple form

$$\gamma_k = cov(y_t, y_{t+k}) = \sigma^2 \sum_{i=0}^{\infty} \alpha_i \alpha_{i+k}, \ k = 0, 1, ..., \infty. \tag{3.2.6}$$

Also, γ_k is finite, which is implied by equation (3.2.4) as

$$|\sum_{i=0}^{\infty}\alpha_i\alpha_{i+k}| \leq \left(\sum_{i=0}^{\infty}|\alpha_i|^2\right).$$

Now define Γ_p to be the $p \times p$ covariance matrix with $(i,j)th$ element $\gamma_{|i-j|}$, $k = |i-j| = 0,1,...,p-1$.

3.2.1 Properties of parameter estimates

Write (3.2.1) in regression form as

$$Y = X_1\Theta_1 + X_2\Theta_2 + \varepsilon = X\Theta + \varepsilon \tag{3.2.7}$$

where $Y = (y_1, y_2, ..., y_T)^{'}$, $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_T)^{'}$, and $\Theta_1 = (\theta_1, ..., \theta_{p_1})^{'}$ and $\Theta_2 = (\theta_{p_2}, ..., \theta_p)^{'}$ and $\Theta = (\theta_1, ..., \theta_p)^{'}$. Also, $X = (X_1, X_2)$, where X_1 is the $T \times p_1$ matrix and X_2 is the $T \times p2$ matrix,

$$X = \begin{bmatrix} y_0 & y_{-1} & \cdots & y_{1-p} \\ y_1 & y_0 & \cdots & y_{2-p} \\ \vdots & \vdots & \vdots & \vdots \\ y_{T-1} & y_{T-2} & \cdots & y_{T-p} \end{bmatrix} = \begin{bmatrix} X_0^{'} \\ X_1^{'} \\ \vdots \\ X_{T-1}^{'} \end{bmatrix}$$

where $X_t = (y_t, ..., y_{t-p+1})^{'}$ and $t = 0, ..., T-1$.

$$X_1 = \begin{bmatrix} y_0 & y_{-1} & \cdots & y_{1-p_1} \\ y_1 & y_0 & \cdots & y_{2-p_1} \\ \vdots & \vdots & \vdots & \vdots \\ y_{T-1} & y_{T-2} & \cdots & y_{T-p_1} \end{bmatrix} = \begin{bmatrix} X_{1,0}^{'} \\ X_{1,1}^{'} \\ \vdots \\ X_{1,T-1}^{'} \end{bmatrix}$$

where $X_{1,t} = (y_t, ..., y_{t-p_1+1})^{'}$ and $t = 0, ..., T-1$.

$$X_2 = \begin{bmatrix} y_{-p_1} & \cdots & y_{1-p} \\ y_{1-p_1} & \cdots & y_{2-p} \\ \vdots & \vdots & \vdots \\ y_{(T-p_1-1)} & \cdots & y_{T-p} \end{bmatrix} = \begin{bmatrix} X_{2,0}^{'} \\ X_{2,1}^{'} \\ \vdots \\ X_{2,T-1}^{'} \end{bmatrix}$$

where $X_{2,t} = (y_{t-p_1}, ..., y_{t-p+1})^{'}$ and $t = 0, ..., T-1$.

The ordinary least squares(OLS) estimate $\hat{\Theta}_F$ based on the full model (3.2.7) are

$$\begin{aligned} \hat{\Theta}_F =& (X^{'}X)^{-1}X^{'}Y \\ =& (X^{'}X)^{-1}X^{'}(X\Theta + \varepsilon) \\ =& \Theta + \left(\sum_{t=1}^{T} X_{t-1}X_{t-1}^{'}\right)^{-1}\left(\sum_{t=1}^{T} X_{t-1}\varepsilon_t\right) \end{aligned} \tag{3.2.8}$$

where the subscript F indicates the use of the full model (3.2.7).

The asymptotic behavior of $\hat{\Theta}_F$ is given by Brockwell and Davis (1986).

Theorem 3.2.1. *(Brockwell and Davis (1986))*

Under Assumption 1, with $\hat{\Theta}_E$ defined as in (3.2.8)*, $\hat{\Theta}_F$ is asymptotically normal distributed (AN), i.e., AN$(\Theta, \frac{\sigma^2}{T}\Gamma_p^{-1})$, where Γ_p is the $p \times p$ covariance matrix with $(\Gamma_p)_{i,j} = \gamma_{|i-j|}$ and $\gamma_{|i-j|}$ is the autocovariance with order $|i-j|$.*

Similarly, we can also write (3.2.2) in regression form as

$$Y = X_2\Theta_2 + \eta, \tag{3.2.9}$$

where $\eta = X_1\Theta_1 + \varepsilon$. The OLS estimate $\hat{\Theta}_{2,S}$ in the simplified model (3.2.9) are

$$\begin{aligned}
\hat{\Theta}_{2,S} =& (X_2'X_2)^{-1}X_2'Y \\
=& (X_2'X_2)^{-1}X_2'(X\Theta + \varepsilon) \\
=& (X_2'X_2)^{-1}X_2'(X_1\Theta_1 + X_2\Theta_2 + \varepsilon) \\
=& \Theta_2 + (X_2'X_2)^{-1}X_2'X_1\Theta_1 + (X_2'X_2)^{-1}X_2'\varepsilon \\
=& \Theta_2 + \left(\sum_{t=1}^{T} X_{2,t-1}X_{2,t-1}'\right)^{-1}\left(\sum_{t=1}^{T} X_{2,t-1}X_{1,t-1}'\Theta_1\right) + \left(\sum_{t=1}^{T} X_{2,t-1}X_{2,t-1}'\right)^{-1}\left(\sum_{t=1}^{T} X_{2,t-1}\varepsilon_t\right)
\end{aligned} \tag{3.2.10}$$

where the subscript S indicates the use of the simplified model (3.2.9).

The following lemma contains some preliminary results, which lead to the limit theory for $\hat{\Theta}_{2,S}$.

Lemma 3.2.1. *Under Assumption 1 as $T \to \infty$,*

(i) $\frac{1}{T}X_2'X_2 \xrightarrow{P} \Gamma_{p_2}$;

(ii) $\frac{1}{T}X_2'X_1 \xrightarrow{P} R$;

(iii) $\frac{1}{T}X_2'\varepsilon\varepsilon'X_2 \xrightarrow{P} \sigma^2\Gamma_{p_2}$

where $\xrightarrow{P}$ means convergence in probability. Γ_{p_2} is the $p_2 \times p_2$ covariance matrix with $(\Gamma_{p_2})_{i,j} = \gamma_{|i-j|}$ and $\gamma_{|i-j|}$ is the autocovariance with order $|i-j|$. R is defined as the $p_2 \times p_1$ matrix of

$$R = \begin{bmatrix} \gamma_{p_1} & \cdots & \gamma_1 \\ \gamma_{(p_1+1)} & \cdots & \gamma_2 \\ \vdots & \vdots & \vdots \\ \gamma_{p-1} & \cdots & \gamma_{p_2} \end{bmatrix}$$

The proof of Lemma 1 is given in the Appendix. The following results follows directly from Lemma 3.2.1.

Theorem 3.2.2. *Under Assumption 1, with* $\hat{\Theta}_{2,S}$ *defined as in* (3.2.10)*,* $\hat{\Theta}_{2,S}$ *is* $AN(\Theta_2+\Gamma_{p_2}^{-1}R\Theta_1, \frac{\sigma^2}{T}\Gamma_{p_2}^{-1})$*, where* Γ_{p_2} *and R are defined in Lemma 3.2.1.*

3.2.2 MSE-based comparison of model predictions

Here we consider to predict y_{T+1} given $\{y_T, ..., y_{T-p+1}\}$ using the FM and SM estimation and then compare their conditional mean squared errors of prediction.

Under Theorem 2.1 and 2.2, we can write

$$\hat{\Theta}_F = \Theta + o_p(\frac{1}{\sqrt{T}})$$

$$\hat{\Theta}_{2,S} = \Theta_2 + \Gamma_{p_2}^{-1}R\Theta_1 + o_p(\frac{1}{\sqrt{T}})$$

Here two types of model predictions are considered

1. SM prediction: $\hat{y}_s = Y_{p2}^{'}\hat{\Theta}_{2,S}$
2. EM prediction: $\hat{y}_e = Y_p^{'}\hat{\Theta}_E$

Here $\hat{y}_s$ and $\hat{y}_f$ are predictors of y_{T+1} by the simplified model and the true model separately. Let $Y_p = (y_T, ..., y_{T-p+1})^{'}$, $Y_{p_1} = (y_T, ..., y_{T-p_1+1})^{'}$ and $Y_{p_2} = (y_{T-p_1}, ..., y_{T-p+1})^{'}$. Now we can compute the conditional mean squared errors(MSE) of predictions of y_{T+1}.

$$MSE(\hat{y}_f | y_T, ..., y_{T-p+1}) = E\{(Y_p^{'}\hat{\Theta}_F - Y_p^{'}\Theta - \varepsilon_{T+1})^2 | y_T, ..., y_{T-p+1}\}$$

$$= \sigma^2 + \frac{\sigma^2}{T}Y_p^{'}\Gamma_p^{-1}Y_p + o(1)$$

$$\approx \sigma^2 + \frac{\sigma^2}{T}Y_p^{'}\Gamma_p^{-1}Y_p$$

$$MSE(\hat{y}_s|y_T,...,y_{T-p+1}) = E\{(Y_{p_2}^{'}\hat{\Theta}_{2,S} - Y_p^{'}\Theta - \varepsilon_{T+1})^2|y_T,...,y_{T-p+1}\}$$

$$= E\{(Y_{p_2}^{'}\hat{\Theta}_{2,S} - Y_{p_1}^{'}\Theta_1 - Y_{p_2}^{'}\Theta_2 - \varepsilon_{T+1})^2|y_T,...,y_{T-p+1}\}$$

$$= \sigma^2 + \frac{\sigma^2}{T}Y_{p_2}^{'}\Gamma_{P_2}^{-1}Y_{p_2} + (Y_{p_1} - R^{'}\Gamma_{p_2}^{-1}Y_{p_2})^{'}\Theta_1\Theta_1^{'}(Y_{p_1} - R^{'}\Gamma_{p_2}^{-1}Y_{p_2}) + o(1)$$

$$\approx \sigma^2 + \frac{\sigma^2}{T}Y_{p_2}^{'}\Gamma_{P_2}^{-1}Y_{p_2} + (Y_{p_1} - R^{'}\Gamma_{p_2}^{-1}Y_{p_2})^{'}\Theta_1\Theta_1^{'}(Y_{p_1} - R^{'}\Gamma_{p_2}^{-1}Y_{p_2})$$

The difference between the conditional mean squared errors of $\hat{y}_f$ and $\hat{y}_s$ is

$$MSE(\hat{y}_f|y_T,...,y_{T-p+1}) - MSE(\hat{y}_s|y_T,...,y_{T-p+1})$$

$$= \frac{\sigma^2}{T}(Y_p^{'}\Gamma_p^{-1}Y_p - Y_{p_2}^{'}\Gamma_{p_2}^{-1}Y_{p_2}) - (Y_{p_1} - R^{'}\Gamma_{p_2}^{-1}Y_{p_2})^{'}\Theta_1\Theta_1^{'}(Y_{p_1} - R^{'}\Gamma_{p_2}^{-1}Y_{p_2})$$

This difference is positive if

$$\frac{\sigma^2}{T}(Y_p^{'}\Gamma_p^{-1}Y_p - Y_{p_2}^{'}\Gamma_{p_2}^{-1}Y_{p_2}) - (Y_{p_1} - R^{'}\Gamma_{p_2}^{-1}Y_{p_2})^{'}\Theta_1\Theta_1^{'}(Y_{p_1} - R^{'}\Gamma_{p_2}^{-1}Y_{p_2}) \geq 0 \qquad (3.2.11)$$

Lemma 3.2.2. *Given any vectors $x, y \in \Re^n$, $x^{'}yy^{'}x \leq x^{'}xy^{'}y$.*

The proof of Lemma 3.2.2 is in the Appendix.

Define the 2-norm $\|\Theta_1\| = \sqrt{\Theta_1^{'}\Theta_1} = \sqrt{\sum_{i=1}^{p_1}\theta_i^2}$. Under Lemma 3.2.2, a sufficient condition for (3.2.11) is

$$\|\Theta_1\|^2 < \frac{\frac{\sigma^2}{T}(Y_p^{'}\Gamma_p^{-1}Y_p - Y_{p_2}^{'}\Gamma_{p_2}^{-1}Y_{p_2})}{(Y_{p_1} - R^{'}\Gamma_{p_2}^{-1}Y_{p_2})^{'}(Y_{p_1} - R^{'}\Gamma_{p_2}^{-1}Y_{p_2})} \qquad (3.2.12)$$

where the 2-norm $\|\Theta_1\| = \sqrt{\sum_{i=1}^{p_1}\theta_i^2}$. This inequality holds when the SM gives better prediction than FM in the sense of smaller mean-squared-errors.

Define the critical value C_p with the form

$$C_p = \sqrt{\frac{\frac{\sigma^2}{T}(Y_p'\Gamma_p^{-1}Y_p - Y_{p_2}'\Gamma_{p_2}^{-1}Y_{p_2})}{(Y_{p_1} - R'\Gamma_{p_2}^{-1}Y_{p_2})'(Y_{p_1} - R'\Gamma_{p_2}^{-1}Y_{p_2})}}.$$

So the inequality (3.2.12) is equivalent of the following inequality

$$\|\Theta_1\| < C_p \tag{3.2.13}$$

If the 2-norm of Θ_1 is less than C_p, the simplified model should be used for predictions in the sense of smaller mean squared errors.

In the real cases, the parameters are unknown, so we must use estimates of parameters to determine if we should use simplified models. Taking the estimate bias into account, we can consider a similar condition with (3.2.13) by replacing the true parameters with estimates and as well as some error correction. Let $\hat{\varepsilon}_t = y_t - \hat{y}_t$, where $\hat{y}_t$ is the fitted value by the full model. To estimate σ^2, we use $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(\hat{\varepsilon}_t - \bar{\varepsilon})^2$, where $\bar{\varepsilon}$ is define as $\bar{\varepsilon} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t$. In order to compute $\hat{\Gamma}_p, \hat{\Gamma}_{p_2}, \hat{R}$, we use the consistent estimates of $\gamma_{|i-j|}$ as

$$\hat{\gamma}_{|i-j|} = \frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})(y_{(t-|i-j|)} - \bar{y}), \quad and \quad \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t.$$

So the following condition totally based on the observed data can be used to make selections between the full models and the simplified ones.

$$\left\|\hat{\Theta}_1\right\| < \tau\hat{C}_p \tag{3.2.14}$$

Where $0 < \tau \leq 1$ and

$$\hat{C}_p = \sqrt{\frac{\frac{\hat{\sigma}^2}{T}(Y_p'\hat{\Gamma}_p^{-1}Y_p - Y_{p_2}'\hat{\Gamma}_{p_2}^{-1}Y_{p_2})}{(Y_{p_1}' - Y_{p_2}'\hat{\Gamma}_{p_2}^{-1}\hat{R})(Y_{p_1}' - Y_{p_2}'\hat{\Gamma}_{p_2}^{-1}\hat{R})'}} \tag{3.2.15}$$

For the constant τ, we will talk about more in section 4.

3.2.3 Example 1. $p_1 = 1$

When $p_1 = 1$, that means we believe θ_1 is close to zero. The simplified model can be written as

$$y_t = \theta_2 y_{t-2} + ... + \theta_p y_{t-p} + \eta_t$$

where $\eta_t = \theta_1 y_{t-1} + \varepsilon_t$ is the stochastic component associated with the model mismatch.

Writing (3.2.1) in the regression form as

$$y_t = X_{t-1}^{'}\theta + \varepsilon_t,\ t = 1, ..., T,$$

where $X_{t-1} = (y_{t-1}, y_{t-2}, ..., y_{t-p})^{'}$, the OLS estimates are

$$\hat{\theta}_F = \theta + \left(\sum_{t=1}^{T} X_{t-1} X_{t-1}^{'} \right)^{-1} \left(\sum_{t=1}^{T} X_{t-1} \varepsilon_t \right) \tag{3.2.16}$$

where the subscript F indicates the use of the truly structured extended model.

Also we write the above simplified model in the regression form as

$$y_t = \tilde{X}_{t-1}^{'}\tilde{\theta} + \eta_t,\ t = 1, ..., T,$$

where $\tilde{X}_{t-1} = (y_{t-2}, ..., y_{t-p})^{'}$ and $\tilde{\theta} = (\theta_2, ..., \theta_p)^{'}$, and the OLS estimates are

$$\hat{\theta}_S = \tilde{\theta} + \theta_1 \left(\sum_{t=1}^{T} \tilde{X}_{t-1} \tilde{X}_{t-1}^{'} \right)^{-1} \left(\sum_{t=1}^{T} \tilde{X}_{t-1} y_{t-1} \right) + \left(\sum_{t=1}^{T} \tilde{X}_{t-1} \tilde{X}_{t-1}^{'} \right)^{-1} \left(\sum_{t=1}^{T} \tilde{X}_{t-1} \varepsilon_t \right). \tag{3.2.17}$$

Under Theorem 2.1 and 2.2, we can write

$$\hat{\theta}_F = \theta + o_p(\frac{1}{\sqrt{T}})$$

$$\hat{\theta}_S = \tilde{\theta} + \theta_1 \Gamma_{p-1}^{-1} R_{p-1} + o_p(\frac{1}{\sqrt{T}})$$

where R_{p-1} is defined as the vector of $(\gamma_1, ..., \gamma_{p-1})'$.

Now we can compute the mean squared errors of predictions of y_{T+1} as

$$MSE(\hat{y}_e | y_T, ..., y_{T-p+1}) \approx \sigma^2 + \frac{\sigma^2}{T} Y_p' \Gamma_p^{-1} Y_p$$

$$MSE(\hat{y}_s | y_T, ..., y_{T-p+1}) \approx \sigma^2 + \frac{\sigma^2}{T} Y_{p-1}' \Gamma_{p-1}^{-1} Y_{p-1} + \theta_1^2 (y_T - Y_{p-1}' \Gamma_{p-1}^{-1} R_{p-1})^2$$

The MSE difference for $\hat{y}_f$ and $\hat{y}_s$ is

$$MSE(\hat{y}_f | y_T, ..., y_{T-p+1}) - MSE(\hat{y}_s | y_T, ..., y_{T-p+1}) =$$

$$\frac{\sigma^2}{T} (Y_p' \Gamma_p^{-1} Y_p - Y_{p-1}' \Gamma_{p-1}^{-1} Y_{p-1}) - \theta_1^2 (y_T - Y_{p-1}' \Gamma_{p-1}^{-1} R_{p-1})^2$$

This difference is positive if

$$|\theta_1| < C_p \tag{3.2.18}$$

where

$$C_p = \sqrt{\frac{\sigma^2 (Y_{p-1}' \Gamma_{p-1}^{-1} Y_{p-1} - Y_{p-1}' \Gamma_{p-1}^{-1} Y_{p-1})}{T (y_T - Y_{p-1}' \Gamma_{p-1}^{-1} R_{p-1})^2}}.$$

If the inequality (3.2.18) holds, the SM will be better than EM in the sense of smaller mean-squared-errors. Note that, since Θ_1 here is of dimension 1, we don't have to use Lemma 2.2 to get

(3.2.18). Therefore, the inequality (3.2.18) is a necessary and sufficient condition for selections between the full model and the simplified model when $p_1 = 1$.

3.3 Simplified Models-AR(p) under Heteroskedasticity

The work here is still based $AR(p)$ models as defined in (3.2.1), but the innovations are heteroscedastic and distributed according to

$$\varepsilon_t = g(\frac{t}{T})u_t, \tag{3.3.1}$$

where $g(\cdot)$ is an unknown non-negative scale function and u_t is stationary with mean 0. Model heteroskedasticity is characterized in $\sigma_t = g(\frac{t}{T})$ as being systematically dependent on the relative position of the observation in terms of the scale function $g(\cdot)$. Since $g(\cdot)$ is taken to be unknown, the formulation is non-parametric. This includes cases where the conditional error variance evolves over time, slowly transitions, or abruptly or periodically changes across the sample.

Assumption 2.(Phillips and Xu (2006))

(i) All the roots of the polynomial $1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_p z^p = 0$ lie outside the unit circle.

(ii) $g(\cdot)$ is non-stochastic, measurable and uniformly bounded on the interval $(-\infty, 1])$, with a finite number of points of discontinuity, $g(\cdot) > 0$ and satisfies a Lipschitz condition except at points of discontinuity.

(iii) $\{u_t\}$ is a strong mixing (α-mixing) martingale difference process with $E(u_t|F_{t-1}) = 0$, $E(u_t^2|F_{t-1}) = 1$, a.s., for all t, with the natural filtration $F_t = \sigma(u_s, s \leq t)$. There exist $\delta > 1$ and $C > 0$, such that $sup_t E u_t^{4\delta} < C < \infty$.

Under Assumption 2(ii), we require the definition of $g(r)$ for $r < 0$ since initial conditions

are in the infinite past and the MA(∞) representation of y_t. Also, $g(\cdot)$ is integrable on the interval [0, 1] up to any finite order, $\int g^m(r)dr = \int g^m$.

3.3.1 Properties of parameter estimates

The limit properties of the OLS estimates of Θ in the full model have been studied by Phillips and Xu (2006). We will use the same simplified model as (3.2.2) and study the asymptotic behavior of the OLS estimates for simplified model later.

Recall the Wold representation (3.2.3) of y_t, we now define Ω_p to be the $p \times p$ matrix with $(i, j) - th$ element $\omega_{|i-j|}$, where

$$\omega_k = \sum_{i=0}^{\infty} \alpha_i \alpha_{i+k},$$

for $k = 0, 1, \ldots, p-1$.

Theorem 3.3.1. *(Phillips and Xu 2006)*
Under Assumption 2, with $\hat{\Theta}_E$ *defined as in* (3.2.8)*,* $\hat{\Theta}_F$ *is* $AN(\Theta, \Sigma)$*,*
where $\Sigma = \frac{\int g^4}{T(\int g^2)^2}\Omega_p^{-1}$.

Lemma 3.3.1 contains some preliminary results, which lead to the limit theory for $\Theta_{2,S}$ under the simplified model.

Lemma 3.3.1. *Under the Assumption 3 as* $T \to \infty$*,*

(a) $\frac{1}{T}\sum_{t=1}^{T} X_{2,t-1} \xrightarrow{P} 0$*;*

(b) $\frac{1}{T}X_2'X_2 \xrightarrow{P} (\int g^2)\Omega_{p_2}$*;*

(c) $\frac{1}{T}X_2'\varepsilon\varepsilon' X_2 \xrightarrow{P} (\int g^4)\Omega_{p_2}$*;*

(d) $\frac{1}{T}X_2'X_1 \xrightarrow{P} (\int g^2)\Omega_1$*;*

(e) $\frac{1}{T}X_2'\varepsilon \xrightarrow{d} N(0,(\int g^4)\Omega_{p_2})$;

where $\xrightarrow{P}$ *means convergence in probability.* Ω_1 *is defined as the* $p_2 \times p_1$ *matrix of*

$$\Omega_1 = \begin{bmatrix} \omega_{p_1} & \cdots & \omega_1 \\ \omega_{(p_1+1)} & \cdots & \omega_2 \\ \vdots & \vdots & \vdots \\ \omega_{p-1} & \cdots & \omega_{p_2} \end{bmatrix}$$

Ω_{p_2} *is the* $p_2 \times p_2$ *matrix with* $(i,j)-th$ *element* $\omega_{|i-j|}$, *where*

$$\omega_k = \sum_{i=0}^{\infty} \alpha_i \alpha_{i+k},$$

for $k = 0,1,\ldots,p_2-1$.

The proof of Lemma 3.3.1 is given in the Appendix. The following theorem follows directly from Lemma 3.3.1.

Theorem 3.3.2. *Under Assumption 2 and 3, with* $\hat{\Theta}_{2,S}$ *defined as in* (3.2.10)*,* $\hat{\Theta}_{2,S}$ *is* $AN(\Theta_2 + \Omega_{p_2}^{-1}\Omega_1\Theta_1, \tilde{\Sigma})$, *where* $\tilde{\Sigma} = \frac{\int g^4}{T(\int g^2)^2}\Omega_{p_2}^{-1}$.

3.3.2 MSE-based comparison of model predictions

By the similar way, the conditional MSE difference for $\hat{y}_f$ and $\hat{y}_s$ is

$$MSE(\hat{y}_f|y_T,...,y_{T-p+1}) - MSE(\hat{y}_s|y_T,...,y_{T-p+1})$$

$$= \frac{\int g^4}{T(\int g^2)^2}(Y_p'\Omega_p^{-1}Y_p - Y_{p_2}'\Omega_{p_2}^{-1}Y_{p_2}) - (Y_{p_1} - \Omega_1'\Omega_{p_2}^{-1}Y_{p_2})'\Theta_1\Theta_1'(Y_{p_1} - \Omega_1'\Omega_{p_2}^{-1}Y_{p_2})$$

Here $\hat{y}_s$ and $\hat{y}_f$ are predictions of y_{T+1} by the simplified model and the full model. And $Y_p = (y_T,...,y_{T-p+1})'$ and $Y_{p_1} = (y_T,...,y_{T-p_1+1})'$ and $Y_{p_2} = (y_{T-p_1},...,y_{T-p+1})'$.

This difference is positive if

$$\frac{\int g^4}{T(\int g^2)^2}(Y_p^{'}\Omega_p^{-1}Y_p - Y_{p_2}^{'}\Omega_{p_2}^{-1}Y_{p_2}) - (Y_{p_1} - \Omega_1^{'}\Omega_{p_2}^{-1}Y_{p_2})^{'}\Theta_1\Theta_1^{'}(Y_{p_1} - \Omega_1^{'}\Omega_{p_2}^{-1}Y_{p_2}) \geq 0 \quad (3.3.2)$$

By Lemma 2.2, a sufficient condition for (3.3.2) is that

$$\|\Theta_1\|^2 < \frac{\frac{\int g^4}{T(\int g^2)^2}(Y_p^{'}\Omega_p^{-1}Y_p - Y_{p_2}^{'}\Omega_{p_2}^{-1}Y_{p_2})}{(Y_{p_1} - \Omega_1^{'}\Omega_{p_2}^{-1}Y_{p_2})^{'}(Y_{p_1} - \Omega_1^{'}\Omega_{p_2}^{-1}Y_{p_2})} \quad (3.3.3)$$

If this inequality holds, the SM will give better predictions than FM in the sense of smaller mean squared errors.

Define the critical value C_p with the form

$$C_p = \sqrt{\frac{\frac{\int g^4}{T(\int g^2)^2}(Y_p^{'}\Omega_p^{-1}Y_p - Y_{p_2}^{'}\Omega_{p_2}^{-1}Y_{p_2})}{(Y_{p_1} - \Omega_1^{'}\Omega_{p_2}^{-1}Y_{p_2})^{'}(Y_{p_1} - \Omega_1^{'}\Omega_{p_2}^{-1}Y_{p_2})}}.$$

$$\|\Theta_1\| < C_p \quad (3.3.4)$$

If $\|\Theta_1\|$ is less than C_p, the simplified model should be used for predictions in the sense of smaller mean-squared errors. In the real case, we must use estimations of parameters to determine if we should use simplified models. Just like what we did in section 2 for the stationary case, we also consider a similar condition here of (3.3.4) by replacing the true parameters with estimates and as well as some error correction. See (3.3.5).

$$\left\|\hat{\Theta}_1\right\| < \tau\hat{C}_p \quad (3.3.5)$$

Where $0 < \tau \leq 1$ and

$$\hat{C}_p = \sqrt{\frac{Y_p'\hat{\Sigma}_p Y_p - Y_{p_2}'\hat{\Sigma}_{p_2} Y_{p_2}}{T(Y_{p_1} - \hat{\Omega}_1'\hat{\Omega}_{p_2}^{-1}Y_{p_2})'(Y_{p_1} - \hat{\Omega}_1'\hat{\Omega}_{p_2}^{-1}Y_{p_2})}} \tag{3.3.6}$$

where

$$\hat{\Sigma}_p = \left(\sum_{t=1}^{T} X_{t-1}X_{t-1}'\right)^{-1}\left(\sum_{t=1}^{T} \hat{\varepsilon}_t^2 X_{t-1}X_{t-1}'\right)\left(\sum_{t=1}^{T} X_{t-1}X_{t-1}'\right)^{-1},$$

$$\hat{\Sigma}_{p_2} = \left(\sum_{t=1}^{T} X_{2,t-1}X_{2,t-1}'\right)^{-1}\left(\sum_{t=1}^{T} \hat{\varepsilon}_t^2 X_{2,t-1}X_{2,t-1}'\right)\left(\sum_{t=1}^{T} X_{2,t-1}X_{2,t-1}'\right)^{-1},$$

$$\hat{\Omega}_1'\hat{\Omega}_{p_2}^{-1} = \left(\sum_{t=1}^{T} X_{1,t-1}X_{2,t-1}'\right)\left(\sum_{t=1}^{T} X_{2,t-1}X_{2,t-1}'\right)^{-1},$$

where $\hat{\varepsilon}_t = y_t - \hat{y}_{t,f}$ and $\hat{y}_{t,f}$ is the fitted value at time t by the full model. Based on Lemma 3.1 and Lemma 2 in Phillips and Xu (2006), it is easy to prove that $\hat{\Gamma}_p$, $\hat{\Gamma}_{p_2}$ and $\hat{\Omega}_1'\hat{\Omega}_{p_2}^{-1}$ are consistent estimates of Γ_p, Γ_{p_2} and $\Omega_1'\Omega_{p_2}^{-1}$.

3.4 Strategy for Assessing Uncertainty about Which Model is Better

Here we will give some empirical simulation examples to illustrate the the selection of models between the full models and simplified models in different situations. There is no theoretical strategy now to select τ. we try to assign some reasonable values to τ and check their performance in the empirical simulation.

3.4.1 Illustrative examples

Example 1:

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \varepsilon_t$$

where $\varepsilon_t \sim i.i.d.N(0,\sigma^2)$. $\theta_2 = 0.5$. θ_1 will take the values of $\{0.4, 0.1, 0.01, 0.001\}$. Generate 1000 datasets using the above AR(2) and each dataset is with sample size n. n can take values of $\{50, 100, 500, 750, 1000\}$. σ^2 can be values of $\{0.1, 1, 10, 50\}$.

Example 2:

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + g(\frac{t}{T})u_t,$$

$$g(r)^2 = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)r^m, \;\; r \in [0,1]$$

where $u_t \sim i.i.d.N(0,1)$. Here the variance of the errors changes continuously from σ_0^2 to σ_1^2. Define $\delta = \sigma_1^2/\sigma_0^2$. In the simulation, we set $m = 2$ and $\delta \in \{0.2, 5\}$, so that positive $(\delta > 1)$ and negative $(\delta < 1)$ trending variances are allowed. Without loss of generality, we set $\sigma_0^2 = 1$ in all cases. The true value of θ_1 is taken from the set $\{0.4, 0.1, 0.05, 0.01, 0.001\}$ and $\theta_2 = 0.5$. Generate 1000 datasets using the above AR(2) and each dataset is with sample size $n \in \{50, 100, 200, 500\}$.

Here four fixed values $\{0.5, 0.8, 0.9, 1\}$ will be assigned to τ separately. Another case where τ is a function of the sample size n will also be considered in the simulation, i.e., $\tau = 1 - \frac{1}{\sqrt{n}}$.

3.4.2 Results and discussions

The effect of different sample sizes to the performance of SMs are recorded in Table 3.1-3.4 and 3.9-18. We can find that when the sample size is small(less than 100), SMs give smaller mean squared prediction errors. In addition, if θ_1 is getting smaller and the data with same size and sample standard deviations, SMs work better than FMs. In Table 3.5-3.8, SMs

outperform FMs when σ^2 is big under the condition that θ_1 is small enough. Also, the smaller θ_1 is, the better the performance of SMs is. For the selection of the value of τ, $\tau = 1 - \frac{1}{\sqrt{n}}$ works in practice. So we can just assign some values to τ as a function of the sample size n. To sum up, there are three factors to decide when SMs work: small sample size, some relatively small coefficient(s) and big variances.

Table 3.1: Performance of different sample size with $\theta_1 = 0.4$, $\sigma^2 = 1$, $\theta_2 = 0.5$

n	20	50	100	500	750	1000
$E(\lvert\hat{\theta}_1\rvert)$	0.3793	0.3914	0.3935	0.3977	0.3979	0.3985
C_p	0.1937	0.1225	0.0866	0.0387	0.0316	0.0274
$\hat{C}_p$	0.2153	0.1263	0.0876	0.0388	0.0317	0.0274
MSE_E	1.1392	1.0532	1.0253	1.0018	1.0009	1.0014
MSE_S	1.2767	1.2502	1.2257	1.2299	1.1978	1.1948
MSE($\tau = 1$)	1.1392	1.0532	1.0253	1.0018	1.0009	1.0014
MSE($\tau = 0.9$)	1.1392	1.0532	1.0253	1.0018	1.0009	1.0014
MSE($\tau = 0.8$)	1.1392	1.0532	1.0253	1.0018	1.0009	1.0014
MSE($\tau = 0.5$)	1.1392	1.0532	1.0253	1.0018	1.0009	1.0014
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	1.1392	1.0532	1.0253	1.0018	1.0009	1.0014

Table 3.2: Performance of different sample size with $\theta_1 = 0.1$, $\sigma^2 = 1$, $\theta_2 = 0.5$

n	20	50	100	500	750	1000
$E(\lvert\hat{\theta}_1\rvert)$	0.0882	0.0961	0.0957	0.0975	0.0976	0.0985
C_p	0.1937	0.1225	0.0866	0.0387	0.0316	0.0274
$\hat{C}_p$	0.2148	0.1280	0.0887	0.0389	0.0317	0.0274
MSE_E	1.1249	1.0507	1.0261	1.0019	1.0013	1.0016
MSE_S	1.0720	1.0413	1.0297	1.0146	1.0134	1.0122
MSE($\tau = 1$)	1.0720	1.0413	1.0261	1.0019	1.0013	1.0016
MSE($\tau = 0.9$)	1.0720	1.0413	1.0261	1.0019	1.0013	1.0016
MSE($\tau = 0.8$)	1.0720	1.0413	1.0261	1.0019	1.0013	1.0016
MSE($\tau = 0.5$)	1.0720	1.0507	1.0261	1.0019	1.0013	1.0016
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	1.0720	1.0413	1.0261	1.0019	1.0013	1.0016

Table 3.3: Performance of different sample size with $\theta_1 = 0.01$, $\sigma^2 = 1$, $\theta_2 = 0.5$

n	20	50	100	500	750	1000
$E(\lvert\hat{\theta}_1\rvert)$	0.00096	0.0082	0.0066	0.0075	0.0076	0.0086
C_p	0.1937	0.1225	0.0866	0.0387	0.0316	0.0274
$\hat{C}_p$	0.2139	0.1280	0.0887	0.0389	0.0317	0.0275
MSE_E	1.1263	1.0507	1.0258	1.0019	1.0013	1.0016
MSE_S	1.0631	1.0281	1.0167	1.0003	1.0004	1.0005
MSE($\tau = 1$)	1.0631	1.0281	1.0167	1.0003	1.0004	1.0005
MSE($\tau = 0.9$)	1.0631	1.0281	1.0167	1.0003	1.0004	1.0005
MSE($\tau = 0.8$)	1.0631	1.0281	1.0167	1.0003	1.0004	1.0005
MSE($\tau = 0.5$)	1.0631	1.0281	1.0167	1.0003	1.0004	1.0016
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	1.0631	1.0281	1.0167	1.0003	1.0004	1.0005

Table 3.4: Performance of different sample size with $\theta_1 = 0.001$, $\sigma^2 = 1$, $\theta_2 = 0.5$

n	20	50	100	500	750	1000
$E(\lvert\hat{\theta}_1\rvert)$	0.0078	0.0008	0.0024	0.0015	0.0014	0.0004
C_p	0.1936	0.1225	0.0866	0.0387	0.0316	0.0274
$\hat{C}_p$	0.2138	0.1279	0.0887	0.0389	0.0317	0.0275
MSE_E	1.1265	1.0507	1.0257	1.0021	1.0013	1.0016
MSE_S	1.0634	1.0280	1.0166	1.0001	1.0002	1.0004
MSE($\tau = 1$)	1.0634	1.0280	1.0166	1.0001	1.0002	1.0004
MSE($\tau = 0.9$)	1.0634	1.0280	1.0166	1.0001	1.0002	1.0004
MSE($\tau = 0.8$)	1.0634	1.0280	1.0166	1.0001	1.0002	1.0004
MSE($\tau = 0.5$)	1.0634	1.0280	1.0166	1.0001	1.0002	1.0004
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	1.0634	1.0280	1.0166	1.0001	1.0002	1.0004

3.5 Conclusions

In this Chapter, we focus on the situations where the simplified models can be used for autoregressive processes. Simplified models have many appealing properties and sometimes give better parameter estimates and model predictions, in sense of mean-squared-error, than the full models, especially when the data are not informative or with a big variation or some parameters are relatively small.

Table 3.5: Performance of different σ^2 with $\theta_1 = 0.4$, $N = 100$, $\theta_2 = 0.5$

σ^2	0.1	1	5	10	50
$E(\lvert\hat{\theta}_1\rvert)$	0.3935	0.3935	0.3935	0.3935	0.3935
C_p	0.0866	0.0866	0.0866	0.0866	0.0866
$\hat{C}_p$	0.0876	0.0876	0.0876	0.0876	0.0876
MSE_E	0.1025	1.0252	5.1262	10.2525	51.2624
MSE_S	0.1226	1.2257	6.1286	12.2572	61.2862
MSE($\tau = 1$)	0.1025	1.0252	5.1262	10.2525	51.2624
MSE($\tau = 0.9$)	0.1025	1.0252	5.1262	10.2525	51.2624
MSE($\tau = 0.8$)	0.1025	1.0252	5.1262	10.2525	51.2624
MSE($\tau = 0.5$)	0.1025	1.0252	5.1262	10.2525	51.2624
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	0.1025	1.0252	5.1262	10.2525	51.2624

Table 3.6: Performance of different σ^2 with $\theta_1 = 0.1$, $N = 100$, $\theta_2 = 0.5$

σ^2	0.1	1	5	10	50
$E(\lvert\hat{\theta}_1\rvert)$	0.0957	0.0957	0.0957	0.0957	0.0957
C_p	0.0866	0.0866	0.0866	0.0866	0.0866
$\hat{C}_p$	0.0887	0.0887	0.0887	0.0887	0.0887
MSE_E	0.1026	1.0261	5.1306	10.2612	51.3061
MSE_S	0.1030	1.0297	5.1484	10.2968	51.4841
MSE($\tau = 1$)	0.1026	1.0261	5.1306	10.2612	51.3061
MSE($\tau = 0.9$)	0.1026	1.0261	5.1306	10.2612	51.3061
MSE($\tau = 0.8$)	0.1026	1.0261	5.1306	10.2612	51.3061
MSE($\tau = 0.5$)	0.1026	1.0261	5.1306	10.2612	51.3061
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	0.1026	1.0261	5.1306	10.2612	51.3061

3.6 Acknowledgements

Chapter 3, in full is currently being prepared for submission for publication of the material. Chen, Jie; Politis, Dimitris N. The dissertation author was the primary investigator and author of this paper.

Table 3.7: Performance of different σ^2 with $\theta_1 = 0.01$, $N = 100$, $\theta_2 = 0.5$

$E(\lvert\hat{\theta}_1\rvert)$	0.0065	0.0065	0.0065	0.0065	0.0065
C_p	0.0866	0.0866	0.0866	0.0866	0.0866
$\hat{C}_p$	0.0887	0.0887	0.0887	0.0887	0.0887
σ^2	0.1	1	5	10	50
MSE_E	0.1026	1.0258	5.1288	10.2575	51.2877
MSE_S	0.1017	1.0167	5.0834	10.1669	50.8343
MSE($\tau = 1$)	0.1017	1.0167	5.0834	10.1669	50.8343
MSE($\tau = 0.9$)	0.1017	1.0167	5.0834	10.1669	50.8343
MSE($\tau = 0.8$)	0.1017	1.0167	5.0834	10.1669	50.8343
MSE($\tau = 0.5$)	0.1017	1.0167	5.0834	10.1669	50.8343
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	0.1017	1.0167	5.0834	10.1669	50.8343

Table 3.8: Performance of different σ^2 with $\theta_1 = 0.001$, $N = 100$, $\theta_2 = 0.5$

σ^2	0.1	1	5	10	50
$E(\lvert\hat{\theta}_1\rvert)$	0.0024	0.0024	0.0024	0.0024	0.0024
C_p	0.0866	0.0866	0.0866	0.0866	0.0866
$\hat{C}_p$	0.0887	0.0887	0.0887	0.0887	0.0887
MSE_E	0.1026	1.0257	5.1286	10.2572	51.2859
MSE_S	0.1017	1.0166	5.0828	10.1656	50.8279
MSE($\tau = 1$)	0.1017	1.0166	5.0828	10.1656	50.8279
MSE($\tau = 0.9$)	0.1017	1.0166	5.0828	10.1656	50.8279
MSE($\tau = 0.8$)	0.1017	1.0166	5.0828	10.1656	50.8279
MSE($\tau = 0.5$)	0.1017	1.0166	5.0828	10.1656	50.8279
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	0.1017	1.0166	5.0828	10.1656	50.8279

Table 3.9: Performance of different N with $\theta_1 = 0.4$, $\theta_2 = 0.5$, $\delta = 0.2$, $m = 2$

n	50	100	200	500
$E(\lvert\hat{\theta}_1\rvert)$	0.3918	0.3938	0.3951	0.3973
C_p	0.1020	0.0857	0.0609	0.0384
$\hat{C}_p$	0.0883	0.0684	0.0476	0.0322
MSE_E	0.7436	0.7389	0.7362	0.7329
MSE_S	0.7865	0.7825	0.7783	0.7796
MSE($\tau = 1$)	0.7436	0.7389	0.7362	0.7329
MSE($\tau = 0.9$)	0.7436	0.7389	0.7362	0.7329
MSE($\tau = 0.8$)	0.7436	0.7389	0.7362	0.7329
MSE($\tau = 0.5$)	0.7436	0.7389	0.7362	0.7329
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	0.7436	0.7389	0.7362	0.7329

Table 3.10: Performance of different n with $\theta_1 = 0.1$, $\theta_2 = 0.5$, $\delta = 0.2$, $m = 2$

n	50	100	200	500
$E(\lvert\hat{\theta}_1\rvert)$	0.0957	0.0959	0.0955	0.0969
C_p	0.1020	0.0856	0.0609	0.0386
$\hat{C}_p$	0.1145	0.0861	0.0625	0.0401
MSE_E	0.7421	0.7386	0.7360	0.7329
MSE_S	0.7399	0.7392	0.7374	0.7354
MSE($\tau = 1$)	0.7399	0.7386	0.7360	0.7329
MSE($\tau = 0.9$)	0.7399	0.7386	0.7360	0.7329
MSE($\tau = 0.8$)	0.7421	0.7386	0.7360	0.7329
MSE($\tau = 0.5$)	0.7421	0.7386	0.7360	0.7329
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	0.7399	0.7386	0.7360	0.7329

Table 3.11: Performance of different N with $\theta_1 = 0.05$, $\theta_2 = 0.5$, $\delta = 0.2$, $m = 2$

n	50	100	200	500
$E(\lvert\hat{\theta}_1\rvert)$	0.0467	0.0464	0.0456	0.0469
C_p	0.1021	0.0857	0.0609	0.0388
$\hat{C}_p$	0.1170	0.0870	0.0626	0.0403
MSE_E	0.7421	0.7386	0.7360	0.7329
MSE_S	0.7378	0.7371	0.7354	0.7332
MSE($\tau = 1$)	0.7378	0.7371	0.7354	0.7329
MSE($\tau = 0.9$)	0.7378	0.7371	0.7354	0.7329
MSE($\tau = 0.8$)	0.7378	0.7371	0.7354	0.7329
MSE($\tau = 0.5$)	0.7378	0.7386	0.7360	0.7329
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	0.7378	0.7371	0.7354	0.7329

Table 3.12: Performance of different N with $\theta_1 = 0.01$, $\theta_2 = 0.5$, $\delta = 0.2$, $m = 2$

n	50	100	200	500
$E(\lvert\hat{\theta}_1\rvert)$	0.0074	0.0068	0.0058	0.0070
C_p	0.1022	0.0856	0.0611	0.0386
$\hat{C}_p$	0.1173	0.0867	0.0627	0.0401
MSE_E	0.7421	0.7385	0.7360	0.7329
MSE_S	0.7372	0.7364	0.7347	0.7325
MSE($\tau = 1$)	0.7372	0.7364	0.7347	0.7325
MSE($\tau = 0.9$)	0.7372	0.7364	0.7347	0.7325
MSE($\tau = 0.8$)	0.7372	0.7364	0.7347	0.7325
MSE($\tau = 0.5$)	0.7372	0.7364	0.7347	0.7325
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	0.7372	0.7364	0.7347	0.7325

Table 3.13: Performance of different N with $\theta_1 = 0.001$, $\theta_2 = 0.5$, $\delta = 0.2$, $m = 2$

n	50	100	200	500
$E(\lvert\hat{\theta}_1\rvert)$	0.0014	0.0021	0.0032	0.0020
C_p	0.1020	0.0857	0.0609	0.0387
$\hat{C}_p$	0.1168	0.0869	0.0625	0.0402
MSE_E	0.7421	0.7385	0.7359	0.7329
MSE_S	0.7372	0.7364	0.7347	0.7325
MSE($\tau = 1$)	0.7372	0.7364	0.7347	0.7325
MSE($\tau = 0.9$)	0.7372	0.7364	0.7347	0.7325
MSE($\tau = 0.8$)	0.7372	0.7364	0.7347	0.7325
MSE($\tau = 0.5$)	0.7372	0.7364	0.7347	0.7325
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	0.7372	0.7364	0.7347	0.7325

Table 3.14: Performance of different N with $\theta_1 = 0.4$, $\theta_2 = 0.5$, $\delta = 5$, $m = 2$

n	50	100	200	500
$E(\lvert\hat{\theta}_1\rvert)$	0.3918	0.3935	0.3958	0.3981
C_p	0.1045	0.0839	0.0599	0.0337
$\hat{C}_p$	0.0986	0.0751	0.0511	0.0346
MSE_E	2.5800	2.4454	2.3899	2.3554
MSE_S	3.5249	3.4061	3.3890	3.4612
MSE($\tau = 1$)	2.5800	2.4454	2.3899	2.3554
MSE($\tau = 0.9$)	2.5800	2.4454	2.3899	2.3554
MSE($\tau = 0.8$)	2.5800	2.4454	2.3899	2.3554
MSE($\tau = 0.5$)	2.5800	2.4454	2.3899	2.3554
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	2.5800	2.4454	2.3899	2.3554

Table 3.15: Performance of different N with $\theta_1 = 0.1$, $\theta_2 = 0.5$, $\delta = 5$, $m = 2$

n	50	100	200	500
$E(\lvert\hat{\theta}_1\rvert)$	0.0964	0.0955	0.0957	0.0980
C_p	0.1071	0.0845	0.0601	0.0339
$\hat{C}_p$	0.1157	0.0860	0.0636	0.0411
MSE_E	2.5675	2.4508	2.3912	2.3571
MSE_S	2.5178	2.4608	2.4294	2.4135
MSE($\tau = 1$)	2.5178	2.4508	2.3912	2.3571
MSE($\tau = 0.9$)	2.5178	2.4508	2.3912	2.3571
MSE($\tau = 0.8$)	2.5675	2.4508	2.3912	2.3571
MSE($\tau = 0.5$)	2.5675	2.4508	2.3912	2.3571
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	2.5178	2.4508	2.3912	2.3571

Table 3.16: Performance of different N with $\theta_1 = 0.05$, $\theta_2 = 0.5$, $\delta = 5$, $m = 2$

n	50	100	200	500
$E(\lvert\hat{\theta}_1\rvert)$	0.0474	0.0459	0.0458	0.0481
C_p	0.1089	0.0845	0.0603	0.0337
hline $\hat{C}_p$	0.1167	0.0860	0.0637	0.0414
MSE_E	2.5669	2.4496	2.3916	2.3571
MSE_S	2.4692	2.4138	2.3818	2.3616
MSE($\tau = 1$)	2.4692	2.4138	2.3818	2.3571
MSE($\tau = 0.9$)	2.4692	2.4138	2.3818	2.3571
MSE($\tau = 0.8$)	2.4692	2.4138	2.3818	2.3571
MSE($\tau = 0.5$)	2.4692	2.4496	2.3916	2.3571
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	2.4692	2.4138	2.3818	2.3571

Table 3.17: Performance of different N with $\theta_1 = 0.01$, $\theta_2 = 0.5$, $\delta = 5$, $m = 2$

n	50	100	200	500
$E(\lvert\hat{\theta}_1\rvert)$	0.0082	0.0063	0.0059	0.0081
C_p	0.1091	0.0841	0.0601	0.0338
$\hat{C}_p$	0.1168	0.0858	0.0631	0.0414
MSE_E	2.5665	2.4486	2.3918	2.3572
MSE_S	2.4544	2.3992	2.3662	2.3443
MSE($\tau = 1$)	2.4544	2.3992	2.3662	2.3443
MSE($\tau = 0.9$)	2.4544	2.3992	2.3662	2.3443
MSE($\tau = 0.8$)	2.4544	2.3992	2.3662	2.3443
MSE($\tau = 0.5$)	2.4544	2.3992	2.3662	2.3443
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	2.4544	2.3992	2.3662	2.3443

Table 3.18: Performance of different N with $\theta_1 = 0.001$, $\theta_2 = 0.5$, $\delta = 5$, $m = 2$

n	50	100	200	500
$E(\lvert\hat{\theta}_1\rvert)$	0.0006	0.0026	0.0030	0.0009
C_p	0.1095	0.0847	0.0597	0.0337
$\hat{C}_p$	0.1170	0.0858	0.0629	0.0412
MSE_E	2.5664	2.4484	2.3919	2.3572
MSE_S	2.4540	2.3987	2.3655	2.3434
MSE($\tau = 1$)	2.4540	2.3987	2.3655	2.3434
MSE($\tau = 0.9$)	2.4540	2.3987	2.3655	2.3434
MSE($\tau = 0.8$)	2.4540	2.3987	2.3655	2.3434
MSE($\tau = 0.5$)	2.4540	2.3987	2.3655	2.3434
MSE($\tau = 1 - \frac{1}{\sqrt{n}}$)	2.4540	2.3987	2.3655	2.3434

Appendix A

Proofs

A.1 Proof of Lemma 3.2.2 in Chapter 3

Define any vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, i.e., $x, y \in \Re^n$. We will give the proof of Lemma 2.2 by induction.

It suffices to show that $(x_1y_1 + \cdots + x_ny_n)^2 \leq (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2)$, since $x^{'}y = y^{'}x = x_1y_1 + \cdots + x_ny_n$, $x^{'}x = x_1^2 + \cdots + x_n^2$ and $y^{'}y = y_1^2 + \cdots + y_n^2$.

Base step: if n = 1, $(x_1y_1)^2 = x_1^2y_1^2$, therefore the base case holds.
Inductive step: assume that $(x_1y_1 + \cdots + x_ny_n)^2 \leq (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2)$ holds, consider the LHS:

$$
\begin{aligned}
&(x_1y_1 + \cdots + x_ny_n + x_{n+1}y_{n+1})^2 \\
&= (x_1y_1 + \cdots + x_ny_n)^2 + (x_{n+1}y_{n+1})^2 + 2(x_{n+1}y_{n+1})(x_1y_1 + \cdots + x_ny_n) \qquad \text{(A.1.1)} \\
&\leq (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2) + (x_{n+1}y_{n+1})^2 + 2(x_{n+1}y_{n+1})(x_1y_1 + \cdots + x_ny_n)
\end{aligned}
$$

Then consider the RHS:

$$
\begin{aligned}
&(x_1^2+\cdots+x_n^2+x_{n+1}^2)(y_1^2+\cdots+y_n^2+y_{n+1}^2) \\
&= (x_1^2+\cdots+x_n^2)(y_1^2+\cdots+y_n^2)+(x_{n+1}y_{n+1})^2+(x_1^2+\cdots+x_n^2)y_{n+1}^2+(y_1^2+\cdots+y_n^2)x_{n+1}^2 \\
&\geq (x_1^2+\cdots+x_n^2)(y_1^2+\cdots+y_n^2)+(x_{n+1}y_{n+1})^2 \\
&\quad +2(x_{n+1}y_{n+1})(x_1y_1+\cdots+x_ny_n) \text{ by Inequality of Arithmetic}
\end{aligned}
\tag{A.1.2}
$$

Under (A.1) and (A.2), the inductive case holds. Now by induction we see that Lemma 2.2 is true.

A.2 Proof of Lemma 3.3.1 in Chapter 3

Proof of part (a).

It suffices to show that $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t \xrightarrow{P} 0$ and $\frac{1}{T}\sum_{t=1}^{T} y_{t-h}\varepsilon_t \xrightarrow{P} 0$ for $p_1+1 \leq h \leq p$.

Note that: $E(\varepsilon_t|F_{t-p_1}) = 0$ and $E(y_{t-h}\varepsilon_t|F_{t-p_1}) = y_{t-h}E(\varepsilon_t|F_{t-p_1}) = 0$.

Based on the Assumption 2 and Lemma A(i) in Phillips and Xu (2006), we have $E(\varepsilon_t^2) < \infty$ and $E(y_{t-h}^2\varepsilon_t^2) \leq \sqrt{Sup_t E(y_{t-h}^4) Sup_t E(\varepsilon_t^4)}) < \infty$. By the law of large numbers for martingale differences, (a) follows directly.

Proof of part (b).

See the proof of Lemma 1 in Phillips and Xu (2006).

Proof of part (c).

It suffices to show the following three convergence results:

(i) $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 \xrightarrow{P} \int g^2$

(ii) $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 y_{t-h} \xrightarrow{P} 0$, for $p_1+1 \le h \ge p$

(iii) $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 y_{t-h}y_{t-h-k} \xrightarrow{P} 0$, for $1 \le h \le p$ and $0 \le k \le p-h$.

Proof (i):

$\{\varepsilon_t^2 - g^2(\frac{t}{T}), F_t\}$ is $\alpha-$mixing by Theorem 14.1 in Davidson (1994) and $E(\varepsilon_t^2 - g^2(\frac{t}{T}))^2 < \infty$ by Assumption 2. By the law of large numbers for L^1- mixingales,

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 = \frac{1}{T}\sum_{t=1}^{T}E(\varepsilon_t^2) + o_p(1) = \frac{1}{T}\sum_{t=1}^{T}g^2(\frac{t}{T}) + o_p(1) \xrightarrow{P} \int g^2.$$

Proof (ii):

Note that $\{\varepsilon_t^2\}$ is mixing and therefore L^2- NED on $\{y_{t-h}\}$ is L^2- NED as shown in the proof of Lemma in Phillips and Xu (2006). So by Theorem 17.9 in Davidson (1994), $\{\varepsilon_t^2 y_{t-h}\}$ is L^1-NED. Moreover, we have

$$E|\varepsilon_t^2 y_{t-h}|^{\delta} \le E(\varepsilon_t^{4\delta})E(y_{t-h})^{2\delta} < \infty$$

by Assumption 2 and Lemma A(i) in Phillips and Xu (2006). So by law of large numbers for L^1-mixingales, we have (ii) proved.

Proof (iii):

In the view of (i) and (ii), note that $\{\varepsilon_t^2 y_{t-h}y_{t-h-k} - g^2(\frac{t}{T}y_{t-h}y_{t-h-k}), F_t\}$ is martingale difference sequence as $E(\varepsilon_t^2 y_{t-h}y_{t-h-k} - g^2(\frac{t}{T}y_{t-h}y_{t-h-k})|F_{t-p_1}) = y_{t-h}y_{t-h-k}E(\varepsilon_t^2 - g^2(\frac{t}{T})|F_{t-p_1}) = 0$

$$\|\varepsilon_t^2 y_{t-h}y_{t-h-k} - g^2(\frac{t}{T}y_{t-h}y_{t-h-k}\|_{\delta} \overset{Minkowski}{\le} \cdots < \infty$$

where the inequalities follow by Minkowski, Cauchy-Schewarz(CS), and Lemma A(i) in Phillips and Xu (2006) respectively. By the law of large numbers, we then have

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 y_{t-h}y_{t-h-k} - \frac{1}{T}\sum_{t=1}^{T}g^2(\frac{t}{T}y_{t-h}y_{t-h-k} \xrightarrow{P} 0$$

Also,

$$\frac{1}{T}\sum_{t=1}^{T} g^2(\frac{t}{T}y_{t-h}y_{t-h-k} = \frac{1}{T}\sum_{t=1}^{T} g^2(\frac{t}{T}E(y_{t-h}y_{t-h-k}) + o_p(1)$$

$$= \frac{1}{T}\sum_{t=1}^{T}\int_{\frac{t}{T}}^{\frac{t+1}{T}} g^2(\frac{[rT]}{T})E(y_{([rT]-h)}y_{([rT]-h-k)})dr + o_p(1)$$

$$= \int_{\frac{T+1}{T}}^{\frac{1}{T}} g^2(\frac{[rT]}{T})E(y_{([rT]-h)}y_{([rT]-h-k)})dr + o_p(1)$$

$$\xrightarrow{P} (\int g^4)\gamma_k$$

So,

$$\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t^2 y_{t-h}y_{t-h-k} = \frac{1}{T}\sum_{t=1}^{T} g^2(\frac{t}{T}y_{t-h}y_{t-h-k} + o_p(1)$$

$$\xrightarrow{P} (\int g^4)\gamma_k$$

and part (c) follows from (i), (ii) and (iii).

Proof of part (d).

Under the prove of (a), $\frac{1}{T}\sum_{t=1}^{T} y_{t-h} \xrightarrow{P} 0$ and $\frac{1}{T}\sum_{t=1}^{T} y_{(t-h)}y_{(t-h-k)} \xrightarrow{P} (\int g^2)\gamma_k$, for $p_1+1 \le h \le p$. So (d) follows.

Proof of part (e).

By the Cramér-Wold device and CTL for martingale differences in the Corollary 5.25 in White (2014), the result in part (e) is direct.

A.3 Proof of Lemma 3.2.1 in Chapter 3

Lemma 3.2.1 is just a special case that $g(\cdot)$ is a positive constant function in Lemma 3.3.1. So the proof can be done with the similar method with Lemma 3.2.1.

Bibliography

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2006). *Volatility and Correlation Forecasting*, pages 777–878. Amsterdam: North-Holland.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, 89(4):701–720.

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2004). Analytical evaluation of volatility forecasts. *International Economic Review*, 45(4):1079–1110.

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2005). Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities. *Econometrica*, 73(1):279–296.

Bandi, F. M., Russell, J. R., and Yang, C. (2008). Realized volatility forecasting and option pricing. *Journal of Econometrics*, 147(1):34–46.

Berkes, I., Horváth, L., and Kokoszka, P. (2004). Testing for parameter constancy in garch (p, q) models. *Statistics & probability letters*, 70(4):263–273.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bose, A. and Mukherjee, K. (2009). Bootstrapping a weighted linear estimator of the arch parameters. *Journal of Time Series Analysis*, 30(3):315–331.

Brockwell, P. J. and Davis, R. A. (1986). *Time Series: Theory and Methods*. Springer-Verlag New York, Inc., New York, NY, USA.

Chen, C. W., Gerlach, R., and Lin, E. M. (2008). Volatility forecasting using threshold heteroskedastic models of the intra-day range. *Computational Statistics & Data Analysis*, 52(6):2990–3010.

Choi, K., Yu, W.-C., and Zivot, E. (2010). Long memory versus structural breaks in modeling and forecasting realized volatility. *Journal of International Money and Finance*, 29(5):857–875.

Dahlhaus, R. (2012). Locally stationary processes. *Handbook of Statistics: Time Series Analysis: Methods and Applications*, 30:351.

Dahlhaus, R. et al. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, 25(1):1–37.

Dahlhaus, R., Rao, S. S., et al. (2006). Statistical inference for time-varying arch processes. *The Annals of Statistics*, 34(3):1075–1114.

Dahlhaus, R., Rao, S. S., et al. (2007). A recursive online algorithm for the estimation of time-varying arch parameters. *Bernoulli*, 13(2):389–422.

Davidson, J. (1994). *Stochastic limit theory: An introduction for econometricians*. OUP Oxford.

Dowla, A., Paparoditis, E., and Politis, D. N. (2003). Locally stationary processes and the local block bootstrap. In Akritas, M. G. and Politis, D. N., editors, *Recent Advances and Trends in Nonparametric Statistics*, pages 437 – 444. Elsevier, Amsterdam.

Dowla, A., Paparoditis, E., and Politis, D. N. (2013). Local block bootstrap inference for trending time series. *Metrika*, 76(6):733–764.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Francq, C., Roy, R., and Zakoïan, J.-M. (2005). Diagnostic checking in arma models with uncorrelated errors. *Journal of the American Statistical Association*, 100(470):532–544.

Francq, C. and Zakoian, J.-M. (2011). *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons.

Fryzlewicz, P., Sapatinas, T., and Rao, S. S. (2006). A Haar-Fisz technique for locally stationary volatility estimation. *Biometrika*, 93(3):687–704.

Fryzlewicz, P., Sapatinas, T., Rao, S. S., et al. (2008). Normalized least-squares estimation in time-varying arch models. *The Annals of Statistics*, 36(2):742–786.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131(1):59–95.

Ghysels, E. and Sohn, B. (2009). Which power variation predicts volatility well? *Journal of Empirical Finance*, 16(4):686–700.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, NJ.

Hansen, P. R. and Lunde, A. (2005). A realized variance for the whole day based on intermittent

high-frequency data. *Journal of Financial Econometrics*, 3(4):525–554.

Hansen, P. R. and Lunde, A. (2006a). Consistent ranking of volatility models. *Journal of Econometrics*, 131(1):97–121.

Hansen, P. R. and Lunde, A. (2006b). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127–161.

Hansen, P. R., Lunde, A., and Nason, J. M. (2003). Choosing the best volatility models: the model confidence set approach. *Oxford Bulletin of Economics and Statistics*, 65(s1):839–861.

Hillebrand, E. (2005). Neglecting parameter changes in garch models. *Journal of Econometrics*, 129(1):121–138.

Kokoszka, P. and Leipus, R. (2000). Change-point estimation in arch models. *Bernoulli*, 6(3):513–539.

Lux, T. and Morales-Arias, L. (2010). Forecasting volatility under fractality, regime-switching, long memory and student-t innovations. *Computational Statistics & Data Analysis*, 54(11):2676–2692.

Mikosch, T. and Starica, C. (2004). Changes of structure in financial time series and the garch model. *Revstat Statistical Journal*, 2(1):41–73.

Montgomery, D. and Runger, G. (2007). *Applied Statistics and Probability for Engineers, 3rd ed*. Wiley India Pvt. Limited.

Pan, L. and Politis, D. N. (2014). Bootstrap prediction intervals for markov processes. *Computational Statistics & Data Analysis*, 100:467–494.

Pan, L. and Politis, D. N. (2016). Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions. *Journal of Statistical Planning and Inference*, 177:1–27.

Paparoditis, E. and Politis, D. N. (2002). Local block bootstrap. *Comptes Rendus Mathematique*, 335(11):959–962.

Patton, A. J. and Sheppard, K. (2009). Optimal combinations of realised volatility estimators. *International Journal of Forecasting*, 25(2):218–238.

Patton, A. J. and Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, 97(3):683–697.

Peng, L. and Yao, Q. (2003). Least absolute deviations estimation for arch and garch models. *Biometrika*, 90(4):967–975.

Phillips, P. C. B. and Xu, K.-L. (2006). Inference in autoregression under heteroskedasticity. *Journal of Time Series Analysis*, 27(2):289–308.

Politis, D. N. (2003). A normalizing and variance–stabilizing transformation for financial time series. In Akritas, M. G. and Politis, D. N., editors, *Recent Advances and Trends in Nonparametric Statistics*, pages 335 – 347. JAI, Amsterdam.

Politis, D. N. (2007). Model-free versus model-based volatility prediction. *Journal of Financial Econometrics*, 5(3):358.

Politis, D. N. (2015). *Model-Free Prediction and Regression: A Transformation-Based Approach to Inference*. Springer, New York.

Politis, D. N. and Thomakos, D. D. (2006). Financial time series and volatility prediction using novas transformations. *Department of Mathematics and Department of Economics, University of California, San Diego*.

Politis, D. N. and Thomakos, D. D. (2013). Novas transformations: flexible inference for volatility forecasting. In *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, pages 489–525. Springer, New York.

Polzehl, J. and Spokoiny, V. (2006). Varying coefficient GARCH versus local constant volatility modeling: comparison of the predictive power. Technical report, SFB 649 discussion paper.

Poon, S.-H. and Granger, C. W. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41(2):478–539.

Priestley, M. (1988). Non-linear and non-stationary time series analysis. *Academic Press, London*.

Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 204–237.

Shephard, N. (1996). *Statistical Aspects of ARCH and Stochastic Volatility Time Series in Econometrics, Finance and Other Fields*. Chapman and Hall, London.

Stărică, C. and Granger, C. (2005). Nonstationarities in stock returns. *Review of Economics and Statistics*, 87(3):503–522.

White, H. (2014). *Asymptotic theory for econometricians*. Academic press.

Wu, S., Harris, T. J., and Mcauley, K. B. (2007). The use of simplified or misspecified models: Linear case. *The Canadian Journal of Chemical Engineering*, 85(4):386–398.