Multiscale GARCH Modeling and Inference

by

Lichen Chen

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2018

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Lars Stentoft Associate Professor and Canada Research Chair in Financial Econometrics Dept. of Economics and Dept. of Statistics and Actuarial Science Western University
Supervisors:	Adam Kolkiewicz Associate Professor Dept. of Statistics and Actuarial Science
	Tony S. Wirjanto Professor Dept. of Statistics and Actuarial Science and School of Accounting and Finance
Internal Members:	Martin Lysy Associate Professor Dept. of Statistics and Actuarial Science
	Greg Rice Assistant Professor Dept. of Statistics and Actuarial Science
Internal-External Member:	Pierre Chaussé Associate Professor Dept. of Economics

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The motivation behind this thesis is the shortage of formal statistical inference methods in the literature for testing whether a time series model is consistent with a sample at multiple sampling frequencies simultaneously. Most existing statistical methods for time series data focuses on a particular frequency of sample. However, in the statistical modeling of financial time series and applications, having a modeling being consistent with data at multiple frequencies can provide better interpretation of the underlying phenomenon and provide convenience in practical applications.

Mantegna and Stanley (1995[49], 1996[50]) and Ghashghaie, et.al.(1996)[27] are among the pioneers in pointing out the distinctive scaling behavior in financial asset return distributions. Mandelbrot, et.al. (1997)[47] explicitly pointed out the need to look at financial time series at multiple frequencies and use the scaling property of the data to help identify a model. Engle and Patton (2001)[22] raised the question of whether a GARCH(1,1) model, acceptable for modeling return volatility at each single time scale from 1-day to 1-week, is consistent across scales.

It is the purpose of this thesis to propose formal statistical inference methods for testing whether a given time series of ARMA and GARCH type is consistent with a sample at multiple frequencies simultaneously. To do so, we first examine the problem of model temporal aggregation. Then, based on temporal aggregation relations, we propose a novel statistical inference methods based on empirical likelihood with estimating equations. The proposed method can be used to formally test hypotheses of the following types: (i) whether a model with a fixed set of parameter values is consistent with sample at multiple frequencies; (ii) whether the model itself is capable of being consistent with the sample at multiple frequencies. Some related problems on GARCH model parameterization and parameter estimation with temporally aggregated data are also addressed.

Acknowledgements

I would like to begin by thanking my PhD supervisors, Professor Adam Kolkiewicz and Professor Tony S. Wirjanto, for their education, their encouragements, and their support. In fact, all the essential ideas in this thesis come from the enormous amount of discussions I had with them. Their broad knowledge and great patience are unparalleled. I also want to thank my PhD thesis committee, Professor Lars Stentoft, Professor Pierre Chaussé, Professor Martin Lysy, and Professor Greg Rice. I benefited hugely from the sharp and constructive suggestions from Prof. Lysy. I had many helpful and pleasant discussions on time series with Prof. Rice during the many walks we had together from school to the complex where we both lived in. Their dedicated inputs significantly improved my thinking and writing.

My special thanks go to Professor Phelim Boyle for his guidance during my entire graduate study years. I have benefited tremulously through taking his courses in financial economics and being given the opportunities to lecturing on stochastic calculus. The social events and "tea times" with him was among the most memorable and fun part of my graduate study at Waterloo. I also deeply indebted to him for the career guidance he provided me with. It is indeed my great fortune having him as my mentor.

I would like to thank my family, in particular my wife Jiahua, my mother and father, Xianrong and Yuenian, for their unconditional love and support throughout my PhD adventure. I especially thank Jiahua for her patience, understanding, and support during the difficult times of my study. Not only did she help me to maintain a positive attitude, she was also always there to listen to me talking about and helping me sorting out my research ideas as a fellow statistics major. I feel blessed having her as my wife and always being with me. Also I would like to thank my extended family members, Dr. Jiafeng Pan and Joseph Chan, for their supports and helps, and the precious brother- and sister-hood.

I thank Professor Peisong Han, Shixiao Zhang, Dr. Xichen She, and Professor Jerry Lawless for helps on the topic of empirical likelihood. I thank Professor Blake LeBaron of Brandeis University for helpful discussions on multiscale modeling of financial time series during his visit to Waterloo. I want to thank Professor Yi Shen for helpful discussion on the topic of scaling and self-similar processes.

Throughout my graduate study in the department of Statistics and Actuarial Science, I had been continuously receiving helps from the Graduate Studies Coordinator Mary Lou Dufton. I am grateful for her invaluable helps on each and every milestones of my graduate study. I also thank Professor Ken Seng Tan and Professor Mu Zhu for providing important supports during my PhD study.

It is my great pleasure to express thanks to my fellow PhD friends over the years, especially Dr. Zhenyu Cui, Dr. Fangyuan Lin, Dr. Kai Liu, Dr. Yuchen Mei, and Dr. Yoshihiro Taniguchi. I learned a lot from the many lunch and dinner time discussions with them. Much of my knowledge gained in graduate school were through spending time with these smart fellows.

During the finishing period of my PhD study, I received generous helps on job seeking and career guidance from Dr. Jit Seng Chen, Dr. Min Chen, Lu Cheng, Dr. Giuseppe DiCesare, Dr. Shannon Kennedy, Jeanine Kwong, and David Melkuev. My heart-felt thanks go to these fellows.

I also want thank the many faculty staff members I had at the Department of Mathematical Sciences at Worcester Polytechnic Institute, where I started my graduate study. Professor Marcel Blais exemplified excellent teaching and taught me much of the fundamental knowledge in Financial Mathematics and Stochastic Calculus. I thank Professor J.J. Malone, Professor Hasanjan Sayit and Professor Domokos Vermes for their guidance on research and recommendations for PhD study. I thank Ellen Mackin, Professor Joseph Petrucelli, Professor Suzanne Weekes for always being there for help. And I thank fellow Waterloo alumni Professor William Martin for his encouragements to me every time he came to Waterloo.

Financial supports from the University of Waterloo, the Department of Statistics and Actuarial Science at the University of Waterloo, the Center of Actuarial Excellence at the University of Waterloo, and the Natural Science and Engineering Research Council of Canada (NSERC) are gratefully acknowledged.

To my family.

Table of Contents

[1]We thank the external examiner, Professor Lars Stentoft, for emphasizing this research direction.

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivations

Over the years, financial time series has become available at increasingly higher frequencies. This increase in information is happening along with, and perhaps is also a driving force behind, the desire to have models that are able to capture the features of data at multiple time scales and fit data consistently across multiple sampling frequencies. This is in contrast to the more traditional modeling approaches which usually focus on one time scale with a single sampling frequency.

The desire for multiscale models is a consequence of observing data at multiple frequencies. Financial time series sampled at different time scales exhibit distinctive behaviors in terms of their statistical properties such as temporal dependency and statistical distribution. Traditionally favored single scale models have been found to be inadequate for modeling the multiscale features in financial time series.

The desire for multiscale models is also driven by practical needs. A brief reflection on the goals of financial modeling would help us appreciate the importance of having models that work well on multiple time scales. In options pricing, the pricing and hedging problems concern the dynamics of the underlying process at two different time scales. Financial risk management requires forecasting volatilities at different horizons. In all of these tasks, we would prefer having only one model for different horizons as using different models for different horizons is inconvenient.

Inspired by the increased popularity of the multiscale perspective in financial time series analysis, we devote this thesis to the study of particular issues in financial time series modeling and inference which are of a multiscale nature, namely, modeling and testing the scaling behavior of the linear dependency structure of ARMA and GARCH models.

1.2 Interests in the Multiscale Behavior of Asset Returns in the Literature and the Focus of This Thesis

The interests in the statistical property of asset price changes over multiple time scales can be traced back at least to Mandelbrot's (1963)[48] modeling of the distribution of cotton price change over multiple time periods. Since then, the multiscale perspective had been an essential part of financial modeling. A surge of interests in multiscale modeling may be linked to a series of publications in the leading scientific journal *Nature*, including Mantegna and Stanley (1995[49], 1996[50]) and Ghashghaie, et.al.(1996)[27]. With larger data sets and a wider range of sample frequencies, the authors presented some empirical patterns of financial asset returns over different scales, also called the *scaling* behavior of returns.

Motivated by these findings, many authors subsequently began to approach the problem of multiscale modeling of financial time series from various perspectives. For instance, Muller, et.al.(1997)[55] discovered asymmetric correlations between long and short horizon volatilities and proposed a Heterogeneous ARCH (HARCH) model to capture this phenomenon. Mandelbrot, et.al. (1997)[47] proposed a model, called a multifractal model of asset returns (MMAR), to capture the moment scaling property in exchange rate returns [1]. Andersen and Bollerslev (1998)[5] investigated the ability of the GARCH(1,1) model to capture volatility persistence from intradaily to weekly scales. LeBaron (2001)[43] gave intuitive insights into capturing the scaling behavior with a stochastic volatility model with three components. Dragulescu and Yakovenko (2002)[15] considered a multiscale goodness-of-fit of the Heston model in terms of its unconditional distributions. Eberlein and Ozkan (2003)[19] considered time consistency of Levy models, also in terms of distributional properties. In summary, a common theme in all of these

[1] See also Augustyniak, et. al. (2018)[2] for a recent development following the ideas of multifractal models.

studies is that the *scaling* behavior of empirical return distribution and/or return volatility dependency can be used to identify the statistical model for asset returns.

Despite the interests in the multiscale modeling of financial time series, there is no clear consensus emerging on a particularly preferred modeling framework or statistical method to focus on. Therefore, instead of asking which model we should use for the purpose of multiscale modeling, we find it more fruitful to use the multiscale perspective to improve the understanding and application of model classes that have already been widely used. In particular, this thesis focuses on ARMA and GARCH type models in studying the problems of temporal aggregation and the formal testing of scaling behavior of their **linear dependency structure** over different time scales.

We focus on the linear dependency structure for the following reasons. Firstly, linear dependency structure is a general concept which is applicable to any stationary non-deterministic process. In fact, the fundamental **Wold Decomposition Theorem** of stationary time series is phrased in term of **linear** dependency. See, for example, Brockwell and Davis (1991)[7], Theorem 5.7.1. Secondly, the scaling relation of the linear dependency structure of a stationary ARMA process can be derived by using well-establish techniques from the temporal aggregation literature. See, for example, Wei (2005)[73] and Silvestrini and Veredas (2008)[66]. In contrast, distributional properties are generally not robust to aggregation. Thirdly, many important concepts relevant to practical applications are based on the linear dependency structure, including autocorrelations, impulse response, and persistence measures. These are closely related to the linear forecasting of financial time series, which is among the ultimate goals of the modeling task.

Within the ARMA and GARCH model frameworks, the scaling property of the linear dependency structure is important for accurate forecasting over different horizons because of the following relation between scale and horizon. Consider an AR(1) process

$$X_t - \phi_1 X_{t-1} = Z_t,$$

where $\{Z_t\} \sim WN(0, \sigma_Z^2)$ is a driving white noise process with $\sigma_Z^2 < \infty$. On the one hand, from standard time series analysis texts, such as Brockwell and Davis(1991)[7], we know that the parameter ϕ_1 is the characteristic root of the autoregressive polynomial. It characterizes the persistence of the process, which is also the rate of decay of the autocorrelation function of the process. The h-step prediction of X_t based on information up to time t in this model is

$$X_{t+h|t} = \phi_1^h X_t.$$

Thus, an accurate estimate of ϕ_1^h for $h = 1, 2, 3, \cdots$ is important for making accurate predictions into the future.

On the other hand, we can derive the h-scale dynamics of the process X_t through a repeated substitution and obtain

$$X_t - \phi_1^h X_{t-h} = Z_t + \phi_1 Z_{t-1} + \cdots + \phi_1^{h-1} Z_{t-h+1}.$$

We observe that, on the h-scale, the process X_t preserves the AR(1) structure (with an additional moving-average part) and the h-scale autoregressive coefficient, ϕ_1^h, coincides with the h-step prediction coefficient of the process on the original scale. Therefore, an accurate estimate of the parameter in the h-scale process is important for making an accurate h-step prediction.

This basic relation between horizon and scale not only highlights the importance of an accurate statistical characterization of the process at multiple time scales for the purpose of multiple step predictions, it also provides us with a way to improve multiple step predictions through examining the model at multiple time scales.

To realize this idea, we need to consider the following two problems:

- **Temporal aggregation:** given a model at a high frequency, what is the statistical representation of the model at lower frequencies?
- **Scale consistency:** when a model is fitted to data at different scales/frequencies, are the fitted models consistent with each other according to their temporal aggregation relation?

To provide the readers with some examples, we give a brief review of studies on each of these problems from the literature.

The Temporal Aggregation Problem

Consider a GARCH(p,q) process with i.i.d. innovations $\{\eta_t\}$:

$$\begin{aligned} \epsilon_t &= \sqrt{h_t}\eta_t, \\ h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}, \end{aligned} \tag{1.1}$$

with the usual assumptions about the model parameters $\alpha_0 > 0$, $\alpha_i \geq 0, i = 1, \cdots, q$, and $\beta_j \geq 0, j = 1, \cdots, p$, in order to ensure that the conditional variance process $\{h_t\}$ remains positive. Here, the observable ϵ_t represents the demeaned asset returns.

In empirical applications, GARCH models have been specified for data at different frequencies, most commonly for daily and weekly data, and are often found to provide a good fit to data at *each* of these frequencies. Drost and Nijman (1993)[16] addressed the question of whether a specification of the above *strong* GARCH process (featuring the *i.i.d* assumption for the innovation terms) is consistent with a strong GARCH specification at the weekly scale. They showed that the strong GARCH specification is not closed under temporal aggregation: the aggregation of a daily scale strong GARCH yields a so-called *weak* GARCH at the weekly scale. In short, $\{h_t\}$ generally no longer has the conditional variance interpretation in the weak GARCH representation, and it only amounts to a *linear projection* of $\{\epsilon_t^2\}$ on the past of the process[2]. Moreover, the orders p and q may change with the level of aggregation. In addition, as we will emphasize, the usual assumptions on the model parameters need to be relaxed as well. Since $\{\epsilon_t^2\}$ in the GARCH model admits an ARMA representation, the problem of temporal aggregation of GARCH processes is closely related to temporal aggregation of ARMA processes.

The Scale Consistency Problem

Drost and Nijman (1993)[16] showed that the parameters in a weekly scale GARCH aggregated from a daily GARCH can be determined from the daily GARCH parameters. They derived a set of formulas in the GARCH(1,1) case which are known as the Drost-Nijman formula. For example, the sum $\alpha_1 + \beta_1$ in a weekly GARCH(1,1) process is shown to be the fifth power of the corresponding sum in

[2] We only give a brief idea about the weak GARCH process here, and the exact definition will be provided later in the relevant chapters.

the daily GARCH(1,1) process. On the one hand, one can estimate a GARCH(1,1) model with a high frequency data, say daily, and then convert the estimated parameters to a lower frequency, say, weekly, using the Drost-Nijman formula. On the other hand, one can estimate a GARCH(1,1) model directly with weekly data. By comparing the converted and the estimated weekly GARCH model parameters, one can examine whether a GARCH(1,1) model consistently describes the data at daily and weekly scales or not. Engle and Patton (2001)[22] found with a sample of Dow Jones Industrial Average (DJIA) index returns that a GARCH(1,1) model fitted to daily returns and a GARCH(1,1) model fitted to several-day returns could give apparently different estimates which vary considerably across sample frequencies. As an important measure of the persistence of the impact of a shock to volatility, volatility half-life is defined as the time taken for the volatility to move halfway back towards its unconditional mean following a deviation from it. In the GARCH(1,1) model, volatility half-life is given by $\tau = \ln(\frac{1}{2})/\ln(\alpha_1+\beta_1)+1$. The estimates of τ obtained in Engle and Patton (2001)[22] using one day to five day returns are 73, 168, 183, 508, and 365 days, respectively[3]. Importantly, these estimated values would be expected to be constant across different sampling frequencies if the GARCH(1,1) model had been consistent with the sample across scales.

The Goal and Objectives of the Thesis

Our **goal** in this thesis is to carry the multiscale modeling perspective to the practice of the ARMA and GARCH modeling. The **objectives** of this thesis are **(i)** to study members in the GARCH family that add to the practical value of modeling dependency in financial asset return volatilities at multiple scales, and **(ii)** to construct statistical tests that are capable of determining whether a weak ARMA structure is compatible with a sample at multiple scales.

To give a brief overview of the tests for scale consistency proposed in this thesis, consider the following second order stationary AR(p) process $\{X_t, t = 0, t \in \mathbb{Z}\}$

[3]The volatility half-life estimates are cited from Engle and Patton (2001)[22]. While the authors do not report interval estimates of the half-life measure and mention that further studies are needed to assess their statistical and forecast significance, there may be some intrinsic difficulties in accurately estimating the half-life measure. These include (i) unknown and possibly intractable distribution of plug-in estimator of half-life using estimated GARCH model parameters, (ii) possibly infinite sample moments due to the construction of the half-life measure, and (iii) intrinsic bias in small samples. These issues may be addressed by using non-parametric method such as bias-corrected bootstrap or a *highest density region* method proposed by Hyndman (1996)[36]. We cite the half-life estimates as an intuitive way of highlighting the scale-inconsistency issue as in Engle and Patton (2001)[22] and thus elaborate on its estimation methods.

given by

$$X_t - \phi_1 X_{t-1} - ... - \phi_p X_{t-p} = Z_t, \tag{1.2}$$

where $\{Z_t\} \sim WN(0, \sigma_Z^2)$ is the driving white noise process with $\sigma_Z^2 < \infty$. Let us assume that one estimates this model with a given daily (d) frequency sample at the highest available frequency (i.e. daily), denoted by scale $m_1 = 1$, and obtain estimates:

$$\hat{\theta}^{(d)} = (\hat{\phi}_1^{(d)}, \cdots, \hat{\phi}_p^{(d)}, \sigma_Z^{2(d)}).$$

Let us assume that for an application purpose, one is interested in the weekly scale dynamics of the process. One thus estimates the model again with weekly (w) data, obtained through a temporal aggregation of the original daily data, and an appropriately aggregated model. One obtains a set of parameters for the weekly scale model, at scale $m_2 = 5$:

$$\hat{\theta}^{(w)} = (\hat{\phi}_1^{(w)}, \cdots, \hat{\phi}_p^{(w)}, \sigma_Z^{2(w)}).$$

As we will show, under the postulated model, the parameters of the model at daily and weekly scales satisfy a certain functional relation which can be derived by using temporal aggregation results. Denote by $f^{d,w}(\theta^{(d)})$ the function that maps the daily scale model parameters to the weekly scale model parameters. If the postulated model is true, then one would have

$$\hat{\theta}^{(w)} = f^{d,w}(\hat{\theta}^{(d)}). \tag{1.3}$$

The relation (1.3) provides us with a theoretical foundation for deciding whether the postulated model is **consistent** with the samples at two different scales. In particular, as we will show, the functional relation between the characteristic roots of the AR polynomials at different scales is a simple power relation. We are going to introduce novel tests which exploit this type of functional relations and utilize the samples from multiple frequencies.

1.3 Contributions and Organization of the Thesis

The contents of this thesis are divided into two parts. Part I, which contains Chapter 2 and Chapter 3, focus on the temporal aggregation problem of ARMA and GARCH models, and related issues of model parameterization. Part II, which contains Chapter 4, Chapter 5, and Chapter 6, focus on our proposed statistical tests of scale consistency based on the framework of empirical likelihood with estimating equations. Chapter 7 concludes the thesis and discusses an avenue for some future research.

In Part I, we make the following contributions. In Chapter 2, we propose a numerical method for computing parameter values in temporally aggregated GARCH(p,q) models with general orders p and q, extending the formula-based method of Drost and Nijman (1993)[16] for GARCH(1,1) model. In Chapter 3, we propose to use the component GARCH models of Ding and Granger (1996)[14] and Engle and Lee (1999)[21] as reparameterization of the GARACH(p,q) model in some situations where the GARCH(p,q) model under general parameter constraints having a singularly shaped parameter space.

In Part II, we propose a novel statistical inference framework based on temporal aggregation and empirical likelihood for testing whether a ARMA or GARCH type model is consistent with a sample at multiple frequencies.

1.4 Notations

Table 1.1: Notations

X	A generic random variable (r.v.)
$\{X_t\}$ or $\{X_t\}_{t\in\mathbb{Z}}$	A generic time series/discrete-time stochastic process or ARMA process, where $\mathbb{Z}$ is the set of integers
$\gamma_X(h)$	Autocovariance coefficient of $\{X_t\}$ at lag h
$\rho_X(h)$	Autocorrelation coefficient of $\{X_t\}$ at lag h
$\{Z_t\}$	A generic white noise (WN) process
$\{\eta_t\}$	Strong white noise process with unit variance
κ_η	Kurtosis coefficient of η_t
$\{\epsilon_t\}$	Innovations with a GARCH process
h_t	Conditional variance
$\nu_t = \epsilon_t^2 - h_t$	Driving white noise process in a GARCH process
h	Time lag or forecast horizon
m	Level of aggregation
$\{\bar{X}_{(m)tm}\}_{t\in\mathbb{Z}}$	Temporally aggregated process at level m of $\{X_t\}$, flow variable case
$\{X_{tm}\}_{t\in\mathbb{Z}}$	Temporally aggregated process at level m of $\{X_t\}$, stock variable case
$\{\bar{Z}_{(m)tm}\}_{t\in\mathbb{Z}}$	Driving WN process in a temporally aggregated ARMA process at level m, flow variable case
$\{Z_{tm}\}_{t\in\mathbb{Z}}$	Driving WN process in a temporally aggregated ARMA process at level m, stock variable case
$\boldsymbol{\Theta}$	Parameter set
$\boldsymbol{\theta}$	Element of the parameter set
$\boldsymbol{\theta}_0$	True parameter value

Chapter 2

Temporal Aggregation of ARMA and GARCH Processes

This chapter contains two parts. In the first part, consisting of Section 2.2 and Section 2.3, we summarize some relevant existing results on temporal aggregation of ARMA and GARCH type processes. The focus is on the structure of the aggregated processes and their statistical representations. In particular, we are interested in the relations between the parameters in the original model and the aggregated model. In the second part, Section 2.4, we make contributions in giving new results on calculating the parameters in the aggregated processes beyond the GARCH(1,1) case as in Drost and Nijman (1993)[16]. Some of the results in this chapter are used later in the thesis.

2.1 Preliminaries

Aggregation schemes

The investigation will involve two types of commonly used aggregation schemes, called *flow* and *stock* variable aggregations, respectively. A *flow* variable is measured over an interval of time and represents the change of a quantity over the interval of time. A *stock* variable is measured at one specific time and represents the existing quantity at that point in time. A simple way to understand the two definitions in the context of modeling financial time series is to take asset returns as an example of a flow variable and asset price as an example of a stock variable.

Let $\{X_t, t = 0, \pm 1, \pm 2, ...\}$ be a generic time series observed at a *high frequency*. We are interested in studying the resulting *aggregated* or *low frequency* series at an aggregation level m, where $m \geq 2$ is an integer. The aggregated processes over m periods are

$$\begin{cases} \bar{X}_{(m)tm} = X_{tm} + X_{tm-1} + ... + X_{tm-m+1}, & \text{flow variables,} \\ X_{(m)tm} = X_{tm}, & \text{stock variables.} \end{cases}$$

Definitions of white noise

White noise sequences are fundamental building blocks of time series models. In many textbooks, for example, Brockwell and Davis(1991)[7], white noise is defined as a sequence of uncorrelated random variable, indexed by time, and with a constant finite second moment. It can be further specified in three alternative ways as in our Definition 1 below. These specifications are often not emphasized in the literature because a clear distinction among these definitions might not be of much interest and importance in the context of modeling the conditional mean as in the ARMA process. It is, however, crucial for the study of temporal aggregation of the GARCH processes.

Definition 1 (White Noise) Let $\{Z_t, t = 0, \pm 1, \pm 2, ...\}$ be a stochastic process with mean 0 and a covariance function

$$\gamma(h) = Cov(Z_t, Z_{t-h}) = \begin{cases} \sigma^2 & \textbf{if } h = 0, \\ 0 & \textbf{if } h \neq 0. \end{cases}$$

The process $\{Z_t\}$ is

1. a *strong white noise* process if the random variables Z_t are independently and identically distributed with mean 0 and variance σ_Z^2;

2. a *semi-strong white noise* process if the random variables Z_t form a martingale difference sequence (m.d.s.) relative to its own past values, i.e. $E(Z_{t+1}|Z_s, -\infty < s \leq t) = 0$;

3. a *weak white noise* process if the random variables Z_t are only assumed to be uncorrelated.

In this thesis, we use $\{\eta_t\}$ to denote a strong white noise sequence with a unit variance.

From the definitions of the strong, semi-strong, and weak white noises, it is clear that we have the following inclusion relation:

$$\{\textbf{Strong WN}\} \subseteq \{\textbf{Semi-Strong WN}\} \subseteq \{\textbf{Weak WN}\}$$

Remark: The definitions of weak and strong white noises are given in some texts on time series analysis, such as Brockwell and Davis(1991)[7] and Francq and Zakoian (2010)[26]. The definition of a semi-strong white noise is not so commonly formulated in textbooks but is of interest in econometric time series analysis due to its link to rational expectations. It will play an important role in the definitions of the ARMA and GARCH processes given shortly below.

Definition 2 (Gaussian Time Series): The process $\{X_t, t = 0, \pm 1, \pm 2, ...\}$ is a Gaussian time series if and only if the distribution functions of $\{X_t\}$ are multivariate normal.

It is well known that for random variables having a joint normal distribution, zero correlation is equivalent to independence. Therefore, for Gaussian white noises, the three definitions (strong, semi-strong, and weak) are equivalent to each other.

2.2 Temporal Aggregation of ARMA Processes

2.2.1 Definitions of strong, semi-strong, and weak ARMA processes

In analogy to the definitions of white noises, we have three definitions for ARMA processes.

Definition 3 (ARMA process): Let $\{X_t, t = 0, \pm 1, \pm 2, ...\}$ be a (second order) stationary stochastic process given by

$$X_t - \phi_1 X_{t-1} - ... - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q}, \tag{2.1}$$

where $\{Z_t\} \sim WN(0, \sigma_Z^2)$ is the driving white noise process with $\sigma_Z^2 < \infty$. We say $\{X_t\}$ is

1. a *strong ARMA(p,q)* process if $\{Z_t\}$ is a strong white noise process;
2. a *semi-strong ARMA(p,q)* process if $\{Z_t\}$ is a semi-strong white noise process;
3. a *weak ARMA(p,q)* process if $\{Z_t\}$ is a weak white noise process.

As a note, $\{X_t\}$ is an ARMA(p,q) process with mean μ if $\{X_t - \mu\}$ is an ARMA(p,q) process (with mean zero).

The definitions of strong/semi-strong/weak ARMA processes are parallel to those of the white noise processes. It is clear that strong/semi-strong/weak ARMA(p,q) process also form an inclusion relation

$$\{\textbf{Strong ARMA}\} \subseteq \{\textbf{Semi-Strong ARMA}\} \subseteq \{\textbf{Weak ARMA}\}$$

Equation (2.1) can be written in a more compact form as

$$\phi(L)X_t = \theta(L)Z_t, t = 0, \pm 1, \pm 2, ..., \tag{2.2}$$

where ϕ and θ are the p^{th} and q^{th} degree polynomials

$$\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p \tag{2.3}$$

and

$$\theta(z) = 1 + \theta_1 z + ... + \theta_q z^q, \tag{2.4}$$

respectively, and L is the lag operator.

We assume that the polynomials $\phi(z)$ and $\theta(z)$ have no common zeros. We assume that $\phi(z)$ has all roots outside the unit circle so that the ARMA process $\{X_t\}$ is a causal function of $\{Z_t\}$, which implies that $\phi(z)$ has no roots on the unit circle and so that $\{X_t\}$ is second-order stationary. Lastly we assume that $\theta(z)$ has all roots outside the unit circle so that the ARMA process $\{X_t\}$ is invertible.

To derive the temporal aggregated process of $\{X_t\}$ at the aggregation level m, we follow the procedure as summarized in Silvestrini and Veredas (2008)[66].

2.2.2 Deriving temporally aggregated ARMA processes

Flow variable aggregation

We start the discussion by expressing the AR polynomial $\phi(L)$ in terms of its inverted roots δ_j's as $\phi(L) = \prod_{j=1}^{p}(1 - \delta_j L)$. Next, we define polynomials

$$T_f(L) = \left[\frac{1 - L^m}{1 - L}\right] \prod_{j=1}^{p} \left[\frac{1 - \delta_j^m L^m}{1 - \delta_j L}\right] \tag{2.5}$$

and

$$T_s(L) = \prod_{j=1}^{p} \left[\frac{1 - \delta_j^m L^m}{1 - \delta_j L} \right] \tag{2.6}$$

to be used for flow and stock variable aggregations, respectively.

Let us first consider the case of the flow variable aggregation. Multiply both sides of the ARMA model in (2.2) by $T_f(L)$, we have

$$\begin{aligned}
T_f(L)\phi(L)X_t &= T_f(L)\theta(L)Z_t \\
\frac{1 - L^m}{1 - L} \cdot \prod_{j=1}^{p} \left[\frac{1 - \delta_j^m L^m}{1 - \delta_j L} \right] \prod_{j=1}^{p} (1 - \delta_j L) X_t &= T_f(L)\Theta(L)Z_t \\
\prod_{j=1}^{p} (1 - \delta_j^m L^m) \cdot \left[\left(\sum_{i=0}^{m-1} L^i \right) X_t \right] &= \prod_{j=1}^{p} \left(\sum_{i=0}^{m-1} \delta_j^i L^i \right) \cdot \sum_{l=0}^{q} \theta_l L^l \cdot \sum_{i=0}^{m-1} L^i \cdot Z_t \\
\left[\bar{\phi}(B) \right] \left[\bar{X}_{(m)tm} \right] &= \bar{\theta}(B) \bar{Z}_{(m)tm}
\end{aligned} \tag{2.7}$$

with $\delta_0 = 0$ and $\theta_0 = 1$, where B is the lag operator on the aggregated scale. For the AR part, we have $\{\bar{X}_{(m)tm}\}$ being the aggregated (flow) variable and $\bar{\phi}(B)$ defines the AR polynomial of the aggregated ARMA model in this flow variable case. For the MA part, we have $\bar{Z}_{(m)tm}$ being a (weak) white noise sequence with respect to the aggregated scale $\{tm, t \in \mathbb{Z}\}$ and $\bar{\theta}(B)$ defines the MA polynomial.

We can see that the aggregated model has the same number of AR lags as the original (or disaggregated) model, with inverted roots being mth power of the corresponding inverted roots of the original AR polynomial.

The MA part of the aggregated process is more complicated than the AR part. By inspection, the lag of the aggregated MA polynomial is (at most) $\lfloor [(p+1)(m-1)+q]/m \rfloor$ where $\lfloor \cdot \rfloor$ indicates the floor function. The exact order of the aggregated processes depends on the values of the parameters and the level of aggregation. Define $\bar{v}_{(m)tm} := \bar{\theta}(B)\bar{Z}_{(m)tm}$. The coefficients of the aggregated MA polynomial can be calculated (in general, numerically) by matching the autocorrelation function of $\{\bar{v}_{(m)tm}\}$ and that of an MA process with the corresponding order. Since the autocovariance function of the MA part is nonzero at only a finite number of lags q^*, the MA coefficients of the aggregated process can be computed by solving a system of q^* nonlinear equations. For example, in the case of the ARMA(2,2) model, this leads to a system of two equations, matching the autocovariances of $\{\bar{v}_{(m)tm}\}$ at lags 1 and 2 (in the aggregated scale) with the corresponding autocovariances of an MA(2) process.

Stock variable aggregation

Next, we consider a stock variable aggregation. By replacing the polynomial T_f with the polynomial T_s and going through the same calculations as in (2.7), we can derive the aggregated ARMA process in the stock variable case as

$$\bar{\phi}(B) \cdot X_{tm} = \bar{\theta}(B) Z_{tm}. \tag{2.8}$$

Compared to the flow variable case, we have a simpler expression in the stock variable case because the polynomial $[1 - L^m]/[1 - L]$ is omitted.

In practice, whether we should consider a flow or stock variable aggregation depends on a specific application problem. For example, a flow variable aggregation is the relevant case for studying asset returns, while a stock variable aggregation is the relevant case for studying asset prices or index levels.

The weak ARMA processes considered in this thesis

As we can see from the definition of weak ARMA processes, the class of weak ARMA processes contains potentially a wide range of processes, including temporal and marginalization of (strong) ARMA and vector ARMA processes, bilinear processes, switching-regime models, and threshold models. We refer readers to Franq and Zakoian (1998)[25] for examples in each of these cases and estimation methods under the general weak ARMA assumption. In this thesis, we restrict our attention to the weak ARMA processes resulting from the temporal aggregation of strong ARMA processes. In particular, the weak ARMA processes we consider are resulting from a temporal aggregation of strong ARMA processes assumed for the highest frequency observed sample.

2.3 Temporal Aggregation of GARCH Processes

In order to have a closed-under-aggregation property, Drost and Nijman (1993)[16] defined a more general class of GARCH models, called the *weak GARCH* models, by generalizing the ARCH and GARCH models as defined in Engle (1982)[20] and Bollerslev (1986)[5], respectively. The commonly used definition of the GARCH model in the literature, the one with independent normalized innovations, is named a *strong GARCH* model by Drost and Nijman (1993)[16]. In addition, they also defined a class of *semi-strong GARCH* models which lie between the strong GARCH model and the weak GARCH model. The definitions of the three classes of GARCH models are given below.

2.3.1 Definitions of strong, semi-strong, and weak GARCH processes

Definition 4 (GARCH Process): Let $\{\epsilon_t, t \in \mathbb{N}\}$ be a second order stationary process with finite fourth moments. Denote $A(L) = 1 + \sum_{i=1}^{q} \alpha_i L^i$ and $B(L) = 1 - \sum_{i=1}^{p} \beta_i L^i$, and let the sequence $\{h_t, t \in \mathbb{N}\}$ be defined as a second order stationary solution of

$$B(L)h_t = \alpha_0 + \{A(L) - 1\}\epsilon_t^2. \tag{2.9}$$

It is assumed that $B(L)$ and $B(L) + 1 - A(L)$ have roots outside the unit circle and hence are invertible. The sequence $\{\epsilon_t, t \in \mathbb{N}\}$ is defined to be generated by
1. a *strong GARCH(p,q)* process if $\alpha_0, A(L)$, and $B(L)$ can be chosen such that

$$z_t := \epsilon_t / \sqrt{h_t} \sim D(0, 1), \tag{2.10}$$

and $\{z_t, t \in \mathbb{N}\}$ is an i.i.d. process. We use $D(0, 1)$ to denote a distribution with mean zero and unit variance;
2. a *semi-strong GARCH(p,q)* proces if $\alpha_0, A(L)$, and $B(L)$ can be chosen such that

$$\begin{aligned} E[\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \ldots] &= 0, \\ E[\epsilon_t^2 | \epsilon_{t-1}, \epsilon_{t-2}, \ldots] &= h_t; \end{aligned} \tag{2.11}$$

3. a *weak GARCH(p,q)* process if $\alpha_0, A(L)$, and $B(L)$ can be chosen such that

$$\begin{aligned} P[\epsilon_t|\epsilon_{t-1}, \epsilon_{t-2}, ...] &= 0, \\ P[\epsilon_t^2|\epsilon_{t-1}, \epsilon_{t-2}, ...] &= h_t, \end{aligned} \tag{2.12}$$

where $P[x_t|\epsilon_{t-1}, \epsilon_{t-2}, ...]$ denotes the best linear predictor of x_t in terms of 1, ϵ_{t-1}, ϵ_{t-2},..., ϵ_{t-1}^2, ϵ_{t-2}^2, ..., i.e.

$$E[(x_t - P[x_t|\epsilon_{t-1}, \epsilon_{t-2}, ...])\epsilon_{t-i}^r] = 0 \text{ for } i \geq 1 \text{ and } r = 0, 1, 2. \tag{2.13}$$

Notice that h_t is driven by its own past values and past values of ϵ_t. If we assume that the conditioning σ-algebra $\mathcal{F}_t$ is generated by the past values of the process plus the initial value of h_t, then h_t is a measurable function with respect to the filtration $\{\mathcal{F}_t\}$. This measurability condition of h_t with respect to the conditioning filtration is important and is needed to show that h_t is the conditional variance of ϵ_t in the strong GARCH definition.

For the strong and semi-strong GARCH, we have

$$E[\epsilon_t\epsilon_{t-k}] = E[E_{t-1}[\epsilon_t\epsilon_{t-k}]] = E[\epsilon_{t-k}E_{t-1}[\epsilon_t]] = E[\epsilon_{t-k} \cdot 0] = 0 \text{ for } k > 0.$$

Thus strong and semi-strong GARCH processes are semi-strong white noise processes. For the weak GARCH process, it follows from definition that they are weak white noises processes.

Just as the definitions for white noise and ARMA processes, strong/semi-strong/weak GARCH process form an inclusion relation: a strong GARCH is also a semi-strong GARCH, and a semi-strong GARCH is also a weak GARCH. Arguably, the most used version of the GARCH model in empirical research is the strong GARCH model.

2.3.2 Examples of semi-strong and weak GARCH processes resulting from temporal aggregation

It is well known since Bollerslev (1986)[5] that a GARCH process admits a ARMA representation of its squared observations ϵ_t^2. The squared observations ϵ_t^2 in a

GARCH(p,q) model can be written as

$$\epsilon_t^2 = \alpha_0 + \sum_{i=1}^{r} (\alpha_i + \beta_i)\epsilon_{t-i}^2 + \nu_t - \sum_{j=1}^{p} \beta_j \nu_{t-j}, \tag{2.14}$$

where $r = \max(p, q)$, $\alpha_i = 0$ for $i > q$, and $\beta_j = 0$ for $j > q$, and $\nu_t = \epsilon_t^2 - h_t$. The process $\{\nu_t\}$ is an uncorrelated sequence for all of the three definitions of GARCH processes (i.e. strong/semi-strong/weak). So a GARCH(p,q) process can be written as an ARMA(r,p) process for its squared observations $\{\epsilon_t^2\}$. The sequence $\{\nu_t\}$ can thus be interpreted as the *driving white noise* of the GARCH process. It can be derived from Definition 4 that, for the strong and semi-strong GARCH processes, h_t is the conditional expectation of ϵ_t^2 with respect to the natural filtration $\{\mathcal{F}_t\}$ where $\mathcal{F}_t := \sigma\{\epsilon_t, \epsilon_{t-1}, ...\}$. Therefore, in the strong and semi-strong GARCH processes, ν_t is a martingale difference sequences (MDS) with respect to the natural filtration $\{\mathcal{F}_t\}$, i.e.

$$E(\nu_t|\mathcal{F}_{t-1}) = 0.$$

In the weak GARCH process, h_t is a linear projection of ϵ_t^2 on the infinite dimensional Hilbert space spanned by all linear combinations of a constant and ϵ_{t-1}, ϵ_{t-2},..., ϵ_{t-1}^2, ϵ_{t-2}^2, ... So $\{\nu_t\}$ is also uncorrelated in weak GARCH but not necessarily an MDS.

The representation in (2.14) is helpful for establishing an analogue between a GARCH process and an ARMA process. Strong and semi-strong GARCH processes are driven by a semi-strong white noise. A weak GARCH process is driven by a weak white noise.

As an illustration of non-aggregation of strong GARCH models, consider the following two cases:

CASE 1: A strong ARCH(1) process aggregates to a semi-strong ARCH(1) process but not a strong ARCH(1) procss under a stock variable aggregation

Consider a strong ARCH(1) model of a stock variable:

$$\begin{aligned} &\epsilon_t = \sqrt{h_t} z_t, \\ &h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2, \text{ where } \alpha_0 > 0, 0 < \alpha_1 < 1, z_t \sim i.i.d.(0,1), E(z_t^4) < \infty. \end{aligned} \tag{2.15}$$

By substitution (2.15) into itself we have

$$\begin{aligned}\epsilon_{2t} &= \sqrt{h_{2t}} z_{2t} \\ &= \sqrt{\alpha_0 + \alpha_1 \epsilon_{2t-1}^2} z_{2t} \\ &= \sqrt{\alpha_0 + \alpha_1 h_{2t-1} z_{2t-1}^2} z_{2t} \\ &= \sqrt{\alpha_0 + \alpha_1(\alpha_0 + \alpha_1 \epsilon_{2t-2}^2) z_{2t-1}^2} z_{2t} \\ &= \sqrt{\alpha_0(1 + \alpha_1 z_{2t-1}^2) + \alpha_1^2 \epsilon_{2t-2}^2 z_{2t-1}^2} z_{2t}.\end{aligned} \tag{2.16}$$

The conditional mean and variance under the aggregated model are

$$E[\epsilon_{2t}|\epsilon_{2t-2}, \epsilon_{2t-4}, ...] = \sqrt{\alpha_0(1 + \alpha_1 z_{2t-1}^2) + \alpha_1^2 \epsilon_{2t-2}^2 z_{2t-1}^2} E[z_{2t}] = 0 \tag{2.17}$$

and

$$\begin{aligned}E[\epsilon_{2t}^2|\epsilon_{2t-2}, \epsilon_{2t-4}, ...] &= \left(\alpha_0(1 + \alpha_1 E[z_{2t-1}^2]) + \alpha_1^2 \epsilon_{2t-2}^2 E[z_{2t-1}^2]\right) E[z_{2t}^2] \\ &= \alpha_0(1 + \alpha_1) + \alpha_1^2 \epsilon_{2t-2}^2,\end{aligned} \tag{2.18}$$

where we have used the fact that z_{2t} and z_{2t-1} are independent of the variables in the conditioning set $\{\epsilon_{2t-2}, \epsilon_{2t-4}, ...\}$. Thus the aggregated process is a semi-strong ARCH(1) process with parameters $\alpha_{(2)0} = \alpha_0(1 + \alpha_1)$ and $\alpha_{(2)1} = \alpha_1^2$.

However, the aggregated process is not a strong ARCH(1) process since the rescaled innovations, $\tilde{z}_{2t} = \epsilon_{2t}/\sqrt{\alpha_0(1 + \alpha_1) + \alpha_1^2 \epsilon_{2t-2}^2}$, are not i.i.d.. Although they have a zero conditional mean and a unit conditional variance (and thus a zero unconditional mean and a unit unconditional variance), the conditional fourth moment is

$$
\begin{aligned}
&E[\tilde{z}_{2t}^4|\epsilon_{2t-2},\epsilon_{2t-4},...]\\
&= E\left[\frac{\epsilon_{2t}^4}{(\alpha_0(1+\alpha_1)+\alpha_1^2\epsilon_{2t-2}^2)^2}|\epsilon_{2t-2},\epsilon_{2t-4},...\right]\\
&= E\left[\frac{h_{2t}^2 z_{2t}^4}{(\alpha_0(1+\alpha_1)+\alpha_1^2\epsilon_{2t-2}^2)^2}|\epsilon_{2t-2},\epsilon_{2t-4},...\right]\\
&= E[z_{2t}^4]E\left[\frac{(\alpha_0+\alpha_1\epsilon_{2t-1}^2)^2}{(\alpha_0(1+\alpha_1)+\alpha^2\epsilon_{2t-2}^2)^2}|\epsilon_{2t-2},\epsilon_{2t-4},...\right]\\
&= E[z_{2t}^4]E\left[\frac{\alpha_0^2+2\alpha_0\alpha_1\epsilon_{2t-1}^2+\alpha_1^2\epsilon_{2t-1}^4}{(\alpha_0(1+\alpha_1)+\alpha_1^2\epsilon_{2t-2}^2}|\epsilon_{2t-2},\epsilon_{2t-4},...\right]\\
&= E[z_{2t}^4]E\left[\frac{\alpha_0^2+2\alpha_0\alpha_1 h_{2t-1}z_{2t-1}^2+\alpha_1^2 h_{2t-1}^2 z_{2t-1}^4}{(\alpha_0(1+\alpha_1)+\alpha_1^2\epsilon_{2t-2}^2)^2}|\epsilon_{2t-2},\epsilon_{2t-4},...\right]\\
&= E[z_t^4]\left(1+\frac{(E[z_t^4]-1)\alpha_1^2(\alpha_0+\alpha_1\epsilon_{2t-2}^2)^2}{(\alpha_0(1+\alpha_1)+\alpha_1^2\epsilon_{2t-2}^2)^2}\right).
\end{aligned}
\tag{2.19}
$$

In order for $E[\tilde{z}_{2t}^4|\epsilon_{2t-2},\epsilon_{2t-4},...]$ to be a constant, we need to require either
(i) $\alpha_1 = 0$, or
(ii) $E[z_t^4] = 1$, or
(iii) $(\alpha_0+\alpha_1\epsilon_{2t-2}^2)/(\alpha_0(1+\alpha_1)+\alpha_1^2\epsilon_{2t-2}^2)$ being a constant.

In case (i), we have no ARCH effect. In case (ii), we have $1 = E[z_t^4] = (E[z_t^2])^2 + Var[z_t^2]$, or $Var[z_t^2] = 0$, or $z_t^2 =$ constant a.s.. In case (iii), we have $\epsilon_{2t-2}^2 =$ constant a.s., which implies, by stationarity of ϵ_t, that $\epsilon_t^2 =$ constant a.s.. Since $z_t^2 = \epsilon_t^2/(\alpha_0+\alpha_1\epsilon_{t-1}^2)$, we have that $z_t^2 =$ constant a.s.. All of these cases lead to the conclusion of either no ARCH effect or that z_t^2 being a constant, which is apparently not true in general. Therefore the aggregated process is not a strong ARCH(1) process. But in this case, it can be seen that the aggregated model is a semi-strong GARCH process. □

CASE 2: A strong ARCH(2) process aggregates to a weak GARCH(1,2) process which is not a semi-strong GARCH(1,2) process

Consider a strong ARCH(2) model of a stock variable under a stock variable aggregation:

$$
\begin{aligned}
&\epsilon_t = \sqrt{h_t}z_t,\\
&h_t = 0.05 + 0.1\epsilon_{t-1}^2 + 0.12\epsilon_{t-2}^2 \text{ where } z_t \sim i.i.d.N(0,1).
\end{aligned}
\tag{2.20}
$$

Writing the process $\{\epsilon_t^2\}$ in the AR(2) form, we have

$$(1-0.4L)(1+0.3L)\epsilon_t^2 = 0.05 + \nu_t, \tag{2.21}$$

where $\nu_t = \epsilon_t^2 - h_t$. Multiplying both sides of the last equation by $(1+0.4L)(1-0.3L)$, we have

$$(1-0.16L^2)(1-0.09L^2)\epsilon_t^2 = 1.4 \times 0.7 \times 0.05 + (1-0.4L)(1+0.3L)\nu_t. \tag{2.22}$$

We omit the details of the next few steps which rely on the procedure to be described in Section 2.3.2. In short, it can be shown that if one aggregates the last AR(2) process at level $m = 2$, one obtains an ARMA(2,1) process given by

$$\epsilon_{(2)t}^2 - 0.25\epsilon_{(2)(t-2)}^2 + 0.144\epsilon_{(2)(t-4)}^2 = 0.49 + 0.0766u_t,$$

where $u_t, u_{t-2}, u_{t-4}, \ldots$ is a white noise sequence at the 2-scale. Thus, the aggregated strong ARCH(2) process at level $m = 2$ admits a weak GARCH(1,2) representation with $\beta_{(2)1} = -0.0766$. However, a (semi-)strong GARCH(1,2) process with a negative β parameter violates the necessary and sufficient conditions for guaranteeing nonnegative conditional variance, as we will explain in Section 3.1. Therefore, parameters in the aggregated ARCH(2) process are not compatible with a (semi-)strong GARCH(1,2) process and thus it can only be a weak GARCH(1,2) process. □

Interpretations of Semi-Strong and Weak GARCH Processes

In the definition of semi-strong GARCH models, squared observations are assumed to follow an ARMA structure in which the error sequence is assumed to be an MDS. However, such models are not closed under aggregation. From the ARMA model literature we know that only weak ARMA models, where innovations are assumed to be serially uncorrelated, are closed under aggregation. (See, for example, Meddahi and Renault(2004)[54]). Therefore, in order to have a model structure closed under temporal aggregation, Drost and Nijman(1993)[16] defined squared observations as following a weak ARMA structure, i.e., ARMA models in which the error sequence is not assumed to be independent, nor even an MDS, but only uncorrelated. However, this relaxed definition achieves closeness under aggregation at the expense of losing the interpretation of h_t as conditional variance; instead, h_t is only the best *linear* prediction of future squared observation based on past observations and past squared observations.

A natural question to ask is how can we construct weak GARCH models? As strong GARCH models are, by definition, also weak GARCH models, it would be more interesting to ask how to obtain *strictly weak* GARCH models which are not strong or semi-strong GARCH models. The answer to this question can be obtained from the last example, which shows that temporally aggregated strong GARCH models are generally weak GARCH models which are not semi-strong. Therefore, one can simulate from a strong GARCH model, then temporally aggregate the simulated values, to obtain a sample from a strictly weak GARCH model. However, the functional form of the conditional distributions (and thus the likelihood function) of this weak GARCH model are generally not available analytically; what is known is the relation between the model parameters of the original high frequency strong GARCH model and the aggregated low frequency weak GARCH model. These model parameters depict the linear dependency structure of the process at each scale.

2.3.3 Deriving temporally aggregated GARCH processes: Drost-Nijman formula for the GARCH(1,1) Process

Given the structure of the temporally aggregated GARCH processes, we are interested in deriving the model parameters in the temporally aggregated models. Drost and Nijman (1993)[16] derived formula for GARCH(1,1). Let $\{\epsilon_t, t \in \mathbb{Z}\}$ be a weak GARCH(1,1) process with symmetric marginal distributions, $h_t = \psi + \beta h_{t-1} + \alpha y_{t-1}^2$, and an unconditional kurtosis coefficient κ_ϵ, then the temporally aggregated process in the case of flow variable aggregation, $\{\bar{\epsilon}_{(m)mt}, t \in \mathbb{Z}\}$, is a symmetric weak GARCH(1,1) process with

$$\bar{h}_{(m)mt} = \bar{\psi}_{(m)} + \bar{\beta}_{(m)}\bar{h}_{(m)m(t-1)} + \bar{\alpha}_{(m)}\bar{\epsilon}^2_{(m)m(t-1)},$$

and a kurtosis coefficient $\bar{\kappa}_{(m)\epsilon}$ where

$$\bar{\psi}_{(m)} = m\psi\frac{1-(\beta+\alpha)^m}{1-(\beta+\alpha)}, \;\; \bar{\alpha}_{(m)} = (\beta+\alpha)^m - \bar{\beta}_{(m)}, \tag{2.23}$$

$$\begin{aligned}\bar{\kappa}_{(m)\epsilon} &= 3 + (\kappa_\epsilon - 3)/m + 6(\kappa_\epsilon - 1) \\ &\times \frac{[m-1-m(\beta+\alpha)+(\beta+\alpha)^m][\alpha - \beta\alpha(\beta+\alpha)]}{m^2(1-\beta-\alpha)^2(1-\beta^2-2\beta\alpha)}.\end{aligned} \tag{2.24}$$

The parameter $\bar{\beta}_{(m)}$ in (2.24) satisfies $|\bar{\beta}_{(m)}| < 1$ and is the solution of a quadratic equation

$$\frac{\bar{\beta}_{(m)}}{1+\bar{\beta}^2_{(m)}} = \frac{a(\beta,\alpha,\kappa_\epsilon,m)(\beta+\alpha)^m - b(\beta,\alpha,m)}{a(\beta,\alpha,\kappa_\epsilon,m)[1+(\beta+\alpha)^{2m}] - 2b(\beta,\alpha,m)}, \tag{2.25}$$

with

$$\begin{aligned}a(\beta,\alpha,\kappa_\epsilon,m) &= m(1-\beta)^2 + 2m(m-1)\frac{(1-\beta-\alpha)^2(1-\beta^2-2\beta\alpha)}{(\kappa_\epsilon-1)[1-(\beta+\alpha)^2]} \\ &+ 4\frac{[m-1-m(\beta+\alpha)+(\beta+\alpha)^m][\alpha-\beta\alpha(\beta+\alpha)]}{1-(\beta+\alpha)^2},\end{aligned} \tag{2.26}$$

$$b(\beta,\alpha,m) = [\alpha - \beta\alpha(\beta+\alpha)]\frac{1-(\beta+\alpha)^{2m}}{1-(\beta+\alpha)^2}. \tag{2.27}$$

From equation (2.23) we see that the it is easy to convert the sum $\alpha+\beta$ from an aggregated (low frequency) model to the corresponding sum in a disaggregated (high frequency) model. This is not true for the individual parameters β or α, as we can see from (2.25) that low frequency β depends on high frequency β through a highly nonlinear relation which also involves other model parameters.

2.3.4 Deriving temporally aggregated GARCH processes: general procedures for GARCH(p,q) Process

Squared observations in a GARCH(p,q) model admit an ARMA(r,p) representation with $r = \max(p, q)$. The AR coefficients of the ARMA(r,p) model are equal to $\beta_i + \alpha_i$ and MA coefficients are equal to $-\beta_i$. In principle, calculating parameters in the aggregated GARCH processes can be done by applying the methods for finding parameters in the aggregated ARMA processes to the ARMA representation of the GARCH processes. However, an additional complication arising in the GARCH case is that the ARMA representation of the GARCH process is in terms of *squared* observations, i.e., $\{\epsilon_t^2\}$, and we need to find an ARMA representation for the squared aggregated process, say, $\{(\epsilon_t + \epsilon_{t-1})^2\}$, instead of $\{\epsilon_t^2 + \epsilon_{t-1}^2\}$. One way to accommodate the cross product terms like $2\epsilon_t\epsilon_{t-1}$ is to group the cross-product terms with the moving average terms. Under the symmetric GARCH assumption, the cross product terms have mean zero and are uncorrelated with the moving average terms.

Next, we first state and prove a theorem regarding the structure of aggregated GARCH(p,q) processes. The theorem we state is a special case of Theorem 1 of Drost and Nijman(1993)[16] and we simplify it to the case of interest to us, i.e. a pure GARCH for the flow variable case. Next, we use a numerical example to show how to derive the parameters in the aggregated GARCH(2,2) process. For a general GARCH(p,q) process, there does not seem to be any straightforward way to find parameters in the aggregated process.

Theorem 1 (Temporal aggregation of the GARCH(p,q) process, flow variable) Let $\{\epsilon_t\}$ be a weak GARCH(p,q) process following (2.9), i.e. $B(L)h_t = \psi + \{A(L) - 1\}\epsilon_t^2$. Then for any integer $m \geq 1$, the aggregated process at level m for a flow variable $\{\bar{\epsilon}_{(m)mt}\}$ is a weak GARCH($\lfloor \frac{(r+1)(m-1)+p}{m} \rfloor, r$) process.
Proof: First we write the GARCH process (2.9) in an ARMA form for its squared observations

$$C(L)\epsilon_t^2 = \psi + B(L)\nu_t,$$

where $C(L) = B(L) - A(L) + 1$, $A(L)$ and $B(L)$ are extended to a common order $r = \max(p, q)$ with $\alpha_i = 0$ for $i > q$ and $\beta_j = 0$ for $j > p$, and $\nu_t = \epsilon_t^2 - h_t$.

Without loss of generality, we assume $p = q = r$. Rewriting $C(L)$ in terms of its inverted roots δ_i's, we have

$$\prod_{i=1}^{p} (1 - \delta_i L)\epsilon_t^2 = \psi + B(L)\nu_t. \tag{2.28}$$

Then, multiplying both side of (2.28) by a polynomial $\prod_{i=1}^{p} (1 + \delta_i L + ... + \delta_i^{m-1} L^{m-1})(1 + L + ... + L^{m-1})$, we get

$$\begin{aligned}
&\prod_{i=1}^{p} (1 - \delta_i^m L^m)(\epsilon_t^2 + \epsilon_{t-1}^2 + ... + \epsilon_{t-m+1}^2) = \\
&\prod_{i=1}^{p} (1 + \delta_i + ... + \delta_i^{m-1}) m\psi \\
&+ \prod_{i=1}^{p} (1 + \delta_i L + ... + \delta_i^{m-1} L^{m-1})(1 + L + ... + L^{m-1}) B(L)\nu_t.
\end{aligned} \tag{2.29}$$

Next, adding $\prod_{i=1}^{p} (1 - \delta_i^m L^m) \cdot 2 \sum_{i=1}^{m-2} \sum_{j=i+1}^{m-1} \epsilon_{t-i}\epsilon_{t-j}$ to both sides of the last equation, we obtain

$$\begin{aligned}
&\prod_{i=1}^{p} (1 - \delta_i^m L^m)(\epsilon_t + \epsilon_{t-1} + ... + \epsilon_{t-m+1})^2 \\
&= \frac{\prod_{i=1}^{p} (1 - \delta_i^m)}{\prod_{i=1}^{p} (1 - \delta_i)} m\psi \\
&+ \prod_{i=1}^{p} (1 + \delta_i L + ... + \delta_i^{m-1} L^{m-1})(1 + L + ... + L^{m-1}) B(L)\nu_t \\
&+ \prod_{i=1}^{p} (1 - \delta_i^m L^m) \cdot 2 \sum_{i=0}^{m-2} \sum_{j=i+1}^{m-1} \epsilon_{t-i}\epsilon_{t-j}.
\end{aligned} \tag{2.30}$$

From the LHS of (2.30) the AR part of the aggregated process has an order p with inverted roots being the mth power of the corresponding inverted roots of the disaggregated process. For the RHS of (2.30), we notice that ν_{t-i} and $\epsilon_{t-i}\epsilon_{t-j}$ are uncorrelated terms. The polynomial multiplying ν_t has an order of $p \cdot (m-1) + (m-1) + p = (p+1)(m-1) + p$ and the polynomial multiplying $\epsilon_t\epsilon_{t-1}$ is of an order of $p \cdot m + 1$. Denote

$$\begin{aligned}
v_t = &\prod_{i=1}^{p} (1 + \delta_i L + ... + \delta_i^{m-1} L^{m-1})(1 + L + ... + L^{m-1}) B(L)\nu_t \\
&+ \prod_{i=1}^{p} (1 - \delta_i^m L^m) \cdot 2 \sum_{i=0}^{m-2} \sum_{j=i+1}^{m-1} \epsilon_{t-i}\epsilon_{t-j}.
\end{aligned} \tag{2.31}$$

By inspection we have that $\{v_t\}$ is an $((p+1)(m-1)+p)$-dependent sequence and thus admits an MA($\lfloor\frac{(r+1)(m-1)+p}{m}\rfloor$) representation on the m-scale. Therefore, a GARCH(p,q) process aggregated at level m admits an ARMA($p, \lfloor\frac{(r+1)(m-1)+p}{m}\rfloor$) representation for its squared observations and thus is a weak GARCH($\lfloor\frac{(r+1)(m-1)+p}{m}\rfloor, p$) process. And since we assume $p = q = r$, the aggregated process can also be denoted by GARCH($\lfloor\frac{(r+1)(m-1)+p}{m}\rfloor, r$). $\square$

2.4 Computing the Coefficients in Aggregated GARCH(p,q) Processes

2.4.1 Fourth moment of GARCH process and autocorrelation function of squared observations of the GARCH process

Existence of fourth moment

In the definition for the weak GARCH process, it is assumed that the GARCH process has a finite unconditional fourth moment. The finite fourth moment condition is also required to define the autocorrelation function of squared observations in a GARCH process. These quantities are used in computing parameters of temporally aggregated GARCH process. In addition, the finite fourth moment condition is needed for the limiting theorems involving squared observations of GARCH processes. Therefore, we summarize some relevant results on the fourth moment and the autocorrelation function of the squared observations of the GARCH processes.

He and Terasvirta (1999a)[34] derived an expression of the fourth moment and autocorrelation function of squared observations of GARCH(p,q) process. Based on the expression of the fourth moment, He and Terasvirta (1999a)[34] gave a necessary condition [1] for the existence of fourth moment. The expressions that He and Terasvirta (1999a)[34] derived involve expectation of products of random matrices with dimensions proportional to the orders of the model. For a GARCH(2,2) model, a relatively simple expression can be derived. But the expressions are difficult to evaluate for higher order GARCH models. Ling and McAleer (2002)[44] provided an necessary and sufficient condition for the existence of even order moments of GARCH(p,q) process with *non-negative* parameters.

Karanasos (1999)[38] derived a system of linear equations involving the fourth moments of GARCH(p,q) processes, which can be solved numerically to obtain the fourth moments. Similar to He and Terasvirta (1999a)[34], Karanasos (1999)[38] gave a necessary condition for the existence of fourth moments. Karanasos (1999)[38] also obtained a formal expression of the autocorrelation function of squared observations.

[1] He and Terasvirta (1999a) called their condition *necessary and sufficient*. However, as pointed out in Ling and McAleer (2002)[44], the conditions given in He and Terasvirta(1999a) was only *necessary*.

In summary, the literature has not provided any ready-to-evaluate expressions of the fourth moment and autocorrelation function of squared observations of GARCH(p,q) process with general orders. Also, the necessary and sufficient condition for the existence of fourth moment, as given in Ling and McAleer (2002)[44], applies only to GARCH(p,q) process with non-negative parameters.

Given the limitations of the literature, we make two contributions. First, we provide a numerical method for computing the fourth moment and the autocorrelation function of squared observations simultaneously by using the method of Brockwell and Davis (1991)[7] for computing variance and autocorrelation function of ARMA processes. Second, in the next chapter, we adapt the proof of Ling and McAleer (2002)[44] and extend their necessary and sufficient for the existence of fourth moment to a class of GARCH models with negative parameters.

Computation of fourth moment and autocorrelation of squared observations

We give a numerical approach to compute the fourth moment structure of GARCH process which is parallel to the method for computing variance and covariance of ARMA process as given in Brockwell and Davis (1991)[7]. Our method is straightforward to implement.

Let $\{\epsilon_t\}$ be a strong GARCH(p,p) process with i.i.d. innovations z_t. Assume that z_t have unit variance and kurtosis coefficient κ_z. As an example, we consider the case when z_t follows a standard normal distribution, $\kappa_z = 3$.

The ARMA representation of the squared observations in a GARCH(p,p) process is an ARMA(p,p) and is given by (2.14), which is rearranged as:

$$\epsilon_t^2 - \sum_{i=1}^{p}(\alpha_i + \beta_i)\epsilon_{t-i}^2 = \alpha_0 + \nu_t - \sum_{j=1}^{p}\beta_j\nu_{t-j}.$$

We can also write the squared process $\{\epsilon_t^2\}$ in an AR(∞) form:

$$\epsilon_t^2 = \alpha_0^* + \sum_{j=0}^{\infty}\psi_j\nu_{t-j}, \tag{2.32}$$

where $\alpha_0^* = \alpha_0/(1 - \sum_{i=1}^{p}(\alpha_i + \beta_i))$.

Multiplying both sides of (2.14) with ϵ_{t-k}^2 and taking expectation, we obtain

$$\begin{aligned}&E[\epsilon_t^2\epsilon_{t-k}^2]-(\alpha_1+\beta_1)E[\epsilon_{t-1}^2\epsilon_{t-k}^2]-\cdots-(\alpha_p+\beta_p)E[\epsilon_{t-p}^2\epsilon_{t-k}^2]\\&=E\left[\left(\alpha_0^*+\sum_{j=0}^{\infty}\psi_j\nu_{t-k-j}\right)\left(\alpha_0+\nu_t-\sum_{j=1}^{p}\beta_j\nu_{t-j}\right)\right]\end{aligned} \tag{2.33}$$

for $k=0,1,2,\cdots$.

Simplifying the expectation on the RHS of the last equation, we have

$$\begin{aligned}&E[\epsilon_t^2\epsilon_{t-k}^2]-(\alpha_1+\beta_1)E[\epsilon_{t-1}^2\epsilon_{t-k}^2]-\cdots-(\alpha_p+\beta_p)E[\epsilon_{t-p}^2\epsilon_{t-k}^2]\\&=\alpha_0^*\alpha_0-\frac{\kappa_z-1}{\kappa_z}E[\epsilon_t^4]\sum_{k\le j\le p}\beta_j\psi_{j-k}\end{aligned} \tag{2.34}$$

for $0\le k\le p$, where we used the fact that $E\nu_t^2=\frac{\kappa_z-1}{\kappa_z}E\epsilon_t^4$, and

$$E[\epsilon_t^2\epsilon_{t-k}^2]-(\alpha_1+\beta_1)E[\epsilon_{t-1}^2\epsilon_{t-k}^2]-\cdots-(\alpha_p+\beta_p)E[\epsilon_{t-p}^2\epsilon_{t-k}^2]=\alpha_0^*\alpha_0 \tag{2.35}$$

for $k\ge p+1$.

For $0\le k\le p$, we have a system of $p+1$ linear equations to solve for $p+1$ unknowns

$$E\epsilon_t^4, E[\epsilon_t^2\epsilon_{t-1}^2],\cdots,E[\epsilon_t^2\epsilon_{t-p}^2],$$

which can be used to further compute $E[\epsilon_t^2\epsilon_{t-k}^2]$ for $k\ge p+1$.

2.4.2 Computing temporal aggregation of higher order GARCH models

Using the methods described in the previous subsection for computing the fourth moment and autocorrelation of squared observations of GARCH processes, we can now compute the parameters of temporally aggregated GARCH processes by following the general procedure as described in Theorem 1. We illustrate this with an example on the temporal aggregation of a GARCH(2,2) process.

Example: Temporal aggregation of a GARCH(2,2) process in the case of a flow variable

We give a numerical example of aggregation of the GARCH(2,2) model. Consider the GARCH(2,2) process defined by

$$\begin{cases} \epsilon_t = \sqrt{h_t} z_t, \ \{z_t\} \overset{i.i.d.}{\sim} (0,1) \\ h_t = 0.08 + 0.05\epsilon_{t-1}^2 + 0.17\epsilon_{t-2}^2 + 0.3h_{t-1} + 0.4h_{t-2}. \end{cases}$$

We show how to derive the aggregate process for a flow variable at aggregation level $m = 2$, i.e. $\{\bar{\epsilon}_{(2)2t}\}$.
The ARMA representation of $\{\epsilon_t^2\}$ is:

$$\epsilon_t^2 = 0.08 + 0.35\epsilon_{t-1}^2 + 0.57\epsilon_{t-2}^2 + \nu_t - 0.3\nu_{t-1} - 0.4\nu_{t-2},$$

or

$$(1 - 0.95L)(1 + 0.6L)\epsilon_t^2 = 0.08 + (1 + 0.5L)(1 - 0.8L)\nu_t.$$

Multiplying this equation by $(1+0.95L)(1-0.6L)(1+L)$, which comes from (2.6), we obtain

$$\begin{aligned} &(1 - 0.9025L^2)(1 - 0.36L^2)(\epsilon_t^2 + \epsilon_{t-1}^2) \\ &= 0.1248 + (1 + 0.95L)(1 - 0.6L)(1 + 0.5L)(1 - 0.8L)(1 + L)\nu_t. \end{aligned}$$

Adding $(1 - 0.9025L^2)(1 - 0.36L^2) \cdot 2\epsilon_t\epsilon_{t-1}$ to both sides of the last equation to make a "complete square", we have

$$\begin{aligned} &(1 - 0.9025L^2)(1 - 0.36L^2)(\epsilon_t + \epsilon_{t-1})^2 \\ &= 0.1248 + (1 + 0.95L)(1 - 0.6L)(1 + 0.5L)(1 - 0.8L)(1 + L)\nu_t \\ &+ (1 - 0.9025L^2)(1 - 0.36L^2) \cdot 2\epsilon_t\epsilon_{t-1}. \end{aligned} \tag{2.36}$$

From the LHS of the last equation we have that

$$\alpha_{(2)1} + \beta_{(2)1} = 0.9025 + 0.36 = 1.2625 \text{ and } \alpha_{(2)2} + \beta_{(2)2} = -0.9025 \times 0.36 = -0.3249.$$

Denote

$$\begin{aligned} v_{2t} &= (1 + 0.95L)(1 - 0.6L)(1 + 0.5L)(1 - 0.8L)(1 + L)\nu_t \\ &+ (1 - 0.9025L^2)(1 - 0.36L^2) \cdot 2\epsilon_t\epsilon_{t-1} \\ &= (1 + 0.05L - 1.075L^2 + 0.031L^3 + 0.228L^4)(1 + L)\nu_t \\ &+ (1 - 1.2625L^2 + 0.3249L^4) \cdot 2\epsilon_t\epsilon_{t-1}. \end{aligned} \tag{2.37}$$

The process $\{v_{2t}\}$ is an MA(2) process, denoted by

$$v_{2t} = u_{2t} - \Theta_1 u_{2(t-1)} - \Theta_2 u_{2(t-2)},$$

where $\{u_{2t}\}$ denotes a white noise process [2], Θ_1 and Θ_2 are the solution of the system of equations

$$\begin{cases} \frac{\Theta_1(1+\Theta_2)}{1+\Theta_1^2+\Theta_2^2} = \frac{Cov(v_{2t},v_{2(t-1)})}{Var(v_{2t})} \\ \frac{\Theta_2}{1+\Theta_1^2+\Theta_2^2} = \frac{Cov(v_{2t},v_{2(t-2)})}{Var(v_{2t})}. \end{cases}$$

In this example,

$$\begin{aligned} Cov(v_{2t}, v_{2(t-1)}) &= [(-1.025)^2 + (-1.0962)^2 + (-0.2655)^2 + (-0.2380)^2] \cdot E\nu_t^2 \\ &+ [(-1.2625)^2 + (-0.4102)^2] \cdot 4 \cdot E[\epsilon_t^2 \epsilon_{t-1}^2], \\ Cov(v_{2t}, v_{2(t-2)}) &= [(0.259)^2 + (0.2394)^2] \cdot E\nu_t^2 + (0.3249)^2 \cdot 4 \cdot E[\epsilon_t^2 \epsilon_{t-1}^2], \\ Var(v_{2t}) &= [1^2 + (1.05)^2 + (-1.025)^2 + (-1.044)^2 + (0.259)^2 + (0.288)^2] \\ &\times E\nu_t^2 + [1^2 + (-1.2625)^2 + (0.3249)^2] \cdot 4 \cdot E[\epsilon_t^2 \epsilon_{t-1}^2], \end{aligned} \tag{2.38}$$

and we can calculate

$$E\nu_t^2 = 3.3174,\ E[\epsilon_t^2 \epsilon_{t-1}^2] = 1.7676.$$

Therefore,

$$\begin{cases} \frac{\Theta_1(1+\Theta_2)}{1+\Theta_1^2+\Theta_2^2} = -0.6119, \\ \frac{\Theta_2}{1+\Theta_1^2+\Theta_2^2} = 0.1177. \end{cases}$$

The solutions Θ_1 and Θ_2 give the GARCH coefficients in the aggregated model, i.e.,

$$\beta_{1(2)} = -\Theta_1 = 1.1070 \text{ and } \beta_{2(2)} = -\Theta_2 = -0.2706.$$

Consequently, the values of the ARCH parameters in the aggregated model, $\alpha_{(2)1}$ and $\alpha_{(2)2}$, can be calculated. The weak GARCH(2,2) representation of the process $\{\bar{\epsilon}_{(2)2t}\}$ is then

$$\bar{\epsilon}^2_{(2)2t} = 0.1248 + 1.2625\bar{\epsilon}^2_{(2)2(t-1)} - 0.3249\bar{\epsilon}^2_{(2)2(t-2)} + u_{2t} - 1.1070u_{2(t-1)} + 0.2706u_{2(t-2)}.$$

Observe that the aggregated model has negative parameters $\alpha_{(2)2} = -0.0543$ and $\beta_{(2)2} = -0.2706$. These results seem to contradict the commonly used non-negative

[2] Here $\{u_{2t}\}$ is a white process defined on the aggregated scale, i.e. the whole sequence is $\{..., u_{2(t-1)}, u_{2t}, u_{2(t+1)}, ...\}$

parameters constraints for GARCH parameters but negative parameters may well be allowed. □

2.5 Summary of Chapter

In this chapter, we explain the problem of temporal aggregation of the ARMA and GARCH processes and show how to derive the parameters of the aggregated low frequency model from the parameters in the original high frequency model. We also extend the results of Drost and Nijman (1993)[16] for computing temporal aggregation of the GARCH(1,1) model parameters to the general GARCH(p,q) model parameters.

Chapter 3

GARCH Models: Non-Negative Conditions, Singular Geometry, and the Component GARCH Models

In this chapter we review results on the parameter constraints for non-negative conditional variance in GARCH processes. We point out the parameter space of a GARCH process under general parameter constraints for non-negative variance may well have singular shapes which could cause serious difficulty to numerical estimation algorithms. To alleviate this problem, we suggest using the component GARCH models as an alternative parameterization of the GARCH processes. The component GARCH models are useful for capturing the dependency structure of empirical return volatility time series and the multiscale-type volatility dynamics.

3.1 Parameter Constraints for GARCH Models

Parameter constraints are among an important aspect of consideration for the estimation of GARCH models. If a GARCH model is estimated by maximum likelihood estimation, then the estimated conditional variances will not be negative *in sample* because any negative conditional variance will cause the likelihood function to reach negative infinity. However, this does not guarantee that the estimation procedure will not generate negative conditional variances *out of sample*.

It is customary to restrict all of the ARCH and GARCH parameters to be nonnegative to ensure that the conditional variance process of the model is nonnegative almost surely. However, as first pointed by Nelson and Cao(1992)[57], this is too restrictive. Nelson and Cao(1992)[57] showed that the necessary and sufficient conditions for a non-negative conditional variance can be substantially weakened, especially in higher order GARCH models. Subsequently, Tsai and Chan(2008)[71] gave more precise characterization of the constraints for a non-negative conditional variance.

More importantly, perhaps, the non-negative parameters constraints exclude the potentially best-fitting model. The class of component GARCH models as proposed by Ding and Granger (1996)[14] generally have negative parameters whereas the conditional variance process remains positive. The component GARCH models are useful for modeling the empirically observed long-memory type volatility dependency and are economically interpretable.

Last but not least, from a temporal aggregation perspective, negative GARCH model parameters may arise from the temporal aggregation of a GARCH model with all positive parameters.

However, simply imposing the necessary and sufficient conditions of Nelson and Cao (1992)[57] may cause difficulty to a numerical estimation algorithms. The Nelson-Cao constraints generally involves nonlinear inequality constraints, and it is well known that numerical optimization under nonlinear constraints can be problematic in this instance. We show with an example in the GARCH(2,2) model case that Nelson-Cao constraints can lead to a singularly shaped parameter space and thus cause a severe problem for numerical estimation algorithms to explore the whole sample space. To overcome this difficulty, we propose to use the component GARCH models as an alternative parameterization.

3.1.1 Necessary and sufficient conditions for a non-negative conditional variance

We first explain how the parameter constraints for a non-negative conditional variance are derived. The basic idea is to express the conditional variance h_t in an ARCH(∞) form. Then, a non-negativity of conditional variance h_t can be ensured by non-negativeness of all of the coefficients in the ARCH representation. As there is an infinite number of ARCH coefficients, it is difficult (if not impossible) to impose a set of infinite number of inequalities in practice. For this, Nelson and Cao(1992)[57] derived an equivalent set of finitely many constraints, which can be used in practice. In the case of GARCH(1,q) and GARCH(2,q) models, Nelson and Cao(1992)[57] showed that their conditions are necessary and sufficient. For the case of GARCH(p,q) where $p > 2$, Nelson and Cao(1992)[57] claimed that the conditions are sufficient. In fact, as shown in Tsai and Chan(2008)[71], the Nelson and Cao(1992)[57] conditions are still necessary and sufficient in the case of $p > 2$.

Consider the strong GARCH(p,q) process

$$\begin{aligned} \epsilon_t &= \sqrt{h_t} z_t \text{ where } \{z_t\} \overset{i.i.d.}{\sim} D(0,1) \\ h_t &= \alpha_0 + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} \end{aligned} \tag{3.1}$$

where $D(0,1)$ means some distribution with zero mean and unit variance.

Define

$$B(z) = 1 - \beta(z) = 1 - \sum_{i=1}^{p} \beta_i z^i \tag{3.2}$$

and denote the roots of $B(z) = 0$ by $\lambda_1, \cdots, \lambda_p$.

We can write the GARCH(p,q) model in an ARCH(∞) form:

$$\begin{aligned} h_t &= \left(1 - \sum_{i=1}^{p} \beta_i L^i\right)^{-1} \left[\alpha_0 + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2\right] \\ &= \alpha_0^* + \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}^2, \end{aligned} \tag{3.3}$$

where L is the lag operator and $\alpha_0^* = \alpha_0/(1 - \sum_{i=1}^{p} \beta_i)$.

Requiring

$$\alpha_0^* \geq 0 \text{ and } \psi_k \geq 0 \text{ for } k \geq 0 \tag{3.4}$$

will be sufficient to guarantee the nonnegativity of h_t almost surely. To make α_0^* and the ψ_k's well defined, we make the following assumptions:

Assumption (A1). The roots of the polynomial $\left(1 - \sum_{i=1}^{p} \beta_i z^i\right)$ (3.5)

are outside the unit circle,

and

Assumption (A2). The polynomials $\left(1-\sum_{i=1}^{p}\beta_i z^i\right)$ and $\sum_{j=1}^{q}\alpha_j z^{j-1}$ (3.6)

have no common roots.

Without loss of generality, assume the following about the roots of $B(z)=0$:

$$1<|\lambda_1|\leq|\lambda_2|\leq\cdots\leq|\lambda_p|. \tag{3.7}$$

Under these assumptions, the ψ_k's are well defined and finite. Denote $M=\max\{p,q\}$, $\alpha_j=0$ for $j>q$ and $\beta_i=0$ for $i>p$. Then $\psi_k, k=1,2,\cdots$, satisfy the system of equations

$$\begin{aligned}\psi_0 &= \alpha_1\\ \psi_1 &= \beta_1\psi_0+\alpha_2\\ \psi_2 &= \beta_1\psi_1+\beta_2\psi_0+\alpha_3\\ &\dots\\ \psi_{M-1} &= \beta_1\psi_{M-2}+\beta_2\psi_{M-3}+\dots+\beta_{M-1}\psi_0+\alpha_M,\ \text{and}\\ \psi_k &= \beta_1\psi_{k-1}+\beta_2\psi_{k-2}+\dots+\beta_M\psi_{k-M}\ \text{for}\ k\geq M.\end{aligned} \tag{3.8}$$

Notice that it is impractical to impose an infinite set of inequality constraints as in (3.8). The contribution of the Nelson-Cao constraints is to reduce the set of *infinite* number of constraints (3.8) to a set of a *finite* number of constraints which are necessary and sufficient for ensuring that the conditional variance process h_t is non-negative. The detailed proofs of this are given in Nelson and Cao (1992)[57] and Tsai and Chan (2008)[71]. We give the Nelson-Cao constraints and only sketch the idea of the proofs here.

The GARCH(1,q) Case: When (3.5) and (3.6) are satisfied, (3.4) holds if and only if

$$\alpha_0\geq 0, \beta\geq 0,\ \text{and}\ \psi_k\geq 0\ \text{for}\ k=0,1,...,q-1.$$

Sketch of proof: In the GARCH(1,q) case, we have $\beta_1=\beta$ and $\beta_i=0$ for $i\geq 2$. Since $\alpha_0^*=\alpha_0/(1-\beta)$ and $|\beta|=1/|\lambda_1|<1$, $\alpha_0^*\geq 0$ is equivalent to $\alpha_0\geq 0$ under the assumption (A1) and $\beta \geq 0$. The coefficients corresponding to (3.8) are

$$
\begin{aligned}
&\psi_0 = \alpha_1 \geq 0, \\
&\psi_1 = \beta\alpha_1 + \alpha_2 \geq 0, \\
&\psi_2 = \beta^2\alpha_1 + \beta\alpha_2 + \alpha_3 \geq 0, \\
&\cdots \\
\psi_{q-1} &= \beta^{q-1}\alpha_1 + \beta^{q-2}\alpha_2 + ... + \beta\alpha_{q-1} + \alpha_q \geq 0, \text{and} \\
&\psi_k = \beta^k\alpha_1 + \beta^{k-1}\alpha_2 + ... + \beta^{k+2-q}\alpha_{q-1} + \beta^{k+1-q}\alpha_q = \beta^{k+1-q}\phi_{q-1} \geq 0 \text{ for } k \geq q.
\end{aligned} \tag{3.9}
$$

The necessity and sufficiency of (3.9) for ensuring $\psi_k \geq 0$ for all k are both apparent. □

The GARCH(p,q) Case where $p \geq 2$: When (3.5) and (3.6) are satisfied,
1) $\alpha_0^* \geq 0$ if and only if $\alpha_0 \geq 0$;
2) Further assuming that the roots of $1-\beta(z) = 0$ are distinct, and that $1 < |\lambda_1| < |\lambda_2|$, (3.4) holds if and only if the following conditions hold:

$$\lambda_1 \text{ is real, and } \lambda_1 > 1, \tag{3.10}$$

$$\alpha(\lambda_1) > 0, \tag{3.11}$$

$$\psi_k \geq 0, \text{ for } k = 1, \cdots, k^*, \tag{3.12}$$

where k^* is the smallest integer greater than or equal to $\max\{0, \gamma\}$, where

$$\gamma^* = \frac{\log r_1 - \log((p-1)r^*)}{\log|\lambda_1| - \log|\lambda_2|},$$

$$r^* = \max_{2 \leq j \leq p} |r_j|,$$

and

$$r_j = -\frac{\alpha(\lambda_j)}{B^{(1)}(\lambda_j)}, 1 \leq j \leq p,$$

in which $B^{(1)}(z)$ is the first derivative of $B(z)$.
Sketch of proof: For part 1), it is true because $\alpha_0^* = \alpha_0/(1-\beta(1))$ and assumption (A1) implies that $1 - \beta(1) > 0$.

To prove the sufficiency of part 2), we start by writing ψ_k with $k \geq \max\{p, q\}$ by using partial fractions formula from Feller (1968)[23](pp. 276) as

$$\psi_k = \frac{r_1}{\lambda_1^{k+1}} + \frac{r_2}{\lambda_2^{k+1}} + \cdots + \frac{r_p}{\lambda_p^{k+1}}.$$

Rearranging the terms in the last equation, we obtain

$$\lambda_1^{k+1}\psi_k \geq r_1 - (p-1)r^* \left(\frac{\lambda_1}{|\lambda_2|}\right)^{k+1}. \tag{3.13}$$

If λ_1 is real and positive, and $r_1 \geq 0$, the first term on the RHS of (3.13) will be positive and it will dominate the rest of the terms on the RHS as $k \to \infty$. So that (3.13) only needs to hold for $k = 1, 2, \cdots, k^*$ where $k^* - 1$ makes (3.13) an equality. Solving this equality for k^*, one can easily see that $k^* = \gamma^*$. The condition $r_1 \geq 0$ can be simplified to $\alpha(\lambda_1) \geq 0$ since $-B^{(1)}(\lambda_1) = \prod_{j=2}^{p}(1 - \lambda_1/\lambda_j)/\lambda_1 > 0$. Assumption (A2) rules out $\alpha(\lambda_1) = 0$. So $r_1 \geq 0$ is equivalent to $\alpha(\lambda_1) > 0$.

The necessity of part 2) can be proved as follows. It is obvious that (3.12) is necessary. To prove the necessity of (3.10) and (3.11), we need to invoke an approximation formula from Feller (1968)[23](pp. 276 - 277) which tells us that, for $k \geq \max\{p, q\}$,

$$\psi_k = \sum_{i=1}^{p} \frac{r_i}{\lambda_i^{k+1}} \sim \frac{r_1}{\lambda_1^{k+1}},$$

in which $\sim$ indicates that the ratio of the two sides tends to 1 as $k \to \infty$. The approximation shows that the term r_1/λ_1^{k+1} will dominate for large k. Therefore, we must have λ_1 be real greater than 0, consequently greater than 1 by assumption (A2), and $r_1 \geq 0$. As shown earlier, $r_1 \geq 0$ is equivalent to $\alpha(\lambda_1) > 0$. $\square$

Remarks 1: In the GARCH(2,q) case, the variable k^* depends on the models parameters α_i's and β_j's since the variable γ^* does so. This means that we have a set of changing number of inequality constraints corresponding to $\psi_1, \cdots, \psi_{k^*}$. This may be practically undesirable.

Remarks 2: Substituting the expression of r_i into equation (3.13), we have

$$\lambda_1^{k+1}\psi_k \geq -\frac{\alpha(\lambda_1)}{B^{(1)}(\lambda_1)} - (p-1)\max_{2\leq i\leq p}\left|\frac{\alpha(\lambda_i)}{B^{(1)}(\lambda_i)}\right|\left(\frac{\lambda_1}{|\lambda_2|}\right)^{k+1}$$
$$= -\frac{1}{B^{(1)}(\lambda_1)}\sum_{j=0}^{q-1}\alpha_{j+1}\lambda_1^j - (p-1)\max_{2\leq i\leq p}\left|\frac{1}{B^{(1)}(\lambda_i)}\sum_{j=0}^{q-1}\alpha_{j+1}\lambda_i^j\right|\left(\frac{\lambda_1}{|\lambda_2|}\right)^{k+1}. \quad (3.14)$$

Observe that the first term on the RHS of the last equation,

$$-\frac{1}{B^{(1)}(\lambda_1)}\sum_{j=0}^{q-1}\alpha_{j+1}\lambda_1^j,$$

is positive, whereas the second term

$$-(p-1)\max_{2\leq i\leq p}\left|\frac{1}{B^{(1)}(\lambda_i)}\sum_{j=0}^{q-1}\alpha_{j+1}\lambda_i^j\right|\left(\frac{\lambda_1}{|\lambda_2|}\right)^{k+1}$$

is negative and its value is increasing (or decreasing in absolute value) as k increases. It is obvious that if $\psi_k \geq 0$ for $k = 1, \cdots, \max\{p,q\}$, $\psi_k \geq 0$ holds for all $k \geq \max\{p,q\}$. Thus, instead of setting $k^* = \max\{0, \gamma^*\}$, we can set $k^* = \max\{p,q\}$. This latter choice of k^* yields Theorem 2 of Nelson and Cao (1992)[57], which, together with the (3.10) and (3.11), specifies necessary and sufficient conditions for a non-negative condition variance for GARCH(2,q) model. Therefore, the necessary and sufficient condition is not unique in that the choice of k^* is not unique. The choice $k^* = \max\{0, \gamma^*\}$ may give a smaller value of k^* but it depends on the model parameter values. In contrast, the choice $k^* = \max\{p,q\}$ does not depend on the model parameter values.

Example 1: GARCH(1,2):
In the case of the GARCH(1,2) model, the Nelson-Cao constraints are

$$\alpha_0 \geq 0, \beta\alpha_1 + \alpha_2 > 0, 0 \leq \beta < 1.$$

□

Example 2: GARCH(2,1):
For the GARCH(2,1) model, the Nelson-Cao conditions are

$$\alpha_0 \geq 0, \alpha_1 \geq 0, \beta_1 \geq 0, \beta_1 + \beta_2 < 1, \beta_1^2 + 4\beta_2 \geq 0.$$

□

Example 3: GARCH(2,2):
Similar to the GARCH(2,1) case, the Nelson-Cao constraints for the GARCH(2,2) model are

$$\begin{aligned}
&\alpha_0 \geq 0, \alpha_1 \geq 0, \beta_1 \geq 0, \beta_1 + \beta_2 < 1, \\
&\beta_1^2 + 4\beta_2 \geq 0, \alpha_1 + \frac{2\alpha_2}{\beta_1 + \sqrt{\beta_1^2 + 4\beta_2}} > 0, \\
&\phi_1 = \beta_1\alpha_1 + \alpha_2 \geq 0, \phi_2 = \beta_1^2\alpha_1 + \beta_1\alpha_2 + \beta_2\alpha_1 \geq 0.
\end{aligned} \tag{3.15}$$

□

We observe that the Nelson-Cao constraints for a non-negative conditional variance become increasingly complicated with higher order GARCH models. In particular, the parameterization of the roots of the polynomial $1 - \beta(z)$ becomes more complicated with an increasing order of p. As we will show in the next section, in the GARCH(2,2) case, the constraints resulting from the constraints on the roots of $1 - \beta(z)$ could result in singularly-shaped geometry in the parameter space under a set of realistic parameter values which are important for the GARCH models to capture a quasi-long memory type dependency in volatility.

3.1.2 Projections of the GARCH(2,2) Parameter Space

In this section, we give some graphical illustration of the GARCH parameter space under the Nelson-Cao constraints. This not only helps illustrating the difference between the Nelson-Cao constraints and the commonly used non-negative parameter constraints, it also highlights the potential difficulties that the Nelson-Cao constraints can pose to estimation algorithms, e.g. MCMC sampling algorithms. To focus attention, ws consider the GARCH(2,2) model. In the GARCH(2,2) case, the dimension of the parameters is five. So we cannot plot the entire parameter space on a single plot. Instead, we can fix α or β and visualize the projection of the parameter space on the other variables.

Fixed α, projection on (β_1, β_2) space
We fix $\alpha = (\alpha_1, \alpha_2)$ at two sets of values. The first set of parameters is $\alpha = [.0573, .2262]$, which is taken from the estimated GARCH(2,2) model with Deutschmark/US Dollar FX return data as reported in Nelson and Cao (1992)[57]. The second set of parameters is $\alpha = [0.1101, -0.1087]$, which is taken from the

estimated two component GARCH model with S&P 500 return data from January 3, 1928 - August 30, 1991, and is reported in the GARCH(2,2) form, as reported in Ding and Granger (1996)[14]. Fig. 3.1 and Fig. 3.2 plot the projections under these two sets of α values, respectively. The (red) stars on the plots indicate the estimated β parameter values.

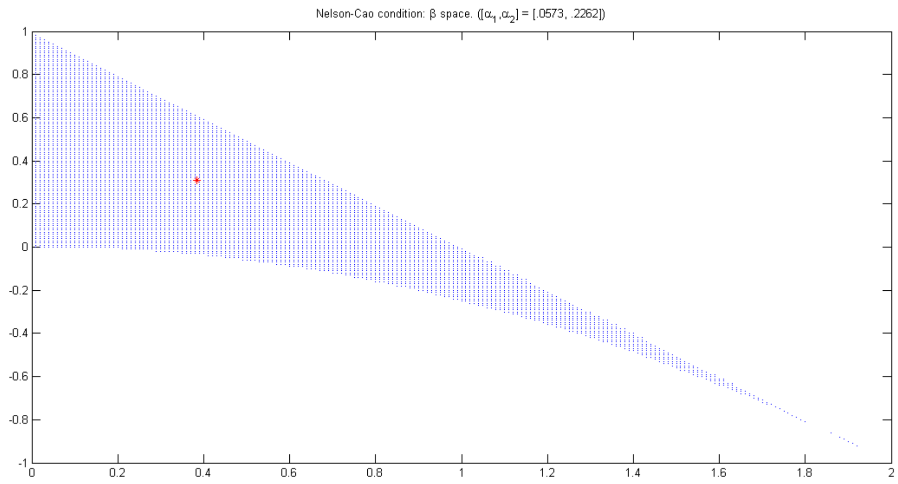

FIGURE 3.1: Projection of the GARCH(2,2) parameter space onto the (β_1, β_2)-plane using the parameter values as estimated in Nelson & Cao(1992)[57] on a sample of Mark/ Dollar exchange rates.

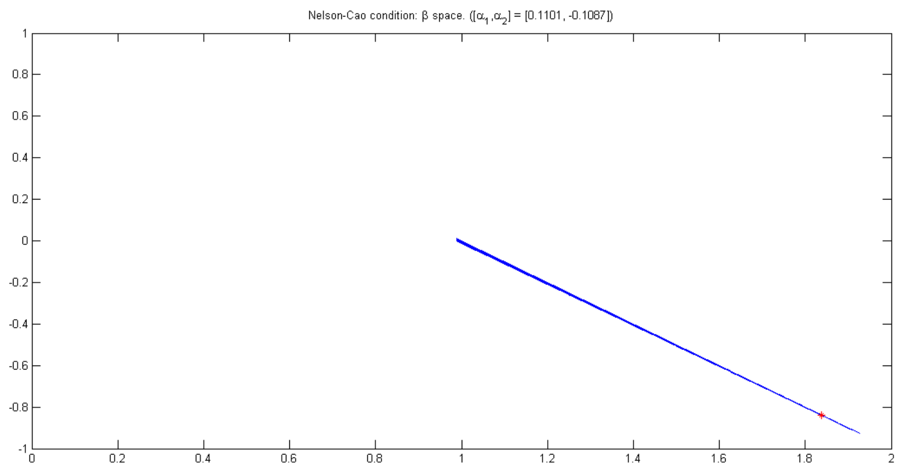

FIGURE 3.2: Projection of the GARCH(2,2) model parameter space onto the (β_1, β_2)-plane using the parameter values as estimated in Ding & Granger (1996)[14] on a sample of the S&P500 index.

Notice that the second set of α parameters has a negative α_2. Comparing the projections on the (β_1, β_2) space with different α values, we see that a negative α_2

parameter could significantly narrow the admissible region for the β parameters and push the 'best-fitting' β parameters to the boundary of the parameter space. This observation generally applies as we have tried other parameter values as well as some analytical analysis (not reported here).

Fixed β, projection on (α_1, α_2)

Similarly, we fix $\beta = (\beta_1, \beta_2)$ at two sets of values corresponding to the two cases for α above. The first set of parameters has $\beta = [.3833, .31]$, corresponding to the estimates with Deutschmark/US Dollar FX return data as reported in Nelson and Cao (1992)[57]. The second set of parameters has $\beta = [1.8380, -0.8394]$, corresponding to the β parameter values of a component-GARCH-equivalent GARCH(2,2) model estimated with S&P 500 return data from January 3, 1928 - August 30, 1991, as reported in Ding and Granger (1996)[14].

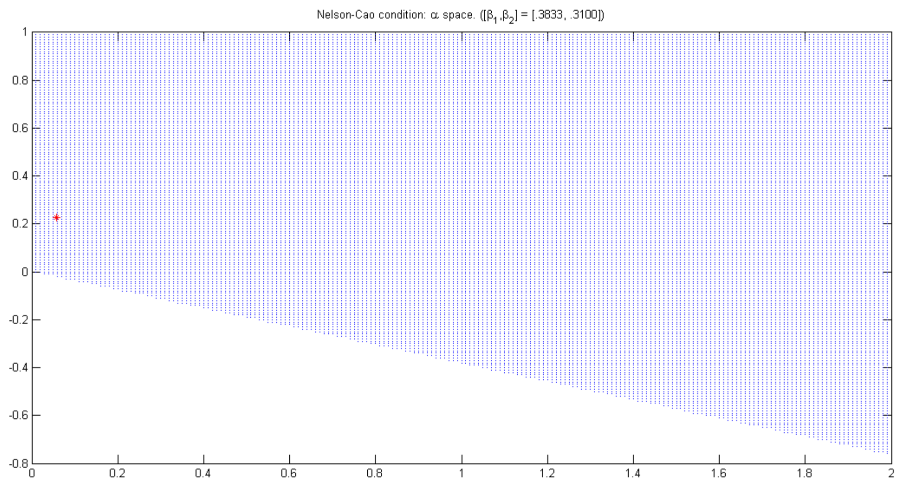

FIGURE 3.3: Projection of the GARCH(2,2) model parameter space onto the (α_1, α_2)-plane using the parameter values as estimated in Nelson & Cao (1992) on a sample of Mark/ Dollar exchange rates.

Fig.3.3 and Fig.3.4 plot the projections under these two sets of β values, respectively. The (red) stars on the plots indicate the estimated α parameter values. From the comparison of the (α_1, α_2) space projections, we see that a negative β_2 parameter does not necessarily narrow the admissible region for the α parameters. However, a negative β_2 also pushes the 'best-fitting' α parameters to the boundary of the parameter space, which is similar to the case with a negative α_2.

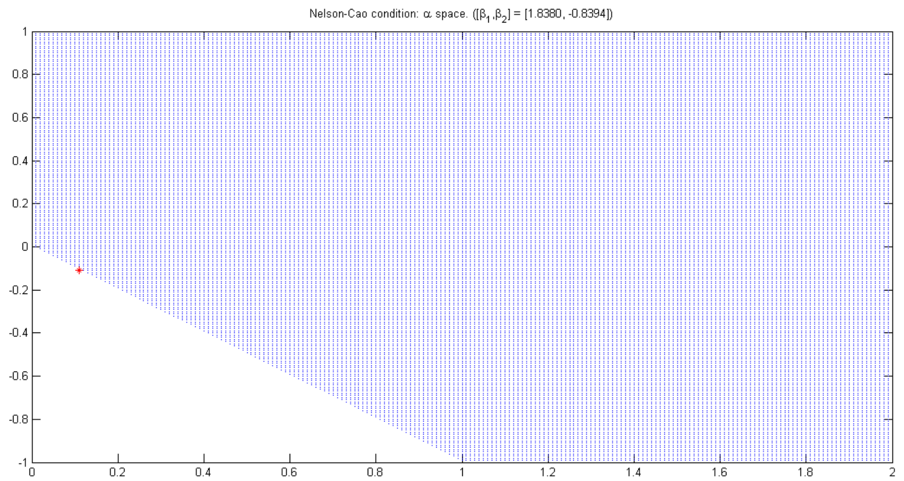

FIGURE 3.4: Projection of the GARCH(2,2) parameter space onto the (α_1, α_2)-plane using parameter values as estimated in Ding & Granger (1996) on a sample of the S&P500 index.

3.2 The Component GARCH Models of Ding & Granger (1996)[14] and Engle & Lee (1999)[21]

In this section, we summarize the component GARCH models proposed in the literature. As we will see, the component GARCH models generally have negative parameter values in their GARCH(p,q) representation. While these negative parameters lie in the corner regions under the GARCH(p,q) parameterization, the corresponding parameterization of the component GARCH model have more regular constraints. Therefore, we propose to use the component GARCH models as alternative parameterizations of the GARCH(p,q) models for exploring the corner regions of the parameter space.

First proposed in Ding and Granger(1996)[14], the class of component GARCH models has been found to be able to provide a better fit to empirical data than the benchmark GARCH(1,1) model in terms of sample ACFs of squared and absolute returns. Ding and Granger(1996)[14] found that the pattern of the sample ACFs of absolute and squared financial returns series are quite different from that of the theoretical ACF of the GARCH(1,1) model or Integrated GARCH(1,1) (IGARCH(1,1)) model. Prior to the introduction of this component GARCH model, Fractionally Integrated GARCH (FIGARCH) models as proposed in Baille, Bollerslev, and Mikelsen (1996)[3] have been used to model the hyperbolically decaying ACF observed in empirical data. However, FIGARCH models are generally difficult to estimate. Meanwhile, there is evidence pointing to volatility components that mean-revert at different speeds, which is known as *multiscale volatility*. See, for example, Fouque, et.al. (2011)[24]. Because of this, one may model volatility with multiple volatility components that evolve at different speeds (or time scales). Some of them may have a short life cycle but large effects on the overall volatility, while others may have a long life cycle but small intermediate effects. The component GARCH models serve both the purpose of capturing slowly decaying autocorrelation function of squared returns and modeling multiscale volatility.

The component GARCH models can be written as restricted GARCH(p,q) models and the corresponding GARCH(p,q) representation generally has negative parameters values. Empirically estimated component GARCH models have parameter values correspond to the GARCH(p,q) model with parameters lying in the corner regions of the parameter space which is difficult to explore under the GARCH(p,q) parameterization. Therefore, the component GARCH models can also be seen as a useful alternative parameterization of the GARCH(p,q) model which serves to mitigate the singular geometry problem caused by general constraints.

3.2.1 Ding and Granger's (1996)[14] parameterization

Two-component case

We start the discussion by considering the case of the two-component GARCH model as proposed in Ding and Granger (1996)[14]:

$$\begin{aligned} \epsilon_t &= \sqrt{h_t} z_t \text{ for } t = 1, ..., T, \\ z_t &\overset{iid}{\sim} N(0,1), \\ h_t &= w h_{1,t} + (1-w) h_{2,t}, \\ h_{1,t} &= \bar{\alpha}_1 \epsilon_{t-1}^2 + (1-\bar{\alpha}_1) h_{1,t-1}, \\ h_{2,t} &= \sigma^2 (1 - \bar{\alpha}_2 - \beta_2) + \bar{\alpha}_2 \epsilon_{t-1}^2 + \bar{\beta}_2 h_{2,t-1}. \end{aligned} \tag{3.16}$$

In this model, the overall variance of returns is modeled as a weighted sum of two components $h_{1,t}$ and $h_{2,t}$ with weights w and $1-w$, respectively. The first component $h_{1,t}$ is an IGARCH(1,1)-type specification and the second component $h_{2,t}$ is a GARCH(1,1)-type specification. Ding and Granger (1996)[14] show that the component type GARCH models are able to reproduce the long-memory type

hyperbolically decaying ACF observed in the empirical return data.
Expanding the two variance components h_{1t} and h_{2t}, we have

$$h_{1,t} = \bar{\alpha}_1 \sum_{k=1}^{\infty} (1 - \bar{\alpha}_1)^{k-1} \epsilon_{t-k}^2, \tag{3.17}$$

and

$$h_{2,t} = \sigma^2 \frac{1 - \bar{\alpha}_2 - \bar{\beta}_2}{1 - \bar{\beta}_2} + \bar{\alpha}_2 \sum_{k=1}^{\infty} \bar{\beta}_2^{k-1} \epsilon_{t-k}^2, \tag{3.18}$$

When $w\bar{\alpha}_1(1-\bar{\alpha}_1)^{k-1} > (1-w)\bar{\alpha}_2\bar{\beta}_2^{k-1}$, the first variance component has a larger effect on the overall variance than the second variance component. Consider, for example, the following set of parameter values which corresponds to the fitted parameters to a sample of S&P500 daily return studied by Ding and Granger(1996)[14]:

$$w = 0.704, \bar{\alpha}_1 = 0.153, \bar{\alpha}_2 = 0.008, \bar{\beta}_2 = 0.991, \text{ and } \sigma = 1.62e - 4.$$

The first component $h_{1,t}$ starts from $0.704 \times 0.153 = 0.1077$ and the second component starts from $0.296 \times 0.008 = 0.0024$. However, the first component decays much faster than the second component (0.847 vs. 0.991). Straightforward calculations show that $w\bar{\alpha}_1(1 - \bar{\alpha}_1)^{k-1} > (1 - w)\bar{\alpha}_2\bar{\beta}_2^{k-1}$ when $k \leq 25$, i.e. the first component $h_{1,t}$ has a larger effect on the overall volatility than the second component $h_{2,t}$ over the 1 day to 25 day horizon. Beyond that, the second component dominates. Therefore, we may interpret the fist component $h_{1,t}$ as capturing the short-run fluctuations of volatility and the second component $h_{2,t}$ as the long-run fluctuations of volatility.

N-component case

Ding and Granger (1996)[14] also generalized the above two-component GARCH to an N-component GARCH:

$$\begin{aligned}
\epsilon_t &= \sqrt{h_t} z_t \text{ for } t = 1, ..., T, \\
z_t &\overset{iid}{\sim} N(0, 1), \\
h_t &= \sum_{i-1}^{N} w_i h_{i,t}, \\
h_{i,t} &= \sigma^2(1 - \bar{\alpha}_i - \bar{\beta}_i) + \bar{\alpha}_i \epsilon_{t-1}^2 + \bar{\beta}_i h_{i,t-1}, \; i = 1, 2, \cdots, N,
\end{aligned} \tag{3.19}$$

where w_i is the weight for volatility component i. Since each volatility component $h_{i,t}$ follows a GARCH(1,1) structure, the parameter constraints $\bar{\alpha}_i > 0$, $\bar{\beta}_i > 0$, $\bar{\alpha}_i + \bar{\beta}_i < 1$ will ensure that $h_{i,t}, i = 1, \cdots, N$ is non-negative and thus the overall conditional variance process h_t is non-negative almost surely. In addition, we may assume that the first $N-1$ components to be integrated, i.e. $\bar{\alpha}_i + \bar{\beta}_i = 1$ for $i = 1, \cdots, N-1$. For an identification of the volatility components, we assume that $\bar{\beta}_1 < \cdots < \bar{\beta}_N$. These constraints follow Ding and Granger (1996)[14].

3.2.2 Engle and Lee's (1999)[21] parameterization

An alternative parameterization of the two-component GARCH model by Engle and Lee(1999)[21] can be motivated by first writing the standard GARCH(1,1) model as:

$$h_t = \sigma^2 + \alpha(\epsilon_{t-1}^2 - \sigma^2) + \beta(h_{t-1} - \sigma^2) \tag{3.20}$$

where σ^2 is the unconditional variance. Engle and Lee(1999)[21], based on previous empirical studies, postulate the existence of a time-varying unconditional variance process. Denote it by q_t, the time-varying unconditional variance component. Then, the conditional variance in the standard GARCH(1,1) model becomes

$$h_t = q_t + \alpha(\epsilon_{t-1}^2 - q_{t-1}) + \beta(h_{t-1} - q_{t-1}), \tag{3.21}$$

where the dynamics of q_t is specified as

$$q_t = \omega + \rho q_{t-1} + \phi(\epsilon_{t-1}^2 - h_{t-1}). \tag{3.22}$$

By writing (3.21) as

$$h_t - q_t = \alpha(\epsilon_{t-1}^2 - q_{t-1}) + \beta(h_{t-1} - q_{t-1}), \tag{3.23}$$

we call $(h_t - q_t)$ the *short-run* or *transitory* volatility component and q_t the *long-run* volatility component. Both volatility components are driven by $(\epsilon_{t-1}^2 - h_{t-1})$. Let $s_t := h_t - q_t$. Then the volatility components can be written in a symmetric form as

$$
\begin{aligned}
h_t &= q_t + s_t, \\
s_t &= \alpha(\epsilon_{t-1}^2 - h_{t-1}) + (\alpha + \beta)s_{t-1}, \\
q_t &= \omega + \phi(\epsilon_{t-1}^2 - h_{t-1}) + \rho q_{t-1}.
\end{aligned} \tag{3.24}
$$

Under this representation, we see that the short-run variance mean-reverts around zero when $0 < (\alpha + \beta) < 1$. The long-run variance dynamics has an AR(1) form when $0 < \rho < 1$ and converges to a constant level $\omega/(1 - \rho)$. Engle and Lee (1999)[21] assume that the long-run variance has a slower mean-reverting rate than the short-run variance, i.e. $0 < (\alpha + \beta) < \rho < 1$. This serves as an identifiability condition. In addition, they expect the immediate impact of the short-run component to be greater than that of the long-run component.[1] So they also impose the restriction that $\alpha > \phi$. Thus, Engle and Lee (1999)[21] used the following set of parameter constraints for their model:

$$
0 < (\alpha + \beta) < \rho < 1, \alpha > \phi > 0, \beta > 0, \phi > 0, \omega > 0. \tag{3.25}
$$

They showed that this set of conditions satisfies the Nelson-Cao constraints and thus is sufficient to guarantee the nonnegativity of conditional variances.

3.2.3 Comparisons of Ding & Granger and Engle & Lee parameterization in the 2-component case

GARCH(2,2) Representation

It will be helpful for our discussion to rewrite the component GARCH models of both Ding and Granger (1996)[14] and Engle and Lee (1999)[21] in the GARCH(2,2) form.

For Ding and Granger's component GARCH model, write the volatility components using the lag operator notation as

$$
\begin{aligned}
h_{1,t} &= \frac{\bar{\alpha}_1}{1 - (1 - \bar{\alpha}_1)L}\epsilon_{t-1}^2, \\
h_{2,t} &= \frac{\sigma^2(1 - \bar{\alpha}_2 - \bar{\beta}_2)}{1 - \bar{\beta}_2} + \frac{\bar{\alpha}_2}{1 - \bar{\beta}_2 L}\epsilon_{t-1}^2.
\end{aligned} \tag{3.26}
$$

[1] This agrees with the fitted model with S&P500 daily returns studied in Ding and Granger (1996)[14].

Then, substituting $h_{1,t}$ and $h_{2,t}$ into h_t, we have

$$\begin{aligned} h_t &= \frac{(1-w)\sigma^2(1-\bar{\alpha}_2-\bar{\beta}_2)}{1-\bar{\beta}_2} + \left[\frac{w\bar{\alpha}_1}{1-(1-\bar{\alpha}_1)L} + \frac{(1-w)\bar{\alpha}_2}{1-\bar{\beta}_2 L}\right]\epsilon_{t-1}^2, \\ &= \frac{(1-w)\sigma^2(1-\bar{\alpha}_2-\bar{\beta}_2)}{1-\beta_2} + \frac{w\bar{\alpha}_1(1-\bar{\beta}_2 L)+(1-w)\bar{\alpha}_2[1-(1-\bar{\alpha}_1)L]}{[1-(1-\bar{\alpha}_1)L](1-\bar{\beta}_2 L)}\epsilon_{t-1}^2, \end{aligned} \tag{3.27}$$

which is a restricted GARCH(2,2) model:

$$\begin{aligned} h_t =& \sigma^2(1-w)\bar{\alpha}_1(1-\bar{\alpha}_2-\bar{\beta}_2) + [w\bar{\alpha}_1+(1-w)\bar{\alpha}_2]\epsilon_{t-1}^2 \\ & - [w\bar{\alpha}_1\bar{\beta}_2+(1-w)(1-\bar{\alpha}_1)\bar{\alpha}_2]\epsilon_{t-2}^2 \\ & + (1-\bar{\alpha}_1+\bar{\beta}_2)h_{t-1} - (1-\bar{\alpha}_1)\bar{\beta}_2 h_{t-2}. \end{aligned} \tag{3.28}$$

The sum of the ARCH and GARCH parameters (which equals $1-(1-w)\bar{\alpha}_1(1-\bar{\alpha}_2-\bar{\beta}_2)$) is bigger than zero and less than one when $0 < w < 1$, $0 < \bar{\alpha}_1 < 1$, and $0 < \bar{\alpha}_2 + \bar{\beta}_2 < 1$. Under these conditions, the process $\{\epsilon_t\}$ is covariance-stationary with

$$Eh_t = Eh_{1,t} = Eh_{2,t} = E\epsilon_t^2 = \sigma^2. \tag{3.29}$$

An interesting point to note is that although the ARCH(2) and GARCH(2) coefficients are negative, the variance processes are still guaranteed to be positive. These negative coefficients are usually not considered in the specification of the GARCH(p,q) models. However, we see in the current case that negative coefficients might arise quite naturally and do not necessarily lead to nonstationarity. More importantly, as pointed out by Ding (2016)[13], restricting the parameters to be positive will likely exclude better fitting models. Similar derivations to Ding and Granger's model show that Engle and Lee's model (3.24) can also be written in the GARCH(2,2) form:

$$\begin{aligned} h_t &= (1-\alpha-\beta)\omega + (\alpha+\phi)\epsilon_{t-1}^2 + [-\phi(\alpha+\beta)-\alpha\rho]\epsilon_{t-2}^2 \\ &\quad + (\rho+\beta-\phi)h_{t-1} + [\phi(\alpha+\beta)-\beta\rho]h_{t-2}. \end{aligned} \tag{3.30}$$

Mapping between parameters in Ding-Granger and Engle-Lee models

As stated in Ding (2016)[13], it is useful to point out that one can establish mappings among the parameters in Ding and Granger (1996)[14] parameterization, in Engle and Lee (1999)[21] parameterization, and that in the GARCH(2,2) parameterization. Denote by a_1, $-a_2$ the ARCH parameters in the GARCH(2,2) representations of the component GARCH models and by b_1, $-b_2$ the GARCH parameters. That is,

$$\begin{aligned} a_1 &= w\bar{\alpha}_1 + (1-w)\bar{\alpha}_2 = \alpha + \phi \\ a_2 &= w\bar{\alpha}_1\bar{\beta}_2 + (1-w)(1-\bar{\alpha}_1)\bar{\alpha}_2 = \alpha\rho + (\alpha+\beta)\phi \\ b_1 &= 1 - \bar{\alpha}_1 + \bar{\beta}_2 = \beta + \rho - \phi \\ b_2 &= (1-\bar{\alpha}_1)\bar{\beta}_2 = \beta\rho - (\alpha+\beta)\phi. \end{aligned} \tag{3.31}$$

For example, given a set of parameters in the GARCH(2,2) parameterization, the corresponding parameters in Engle and Lee's parameterization are given by

$$\begin{aligned} \rho &= \frac{1}{2}\left((a_1+b_1) \pm \sqrt{(a_1+b_1)^2 - 4(a_2+b_2)}\right) \\ \alpha+\beta &= \frac{1}{2}\left((a_1+b_1) \mp \sqrt{(a_1+b_1)^2 - 4(a_2+b_2)}\right) \\ \phi &= \frac{a_2 - a_1\rho}{\alpha+\beta-\rho} \\ \alpha &= a_1 - \phi \\ \beta &= \frac{1}{2}\left((a_1+b_1) \mp \sqrt{(a_1+b_1)^2 + 4(a_2+b_2)}\right) - \alpha. \end{aligned} \tag{3.32}$$

3.2.4 Modeling multiscale volatility

The component volatility models are useful for modeling the so-called *multiscale volatility*, which is an empirical phenomenon as elaborated in Chapter 3 of Fouque, et.al. (2011)[24]. Intuitively, it refers to the empirical observations that financial volatilities have a mean-reverting behavior at two or more time scales with different and, often, well-separated rates. Component type volatility models are natural candidates for modeling multiscale volatility. See, for example, Barndorff-Nielsen and Shephard (2001)[4], Chernov, et.al. (2003)[10], Christoffersen, et.al. (2008)[11], as well as Fouque, et.al. (2011)[24]. In these models, the temporal dependency structure of volatility is parameterized with multiple parameters representing the multiple decaying rates of the components.

The two component GARCH models is one of the simplest multiscale volatility model. From the estimated parameters on a S&P500 index sample as reported in Ding and Granger (1996)[14], we observe that innovations to the volatility components have well-separated decay rates $\bar{\beta}_1 = 1 - \bar{\alpha}_1 = .847$ and $\bar{\beta}_2 = .991$, respectively. Similar estimates from another S&P 500 index sample and from individual stock returns can be found in Engle and Lee (1999)[21]. In our empirical study of the component GARCH models presented in the next section, we find that estimated component GARCH models on the DJIA return sample as used in Engle and Patton (2001)[22] also have components with well-separated decay rates with respect to innovations.

3.3 An empirical study of the component GARCH models on a Dow Jones Industrial Average (DJIA) index return sample

In this section, we provide an empirical illustration of the estimation of the component GARCH models on a DJI return sample as used in Engle and Patton (2001)[22]. Although there exists several statistical software packages for estimating GARCH(p,q) models (with non-negative parameter constraints), to the best of our knowledge, not much attention has been paid to GARCH models with general parameter constraints, nor is there attention to the component GARCH models. The **R** package **rugarch** implements the two-component GARCH of Engle and Lee (1999)[21].

The **R** package **RStan** provides an efficient and convenient framework to estimate the component GARCH models using Bayesian MCMC estimation. It requires only a simple model specification and conducts an automated efficient MCMC sampling.

We use the sample of Dow Jones Industrial Average (DJIA) index daily returns from Aug. 23, 1988 through Aug. 22, 2000, yielding a total of 3,131 observations of daily adjusted closing prices [2]. The sample size of daily log returns is T = 3,130. Figure 3.5 plots the return sample.

We fit three models from the component GARCH family to the sample: 1. a GARCH(1,1) model, 2. a two component GARCH model, and 3. a three component GARCH model. We assume a normal distribution for the innovations in

[2] A stock's adjusted closing price is the daily close price amended to include any distributions and corporate actions that occurred at any time prior to the next day's open. In the analysis of historical returns, adjusted closing price are often used instead of closing price.

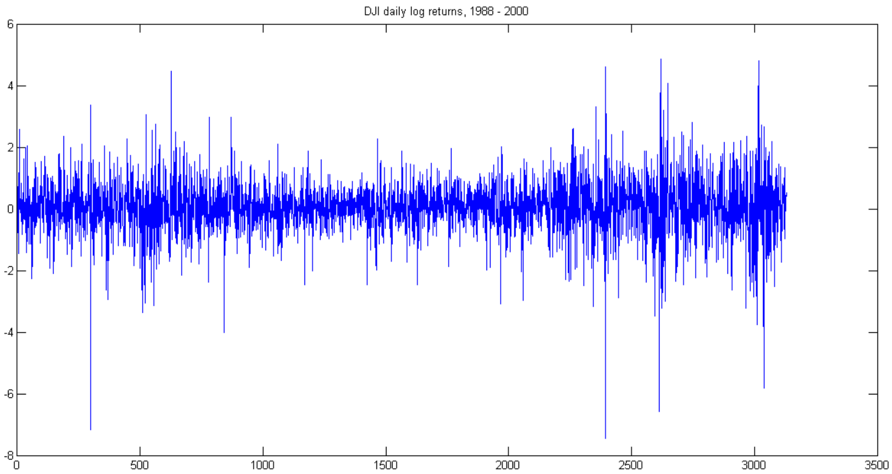

FIGURE 3.5: Sample returns on the Dow Jones Industrial Average index.

each of the three models as in Engle and Patton (2001)[22]. As to the prior distribution, we use normal priors with mean 0 and variance of 100 for all of the model parameters. Such choices lead to a very weak prior impact.

Table 3.1 reports posterior means and 95% posterior credible intervals (i.e. 2.5% and 97.5% quantiles of the posterior distribution) of the parameters in the three GARCH models considered. We observe that the two component model have cleanly separated parameter values $\bar{\alpha}_1$ and $\bar{\alpha}_2$, which means that the two volatility components are cleanly separated in terms of persistence. Similarity, the $\bar{\alpha}_1$, $\bar{\alpha}_2$, and $\bar{\alpha}_3$ parameters in the three component GARCH model are also well-separated in terms of their posterior means. The posterior interval estimates of the parameters in the three component model do have overlaps, which implies that the sample under study does not contain enough information to separate all the three components as cleanly as in the one- and two-component models. We therefore do not pursue estimating models with larger numbers of components, although the estimation of higher order models can be carried out similarly in principle on longer samples.

Last but not least, we conduct white noise tests on the standardized residuals resulting from the three estimated GARCH models in order to assess whether they have adequately captured the dependency in volatility. Following Engle and Patton (2001)[22], we use the Ljung-Box Q-statistics at lags up to lags 60. For all of the three models, the p-values of the tests are above 0.5, meaning that there

TABLE 3.1: Posterior inference results of the parameters in the GARCH models

	GARCH(1,1)	Two-component	Three-component
μ	0.061 (.032, .088)	0.063 (.035, .090)	0.063 (.035, .091)
w_1		0.43 (.28, .57)	0.18 (.0073, .46)
w_2			0.30 (.023, .56)
σ^2	1.07 (.70, 2.51)	2.45 (.61, 8.87)	1.81 (.60, 5.38)
$\bar{\alpha}_1$	0.04 (.028, .057)	0.14 (.083, .22)	0.26 (.10, .56)
$\bar{\alpha}_2$		0.0067 (.0037, .013)	0.091 (.0062, .18)
$\bar{\alpha}_3$			0.0063 (.0025, .012)
$\bar{\beta}_N$	0.95 (.93, .97)	0.993 (.986, .996)	0.9925 (.985, .996)

$(\cdot, \cdot)$ are 95% credible intervals.

is no evidence against the null hypothesis of a zero correlation in the standardized residuals. Therefore, all the three models estimated in our study adequately capture dependency in the DJI volatility.

3.4 Summary of Chapter

In this chapter we have discussed the issue of parameter constraints for GARCH processes. In particular, we have shown that negative parameter values, which may result from temporal aggregation, could lead to a singular geometry in the parameter space and constitute practical difficulties to parameter estimation. As a solution to the potential singular geometry problem, we propose to use the component GARCH models as a re-parameterization tool for some practically relevant cases, which also have a multiscale volatility interpretation.

Chapter 4

Empirical Likelihood-Based Two-scale Tests of Scale-Consistency for ARMA Processes

One of the most important themes of this chapter is that the empirical likelihood framework provides us with a way of formally testing whether a time series model is compatible with data at two different sampling frequencies.

4.1 Chapter Introduction

In this chapter, we introduce a class of tests for the *scaling* property of the **linear dependency structure** of ARMA processes, called the *Multiscale Tests of Scale-Consistency*. The tests exploit the temporal aggregation relation of time series models and **are designed to test whether a given weak ARMA structure is consistent with a time series sampled at multiple frequencies**. An important application of the test is to test whether a high frequency volatility model is also consistent with a low frequency return sample[1]. The proposed tests are based on the weak ARMA structure which is a general stationary non-deterministic process. The linear dependency structures are tested without relying

[1]As we will see, our current version of the tests uses low frequency model representations derived from the high frequency model. The low frequency model representations are only *necessary conditions* for testing whether the low frequency model is also fully sufficient for the low frequency data.

on any parametric distributional assumptions. The resulting tests **exploit information from multiple frequencies of samples and are generally more powerful than the corresponding tests based on one frequency of sample. They can be used to complement the usual statistical tests for ARMA type processes, such as white noise tests for residuals.**

The idea of using information in *samples at multiple frequencies* to test model specification has been found useful by many authors. Lo and MacKinlay (1988)[46] proposed a variance ratio test of a random walk process which is based on the ratio of variances of a series at different scales. Mandelbrot, et.al.(1997)[47] suggested that reliance upon a single time scale may lead to forecasts which vary with sample frequencies and proposed a model called a Multifractal Model of Asset Returns to capture the moment scaling property observed in exchange rate returns. More recently, Ohanissian, et.al.(2008)[60] proposed a test of long-memory which is based on the invariance property of the fractional integration parameter with respect to temporal aggregation. Our proposed tests show that the exploration of information from multiple frequency samples can also be beneficial for a search for the specification of ARMA-type models.

According to the Wold Decomposition Theorem of time series, the weak ARMA structure is a basic structure applicable to any stationary non-deterministic process. It also exists in some commonly used nonlinear processes, such as GARCH processes. In particular, the squared observations in a ARCH(p) process is a weak AR(p) process, and the squared observations in a GARCH(p,q) process is a weak ARMA process.

Our proposed test is based on the framework of empirical likelihood (EL) which is implemented through a set of estimating equations. The empirical likelihood framework allows us to carry out **likelihood type inference** without specifying a distributional model. It leads to test statistics with asymptotic distributions analogous to their fully parametric likelihood counterparts and data-determined confidence regions. The empirical likelihood framework can also be conveniently implemented through a set of estimating equations, which is a very general way of estimating parameters of statistical and time-series models.

We form estimating equations over multiple sampling frequencies based on the temporal aggregation relations. The basic estimating equations corresponding to a single sampling frequency may already be enough to identify a model under the usual single scale inference procedures. However, we add auxiliary estimating equations corresponding to a second sampling frequency to construct more powerful tests which test, in particular, whether a model is compatible with data simultaneously at two different scales. If the null hypothesis, that a model is compatible with data simultaneously at two different scales, is rejected, our proposed test could suggest to the user of the model to increase the autoregressive and/or moving average orders, which could increase the flexibility of the model and thus to better capture the scaling property of the data in terms of its linear dependency structure.

A major motivation for the multiscale tests of scale-consistency is the "volatility half-life puzzle" pointed out in Engle and Patton (2001)[22]. Our proposed test can be used to formally address the following questions:

- Is the QMLE based on a high frequency sample also consistent with data at lower frequencies?
- Is a particular model, like a GARCH model, able to simultaneously fit data at two different time scales?

The proposed testing procedure can be applied to general ARMA and GARCH type processes where temporal aggregation is performed.

Fig. 4.1 provides an overview of our proposed testing procedure.

The rest of the chapter is organized as follows. In section 4.2 we show how to form the corresponding estimating equations. In section 4.3 we give a general description of our proposed tests for the ARMA(p,q) processes. In section 4.4 we give some asymptotic results for the proposed tests. Section 4.5 discusses some computational details. In section 4.6 we present some simulation studies to assess the finite sample performances of our proposed tests. Section 4.7 concludes.

Some background materials on the empirical likelihood are given in Appendix A.

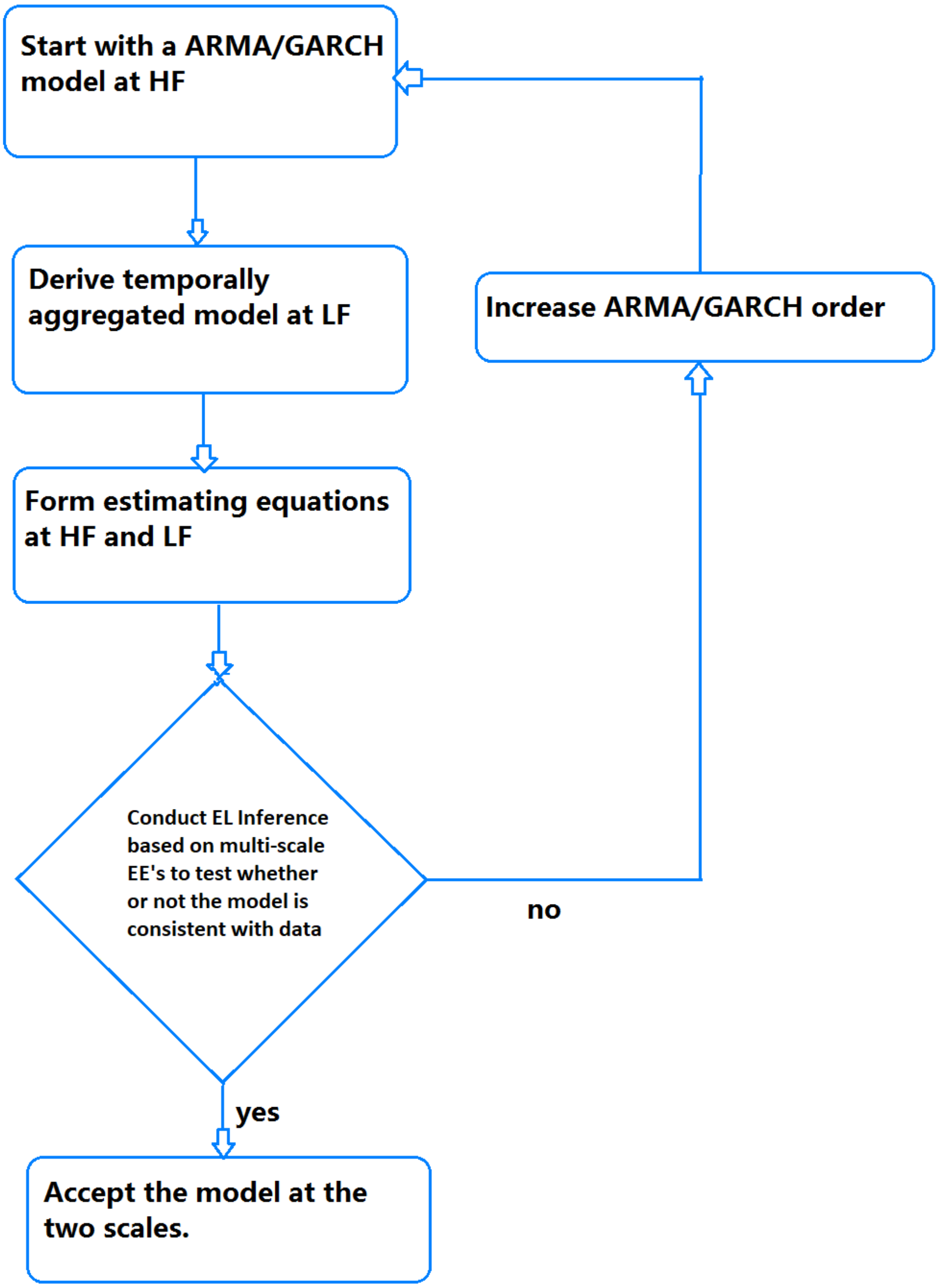

Figure 4.1: Flowchart of multiscale testing procedure.

4.2 Estimating Equations for ARMA(p,q) Processes at Different Time Scales

Let us consider a stationary strong ARMA(p,q) process with mean $\mu = \phi_0/(1 - \phi_1 - \cdots - \phi_p)$:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \dots \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \tag{4.1}$$

where $\{Z_t\}$ is a sequence of *i.i.d.* random variables defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with zero mean and common variance σ_Z^2, denoted as $D(0, \sigma_Z^2)$. We also assume that the AR polynomials $\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and MA polynomials $\Theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ have no common zeros. We assume that $\Phi(z)$ has all roots outside the unit circle so that the ARMA process $\{X_t\}$ is causal and second-order stationary.

We denote by m_1 the level of aggregation of the (highest frequency) observations with respect to the data generating level at which the model (4.1) is assumed. For example, if the true data generating process (DGP) (4.1) operates on the daily scale and we only have weekly observations, then $m_1 = 5$. We further denote by m_2 a second level of aggregation which is higher than m_1, i.e. $m_2 > m_1$, where m_2 is an integer multiple of m_1. Using these notations, $m_1 = 1$ if we have observations at the data generating level. We call m_1 a high frequency (HF) and m_2 a low frequency (LF).

To facilitate our presentation, we assume that $m_1 = 1$ unless otherwise specified. We also use the notation m to denote a generic level of aggregation. We would mostly consider $m_1 = 1$ and $m_2 = m$. But we keep the m_1 and m_2 notations to allow for a further generalization. The cases of general m_1 can be derived relatively straightforwardly.

An **estimating equations (EE)** approach defines how the parameters of a statistical model should be estimated. Consider a random vector $X \in \mathbb{R}^d$ following a distribution function F with an unknown s-dimensional parameter $\boldsymbol{\theta} \in \mathbb{R}^s$, and a real, r-dimensional vector-valued function $g(x, \boldsymbol{\theta}) \in \mathbb{R}^r$ given by

$$g(x, \boldsymbol{\theta}) = [g_1(x, \boldsymbol{\theta}), \cdots, g_r(x, \boldsymbol{\theta})]'. \tag{4.2}$$

Suppose that information about both data and F is contained in the following equation

$$E(g(X, \boldsymbol{\theta})) = 0, \tag{4.3}$$

where the expectation is taken with respect to the random vector X with distribution parameterized by θ.

We assume that $g_1(X, \boldsymbol{\theta}), \cdots, g_r(X, \boldsymbol{\theta})$ have non-degenerate and invertible variance-covariance matrix.

The parameter $\boldsymbol{\theta}$ can be estimated by solving

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i, \hat{\boldsymbol{\theta}}) = 0 \tag{4.4}$$

for $\hat{\boldsymbol{\theta}}$, where $X_1, ..., X_n$ is a random sample from $F_{\boldsymbol{\theta}}$.

Equation (4.4) is called an *estimating equation* and $g(x, \boldsymbol{\theta})$ is called an *estimating function*.

We need at least as many equations as the number of parameters, i.e. $r \geq s$. When $r = s$, and under the following conditions on $g(x, \boldsymbol{\theta})$ and the distribution F given by Godambe (1960)[28], equation (4.3) has a solution with respect to $\boldsymbol{\theta}$. We shall denote this root as $\boldsymbol{\theta}_0$ and refer to it as the "true value".

Conditions on the estimating equations:
(i) $E[g(X, \boldsymbol{\theta}) : \boldsymbol{\theta}] = 0$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$;
(ii) for almost all x, $\partial g/\partial \boldsymbol{\theta}$ exists for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$;
(iii) $\int g(x, \boldsymbol{\theta}) p(x, \boldsymbol{\theta}) dx$ is differentiable under the integral sign where $p(x, \boldsymbol{\theta})$ is the density function of X;
(iv) $[E(\partial g/\partial \boldsymbol{\theta}(X) : \boldsymbol{\theta})]^2 > 0$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Following Wirjanto (1997)[74] and Smith (2011)[68], among others, we assume that the focus is on a unique $\boldsymbol{\theta}_0$ which satisfies $E[g(X, \boldsymbol{\theta}_0)] = 0$[2]. For general methods of dealing with the potential problem of multiple root problems in estimation, we refer readers to Small, et.al. (2000)[67].

Well-known examples of estimating equations include those corresponding to the method of moments and the maximum likelihood estimator. Since we will be using estimating equations based on temporally aggregated models, and likelihood

[2] Otherwise, we would focus on the most practically meaningful $\boldsymbol{\theta}_0$.

assumptions are generally not closed under temporal aggregation, we consider estimating equations corresponding to the method of moments.

In general, we seek estimating equations that are not linearly dependent, so that the variance-covariance matrix of the estimating functions is invertible, and in consequence the variance-covariance of the limiting distribution of the estimator is well-defined.

For more comprehensive treatments of the topic of estimating equations, we refer the reader to Godambe (1991)[29], Godambe and Heyde (1987)[30], and McLeish and Small(1988)[51].

Example 4.1: Single-scale estimating equations for the ARMA(1,1) process

Consider the model (4.1) with $p = 1$ and $q = 1$. A possible choice of estimating functions $g(X, \boldsymbol{\theta})$ satisfying $E[g(X, \boldsymbol{\theta})] = 0$ for estimating the parameter vector $\boldsymbol{\theta} := (\phi_0, \phi_1, \theta_1, \sigma_Z^2)$ can be based on the following equalities:

$$\begin{aligned} E[X_t - \phi_0 - \phi_1 X_{t-1}] = 0, \\ E[(X_t - \phi_0 - \phi_1 X_{t-1})X_{t-2}] = 0, \\ E[(X_t - \phi_0 - \phi_1 X_{t-1})^2] - (1+\theta_1^2)\sigma_Z^2 = 0 \\ E[(X_t - \phi_0 - \phi_1 X_{t-1})(X_{t-1} - \phi_0 - \phi_1 X_{t-2})] - \theta_1 \sigma_Z^2 = 0. \end{aligned} \tag{4.5}$$

The corresponding estimating equations are

$$\begin{aligned} \sum_{i=3}^{n} X_i - \phi_0 - \phi_1 X_{i-1} = 0, \\ \sum_{i=3}^{n} (X_t - \phi_0 - \phi_1 X_{t-1})X_{t-2} = 0, \\ \sum_{i=3}^{n} (X_t - \phi_0 - \phi_1 X_{t-1})^2 - (1+\theta_1^2)\sigma_Z^2 = 0 \\ \sum_{i=3}^{n} (X_t - \phi_0 - \phi_1 X_{t-1})(X_{t-1} - \phi_0 - \phi_1 X_{t-2}) - \theta_1 \sigma_Z^2 = 0. \end{aligned} \tag{4.6}$$

Here we have a just determined case with equal number of equations as the number of parameters.

By the ARMA(1,1) model assumption, the residual sequence is

$$X_t - \phi_0 - \phi_1 X_{t-1} = Z_t + \theta_1 Z_{t-1}.$$

The first EE corresponds to the fact that the residuals have mean zero. The second EE is based on the fact X_{t-2} is a function of $\{Z_{t-\tau}, \tau \geq 2\}$ by the causality assumption, and, since $\{Z_t\}$ is a temporally uncorrelated sequence, we have that $X_t - \phi_0 - \phi_1 X_{t-1}$ is orthogonal to X_{t-2}. Likewise, other choices of X_τ with $\tau \geq 2$ may also be used to form orthogonality conditions. The third and fourth equations result from matching the variance and first-order auto-covariance of the residual variance. $\square$

In our proposed method of multiscale tests, we construct estimating equations at two different time scales by following two main steps:

Step 1 (S1) - we start with the basic estimating equations, which are the same as one would have in the usual single scale estimation. This first set of estimating equations correspond to the high frequency (HF) and thus we name it the HF estimating equations;

Step 2 (S2) - we add one or more auxiliary estimating equations from a second time scale, corresponding to the temporally aggregated model. We name this second set of estimating equations the low frequency (LF) estimating equations. The parameters in the LF estimating equations are parameterized independently of those in the HF estimating equations. In this way, no prior constrains are imposed on the relations between the parameters in the HF and LF estimating equations. Whether the temporal aggregation relation are satisfied by the HF and LF estimating equations will be tested with at a specified confidence level.

The final set of estimating equations should satisfy the general rules for estimating equations. The relation between parameters at the two different time scales is very important to the proposed test and will be explained later in detail.

Example 4.2 Two scale estimating equations for the ARMA(1,1) process based on flow aggregation

In the proposed multiscale test, we form estimating equations at two scales $m_1 = 1$ and $m_2 = m$, respectively, based on the forms of the model at the corresponding levels of aggregation. Let us consider the case of a flow variable.

(S1) We start with the natural choice of estimating equations at the HF $m_1 = 1$ given by (4.6).

(S2) In addition to (4.6), we add the counterparts of (4.6) at the LF $m_2 = m$.

Step 1 and step 2 lead to:

$$\begin{aligned} E[X_t - \phi_0 - \phi_1 X_{t-1}] &= 0, \\ E[(X_t - \phi_0 - \phi_1 X_{t-1})X_{t-2}] &= 0, \\ E[(X_t - \phi_0 - \phi_1 X_{t-1})^2] - (1+\theta_1^2)\sigma_Z^2 &= 0, \\ E[(X_t - \phi_0 - \phi_1 X_{t-1})(X_{t-1} - \phi_0 - \phi_1 X_{t-2})] - \theta_1\sigma_Z^2 &= 0, \\ E[\bar{X}_{(m)t} - \bar{\phi}_{(m)0} - \bar{\phi}_{(m)1}\bar{X}_{(m)t-m}] &= 0, \\ E[(\bar{X}_{(m)t} - \bar{\phi}_{(m)0} - \bar{\phi}_{(m)1}\bar{X}_{(m)t-m})\bar{X}_{(m)t-2m}] &= 0, \\ E[(\bar{X}_{(m)t} - \bar{\phi}_{(m)0} - \bar{\phi}_{(m)1}\bar{X}_{(m)t-m})^2] - \bar{\sigma}^2_{(m)Z} &= 0, \\ E[(\bar{X}_{(m)t} - \bar{\phi}_{(m)0} - \bar{\phi}_{(m)1}\bar{X}_{(m)t-m})(\bar{X}_{(m)t-m} - \bar{\phi}_{(m)0} - \bar{\phi}_{(m)1}\bar{X}_{(m)t-2m})] - \bar{\gamma}^{(1)}_{(m)Z} &= 0, \end{aligned} \tag{4.7}$$

where $\bar{\phi}_{(m)0}$, $\bar{\phi}_{(m)1}$, $\bar{\sigma}^2_{(m)Z}$, $\bar{\gamma}^{(1)}_{(m)Z}$ represent the LF intercept, the AR coefficient, the LF residual variance and lag-1 auto-covariance, respectively, and are assumed to be independent of the HF model parameters. These new parameters are introduced so that the system of equations (4.7) is just-determined and thus always has a unique solution. We will refer to $(\phi_0, \phi_1, \theta_1, \sigma_Z^2)$ as the HF model parameters, abbreviated as $\boldsymbol{\theta}^{HF}$, and to $(\bar{\phi}_{(m)0}, \bar{\phi}_{(m)1}, \bar{\sigma}^2_{(m)Z}, \bar{\gamma}^{(1)}_{(m)Z})$ as the LF model parameters, abbreviated as $\boldsymbol{\theta}^{LF}$. The parameter vector of the system of estimating equations is $\boldsymbol{\theta} := (\boldsymbol{\theta}^{HF}, \boldsymbol{\theta}^{LF})$.

In order to test scale-consistency, we need to establish a relation between the two sets of parameters. Under the assumed ARMA(1,1) model (i.e. (4.1) with $p = 1$ and $q = 1$), an application of temporal aggregation as explained in Chapter 2 tells us that

$$\bar{\phi}_{(m)0} = (1 + \phi_1 + \cdots + \phi_1^{(m-1)}) \cdot m \cdot \phi_0,$$

$$\bar{\phi}_{(m)1} = \phi_1^m,$$

and $\bar{\sigma}^2_{(m)Z}$ and $\bar{\gamma}^{(1)}_{(m)Z}$ are respectively equal to the variance and lag-1 autocovariance implied by temporal aggregation of the HF ARMA(1,1) model.

Again, from temporal aggregation we know that

$$\begin{aligned}
&\bar{X}_{(m)t} - \bar{\phi}_{(m)0} - \bar{\phi}_{(m)1}\bar{X}_{(m)t-m} \\
&= \frac{1-\phi_1^m L^m}{1-\phi_1 L}(1 + L + \cdots + L^{m-1})(1-\phi_1 L)X_t \\
&= (1 + \phi_1 L + \cdots + \phi_1^{m-1}L^{m-1})(1 + L + \cdots + L^{m-1})(1+\theta_1 L)Z_t.
\end{aligned} \tag{4.8}$$

This tells us that $\bar{\sigma}^2_{(m)Z}$ and $\bar{\gamma}^{(1)}_{(m)Z}$ can be calculated based on the LF residual sequence

$$(1 + \phi_1 L + \cdots + \phi_1^{m-1}L^{m-1})(1 + L + \cdots + L^{m-1})(1+\theta_1 L)Z_t.$$

For example, when $m = 2$, we have, under the true model,

$$\begin{aligned}
\bar{\sigma}^2_{(m)Z} &= \left[1 + (1+\phi_1+\theta_1)^2 + (\phi_1+\theta_1+\phi_1\theta_1)^2 + \phi_1^2\theta_1^2\right]\sigma_Z^2, \\
\bar{\gamma}^{(1)}_{(m)Z} &= \left[(\phi_1+\theta_1+\phi_1\theta_1) + \phi_1\theta_1(1+\phi_1+\theta_1)\right]\sigma_Z^2.
\end{aligned} \tag{4.9}$$

(4.10)

The explicit expression of $\bar{\sigma}^2_{(m)Z}$ and $\bar{\gamma}^{(1)}_{(m)Z}$ for general m are complicated but it is straightforward to compute them numerically.

In addition, when $\{X_t\}$ follows a ARMA(1,1) model, the maximal lag of Z_t in the LF residual is $2m-1$. Since $\bar{X}_{(m)t-2m}$ can be written as a function of $\{Z_{t-\tau}, \tau \geq 2m\}$ and $\{Z_t\}$ is a temporally uncorrelated sequence, $\bar{X}_{(m)t} - \bar{\phi}_{(m)0} - \bar{\phi}_{(m)1}\bar{X}_{(m)t-m}$ and $\bar{X}_{(m)t-2m}$ are orthogonal, it is legitimate to use the lagged variable $\bar{X}_{(m)t-2m}$ in forming the orthogonality condition the LF as of the sixth equation in (4.7). $\square$

4.3 The Two-scale Test for ARMA Processes

In this section we present our proposed test for testing scale-consistency of an ARMA(p,q) process. The test is designed to test whether the linear dependency structure of a given stationary ARMA(p,q) process is consistent with data at multiple frequencies. The test does not hinge on any assumption about the distribution of the innovation process and hence it focuses on testing the linear dependency structure. The test is based on the temporal aggregation relation of the model under the null hypothesis at different levels of aggregation. Models that are rejected by the multiscale test are considered not to adequately capture the linear dependency in the data to the extent that estimating the process at different scales may lead to inconsistent estimates. Therefore, in the case of rejection of the null hypothesis, the user may consider extending the model to higher orders in order to increase its flexibility and to better capture the linear dependency structure in the data.

The main idea in the construction of the test is to cast the testing problem in the framework of vector empirical likelihood inference. Samples at multiple scales are used to form vectors of observations and then used to construct estimating equations at multiple scales. With a proper parameterization, the multiscale estimating equations constitute a just determined system. Under the null hypothesis, the parameters must satisfy certain functional relation, based on which the test statistic is constructed.

4.3.1 The null and the alternative hypotheses

As discussed in the introduction, the purpose of the test is to test whether the linear dependency structure of a postulated ARMA process is consistent with data at multiple scales. In an ARMA process, the linear dependency structure is determined by both the AR and the MA coefficients. Therefore, the quantities of interests are the AR and MA parameters $(\phi_0, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma_Z^2)$ and their counterparts in the aggregated processes.

For a given ARMA(p,q) process, our proposed two-scale test tests the following hypothesis:

$$\begin{aligned} H_0 &: f_{(m)}(\boldsymbol{\theta}^{HF}, \boldsymbol{\theta}^{LF}) = 0, \\ H_A &: f_{(m)}(\boldsymbol{\theta}^{HF}, \boldsymbol{\theta}^{LF}) \neq 0, \end{aligned} \tag{4.11}$$

where $\boldsymbol{\theta}^{HF} := (\phi_0, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma_Z^2)$ denotes the HF model parameters, $\boldsymbol{\theta}^{LF} := (\bar{\phi}_{(m)0}, \bar{\phi}_{(m)1}, \cdots, \bar{\phi}_{(m)p}, \bar{\sigma}^2_{(m)Z}, \bar{\gamma}^{(1)}_{(m)Z}, \cdots, \bar{\gamma}^{(q)}_{(m)Z})$ denotes the LF model parameters, and the function $f_{(m)}$ is determined by the temporal aggregation relation between the high and low frequency model parameters in the ARMA(p,q) process at levels $m_1 = 1$ and $m_2 = m$, as in equation (4.7) of Example 4.2.

Notice that the parameter vector in the hypothesis is $\boldsymbol{\theta} := (\boldsymbol{\theta}^{HF}, \boldsymbol{\theta}^{LF})$. We use the notation $f_{(m)}(\boldsymbol{\theta}^{HF}, \boldsymbol{\theta}^{LF})$, instead of $f_{(m)}(\boldsymbol{\theta})$, in order to emphasize the partition of the parameter vector into the parts associated with the HF and the LF. Later on, we may also use the more succinct notation of $f_{(m)}(\boldsymbol{\theta})$ for the sake of brevity.

Example 4.2 (continued) A specialization of the multiscale test in the case of ARMA(1,1) process with $m_2 = 2$ tests the following hypothesis

$$\begin{aligned} H_0 &: f_{(2)}(\boldsymbol{\theta}^{HF}, \boldsymbol{\theta}^{LF}) = 0, \\ H_A &: f_{(2)}(\boldsymbol{\theta}^{HF}, \boldsymbol{\theta}^{LF}) \neq 0, \end{aligned} \tag{4.12}$$

where $\boldsymbol{\theta}^{HF} := (\phi_0, \phi_1, \theta_1, \sigma_Z^2)$, $\boldsymbol{\theta}^{LF} := (\bar{\phi}_{(2)0}, \bar{\phi}_{(2)1}, \bar{\sigma}^2_{(2)Z}, \bar{\gamma}^{(1)}_{(2)Z})$, and

$$\begin{aligned} & f_{(2)}(\boldsymbol{\theta}^{HF}, \boldsymbol{\theta}^{LF}) \\ := & \begin{pmatrix} (1+\phi_1)\cdot 2\cdot \phi_0 - \bar{\phi}_{(2)0} \\ \phi_1^2 - \bar{\phi}_{(2)1} \\ [1 + (1+\phi_1+\theta_1)^2 + (\phi_1 + (1+\phi_1)\theta_1)^2 + \phi_1^2\theta_1^2]\sigma_Z^2 - \bar{\sigma}^2_{(2)Z} \\ [(\phi_1 + \theta_1 + \phi_1\theta_1) + \phi_1\theta_1(1+\phi_1+\theta_1)]\,\sigma_Z^2 - \bar{\gamma}^{(1)}_{(2)Z} \end{pmatrix}. \end{aligned} \tag{4.13}$$

The elements in the vector in (4.13) are the differences between the LF parameters implied by the HF parameters and the directly estimated LF parameters. The four elements correspond to the intercept term, the AR coefficient, the residuals variances, and the first-order residual auto-covariance, respectively. All of the elements in matrix (4.13) are equal to zero under the null hypothesis. □

We will consider two types of testing problems:
(i) testing a model with a particular set of parameters;
(ii) testing the model.

From a computational perspective, a major difference between the two types of testing problems is that in the former type of test, the model parameters are fixed at the hypothesized values whereas, in the later type of test, the model parameters are estimated.

To provide the reader with some intuition about the usefulness of each of the two types of the proposed tests, we describe two corresponding plausible cases in which the functional relation $f_{(m)}(\boldsymbol{\theta}^{HF}, \boldsymbol{\theta}^{LF}) = 0$ may be violated:

1. **Discrepancy between the true and estimated value of the autoregressive parameter ϕ_1.** Consider, for example, the ARMA(1,1) model. In this case, a small estimation bias in ϕ_1 will be magnified through the power function relation when we examine the process at an temporally aggregated level.

2. **Data coming from a model with a different dependency structure.** For example, data may be generated from a higher order autoregressive and/or moving average component. In this case, both the relation regarding the AR coefficient and that regarding the residual variances at scales m_1 and m_2 will likely deviate from the relation $f_{(m)}(\boldsymbol{\theta}^{HF}, \boldsymbol{\theta}^{LF}) = 0$.

In practice, there can be many more possible ways of deviations from the null hypothesis. The possibilities we consider here are motivated by cases encountered in financial return volatility modeling.

4.3.2 Construction of two-frequency samples, estimating equations, and empirical likelihood testing for ARMA(p,q) model

In this section, we give the steps in constructing the two-scale tests.

Constructing two-frequency samples as vectors of observations

In order to formulate the empirical likelihood inference method using samples from two time scales, we need to formulate samples corresponding to the two scales.

Denote

$$\begin{aligned}\bar{e}_{(m)t}(\boldsymbol{\theta}) &:= \left[1 - \sum_{j=1}^{p} \phi_{(m)j} L^m\right] \left(\sum_{i=1}^{m-1} L^i\right) X_t - \bar{\phi}_{(m)0} \\ &= \left[1 - \sum_{j=1}^{p} \phi_{(m)j} L^m\right] \bar{X}_{(m)t} - \bar{\phi}_{(m)0},\end{aligned} \tag{4.14}$$

where L denotes a lag operator. We call $\{e_t, t \in \mathbb{Z}^+ \text{ and } t \leq T\}$ and $\{\bar{e}_{(m)t}, t \in \mathbb{Z}^+ \text{ and } t \leq T\}$ the high frequency (HF) and the low frequency (LF) samples, respectively. Stacking the HF and LF samples into a vector, we have $\{(e_t, \bar{e}_{(m)t})', t \in \mathbb{Z}^+ \text{ and } t \leq T\}$ as our vector-valued observations. By doing so, we can cast the inference problem into a vector-valued (block) empirical likelihood framework. More detailed examples will be provided in the following subsection.

Constructing the estimating equations

To test the null hypothesis given in (4.11) based on a sample of data, we now formulate the corresponding estimating equations to be used for the empirical likelihood inference.

Using the given observations, we form estimating equations which are satisfied by the postulated ARMA(p,q) process at both of the scales $m_1 = 1$ and $m_2 = m$. We use estimating equations analogous to the ones used in Example 4.2 for the ARMA(1,1) model. These are straightforward generalizations of the estimating equations for the ARMA(1,1) process. At the HF, $m_1 = 1$, we have a generalization of (4.6) :

$$\sum_t g_{HF,t}(\phi_0, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma_Z^2) = 0 \tag{4.15}$$

where

$$g_{HF,t}(\phi_0, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma_Z^2) = \begin{pmatrix} e_t(\boldsymbol{\theta}) \\ e_t(\boldsymbol{\theta})X_{t-p-1} \\ e_t(\boldsymbol{\theta})X_{t-p-2} \\ \cdots \\ e_t(\boldsymbol{\theta})X_{t-2p} \\ e_t(\boldsymbol{\theta})^2 - (1+\theta_1^2+\cdots+\theta_q^2)\sigma_Z^2 \\ e_t(\boldsymbol{\theta})e_{t-1}(\boldsymbol{\theta}) - (\theta_1+\theta_2\theta_1\cdots+\theta_q\theta_{q-1})\sigma_Z^2 \\ \cdots \\ e_t(\boldsymbol{\theta})e_{t-q+1}(\boldsymbol{\theta}) - (\theta_{t-q+1}+\theta_q\theta_1)\sigma_Z^2 \end{pmatrix}. \tag{4.16}$$

Next, we add estimating equations from the LF. In the case of a flow aggregation, they are generalizations of LF estimating equations in (4.7):

$$\sum_t g_{LF,t}(\bar{\phi}_{(m)0}, \bar{\phi}_{(m)1}, \cdots, \bar{\phi}_{(m)p}, \bar{\sigma}^2_{(m)Z}, \bar{\gamma}^{(1)}_{(m)Z}, \cdots, \bar{\gamma}^{(q)}_{(m)Z}) = 0 \tag{4.17}$$

where

$$g_{LF,t}(\bar{\phi}_{(m)0}, \bar{\phi}_{(m)1}, \cdots, \bar{\phi}_{(m)p}, \bar{\sigma}^2_{(m)Z}, \bar{\gamma}^{(1)}_{(m)Z}, \cdots, \bar{\gamma}^{(q)}_{(m)Z}) := \begin{pmatrix} \bar{e}_{(m)t}(\boldsymbol{\theta}) \\ \bar{e}_{(m)t}(\boldsymbol{\theta})\bar{X}_{(m)t-p(m-1)-q-m} \\ \bar{e}_{(m)t}(\boldsymbol{\theta})\bar{X}_{(m)t-p(m-1)-q-m-1} \\ \cdots \\ \bar{e}_{(m)t}(\boldsymbol{\theta})\bar{X}_{(m)t-p(m-1)-q-m-(p-1)} \\ \bar{e}_{(m)t}(\boldsymbol{\theta})^2 - \bar{\sigma}^2_{(m)Z} \\ \bar{e}_{(m)t}(\boldsymbol{\theta})\bar{e}_{(m)t-m}(\boldsymbol{\theta}) - \bar{\gamma}^{(1)}_{(m)Z} \\ \cdots \\ \bar{e}_{(m)t}(\boldsymbol{\theta})\bar{e}_{(m)t-qm}(\boldsymbol{\theta}) - \bar{\gamma}^{(q)}_{(m)Z} \end{pmatrix}, \tag{4.18}$$

Stacking the HF and LF estimating equations into a single vector, we have the final vector of estimating equations given by

$$\sum_t g_t(\phi_0, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma_Z^2, \bar{\phi}_{(m)0}, \bar{\phi}_{(m)1}, \cdots, \bar{\phi}_{(m)p}, \bar{\sigma}^2_{(m)Z}, \bar{\gamma}^{(1)}_{(m)Z}, \cdots, \bar{\gamma}^{(q)}_{(m)Z}) = 0, \tag{4.19}$$

where

$$
g_t(\phi_0, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma_Z^2, \bar{\phi}_{(m)0}, \bar{\phi}_{(m)1}, \cdots, \bar{\phi}_{(m)p}, \bar{\sigma}^2_{(m)Z}, \bar{\gamma}^{(1)}_{(m)Z}, \cdots, \bar{\gamma}^{(q)}_{(m)Z})
:= \begin{pmatrix}
e_t(\boldsymbol{\theta}) \\
e_t(\boldsymbol{\theta})X_{t-p-1} \\
e_t(\boldsymbol{\theta})X_{t-p-2} \\
\cdots \\
e_t(\boldsymbol{\theta})X_{t-2p} \\
e_t(\boldsymbol{\theta})^2 - (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma_Z^2 \\
e_t(\boldsymbol{\theta})e_{t-1}(\boldsymbol{\theta}) - (\theta_1 + \theta_2\theta_1 \cdots + \theta_q\theta_{q-1})\sigma_Z^2 \\
\cdots \\
e_t(\boldsymbol{\theta})e_{t-q+1}(\boldsymbol{\theta}) - (\theta_{t-q+1} + \theta_q\theta_1)\sigma_Z^2 \\
\bar{e}_{(m)t}(\boldsymbol{\theta}) \\
\bar{e}_{(m)t}(\boldsymbol{\theta})\bar{X}_{(m)t-p(m-1)-q-m} \\
\bar{e}_{(m)t}(\boldsymbol{\theta})\bar{X}_{(m)t-p(m-1)-q-m-1} \\
\cdots \\
\bar{e}_{(m)t}(\boldsymbol{\theta})\bar{X}_{(m)t-p(m-1)-q-m-(p-1)} \\
\bar{e}_{(m)t}(\boldsymbol{\theta})^2 - \bar{\sigma}^2_{(m)Z} \\
\bar{e}_{(m)t}(\boldsymbol{\theta})\bar{e}_{(m)t-m}(\boldsymbol{\theta}) - \bar{\gamma}^{(1)}_{(m)Z} \\
\cdots \\
\bar{e}_{(m)t}(\boldsymbol{\theta})\bar{e}_{(m)t-qm}(\boldsymbol{\theta}) - \bar{\gamma}^{(q)}_{(m)Z}
\end{pmatrix}. \tag{4.20}
$$

Conducting empirical likelihood inference

With the estimating equation defined above, an empirical likelihood inference following the lines of Qin and Lawless (1994)[65] can be applied. Since the vector-valued two-frequency samples are temporal dependent as the process X_t is temporally dependent, we apply the block empirical likelihood inference framework of Kitamura (1997)[40], which is a modification of the framework of Qin and Lawless (1994)[65] to account for temporally dependency in the data. The basic idea of the blocking technique is to construct new observations which nonparametrically preserve the dependence structure of the original series and thus deliver valid asymptotic inference results based on the blocked observations. Alternative procedures to the blocking technique, such as kernel smoothing techniques, may be considered. However, we focus our attention on multiscale inference by only considering the blocking technique in this thesis, and reserve the investigation of using alternative techniques as a future research topic.

A block empirical likelihood inference is based on the following profile empirical likelihood function

$$\mathcal{R}_B(\boldsymbol{\theta}) = \sup_{w_i^B} \left\{ \prod_{i=1}^{Q} Q w_i^B | w_i^B > 0, \sum_{i=1}^{Q} w_i^B = 1, \sum_{i=1}^{Q} w_i^B T_i(\boldsymbol{\theta}) = 0 \right\}, \tag{4.21}$$

where

$$\boldsymbol{\theta} = (\phi_0, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma_Z^2, \bar{\phi}_{(m)0}, \bar{\phi}_{(m)1}, \cdots, \bar{\phi}_{(m)p}, \bar{\sigma}^2_{(m)Z}, \bar{\gamma}^{(1)}_{(m)Z}, \cdots, \bar{\gamma}^{(q)}_{(m)Z})$$

is the parameter vector, and $T_i(\boldsymbol{\theta}) = \frac{1}{M} \sum_{j=1}^{M} g_{(i-1)L+j}(\boldsymbol{\theta}), i = 1 \cdots, Q$ are the blocked observations in which M denotes the block length and L is the separation between block starting points.

The decision rule for rejecting the null hypothesis H_0

According to the asymptotic results presented in the following section (i.e. Theorem 1 and Theorem 2 of Section 4.4), under the true model, we have that the log profile empirical likelihood statistic

$$W_B(\boldsymbol{\theta}_0) = -2A_n^{-1} \log \mathcal{R}_B(\boldsymbol{\theta}_0), \tag{4.22}$$

where $A_n = QM/n$, converges to a $\chi^2_{2(p+q+2)}$ distribution under the null hypothesis and

$$W_B(\tilde{\boldsymbol{\theta}}) = -2A_n^{-1} \log \mathcal{R}_B(\tilde{\boldsymbol{\theta}}) \tag{4.23}$$

converges to a $\chi^2_{(p+q+2)}$ distribution under the null hypothesis, in which $\tilde{\boldsymbol{\theta}}$ maximizes the profile empirical likelihood function (4.21).

Testing $H_0 : f_{(m)}(\boldsymbol{\theta}) = 0$ **at** $\boldsymbol{\theta} = \boldsymbol{\theta}_0$
The decision rule is the following: if the value of $W_B(\boldsymbol{\theta}_0)$ is greater than the $(1-\alpha)$-quantile of a $\chi^2_{2(p+q+2)}$ distribution, then we reject the null hypothesis $H_0 : f_{(m)}(\boldsymbol{\theta}_0) = 0$ at the level of significance α.

Testing $H_0 : f_{(m)}(\boldsymbol{\theta}) = 0$
The decision rule is as follows: if the minimal value of $W_B(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ over the subset of the parameter space of $\boldsymbol{\theta}$ defined by $f_{(m)}(\boldsymbol{\theta}) = 0$, denoted as $W_B(\tilde{\boldsymbol{\theta}})$, is greater than the $(1-\alpha)$-quantile of a $\chi^2_{(p+q+2)}$ distribution, then we reject the null hypothesis $H_0 : f_{(m)} = 0$ at the level of significance α.

Example 4.2 (continued) In the case of the ARMA(1,1) process, the estimating equations (4.7) contain a total of 8 estimating equations which are linearly independent. Therefore, $W_B(\boldsymbol{\theta}_0)$ converges to a $\chi^2_{(8)}$ distribution under the null hypothesis and $W_B(\tilde{\boldsymbol{\theta}})$ converges to a $\chi^2_{(4)}$ distribution under the null hypothesis. Decision rules for rejecting the null hypothesis can be made accordingly. □

Intuitions behind the proposed tests

Here we provide some intuitive explanation of our proposed multiscale tests using a simple AR(1) setting. Consider testing the scale consistency of the following AR(1) process:

$$X_t = \phi X_{t-1} + Z_t,$$

where we assume that the intercept term is zero for the sake of simplicity and that the parameter space for ϕ is $\{\phi; \phi \in (-1, 1)\}$. If this AR(1) model is consistent with data at the two scales, say, $m_1 = 1$ and $m_2 = 2$, then the AR(1) coefficient at the scales m_1 and m_2, denoted respectively as ϕ and $\bar{\phi}_{(2)}$ shall follow the temporal aggregation relation of $\bar{\phi}_{(2)} = \phi^2$. This quadratic relation defines a subset of the parameter space. We may evaluate the log-empirical likelihood ratio statistic over this quadratic subset and thus evaluate how likely this quadratic relation holds for a given data set. Graphically, as illustrated in Fig. 4.2, the highest likelihood over the quadratic subset lies between the confidence levels of 95% and 99%. Consequently, we do not reject the null hypothesis that the hypothesized AR(1) relation holds at the scales $m_1 = 1$ and $m_2 = 2$ with a two-scale test at 5% level, but we reject the same null hypothesis at 1% level.

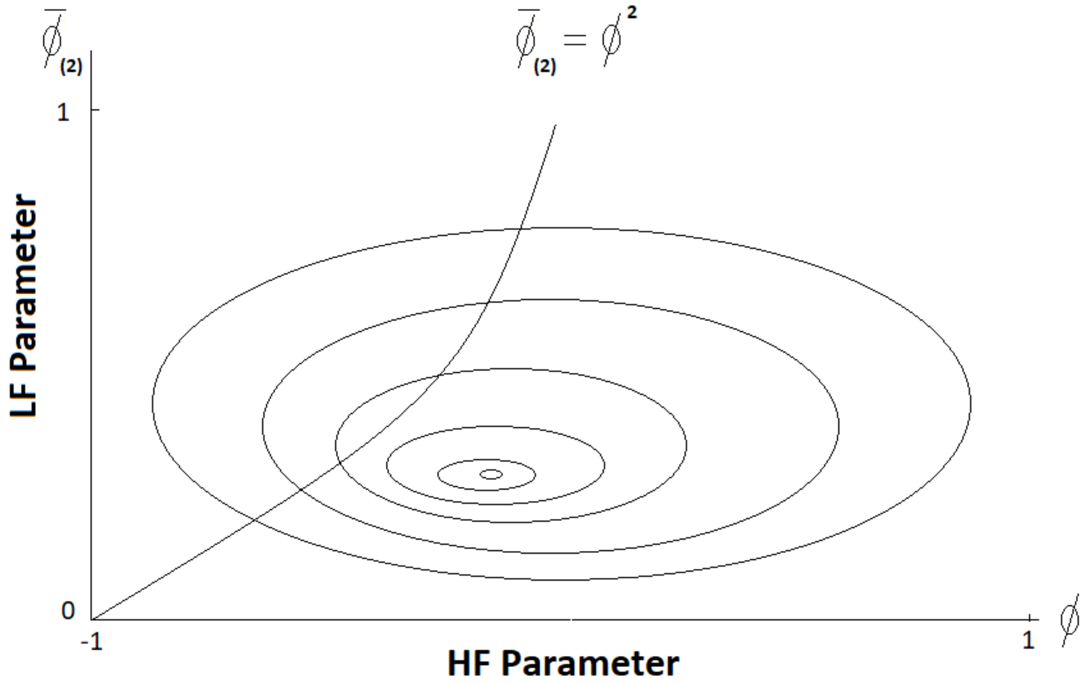

FIGURE 4.2: Intuition behind the multiscale tests. The contours are empirical likelihood confidence regions for the parameters ϕ and $\bar{\phi}_2$, and the contours correspond to confidence levels of 50%, 90%, 95%, 99%, 99.9%, 99.99%.

4.4 Asymptotic Results

In this section, we give a proof of the asymptotic results for our proposed tests. The proof follows closely the one for i.i.d. observations given in Owen (2001)[62] and is adapted to the setting of dependent processes of our interest. It also borrow heavily from the proof in Kitamura (1997)[40] for general dependent processes. We fill in the steps which are omitted in Kitamura (1997)[40] and point the readers to Kitamura (1997)[40] where the needed steps are given there.

We make the following assumptions:
(i) The process $\{X_t\}$ is strictly stationary and ergodic;
(ii) The process $\{X_t\}$ is α-mixing with a mixing coefficient $\alpha_X(k)$ satisfying

$$\sum_{k=1}^{\infty} \alpha_X(k)^{1-1/c} < \infty$$

where c is some constant;
(iii) The process $\{X_t\}$ has enough moments such that $g(X_t, \theta_0)$ has a finite $2c$ moment.

Let $W_B(\boldsymbol{\theta}) := -2A_n^{-1} \log \mathcal{R}_B(\boldsymbol{\theta})$ where $A_n = QM/n$.

Theorem 1 (Block Empirical Likelihood Theorm for Testing a Model with a Particular Set of Parameters) Under the true model, $W_B(\boldsymbol{\theta}_0)$ converges, as n goes to ∞, to a χ_r^2 distribution where r is the number of estimating equations.

Proof: When 0 is inside the convex hull of the $T_i(\theta)$'s, there is a unique set of weights $w_i^B > 0$ with $\sum_{i=1}^{Q} w_i^B = 1$ and $\sum_{i=1}^{Q} w_i^B T_i(\theta_0) = 0$ for which $\prod_{i=1}^{Q} Q w_i^B$ is maximized. By a Lagrange multiplier argument, the maximizing weights can be written as

$$w_i^B = \frac{1}{Q} \frac{1}{1 + \lambda' T_i(\theta_0)},$$

where the vector $\lambda = \lambda(\theta_0) \in \mathbb{R}^r$ satisfies r equations given by

$$l(\lambda) \equiv \frac{1}{Q} \sum_{i=1}^{Q} \frac{T_i(\theta_0)}{1 + \lambda' T_i(\theta_0)} = 0. \tag{4.24}$$

Let $\lambda = ||\lambda||\xi$ where $||\cdot||$ denotes the Euclidean norm and ξ is a unit vector. Next we introduce the following equation

$$Y_i = \lambda' T_i(\theta_0), \text{ and } Z_Q^* = \underset{1 \le i \le Q}{max} ||T_i(\theta_0)||.$$

Substituting $1/(1 + Y_i) = 1 - Y_i/(1 + Y_i)$ into $\xi' l(\lambda) = 0$ and multiplying both sides of the equation by M, we obtain

$$||\lambda||\xi' \tilde{S} \xi = M \xi' \frac{1}{Q} \sum_{i=1}^{Q} T_i(\theta_0), \tag{4.25}$$

where

$$\tilde{S} = \frac{M}{Q} \sum_{i=1}^{Q} \frac{T_i(\theta_0) T_i(\theta_0)'}{1 + T_i(\theta_0)}. \tag{4.26}$$

Denote

$$S = \frac{M}{Q} \sum_{i=1}^{Q} T_i(\theta_0) T_i(\theta_0)'.$$

Since the weights $w_i^B > 0$, we have $1 + T_i(\theta_0) > 0$. Therefore we have

$$\begin{aligned} ||\lambda||\xi' S \xi &\le ||\lambda||\xi' \tilde{S} \xi (1 + ||\lambda|| Z_Q^*) \\ &= M \xi' \frac{1}{Q} \sum_{i=1}^{Q} T_i(\theta_0)(1 + ||\lambda|| Z_Q^*). \end{aligned} \tag{4.27}$$

So

$$||\lambda||\left(\xi' S\xi - Z_Q^* M\xi'\frac{1}{Q}\sum_{i=1}^{Q} T_i(\theta_0)\right) \leq M\xi'\frac{1}{Q}\sum_{i=1}^{Q} T_i(\theta_0).$$

By the CLT for an α-mixing process of Ibragimov and Linnik (1971)[37] Theorem 18.5.3, we have $\sum_{i=1}^{Q} T_i(\theta_0)/Q = O_p(n^{-1/2})$. By Lemma 3.2 of Kunsch (1989)[42], it can be shown that $Z_Q^* = o(n^{1/2}M^{-1})$. Finally, a central limit theorem applied to S shows that $\xi' S\xi = O_p(1)$. It follows that $||\lambda|| = O_p(Mn^{-1/2})$. Since $M = o(n^{1/2})$, this proves that $\lambda(\theta_0)$ converges to 0 in probability.

Having established an order bound for $||\lambda||$, we can use Lemma 3.2 of Kunsch (1989)[42] to show that

$$\max_{1\leq i\leq Q}|Y_i| = O_p(Mn^{-1/2})o(n^{1/2}M^{-1}) = o_p(1). \tag{4.28}$$

Using $l(\lambda) = 0$ again we obtain

$$\begin{aligned} 0 &= \frac{M}{Q}\sum_{i=1}^{Q} T_i(\theta_0)\left(1 - Y_i + \frac{Y_i^2}{1-Y_i}\right) \\ &= \frac{M}{Q}\sum_{i=1}^{Q} T_i(\theta_0) - \frac{M}{Q}\sum_{i=1}^{Q} T_i(\theta_0)T_i(\theta_0)'\lambda + \frac{M}{Q}\sum_{i=1}^{Q}\frac{T_i(\theta_0)Y_i^2}{1-Y_i}. \end{aligned} \tag{4.29}$$

The final term in the last expression has a norm bounded by

$$\frac{M}{Q}\sum_{i=1}^{Q}||T_i(\theta_0)||^3||\lambda||^2|1-Y_i|^{-1} = Mo(n^{1/2}M^{-1})O_p(n^{-1/2})O_p(M^2n^{-1})O_p(1) = o_p(1)$$

where we used $M = o(n^{1/2})$. Thus, we have

$$\lambda = S^{-1}\frac{M}{Q}\sum_{i=1}^{Q} T_i(\theta_0) + \beta$$

where $\beta = o_p(1)$.

By (4.28) we may write

$$\log(1+Y_i) = Y_i - \frac{1}{2}Y_i^2 + \eta_i,$$

where for some finite $B > 0$

$$Pr(|\eta_i| \leq B|Y_i|^3, 1 \leq i \leq Q) \to 1,$$

as $n \to \infty$.

Now we may write

$$\begin{aligned}
-2A_N^{-1}\log\mathcal{R}_B(\theta_0) &= -2A_N^{-1}\sum_{i=1}^{Q}\log(Qw_i^B) \\
&= 2A_N^{-1}\sum_{i=1}^{Q}\log(1+Y_i) \\
&= 2A_N^{-1}\sum_{i=1}^{Q}Y_i - A_N^{-1}\sum_{i=1}^{Q}Y_i^2 + 2A_N^{-1}\sum_{i=1}^{Q}\eta_i \\
&= A_N^{-1}QM\bar{T}(\theta_0)S^{-1}\bar{T}(\theta_0) - A_N^{-1}\frac{Q}{M}\beta' S\beta + 2A_N^{-1}\sum_{i=1}^{Q}\eta_i
\end{aligned} \tag{4.30}$$

where $\bar{T}(\theta_0) = \sum_{i=1}^{Q} T_i(\theta_0)/Q$.

In the limit as $Q \to \infty$, we have

$$A_N^{-1}QM\bar{T}(\theta_0)S^{-1}\bar{T}(\theta_0) = n\bar{T}(\theta_0)S^{-1}\bar{T}(\theta_0) \to \chi^2_{(r)}$$

in distribution,

$$A_N^{-1}\frac{Q}{M}\beta' S\beta = \frac{n}{M^2}o_p(1)O_p(1)o_p(1) = o_p(1),$$

and

$$\begin{aligned}
A_N^{-1}|\sum_{i=1}^{Q}\eta_i| \leq A_N^{-1}B||\lambda||^3\sum_{i=1}^{Q}||T_i(\theta_0)||^2 &= \frac{n}{QM}O_p(M^3n^{-3/2})o(n^{1/2}M^{-1})QO_p(n^{-1/2}) \\
&= o_p(1).
\end{aligned}$$

Therefore $W_B(\boldsymbol{\theta}_0) = -2A_N^{-1}\log\mathcal{R}_B(\theta_0) \to \chi^2_r$ in distribution. □

Theorem 2 (Block Empirical Likelihood Theorem for Testing a Model) Under the true model, $W_B(\tilde{\boldsymbol{\theta}})$ converges to a $\chi^2_{(r-s)}$ distribution.
Proof: First, we establish the asymptotic consistency and asymptotic normality of $\tilde{\theta}_n$ and $\tilde{\lambda}_n = \lambda(\tilde{\theta}_n)$ which correspond to the maximizer of the log empirical likelihood ratio function. This can be done by checking the assumptions of Theorem 1 of Kitamura (1997)[40] and thus applying the theorem.

Then, we can establish the asymptotic distribution of the block empirical likelihood statistic for testing a model by following the argument of Theorem 2 (i) of

Kitamura (1997) [40]. □

4.5 Computation of the test statistic

In this section, we give some details about computing the values of the test statistics.

Existence of optimum

As explained in Owen (2001)[62], section 2.9, the objective function of the empirical likelihood optimization problem $\sum_{i=1}^{n_m} \log(n_m w_i)$ is a strictly concave function on a convex set of weight vectors. Therefore, a unique global minimum exists. Moreover, the minimum does not have any $w_i = 0$, so it is an interior point of the domain.

Dealing with the parameter constraints

To compute the value of the test statistic under the null hypothesis H_0, we need to conduct a constrained optimization where the constraints are imposed by the temporal aggregation relation between the HF and LF parameters from the temporal aggregation relation. Such a constrained EL testing problem had been considered in Qin (1992)[64], Chapter 3, who showed that there are two approaches to deal with this problem which are first order equivalent.

The first approach is to express the LF parameters in terms of the HF parameters and then optimize the test statistic with respect to the HF parameters. In this case, the optimized parameters always obey the functional relation under H_0, and there are more constraints (or estimating equations) than the number of parameters. the Empirical Likelihood Theorem (ELT) for testing a model can be applied to derive the limiting distribution of the test statistics.

A second approach is to treat the constraints among the parameters as additional constraints upon the estimating equations (or moment constraints). This second approach has the advantage over the first approach when we cannot express some of the parameter explicitly as functions of the others. Qin (1992)[64], Chapter 3, derived the asymptotic distribution of the test statistics subject to parameter constraints, which has a χ^2 distribution with the same degrees of freedom as the over-identification test in the first approach.

In our multiscale test, the LF parameters can naturally be expressed as functions of the HF parameters. Therefore, it is computationally more straightforward to use the over-identification test approach. That is, when computing the value of the test statistic $W(\boldsymbol{\theta})$, we parameterize the LF parameters as functions of the HF parameters and optimize with respect to the HF parameters.

Choosing block length in the BEL

As pointed out in Nordman and Lahiri (2014)[59], theoretical results on optimal block length selection remains an open research question. There are two main types of strategies for determining the block length. The first strategy borrows the idea from spectral density estimation. As pointed out in Kitamura (1997)[40], the block-based variance estimator in BEL can be seen as a spectral density estimator based on Bartlett's kernel. Thus, rules for a kernel bandwidth selection have been used to select the block length. However, as discussed in Nordman and Lahiri (2014)[59], different approaches may lead to different choices of block length, which are not guaranteed to be theoretically optimal. Another strategy, from Politis, et.al. (1999)[63], is to choose a block size based on the principle that approximately correct block lengths for inference might be characterized by a stable behavior of confidence regions with respect to the block length. One can use a visual inspection to determine an appropriate block size based on plots of confidence regions (or equivalent measures) against the block sizes. We use this second strategy in our simulation studies.

4.6 Simulation Study

In this section, we conduct simulation studies to investigate the finite sample properties of our proposed test for the ARMA models.

Corresponding to the intended uses of our proposed tests, we conduct two types of tests. The first type tests a model with a particular set of parameters, and the second type tests the model. For each type of tests, we study their empirical size properties and demonstrate the empirical power properties against particular alternatives motivated by practical situations.

As for the alternatives in the first type of tests, we consider data generated from the true model but with slightly different parameters. This alternative mimics the practical situation in which one may have a biased estimate of model parameters resulting from, say, a misspecified innovation distribution. For the second type of tests, we generate data from some higher order models, which mimics the practical situation of an under-specified model due to the existence of a multiscale phenomenon as observed in financial time series.

For every data generating process, we simulate from a strong model with i.i.d. normal innovations. The HF data corresponds to $m_1 = 1$, i.e. we use all of the simulated data. The LF m_2 is chosen at various values.

4.6.1 Testing a model with a particular set of parameters

AR process

Size of the test

We generate data from an AR(1) process with parameter $\boldsymbol{\theta}_0 = (\phi_0, \phi_1, \sigma_Z^2) = (0, 0.95, 1)$ and test the estimating equations (4.7) with the true parameter $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ using the block EL inference. In the case of the AR(1) process, (4.7) contains a set of 6 estimating equations [3]. We study the empirical sizes of the test with various values of sample size and levels of aggregation.

[3] Because the AR(1) process does not have the θ_1 parameter, we have two fewer equations compared to the eight equations as in (4.7).

To determine the choice of the block length, we vary block length from 1 to some values large enough. For each level of aggregation, we plot the empirical sizes against the block lengths to determine the appropriate block length to be used.

Fig. 4.3 is an example plot of empirical sizes against the block lengths for aggregation level $m_2 = 5$. In this plot, the curves in different line types from top to bottom correspond to empirical sizes at significance levels 99%, 95%, and 90%, respectively. The corresponding horizontal lines represent the nominal significance levels. When a curve intersects the horizontal line of the same type, it means that the block length corresponding to the intersection point yields a test statistic with an empirical size matching its nominal value. In this case, we choose a block length of $M = 15$ as the empirical size of the test statistics become stable at the nominal levels with respect to the block length starting from $M = 15$. In the cases where there is no intersection of a curve with a horizontal line of the same type, we pick points where the two are closest to each other.

Admittedly, the method that we use to choose the block lengths is more of an intuitive one which is described in Nordman and Lahiri (2014)[59]. In that paper, the authors provided a comprehensive review of EL methods for time series data. In particular, Nordman et.al. (2013)[58] proposed the expansive block empirical likelihood (EBEL) method, which uses data blocks of every possible length, may be used to avoid the problem of choosing a particular block length. We leave the exploration of block length selection methods for future research.

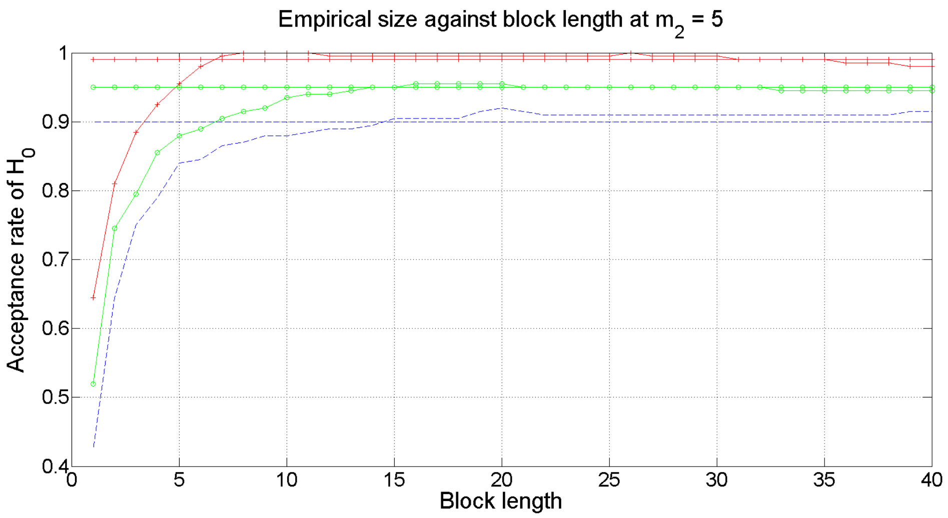

FIGURE 4.3: Empirical sizes of the test for AR(1) model against block length for aggregation level $m_2 = 5$.

Table 4.1 shows the percentage of $W_B(\boldsymbol{\theta}_0) \geq \chi^{-1}_{(4)}(1-\alpha)$ ($\alpha = 0.1, 0.05, 0.01$) based on 1,000 replications for the AR(1) process with parameter vector $(\phi_1, \sigma_Z^2) = (0.95, 1)$.

TABLE 4.1: Empirical size of the two-scale tests when the observations are generated from the AR(1) model with $(\phi_1, \sigma_Z^2) = (0.95,\ 1)$.

m_2	$T=$	1,500		m_2	$T=$	3,000		m_2	$T=$	6,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
2	11.8	6.7	0.7	2	12.0	6.1	1.7	2	10.7	6.2	1.7
5	11.2	6.0	0.9	5	10.4	4.9	1.3	5	10.0	5.3	1.3
10	15.8	8.5	2.1	10	13.4	6.8	1.4	10	12.5	6.4	1.4
20	17.8	11.8	4.2	20	14.4	8.9	2.0	20	12.8	7.8	2.0
m_2	$T=$	12,000		m_2	$T=$	30,000		m_2	$T=$	60,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
2	10.6	6.2	1.4	2	10.4	5.1	1.3	2	12.2	6.3	1.3
5	12.1	5.9	1.1	5	10.3	5.8	0.8	5	7.9	3.9	0.9
10	13.7	7.4	2.1	10	12.9	6.1	1.3	10	12.8	6.8	1.8
20	14.0	7.7	2.2	20	13.2	7.4	1.4	20	12.9	8.0	1.4

As one can see from the empirical sizes in Table 4.1, tests using LF EE based on a lower level of aggregations (i.e. $m_2 = 2, 5$) have accurate empirical sizes. However, it requires a relatively large sample size of $T = 6,000$ in order for the tests using higher levels of LF aggregations (i.e. $m_2 = 10, 20$) to be well approximated by the theoretical asymptotic distribution. We also notice that the sizes of the tests corresponding to $m_2 = 5$ seem not to follow the pattern of changes based on the neighboring scales. By trying some slightly different choices of block lengths, we found that the sizes of the $m_2 = 5$ tests can be made in line with the patterns with more tailored choices of block lengths. However, we present here and in the rest of the thesis with size and power results based on a simple fixed choice of block lengths. We leave the deeper exploration of the issue of block length selection as part of the future research questions.

Power of the test against small deviations in the parameters

To demonstrate the empirical power of the test, we test against a small deviation from the true parameter. In order to demonstrate the advantage of using a multiscale sample, we compare the powers of the multiscale test with the corresponding single-scale test where only HF estimating equations are used.

Table 4.2 shows the percentage of $W_B(\boldsymbol{\theta}_0) \geq \chi^{-1}_{(b)}(1-\alpha)$ $(\alpha = 0.1, 0.05, 0.01)$ based on 1,000 replications where b is the relevant degrees of freedom of the limiting distribution. Specifically, $b = 4$ in the multiscale test and $b = 2$ in the corresponding single scale test. In this case data is generated from an AR(1) process with $(\phi_0, \phi_1, \sigma^2_Z) = (0,\ 0.95,\ 1)$ and a slightly disturbed parameter of $\phi_1 = 0.94$ is tested.

From Table 4.2 we observe that the multiscale test has a good power property and is consistently more powerful than the corresponding single scale test.

TABLE 4.2: Empirical power comparison of one- and two-scale tests (in %) when the observations are from the AR(1) model with parameter vector $(\phi_0, \phi_1, \sigma^2_Z) = (0,\ 0.95,\ 1)$.
Testing parameter $(0, 0.94, 1)$.

						1 Scale					
	$T =$	1,500			$T =$	3,000			$T =$	6,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
	18.1	10.9	3.0		26.7	17.9	6.5		39.1	27.5	13.6
	$T =$	12,000			$T =$	30,00			$T =$	60,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
	65.9	54.9	33.3		89.6	84.8	66.8		98.9	99.9	99.9
						2 Scales					
m_2	$T =$	1,500		m_2	$T =$	3,000		m_2	$T =$	6,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
2	26.8	18.3	7.1	2	41.5	32.0	16.1	2	64.3	53.0	34.0
5	30.1	21.1	8.6	5	41.0	31.7	17.7	5	66.6	54.6	35.5
10	29.5	19.3	7.7	10	42.0	30.9	14.6	10	62.0	50.9	32.6
20	27.8	18.6	6.6	20	38.5	27.3	13.1	20	61.0	48.2	27.2
m_2	$T =$	12,000		m_2	$T =$	30,000		m_2	$T =$	60,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
2	88.8	84.2	69.4	2	99.8	99.5	98.2		100	100	100
5	90.4	83.6	70.4	5	99.8	99.7	98.3		100	100	100
10	86.6	81.0	65.8	10	99.6	99.2	97.7		100	100	100
20	85.9	78.4	61.5	20	99.4	99.2	96.9		100	100	100

To provide some intuition behind the increased power brought about by the multiscale scale test, we plot in Fig.4.4 the sample version of the two scale EEs. Correctly specified EEs would result in histograms which centered closely around 0. In contrast, violation of one or more of the EEs would result in deviation from 0 of the corresponding histogram(s). A careful observation of the that the small deviation in the AR coefficient at HF is magnified through the LF estimating equations, and thus leads to a higher power of the two-scale test compared to a single scale test based on the HF sample only.

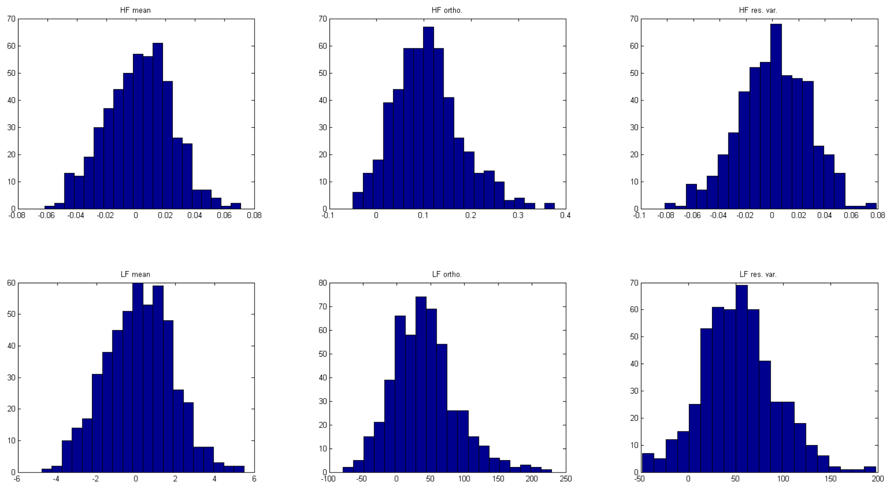

FIGURE 4.4: Sample moment conditions at two scales for the AR(1) model with LF at $m_2 = 10$.

ARMA processes

Size of the test

Parallel to the study of the AR(1) process example above, we conduct another study using an ARMA(1,1) process.

Table 4.3 shows the percentage of $W_B(\boldsymbol{\theta}_0) \geq \chi^{-1}_{(8)}(1-\alpha)$ ($\alpha = 0.1, 0.05, 0.01$) based on 1,000 repetitions for the ARMA(1,1) process with $(\phi_0, \phi_1, \theta_1, \sigma_Z^2)$ = (0.082, 0.9904, -0.9505, 1).

TABLE 4.3: Empirical size of the two-scale tests when the observations are generated from the ARMA(1,1) model with parameter vector $(\phi_0, \phi_1, \theta_1, \sigma_Z^2) =$ (0.082, 0.9904, -0.9505, 1).

m_2	$T=$	1,500		m_2	$T=$	3,000		m_2	$T=$	6,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
2	11.3	6.8	1.3	2	9.9	5.6	1.2	2	10.1	5.6	1.4
5	9.4	5.5	1.4	5	7.8	3.2	0.5	5	8.9	4.7	0.7
10	10.7	6.3	1.5	10	10.2	4.6	0.8	10	7.6	4.0	0.8
20	12.7	7.7	2.6	20	10.7	5.5	1.4	20	10.4	5.2	1.0
m_2	$T=$	12,000		m_2	$T=$	30,000		m_2	$T=$	60,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
2	12.0	6.5	1.1	2	9.9	4.9	1.4	2	9.0	5.8	1.8
5	9.9	5.3	0.9	5	9.1	5.0	0.8	5	9.5	4.0	1.0
10	8.6	4.5	0.7	10	8.8	4.3	0.8	10	9.2	4.8	0.6
20	9.3	5.6	0.9	20	9.4	4.9	1.0	20	10.5	4.9	1.3

As one can see from the empirical sizes in Table 4.3, the test has good empirical sizes.

Power of the test against small deviations in the parameters

To demonstrate the empirical power of the test, we test against a small deviation from the true parameter. We compare the power ot the multiscale with the corresponding single scale test where only HF estimating equations are used to demonstrate the advantage of using multiscale information.

Table 4.4 shows the percentage of $W_B(\boldsymbol{\theta}_0) \geq \chi^{-1}_{(8)}(1-\alpha)$ ($\alpha = 0.1, 0.05, 0.01$) based on 1,000 replications. In this case the data is generated from an ARMA(1,1) process with a parameter vector $(\phi_0, \phi_1, \theta_1, \sigma_Z^2) =$ (0.0082, 0.9904, - 0.9505, 1), and a slightly perturbed parameter of $\phi_1 = 0.9854$ is tested with the other parameters fixed at the true value.

We compare the power of tests based on a single scale EE (i.e. HF EE) and that based on the multiscale EEs with various LF levels. We observe that the tests based on the single scale EEs has a low power against the small deviation in the AR parameter. In contrast, multiscale EEs are consistently more powerful and the power increases quickly with increasing levels of aggregations and sample sizes.

TABLE 4.4: Empirical power comparison of one- and two-scale tests (in %) when the observations are generated from the ARMA(1,1) model with parameter vector $(\phi_0, \phi_1, \theta_1, \sigma_Z^2) =$ (0.0082, 0.9904, -0.9505, 1). Testing parameter (0.0082, 0.9854, -0.9505, 1).

					1	**Scale**					
	$T =$	1,500			$T =$	3,000			$T =$	6,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
	9.7	5.0	1.4		11.3	5.2	0.8		9.1	5.7	1.2
	$T =$	12,000			$T =$	30,00			$T =$	60,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
	11.3	5.8	1.5		12.5	6.7	1.4		14.8	8.2	2.2
					2	**Scales**					
m_2	$T =$	1,500		m_2	$T =$	3,000		m_2	$T =$	6,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
2	11.5	7.1	1.2	2	10.7	6.0	1.3	2	11.9	6.4	1.5
5	11.0	5.5	1.7	5	10.9	5.1	0.6	5	13.7	8.0	1.8
10	17.9	9.6	3.0	10	23.5	13.2	3.1	10	45.8	29.4	10.8
20	40.5	29.2	11.3	20	68.7	56.1	30.7	20	95.9	91.0	74.7
m_2	$T =$	12,000		m_2	$T =$	30,000		m_2	$T =$	60,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
2	14.1	7.3	1.8	2	16.2	9.4	2.6	2	24.9	14.1	4.3
5	25.1	12.9	3.9	5	61.2	44.2	16.0	5	96.7	92.1	66.9
10	86.3	73.6	38.8	10	100	100	99.6	10	100	100	100
20	100	100	99.7	20	100	100	100	20	100	100	100

4.6.2 Testing a model

In this part, we conduct simulation studies to investigate the empirical size and power properties of our proposed tests for testing a model. To investigate the empirical size property, we generate the data from an ARMA(1,1) model and test the model specified through the two-scale estimating equations (4.7) optimized under the constraints of temporal aggregation relations. To investigate the empirical power property, we generate the data from a particular ARMA(2,2) model.

Size of the test

Table 4.5 shows the percentage of $W_B(\tilde{\boldsymbol{\theta}}) \geq \chi^{-1}_{(4)}(1-\alpha)$ $(\alpha = 0.1, 0.05, 0.01)$ based on 1,000 repetitions for the ARMA(1,1) process with parameter vector $(\phi_0, \phi_1, \theta_1, \sigma_Z^2)$ = (0.082, 0.9904, -0.9505, 1).

TABLE 4.5: Empirical size of the two-scale tests when the observations are generated from the ARMA(1,1) model with parameter vector $(\phi_0, \phi_1, \theta_1, \sigma_Z^2)$ = (0.082, 0.9904, -0.9505, 1).

m_2	$T =$	3,000		m_2	$T =$	6,000		m_2	$T =$	12,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
10	5.3	3.8	2.0	10	7.7	5.6	2.4	10	8.3	6.5	2.1
20	4.8	2.7	1.2	20	6.0	3.3	1.0	20	7.2	4.6	1.4
30	6.2	4.4	2.4	30	5.5	3.4	1.3	30	6.7	3.8	0.8

As one can see from the empirical sizes in Table 4.5, the multiscale test generally has conservative empirical sizes in this case.

Power of the test against multiscale-type higher order data generating mechanism

To demonstrate the empirical power of the test, we test against an ARMA(2,2) process with a choice of parameter values motivated by the ARMA representation of two-component GARCH models, which has a multiscale volatility interpretation.

Table 4.6 shows the percentage of $W_B(\tilde{\boldsymbol{\theta}}) \geq \chi^{-1}_{(4)}(1-\alpha)$ $(\alpha = 0.1, 0.05, 0.01)$ based on 1,000 replications. In this case the data is generated from an ARMA(2,2) process with parameter vector $(\phi_0, \phi_1, \phi_2, \theta_1, \theta_2, \sigma_Z^2)$ = (0.00008, 1.9279, - 0.9280, -1.8644, 0.8652, 1).

From Table 4.6 we observe that the multiscale test has increasing powers along an increasing levels of aggregations and sample sizes.

TABLE 4.6: Empirical power of the two-scale tests (in %) when the observations are generated from the ARMA(2,2) model with parameter vector $(\phi_0, \phi_1, \phi_2, \theta_1, \theta_2, \sigma_Z^2)$ = (0.00008, 1.9279, - 0.9280, -1.8644, 0.8652, 1). Testing the ARMA(1,1) model.

m_2	$T=$	3,000		m_2	$T=$	6,000		m_2	$T=$	12,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
10	13.8	8.0	3.0	10	22.4	15.4	6.8	10	37.8	29.8	13.4
20	54.4	45.4	27.8	20	80.8	76.0	59.2	20	93.8	93.8	88.0
30	73.2	67.8	52.8	30	92.0	90.2	81.8	30	98.0	98.0	94.2

4.7 Section Conclusion

In this chapter, we present a class of novel tests of scale-consistency for ARMA models. We use the empirical likelihood framework to combine information from samples from different scales. The proposed testing framework can be used to test against deviations from the true parameters or departure from the true model

Simulation studies show that the proposed tests have good empirical size properties and superior power properties compared to the corresponding tests based on only HF sample in terms of detecting a small bias in the ARMA model parameters. The tests also have good empirical size properties and powers against particular higher order models with a multiscale type behavior.

Chapter 5

Empirical Likelihood-Based Two-scale Tests of Scale-Consistency for GARCH Processes

5.1 Chapter Introduction

In this chapter, we extend the two-scale test for the ARMA processes proposed in Chapter 4 to test scale consistency of GARCH processes.

Due to the fact that the squared observations in the GARCH models have ARMA representations, and the ARMA representation can be used to recover the GARCH parameters, we can test the scaling property of the GARCH processes through their linear ARMA representations.

5.2 Model, Testing Framework, and Adaptations

Let us consider a stationary strong GARCH(p,q) process:

$$\begin{aligned} \epsilon_t &= \sqrt{h_t} z_t \\ h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}^2, \end{aligned} \tag{5.1}$$

where $\{z_t\}$ is a sequence of *i.i.d.* random variables with mean zero and unit variance.

The squared observations of a GARCH(p,q) process admit the following restricted ARMA(r,p) representation:

$$\epsilon_t^2 = \alpha_0 + \sum_{i=1}^{r} (\alpha_i + \beta_i) \epsilon_{t-i}^2 + \nu_t - \sum_{j=1}^{p} \beta_j \nu_{t-j}, \tag{5.2}$$

where $r = \max(p,q)$, and $\nu_t = \epsilon_t^2 - h_t$ is an MDS with respect to the natural filtration of the process $\{\epsilon_t\}$. In particular, the $\{\nu_t\}$ is a temporally uncorrelated sequence. Therefore, the testing framework for the ARMA processes proposed in Chapter 4 can, in principle, be adapted to test scale-consistency of the GARCH processes by replacing the $\{X_t\}$ sequence of the ARMA(p,q) process by the $\{\epsilon_t^2\}$ sequence of the GARCH process. Other than this replacement, the implementation of the two-scale test for the GARCH(p,q) processes is exactly the same as the corresponding test for the ARMA(p,q) process. Specifically, in the case of the GARCH(1,1) process, which is considered in our simulation studies, we use estimating equations of the form of (4.7) with X_t replaced by ϵ_t^2 and the aggregated variables of the form

$$\bar{X}_{(m)t} = \epsilon_t^2 + \epsilon_{t-1}^2 + \cdots + \epsilon_{t-m+1}^2.$$

One issue needs to be emphasized when we apply the test to test the scaling behavior of return volatilities. When we model financial asset (log-)return volatility with GARCH processes, our multiscale test of scale consistency is applied to test whether an assumed GARCH model is appropriate for modeling both the volatility of HF returns ϵ_t and the volatility of m-period LF returns, which is denoted as

$$\bar{\epsilon}_{(m)t} := \epsilon_t + \epsilon_{t-1} + \cdots \epsilon_{t-m+1}.$$

From the temporal aggregation results given in Chapter 2, we can derive the dynamic of $\bar{\epsilon}_{(m)t}$ from the original HF model governing the dynamics of ϵ_t. Without loss of generality, consider ϵ_t following a ARCH(1) model, which implies

$$(1 - \alpha_1 L)\epsilon_t^2 = \alpha_0 + \nu_t,$$

where $\nu_t = \epsilon_t^2 - h_t$ is the driving white noise sequence in the AR representation of the ARCH model. An application of the temporal aggregation techniques yields the dynamics of the squared m-period returns:

$$\begin{aligned}(1 - \alpha_1^m L^m)\bar{\epsilon}_{(m)t}^2 &= (1 - \alpha_1^m L^m) \left(\sum_{i=1}^{m-1} L^i \epsilon_t \right)^2 \\ &= \left(\frac{1 - \alpha_1^m L^m}{1 - \alpha_1 L} \sum_{i=1}^{m-1} L^i \right) (\alpha_0 + \nu_t) - 2(1 - \alpha_1^m L^m) \sum_{2 \le i \le m-1} \sum_{1 \le j < i} \epsilon_{t-i}\epsilon_{t-j}.\end{aligned} \tag{5.3}$$

We may formally derive LF estimating equations based on (5.3) as we did for AR models. However, unlike AR processes, the cross-product terms $\sum_{2 \le i \le m-1} \sum_{1 \le j < i} \epsilon_{t-i}\epsilon_{t-j}$ create an additional complication when we calculate the variance of the residuals (i.e. the RHS of (5.3)). In particular, without the assumption that the distribution of ϵ_t is symmetric around zero, the cross products terms give rise to non-zero covariances with the ν_t terms, which can only be calculated with a further specification of the exact distributions of the ϵ_t's.

However, we may avoid dealing with the cross-product terms by adding the term

$$2(1 - \alpha_1^m L^m) \sum_{2 \le i \le m-1} \sum_{1 \le j < i} \epsilon_{t-i}\epsilon_{t-j}$$

to both sides of (5.3) to obtain the following equation:

$$(1 - \alpha_1^m L^m) \left(\sum_{i=1}^{m-1} L^i \epsilon_t^2 \right) = \left(\frac{1 - \alpha_1^m L^m}{1 - \alpha_1 L} \sum_{i=1}^{m-1} L^i \right) (\alpha_0 + \nu_t). \tag{5.4}$$

Equation (5.4) may be called the dynamics of sum of squared returns whereas equation (5.3) is squared LF returns (or sum of returns squared). Our original interest, which is in testing the dynamics of squared LF returns, can be tested similarly through the dynamics of sum of returns squared, which is mathematically simpler. Specifically, since the AR polynomials in (5.3) and in (5.4) are the same, the tests are equivalent in testing the scaling behavior of the AR part. The proposed approach can be justified in several ways. The AR part is generally of more interest than the MA part. Statistically, the AR part determines the asymptotic behavior of the autocorrelation function whereas the MA polynomial only affects the first few lags of the autocorrelation function. In addition, in the econometrics literature, some popular measures of persistence of an ARMA type process are defined in terms of the AR coefficients, like the sum of AR coefficients or the largest AR root in terms of absolute value. See, for example, Stock (1991)[70], Dias and Marques (2010)[12], and Hansen and Lunde (2014)[33]. Thus, our modification of the test does not affect the main interest in terms of testing the scaling behavior of the linear dependency structure, although there is some issues associated with a strict interpretation of the quantities involved.

Last but not least, the sacrifice of a strict interpretability for mathematical convenience may be necessary when temporal aggregation needs to be performed. An example can be found in Ohanissian, et.al.(2008)[60]. In that case, the log returns on a financial asset, r_t, is modeled by

$$r_t = \sigma \exp(Y_t/2)e_t, \tag{5.5}$$

where $\{Y_t\}$ is a stationary Gaussian long memory process independent of the i.i.d. mean-zero random variables $\{e_t\}$. In order to estimate the long memory parameter associated with the process Y_t, a transformation

$$W_t \equiv \log(r_t^2) = \log(\sigma^2) + Y_t + \log(e_t^2)$$

is taken to linearize the process. Then, to estimate the long-memory parameter of low frequency returns, (flow variable) temporal aggregation is performed on the transformed variable W_t, resulting in summations $\log(r_t^2)$, which does not have a straightforward financial interpretation.

5.3 Simulation Study

5.3.1 Testing a model with a particular set of parameters

In this section, we conduct simulation studies to investigate the finite sample properties of our proposed test in the case of GARCH models. The design of the simulations studies are parallel to that for the ARMA processes in Chapter 4. Namely, we conduct two types of tests: testing a model for a given set of parameters and testing the model itself. The alternative data generating mechanisms used in the empirical power study also parallels that in the ARMA case.

For every data generating process, we simulate from a strong model with i.i.d. normal innovations. The HF data corresponds to $m_1 = 1$, i.e. we use all of the simulated data. The LF m_2 is chosen at various values.

Size of the test

Table 5.1 shows the percentage of $W_B(\boldsymbol{\theta}_0) \geq \chi^{-1}_{(8)}(1-\alpha)$ ($\alpha = 0.1, 0.05, 0.01$) based on 1,000 repetitions for the GARCH(1,1) process with parameter vector $(\alpha_0, \alpha_1, \beta_1) = (0.0082,\ 0.0399,\ 0.9505)$.

TABLE 5.1: Empirical size of the two-scale tests when the observations are from the GARCH(1,1) model with parameter vector $(\alpha_0, \alpha_1, \beta_1) = (0.0082,\ 0.0399,\ 0.9505)$.

m_2	$T=$ 1,500			m_2	$T=$ 3,000			m_2	$T=$ 6,000		
	10%	5%	1%		10%	5%	1%		10%	5%	1%
2	8.6	5.2	0.4	2	7.3	4.6	0.4	2	9.8	4.7	0.7
5	11.3	6.5	2.3	5	9.8	6.4	1.5	5	8.9	4.6	1.1
10	8.8	6.3	2.4	10	8.5	3.9	1.4	10	9.2	4.1	1.2
20	13.5	5.6	1.9	20	11.8	6.5	1.9	20	10.1	5.1	1.6
m_2	$T=$ 12,000			m_2	$T=$ 30,000			m_2	$T=$ 60,000		
	10%	5%	1%		10%	5%	1%		10%	5%	1%
2	8.7	4.4	0.7	2	9.0	4.5	1.2	2	9.2	4.8	0.8
5	8.6	4.8	1.0	5	9.2	5.2	1.4	5	9.6	5.2	1.7
10	9.0	5.2	1.6	10	8.9	4.6	1.2	10	9.4	4.8	1.4
20	10.6	5.6	3.8	20	10.7	5.8	2.0	20	10.6	5.4	1.6

As one can see from the empirical sizes in Table 5.1, the test has empirical sizes in broad agreement with the nominal sizes as sample size increases.

Power of the test against small deviations in the parameters

To demonstrate the empirical power of the test, we test against a small deviation from the true parameter. We compare the power of the multiscale test with the corresponding single scale test where only HF estimating equations are used to demonstrate the advantage of using multiscale information.

Table 5.2 shows the percentage of $W_B(\boldsymbol{\theta}_0) \geq \chi^{-1}_{(8)}(1-\alpha)$ $(\alpha = 0.1, 0.05, 0.01)$ based on 1,000 replications. In this case the data is generated from a GARCH(1,1) process with parameter vector $(\alpha_0, \alpha_1, \beta_1) =$ (0.0082, 0.0399, 0.9505) and a slightly perturbed parameter value of $\alpha_1 = 0.0349$ is tested with the other parameters fixed at true value.

We compare the power of tests based on a single scale EE (i.e. HF EE) with that based on multiscale EEs with various LF levels. We observe that the tests based on a single scale EEs has virtually no power against the small deviation in the parameter value of α_1. In contrast, multiscale tests using aggregation levels of $m_2 = 5$ or higher are consistently more powerful, and the powers grow quickly with both the sample size and the levels of aggregation.

TABLE 5.2: Empirical power comparison of one- and two-scale tests (in %) when the observations are generated from the GARCH(1,1) model with parameter vector $(\alpha_0, \alpha_1, \beta_1)$ = (0.0082, 0.0399, 0.9505).
Testing parameter vector $(0.0082, 0.0349, 0.9505)$.

					1 Scale						
	$T=$	1,500			$T=$	3,000			$T=$	6,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
	1.1	0.6	0		1.1	0.6	0		0.7	0.2	0
	$T=$	12,000			$T=$	30,000			$T=$	60,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
	0.3	0.2	0.2		1.0	0.5	0		1.3	0.6	0
					2 Scales						
m_2	$T=$	1,500		m_2	$T=$	3,000		m_2	$T=$	6,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
2	2.5	1.3	0.4	2	2.0	1.5	0.4	2	1.4	0.6	0.3
5	13.4	9.0	3.6	5	17.8	12.0	5.7	5	30.3	19.6	7.0
10	26.7	17.4	6.7	10	53.0	40.2	19.8	10	89.4	81.9	60.4
20	64.0	54.1	32.6	20	91.6	87.3	77.1	20	99.7	99.6	97.8
m_2	$T=$	12,000		m_2	$T=$	30,000		m_2	$T=$	60,000	
	10%	5%	1%		10%	5%	1%		10%	5%	1%
2	1.9	0.4	0.3	2	4.5	3.2	0.4	2	8.7	4.8	1.0
5	78.6	61.7	26.8	5	100	100	99.8	5	100	100	100
10	100	100	98.6	10	100	100	100	10	100	100	100
20	100	100	100	20	100	100	100	20	100	100	100

5.3.2 Testing a model

In this part, we conduct simulation studies to investigate the empirical size and power of our proposed tests when testing a model. To investigate the empirical size property, we generate data from a GARCH(1,1) model and test the model specified through the two-scale estimating equations (4.7) optimized under the constraints of temporal aggregation relations. To investigate the empirical power property, we generate data from a particular GARCH(2,2) model.

Size of the test

Table 5.3 shows the percentage of $W_B(\tilde{\boldsymbol{\theta}}) \geq \chi^{-1}_{(4)}(1-\alpha)$ $(\alpha = 0.1, 0.05, 0.01)$ based on 1,000 repetitions for the GARCH(1,1) process with parameter vector $(\alpha_0, \alpha_1, \beta_1)$ = (0.0082, 0.9904, 0.9505).

As one can see from the empirical sizes in Table 5.3, the multiscale test generally has a smaller number of rejections of the null hypothesis than the nominal values.

TABLE 5.3: Empirical size of the two-scale tests when the observations are from the GARCH(1,1) model with parameter vector $(\alpha_0, \alpha_1, \beta_1)$ = (0.0082, 0.9904, 0.9505).

m_2	$T=$ 6,000			m_2	$T=$ 12,000			m_2	$T=$ 18,000		
	10%	5%	1%		10%	5%	1%		10%	5%	1%
10	3.3	1.9	0.6	10	2.3	1.5	0.4	10	5.1	2.2	0.8
20	6.0	3.6	1.7	20	6.9	3.3	1.0	20	6.2	3.9	1.0
30	5.4	3.7	0.9	30	6.2	3.9	1.0	30	7.5	4.8	0.6

Power of the test against multiscale-type higher order data generating mechanism

To demonstrate the empirical power of the test, we generate data from a particular GARCH(2,2) model corresponding to a two-component GARCH model estimated with real data, which has a multiscale volatility interpretation.

Table 5.4 shows the percentage of $W_B(\tilde{\boldsymbol{\theta}}) \geq \chi^{-1}_{(4)}(1-\alpha)$ $(\alpha = 0.1, 0.05, 0.01)$ based on 1,000 replications. In this case data is generated from a GARCH(2,2) process with parameter vector $(\alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2)$ = (7e-5, 0.0635, -0.0628, 1.8644, -0.8652) and a GARCH(1,1) model is tested.

TABLE 5.4: Empirical power of the two-scale tests (in %) when the observations are generated from the GARCH(2,2) model with parameter vector $(\alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2)$ = (7e-5, 0.0635, -0.0628, 1.8644, -0.8652). Testing the GARCH(1,1) model.

m_2	$T=$ 6,000			m_2	$T=$ 12,000			m_2	$T=$ 18,000		
	10%	5%	1%		10%	5%	1%		10%	5%	1%
10	7.3	4.6	3.4	10	6.6	3.1	1.6	10	10.5	6.9	4.1
20	32.1	23.5	9.7	20	55.2	46.3	30.9	20	76.8	66.8	47.8
30	47.6	38.4	21.5	30	76.3	67.2	50.2	30	92.9	87.3	73.9

From table 5.4 we observe that a fairly large sample (i.e. $T = 18,000$) size and a high level of aggregation at the LF is needed in order for the two-scale test to have a significant power.

5.4 Empirical Study

In this section, we apply the two-scale test to the Dow Jones Industrial Average (DJIA) index return sample used in Engle and Patton (2001)[22]. The sample data used in Engle and Patton (2001)[22] is the DJIA daily percentage returns from August 23, 1988 to August 22, 2000. It contains a total of 3,130 observations. Engle and Patton (2001)[22] used quasi-maximum likelihood estimation (QMLE) based on a Gaussian innovation distribution. They found that the Schwarz Information Criterion favors the GARCH(1,1) model in the GARCH(p,q) class for $p \in [1, 5]$ and $q \in [1, 2]$. The resulting QMLE estimates of the model parameters are

$$\alpha_0 = 0.0082, \alpha_1 = 0.0399, \beta_1 = 0.9505.$$

In addition, squared returns normalized by conditional variance filtered using a GARCH(1,1) model with this set of parameters pass the Ljung-Box Q test of white noise, suggesting that the persistence in the variance of returns has been adequately captured and the standardized residuals are white noise.

In the following subsections, we conduct two-scale tests of consistency to test whether (i) the GARCH(1,1) model as estimated in Engle and Patton (2001) is consistent with the DJIA sample at different scales and (ii) whether the GARCH(1,1) model itself is consistent with the DJIA sample at different scales.

5.4.1 Two-scale testing of QMLE estimates

We test the GARCH(1,1) model with the set of QMLE as estimated in Engle and Patton (2001)[22] using our proposed two-scale test for its consistency with data at two scales. This corresponds to the situation of testing a model with a particular set of parameters. For the purpose of comparison, we also test the QMLE at a single scale (i.e. daily). Since the QMLE does not contain an estimate of the variance of the innovations $\{\nu_t\}$ in the ARMA representation of squared observation in the GARCH process, σ_ν^2, we first estimate the parameter σ_ν^2. We choose the value of σ_ν^2 such that the profile empirical likelihood based on the HF data is maximized. This yields an estimate of $\hat{\sigma}_\nu^2 = 4.34$.

While the choice of block size is based on calibrating the size of the tests to the nominal values in the simulation studies, there seems to be no rule for choosing a block size in the empirical studies. We thus conduct tests using a range of values for the block size. As exemplified in Fig.5.1, the value of the EL test statistic is usually a slowly changing and smooth function of the block size and the conclusion of the test is consistent with respect to a range of values of the block size.

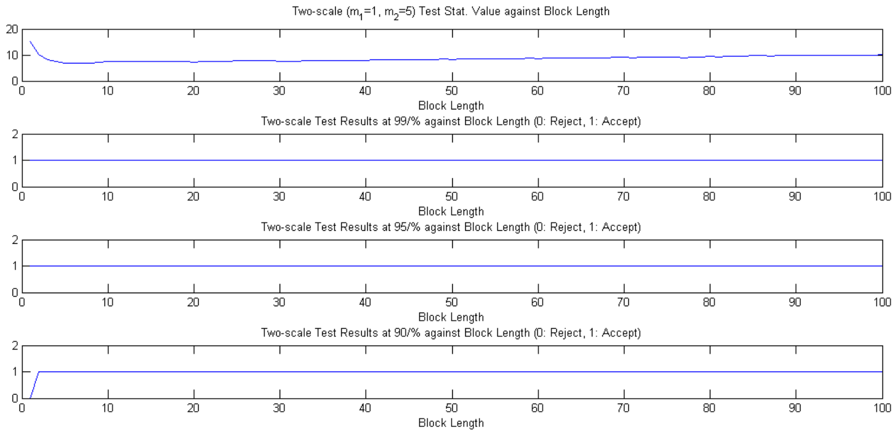

FIGURE 5.1: Two-scale test statistic values and test results against block length for the DJIA sample data.

Table 5.5 summarizes the test results of single and two-scale tests for the DJIA sample data. First of all, the single scale test based on only daily returns indicates that the GARCH(1,1) model with the parameter vector $(\alpha_0, \alpha_1, \beta_1) = (0.0082,$ $0.0399, 0.9505)$ is not rejected at the daily scale. Next, the GARCH(1,1) model is rejected by the two-scale tests using daily and two-day returns, daily and three-day returns, and daily and 60-day returns at higher confidence levels. A closer look at the values of two-scale test statistics associated with 2-day, 3-day, and 60-day tests indicates that these values are not too large above the corresponding threshold values. Thus we interpret the test results as that the GARCH(1,1) model with the parameter vector $(\alpha_0, \alpha_1, \beta_1) = (0.0082, 0.0399, 0.9505)$ not being inconsistent at the pairs of scales $m_1 = 1$ and m_2 with $m_2 = 4, 5, 10, 20, 30$, while slightly inconsistent at the pairs of scales $m_1 = 1$ and $m_2 = 2, 3, 60$.

The rejections of the estimated GARCH(1,1) model at scales $m_2 = 2$ and $m_2 = 3$ do not seem to fit into the patterns we have observed in the simulation studies where rejections at lower frequencies are usually followed by rejections at higher frequencies. Therefore, we suspect that there are features in the data at 2-day and 3-day scales which have caused violations of scale-consistency in ways other than those considered in our simulation studies (i.e. small deviation in the parameters or multiscale-type data generating process). In the literature, 2-day and 3-day returns are much less studied compared to returns over horizons with calendar meaning such as 5-day (weekly), 10-day (bi-weekly) and so on. Therefore we do not pursue further investigations into the reasons why the estimated GARCH(1,1) model is rejected at 2- and 3-day scales.

Overall, the GARCH(1,1) model estimated with a daily return sample of size 3,130 daily return observations as in Engle and Patton (2001) is not rejected by our two-scale consistency test over a range of scale up to 20-day scale, except on 2-day and 3-day scales. We would also like to emphasis here some of the limitations that the above conclusions may subject to. First of all, given the sample size of 3,130, the empirical power property of the two-scale test with the LF being less than 20 may not be high enough as one can see from the simulation studies (i.e. Table 5.2). Secondly, as part of the nature of two-scale tests, they test the pair-wise consistency of the model at the HF scale (i.e. daily) and one particular LF scale instead of the simultaneous consistency at an arbitrary set of scales. In the next chapter, we extend the two-scale tests to multiscale tests which address both limitations.

Table 5.5: Empirical likelihood testing results of GARCH(1,1) model with parameter vector $(\alpha_0, \alpha_1, \beta_1) = (0.0082,\ 0.0399,\ 0.9505)$ and $\hat{\sigma}_\nu^2 = 4.34$ on the DJIA return sample as in Engle and Patton (2001).

One Scale (daily)			
	10%	5%	1%
Test statistics $\chi^2(4)$			
Critical values	7.78	9.49	13.28
Test statistics value:	5.79		
Test Conclusion	A	A	A
Two Scales			
	10%	5%	1%
Test statistics $\chi^2(8)$			
Critical values	13.36	15.51	20.09
Test statistics value (average):			
$m_2 = 2$	19.89		
Test Conclusion	R	R	A
$m_2 = 3$	15.41		
Test Conclusion	R	A	A
$m_2 = 4$	10.11		
Test Conclusion	A	A	A
$m_2 = 5$	8.61		
Test Conclusion	A	A	A
$m_2 = 10$	6.35		
Test Conclusion	A	A	A
$m_2 = 20$	7.76		
Test Conclusion	A	A	A
$m_2 = 30$	8.40		
Test Conclusion	A	A	A
$m_2 = 60$	17.73		
Test Conclusion	R	R	A

A: accept; R: reject.

(Test statistic values is averaged over a range of choices of block length.)

5.4.2 Two-scale testing of the GARCH(1,1) model

In this section, we conduct two-scale test of the GARCH(1,1) model using the DJIA sample data. This corresponds to testing the model itself. We have already seen from the preceding subsection that the GARCH(1,1) model with the set of QMLE as estimated in Engle and Patton (2001)[22] is not rejected by the two-scale tests for a range of scales, except at scales 2, 3, 30 and 60. The tests in this subsection helps to answer whether the GARCH(1,1) model can be consistent for the data at the daily and LF scales at 2, 3, 30 or 60.

Table 5.6 shows the results of two-scale testing. We observe that the conclusions of two-scale tests of the GARCH(1,1) model for the DJIA index return sample generally follows the same pattern of the corresponding tests of the GARCH(1,1) with a particular set of QMLE. This indicates that the set of QMLE as estimated in Engle and Patton (2001)[22] does a fairly good job in enabling the GARCH(1,1) model to fit the sample data at different scales as the model itself is capable of.

TABLE 5.6: Log profile empirical likelihood statistics value for testing the GARCH(1,1) model using sample from HF and various LFs on the DJIA return sample as in Engle and Patton (2001)[22].

Two-Scale Test of Model	10%	5%	1%
Test statistics $\chi^2(4)$			
Critical values	7.78	9.49	13.28
Test statistics value (average):			
$m_2 = 2$	4.49		
Test Conclusion	A	A	A
$m_2 = 3$	11.03		
Test Conclusion	R	R	A
$m_2 = 4$	1.45		
Test Conclusion	A	A	A
$m_2 = 5$	0.97		
Test Conclusion	A	A	A
$m_2 = 10$	1.42		
Test Conclusion	A	A	A
$m_2 = 20$	6.60		
Test Conclusion	A	A	A
$m_2 = 30$	7.17		
Test Conclusion	A	A	A
$m_2 = 60$	9.48		
Test Conclusion	R	A	A

A: accept; R: reject.

(Test statistic values is averaged over a range of choices of block length.)

5.5 Section Conclusion

In this section, we have presented the two-scale tests of consistency for GARCH processes. The theoretical development of the two-scale tests for GARCH processes follows closely the corresponding theory for the ARMA processes with some minor adaptation.

Simulation studies show that the multiscale test is useful for detecting a small bias in models parameters which is otherwise difficult to detect by using only a high frequency sample. It is also useful for detecting a certain type of model misspecification.

Empirical study on a sample of Dow Jones Industrial Average index return sample indicates that a GARCH(1,1) model accepted at the high frequency data may not be adequate for data at certain lower frequencies.

Chapter 6

Empirical Likelihood-Based Multiscale Tests of Scale-Consistency

6.1 Chapter Introduction

The two-scale tests described in the previous chapters can be generalized straightforwardly to using a number of S scales where $S \geq 3$ [1]. Such a generalization is both natural and of practical values. Firstly, it can lead to an increase in the power of the tests. Secondly, it allows one to test the model at a selection of scales over which the model may be used.

As in the two-scale tests, we start with a set of high frequency (HF) estimating equations at scale m_1 and then add corresponding estimating equations from low frequencies (LF). Instead of having only one LF scale m_2 as in the two-scale test, we add estimating equations from a set of LF scales m_2, ..., m_S.

6.2 The Multiscale Tests

We present the S-scale version of the test in the context of the ARMA(p,q) process (4.1). Adaption of the test to GARCH processes follow the same steps outlined in Chapter 5.

[1] The two-scale tests correspond to the case of $S = 2$.

6.2.1 S-scale Estimating Equations

The same rules for choosing estimating equations described in section 4.2 apply to the construction of multiscale estimating equations. Specifically, we can construct multiscale estimating equations analogous to the two-scales estimating equations in Example 4.2 as follows.

Example 6.1 S-scale estimating equations for the ARMA(1,1) process based on flow aggregation

$$
\begin{aligned}
E[X_t - \phi_0 - \phi_1 X_{t-1}] &= 0,\\
E[(X_t - \phi_0 - \phi_1 X_{t-1})X_{t-2}] &= 0,\\
E[(X_t - \phi_0 - \phi_1 X_{t-1})^2] - (1+\theta_1^2)\sigma_Z^2 &= 0,\\
E[(X_t - \phi_0 - \phi_1 X_{t-1})(X_{t-1} - \phi_0 - \phi_1 X_{t-2})] - \theta_1\sigma_Z^2 &= 0,\\
E[\bar{X}_{(m_2)t} - \bar{\phi}_{(m_2)0} - \bar{\phi}_{(m_2)1}\bar{X}_{(m_2)t-m_2}] &= 0,\\
E[(\bar{X}_{(m_2)t} - \bar{\phi}_{(m_2)0} - \bar{\phi}_{(m_2)1}\bar{X}_{(m_2)t-m_2})\bar{X}_{(m_2)t-m_2-1}] &= 0,\\
E[(\bar{X}_{(m_2)t} - \bar{\phi}_{(m_2)0} - \bar{\phi}_{(m_2)1}\bar{X}_{(m_2)t-m_2})^2] - \bar{\sigma}^2_{(m_2)Z} &= 0,\\
E[(\bar{X}_{(m_2)t} - \bar{\phi}_{(m_2)0} - \bar{\phi}_{(m_2)1}\bar{X}_{(m_2)t-m_2}) & \\
\times(\bar{X}_{(m_2)t-m_2} - \bar{\phi}_{(m_2)0} - \bar{\phi}_{(m_2)1}\bar{X}_{(m_2)t-2m_2})] - \bar{\gamma}^{(1)}_{(m_2)Z} &= 0,\\
\cdots & \\
E[\bar{X}_{(m_S)t} - \bar{\phi}_{(m_S)0} - \bar{\phi}_{(m_S)1}\bar{X}_{(m_S)t-m_S}] &= 0,\\
E[(\bar{X}_{(m_S)t} - \bar{\phi}_{(m_S)0} - \bar{\phi}_{(m_S)1}\bar{X}_{(m_S)t-m_S})\bar{X}_{(m_S)t-m_S-1}] &= 0,\\
E[(\bar{X}_{(m_S)t} - \bar{\phi}_{(m_S)0} - \bar{\phi}_{(m_S)1}\bar{X}_{(m_S)t-m_S})^2] - \bar{\sigma}^2_{(m_S)Z} &= 0,\\
E[(\bar{X}_{(m_S)t} - \bar{\phi}_{(m_S)0} - \bar{\phi}_{(m_S)1}\bar{X}_{(m_S)t-m_S}) & \\
\times(\bar{X}_{(m_S)t-m_S} - \bar{\phi}_{(m_S)0} - \bar{\phi}_{(m_S)1}\bar{X}_{(m_S)t-2m_S})] - \bar{\gamma}^{(1)}_{(m_S)Z} &= 0,
\end{aligned}
\tag{6.1}
$$

where $\bar{\phi}_{(m)0}$, $\bar{\phi}_{(m)1}$, $\bar{\sigma}^2_{(m)Z}$, $\bar{\gamma}^{(1)}_{(m)Z}$ represent the LF intercept, AR coefficient, and the LF residual variance and lag-1-auto-covariance at scales $m = m_2, \cdots, m_S$, respectively. The parameters in the LF estimating equations are assumed to be independent of the HF model parameters as in the two-scale tests. These new parameters are introduced so that the system (6.1) is just-determined and thus always has a unique solution. We will refer to $(\phi_0, \phi_1, \theta_1, \sigma_Z^2)$ as the HF model parameters, abbreviated as $\boldsymbol{\theta}^{HF}$, and to $(\bar{\phi}_{(m_2)0}, \bar{\phi}_{(m_2)1}, \bar{\sigma}^2_{(m)Z}, \bar{\gamma}^{(1)}_{(m_2)Z}, \cdots, \bar{\phi}_{(m_S)0}, \bar{\phi}_{(m_S)1}, \bar{\sigma}^2_{(m)Z}, \bar{\gamma}^{(1)}_{(m_S)Z})$ altogether as the LF model parameters, abbreviated as $\boldsymbol{\theta}^{LF}$.

6.2.2 The null and the alternative hypotheses

For a given ARMA(p,q) process, our proposed multiscale test tests the following hypothesis:

$$\begin{aligned} H_0 &: f_{(m_2,\cdots,m_S)}(\boldsymbol{\theta}^{HF},\boldsymbol{\theta}^{LF}) = 0, \\ H_A &: f_{(m_2,\cdots,m_S)}(\boldsymbol{\theta}^{HF},\boldsymbol{\theta}^{LF}) \neq 0, \end{aligned} \tag{6.2}$$

where the function $f_{(m_2,\cdots,m_S)}$ is determined by the temporal aggregation relation between the high and low frequency model parameters in the ARMA(p,q) process at the HF $m_1 = 1$ and each of the LFs m_2 through m_S. We give an example in the case of $S = 3$ below.

Example 6.1 (continued) A specialization of the multiscale test in the case of the ARMA(1,1) process with $S = 3$, i.e. a HF scale and two LF scales, $m_2 = 2$ and $m_3 = 3$, tests the following hypothesis

$$\begin{aligned} H_0 &: f_{(m_2,m_3)}(\boldsymbol{\theta}^{HF},\boldsymbol{\theta}^{LF}) = 0, \\ H_A &: f_{(m_2,m_3)}(\boldsymbol{\theta}^{HF},\boldsymbol{\theta}^{LF}) \neq 0, \end{aligned} \tag{6.3}$$

where

$$\boldsymbol{\theta}^{HF} := (\phi_0, \phi_1, \theta_1, \sigma_Z^2),$$

$$\boldsymbol{\theta}^{HF} := (\bar{\phi}_{(m_2)0}, \bar{\phi}_{(m_2)1}, \bar{\sigma}^2_{(m_2)Z}, \bar{\gamma}^{(1)}_{(m_2)Z}, \bar{\phi}_{(m_3)0}, \bar{\phi}_{(m_3)1}, \bar{\sigma}^2_{(m_3)Z}, \bar{\gamma}^{(1)}_{(m_3)Z}),$$

and

$$f_{(m_2,m_3)}(\boldsymbol{\theta}^{HF},\boldsymbol{\theta}^{LF}) := \begin{pmatrix} (1+\phi_1)\cdot 2\cdot\phi_0 - \bar{\phi}_{(m_2)0} \\ \phi_1^2 - \bar{\phi}_{(m_2)1} \\ [1+(1+\phi_1+\theta_1)^2 + (\phi_1+(1+\phi_1)\theta_1)^2 + \phi_1^2\theta_1^2]\sigma_Z^2 - \bar{\sigma}^2_{(m_2)Z} \\ [(\phi_1+\theta_1+\phi_1\theta_1) + \phi_1\theta_1(1+\phi_1+\theta_1)]\,\sigma_Z^2 - \bar{\gamma}^{(1)}_{(m_2)Z} \\ (1+\phi_1+\phi_1^2)\cdot 3\cdot\phi_0 - \bar{\phi}_{(m_3)0} \\ \phi_1^3 - \bar{\phi}_{(m_3)1} \\ [1+(1+\phi_1+\theta_1)^2 + [(1+\phi_1)(1+\theta_1)+\phi_1^2]^2 + [\phi_1(1+\phi_1)(1+\theta_1)+\theta_1]^2 \\ +[\phi_1\theta_1+\phi_1^2(1+\theta_1)]^2 + \phi_1^4\theta_1^2]\sigma_Z^2 - \bar{\sigma}^2_{(m_3)Z} \\ [[\phi_1(1+\phi_1)(1+\theta_1)+\theta_1] \\ +[\phi_1\theta_1+\phi_1^2(1+\theta_1)](1+\phi_1+\theta_1) + \phi_1^2\theta_1[(1+\phi_1)(1+\theta_1)+\phi_1^2]]\sigma_Z^2 - \bar{\gamma}^{(1)}_{(m_3)Z} \end{pmatrix}. \tag{6.4}$$

The elements in the vector in (6.4) are the differences between (1) the LF parameters at each of the LF scale implied by the HF parameters and (2) the directly estimated LF parameters. The first four elements correspond to the intercept term, the AR coefficient, residual variance, and first-order residual auto-covariance at scale $m_2 = 2$. The next four are analogous elements for the scale $m_3 = 3$. □

Similar to the two-scale tests, we consider two types of testing problems: (i) testing a model with a particular set of parameters and (ii) testing the model itself.

6.2.3 Constructing S-frequency samples as a vector of observations

In order to formulate the empirical likelihood inference problem using samples from S time scales, we need to formulate samples corresponding to the S scales.

Denote

$$\begin{aligned}\bar{e}_{(m_k)t}(\boldsymbol{\theta}) &:= \left[1 - \sum_{j=1}^{p} \bar{\phi}_{(m_k)j} L^{m_k}\right] \left(\sum_{i=1}^{m_k-1} L^i\right) X_t - \bar{\phi}_{(m_k)0} \\ &= \left[1 - \sum_{j=1}^{p} \bar{\phi}_{(m_k)j} L^{m_k}\right] \bar{X}_{(m_k)t} - \bar{\phi}_{(m_k)0}\end{aligned} \tag{6.5}$$

for $k = 1, \cdots, S$. Then, $\{e_t, t \in \mathbb{Z}^+ \text{ and } t \leq T\}$ is the HF sample and $\{\bar{e}_{(m_k)t}, t \in \mathbb{Z}^+ \text{ and } t \leq T\}$, $k = 2, \cdots, S$ are the LF samples. Stacking the HF and LF samples into a vector, we have $\{(e_t, \bar{e}_{(m_2)t}, \cdots, \bar{e}_{(m_S)t})', t \in \mathbb{Z}^+ \text{ and } t \leq T\}$ as our vector-valued observations. By doing so, we can cast the inference problem as a vector-valued (block) empirical likelihood framework. More detailed examples will be provided in the following subsection.

6.2.4 Empirical likelihood inference based on multiscale estimating equations

To test the null hypothesis given in (6.2) based on a sample data, we now formulate the corresponding estimating equations to be used for the empirical likelihood inference.

Using the given observations, we form estimating equations which are satisfied by the postulated ARMA(p,q) process at every scales from $m_1 = 1$ through m_S.

We use estimating equations analogous to the ones used in Example 6.1. These are straightforward generalizations of the estimating equations for the ARMA(1,1) process. At the HF, $m_1 = 1$, we have:

$$\sum_t g_{HF,t}(\phi_0, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma_Z^2) = 0 \tag{6.6}$$

where

$$g_{HF,t}(\phi_0, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma_Z^2) := \begin{pmatrix} e_t(\boldsymbol{\theta}) \\ e_t(\boldsymbol{\theta})X_{t-p-1} \\ e_t(\boldsymbol{\theta})X_{t-p-2} \\ \cdots \\ e_t(\boldsymbol{\theta})X_{t-2p} \\ e_t(\boldsymbol{\theta})^2 - (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma_Z^2 \\ e_t(\boldsymbol{\theta})e_{t-1}(\boldsymbol{\theta}) - (\theta_1 + \theta_2\theta_1 \cdots + \theta_q\theta_{q-1})\sigma_Z^2 \\ \cdots \\ e_t(\boldsymbol{\theta})e_{t-q+1}(\boldsymbol{\theta}) - (\theta_{t-q+1} + \theta_q\theta_1)\sigma_Z^2 \end{pmatrix}. \tag{6.7}$$

Next, we add estimating equations from the LFs. In the case of the flow aggregation, they are generalizations of LF estimating equations in (4.7). For the LF scale m_k where $2 \le k \le S$, we have

$$\sum_t g_{LF,t}(\bar{\phi}_{(m_k)0}, \bar{\phi}_{(m_k)1}, \cdots, \bar{\phi}_{(m_k)p}, \bar{\sigma}^2_{(m_k)Z}, \bar{\gamma}^{(1)}_{(m_k)Z}, \cdots, \bar{\gamma}^{(q)}_{(m_k)Z}) = 0 \tag{6.8}$$

where

$$g_{LF,t}(\bar{\phi}_{(m_k)0}, \bar{\phi}_{(m_k)1}, \cdots, \bar{\phi}_{(m_k)p}, \bar{\sigma}^2_{(m_k)Z}, \bar{\gamma}^{(1)}_{(m_k)Z}, \cdots, \bar{\gamma}^{(q)}_{(m_k)Z}) := \begin{pmatrix} \bar{e}_{(m_k)t}(\boldsymbol{\theta}) \\ \bar{e}_{(m_k)t}(\boldsymbol{\theta})\bar{X}_{(m_k)t-p(m_k-1)-q-m_k} \\ \bar{e}_{(m_k)t}(\boldsymbol{\theta})\bar{X}_{(m_k)t-p(m_k-1)-q-m_k-1} \\ \cdots \\ \bar{e}_{(m_k)t}(\boldsymbol{\theta})\bar{X}_{(m_k)t-p(m_k-1)-q-m_k-(p-1)} \\ \bar{e}_{(m_k)t}(\boldsymbol{\theta})^2 - \bar{\sigma}^2_{(m_k)Z} \\ \bar{e}_{(m_k)t}(\boldsymbol{\theta})\bar{e}_{(m_k)t-m_k}(\boldsymbol{\theta}) - \bar{\gamma}^{(1)}_{(m_k)Z} \\ \cdots \\ \bar{e}_{(m_k)t}(\boldsymbol{\theta})\bar{e}_{(m_k)t-qm_k}(\boldsymbol{\theta}) - \bar{\gamma}^{(q)}_{(m_k)Z} \end{pmatrix}. \tag{6.9}$$

Stacking the HF and LF estimating equations into a single vector, we have the final vector of estimating equations as

$$\sum_t g_t(\phi_0, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma_Z^2, \bar{\phi}_{(m_2)0}, \bar{\phi}_{(m_2)1}, \cdots, \bar{\phi}_{(m_2)p}, \bar{\sigma}^2_{(m_2)Z}, \bar{\gamma}^{(1)}_{(m_2)Z}, \cdots,$$
$$\bar{\gamma}^{(q)}_{(m_2)Z} \cdots, \bar{\phi}_{(m_S)0}, \bar{\phi}_{(m_S)1}, \cdots, \bar{\phi}_{(m_S)p}, \bar{\sigma}^2_{(m_S)Z}, \bar{\gamma}^{(1)}_{(m_S)Z}, \cdots, \bar{\gamma}^{(q)}_{(m_S)Z}) = 0, \tag{6.10}$$

where

$$
\begin{aligned}
&g_t(\phi_0, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma_Z^2, \bar{\phi}_{(m_2)0}, \bar{\phi}_{(m_2)1}, \cdots, \bar{\phi}_{(m_2)p}, \bar{\sigma}^2_{(m_2)Z}, \bar{\gamma}^{(1)}_{(m_2)Z}, \cdots, \bar{\gamma}^{(q)}_{(m_2)Z} \\
&\quad \cdots, \bar{\phi}_{(m_S)0}, \bar{\phi}_{(m_S)1}, \cdots, \bar{\phi}_{(m_S)p}, \bar{\sigma}^2_{(m_S)Z}, \bar{\gamma}^{(1)}_{(m_S)Z}, \cdots, \bar{\gamma}^{(q)}_{(m_S)Z})
\end{aligned}
$$

$$
:= \begin{pmatrix}
e_t(\boldsymbol{\theta}) \\
e_t(\boldsymbol{\theta})X_{t-p-1} \\
e_t(\boldsymbol{\theta})X_{t-p-2} \\
\cdots \\
e_t(\boldsymbol{\theta})X_{t-2p} \\
e_t(\boldsymbol{\theta})^2 - (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma_Z^2 \\
e_t(\boldsymbol{\theta})e_{t-1}(\boldsymbol{\theta}) - (\theta_1 + \theta_2\theta_1 \cdots + \theta_q\theta_{q-1})\sigma_Z^2 \\
\cdots \\
e_t(\boldsymbol{\theta})e_{t-q+1}(\boldsymbol{\theta}) - (\theta_{t-q+1} + \theta_q\theta_1)\sigma_Z^2 \\
\bar{e}_{(m_2)t}(\boldsymbol{\theta}) \\
\bar{e}_{(m_2)t}(\boldsymbol{\theta})\bar{X}_{(m_2)t-p(m_2-1)-q-m_2} \\
\bar{e}_{(m_2)t}(\boldsymbol{\theta})\bar{X}_{(m_2)t-p(m_2-1)-q-m_2-1} \\
\cdots \\
\bar{e}_{(m_2)t}(\boldsymbol{\theta})\bar{X}_{(mm_2)t-p(m_2-1)-q-m_2-(p-1)} \\
\bar{e}_{(m_2)t}(\boldsymbol{\theta})^2 - \bar{\sigma}^2_{(m_2)Z} \\
\bar{e}_{(m_2)t}(\boldsymbol{\theta})\bar{e}_{(m_2)t-m_2}(\boldsymbol{\theta}) - \bar{\gamma}^{(1)}_{(m_2)Z} \\
\cdots \\
\bar{e}_{(m_S)t}(\boldsymbol{\theta})\bar{e}_{(m_S)t-qm_S}(\boldsymbol{\theta}) - \bar{\gamma}^{(q)}_{(m_S)Z} \\
\cdots \\
\bar{e}_{(m_S)t}(\boldsymbol{\theta}) \\
\bar{e}_{(m_S)t}(\boldsymbol{\theta})\bar{X}_{(m_S)t-p(m_S-1)-q-m_S} \\
\bar{e}_{(m_S)t}(\boldsymbol{\theta})\bar{X}_{(m_S)t-p(m_S-1)-q-m_S-1} \\
\cdots \\
\bar{e}_{(m_2)t}(\boldsymbol{\theta})\bar{X}_{(m_S)t-p(m_S-1)-q-m_S-(p-1)} \\
\bar{e}_{(m_S)t}(\boldsymbol{\theta})^2 - \bar{\sigma}^2_{(m_S)Z} \\
\bar{e}_{(m_S)t}(\boldsymbol{\theta})\bar{e}_{(m_S)t-m_S}(\boldsymbol{\theta}) - \bar{\gamma}^{(1)}_{(m_S)Z} \\
\cdots \\
\bar{e}_{(m_S)t}(\boldsymbol{\theta})\bar{e}_{(m_S)t-qm_S}(\boldsymbol{\theta}) - \bar{\gamma}^{(q)}_{(m_S)Z}
\end{pmatrix}. \tag{6.11}
$$

Block empirical likelihood inference is then based on the following profile empirical likelihood function

$$\mathcal{R}_B(\boldsymbol{\theta}) = \sup\left\{\prod_{i=1}^{Q} Qw_i^B | w_i^B > 0, \sum_{i=1}^{Q} w_i^B = 1, \sum_{i=1}^{Q} w_i^B T_i(\boldsymbol{\theta}) = 0\right\}, \tag{6.12}$$

where

$$\begin{aligned}\boldsymbol{\theta} = (&\phi_0, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma_Z^2, \bar{\phi}_{(m_2)0}, \bar{\phi}_{(m_2)1}, \cdots, \bar{\phi}_{(m_2)p}, \bar{\sigma}^2_{(m_2)Z}, \bar{\gamma}^{(1)}_{(m_2)Z}, \cdots, \\ &\bar{\gamma}^{(q)}_{(m_2)Z} \cdots, \bar{\phi}_{(m_S)0}, \bar{\phi}_{(m_S)1}, \cdots, \bar{\phi}_{(m_S)p}, \bar{\sigma}^2_{(m_S)Z}, \bar{\gamma}^{(1)}_{(m_S)Z}, \cdots, \bar{\gamma}^{(q)}_{(m_S)Z})\end{aligned}$$

is the parameter vector, $T_i(\boldsymbol{\theta}) = \frac{1}{M}\sum_{j=1}^{M} g_{(i-1)L+j}(\boldsymbol{\theta})$ is the blocked observations.

Decision rule for rejecting the null hypothesis H_0

The same asymptotic results for the two-scale tests in Chapter 4 apply to the multiscale tests. According to these results, we have that the log profile empirical likelihood statistics $W(\boldsymbol{\theta}_0)$ converges to a $\chi^2_{S(p+q+2)}$ distribution and $W(\tilde{\boldsymbol{\theta}})$ converges to a $\chi^2_{(S-1)(p+q+2)}$ distribution under the null hypothesis.

Testing $H_0 : f_{(m)}(\boldsymbol{\theta}) = 0$ **at** $\boldsymbol{\theta} = \boldsymbol{\theta}_0$
The decision rule for the test is the following: if the value of $W(\boldsymbol{\theta}_0)$ is greater than the $(1-\alpha)$-quantile of a $\chi^2_{S(p+q+2)}$ distribution, then we reject the null hypothesis $H_0 : f_{(m)}(\boldsymbol{\theta}_0) = 0$ at the level of significance α.

Testing $H_0 : f_{(m)}(\boldsymbol{\theta}) = 0$
The decision rule for the test is the following: if the minimal value of $W(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, denoted as $W(\tilde{\boldsymbol{\theta}})$, is greater than the $(1-\alpha)$-quantile of a $\chi^2_{(S-1)(p+q+2)}$ distribution, then we reject the null hypothesis $H_0 : f_{(m)} = 0$ at the level of significance α.

Example 6.1 (continued) In the case of the ARMA(1,1) process, the estimating equations (4.7) contain a total of 12 estimating equations which are not linearly dependent. Therefore, $W(\boldsymbol{\theta}_0)$ converges to a $\chi^2_{(12)}$ distribution and $W(\tilde{\boldsymbol{\theta}})$ converges to a $\chi^2_{(8)}$ distribution under the null hypothesis. Decision rules for rejecting the null hypothesis can be made accordingly. □

6.3 Simulation Study for ARMA Processes

In this section, we conduct simulation studies to investigate the finite sample properties of the S-scale tests of ARMA models.

Corresponding to the design of the simulation studies for the two-scale tests, we conduct two types of tests. The first type tests a model for a given set of parameters, and the second type tests the model itself. For each type of tests, we study their empirical size properties and demonstrate the empirical power properties against particular alternatives motivated by practical situations.

As for the alternatives in the first type of tests, we consider data generated from the true model but with slightly different parameters. For the second type of tests, we generate data from some higher order models, which mimics the practical situation of an under-specified model due to the existence of multiscale phenomenon as observed in financial volatility of return time series.

For every data generating process, we simulate from a strong model with i.i.d. normal innovations. The HF data corresponds to $m_1 = 1$, i.e. we use all of the simulated data. The LF m_2 is chosen at various values.

6.3.1 Testing a model with a particular set of parameters

Size of the test

We generate data from an ARMA(1,1) process with parameter vector $\boldsymbol{\theta}_0 = (\phi_0, \phi_1, \theta_1, \sigma_Z^2) = (0.0082, 0.9904, -0.9505, 1)$ and test the estimating equations (6.1) with the true parameter $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ using the block EL inference. In the case of the ARMA(1,1) process, (6.1) contains a set of 12 estimating equations. We study the empirical sizes of the test with various values of sample size and levels of aggregation.

To determine the choice of the block length, we vary the block length from 1 to some values large enough. For each level of aggregation, we plot the empirical sizes against the block lengths to determine the appropriate block length to be used.

Table 6.1 shows the percentage of $W_B(\boldsymbol{\theta}_0) \geq \chi^{-1}_{(4S)}(1-\alpha)$ $(\alpha = 0.1, 0.05, 0.01)$ based on 1,000 repetitions for the ARMA(1,1) process.

Table 6.1: Empirical size of the multiscale tests when the observations are from the ARMA(1,1) model with parameter vector $(\phi_0, \phi_1, \theta_1, \sigma_Z^2)$ = (0.0082, 0.9904, -0.9505, 1).

$m_2 - m_S$	$T =$	1,500		$m_2 - m_S$	$T =$	3,000	
	10%	5%	1%		10%	5%	1%
2, 5	6.6	4.0	2.3	2, 5	8.7	5.0	1.9
2, 10	9.5	4.1	2.0	2, 10	9.1	4.8	2.0
5, 10	7.3	4.6	1.8	5, 10	9.6	5.0	1.6
2, 5, 10	8.2	5.2	1.9	2, 5, 10	9.3	5.1	1.4
$m_2 - m_S$	$T =$	6,000		$m_2 - m_S$	$T =$	12,000	
	10%	5%	1%		10%	5%	1%
2, 5	9.1	4.5	1.9	2, 5	8.5	4.4	1.2
2, 10	10.6	5.2	1.8	2, 10	10.2	5.2	1.2
5, 10	10.4	4.6	1.4	5, 10	9.6	4.6	1.0
2, 5, 10	9.2	5.2	1.3	2, 5, 10	9.1	3.5	1.3

As one can see from the empirical sizes in Table 6.1, the three-scale test has good empirical sizes.

Power of the test against small deviations in the parameters

To demonstrate the empirical power of the test, we test against a small deviation from the true parameter.

Table 6.2 shows the percentage of $W_B(\boldsymbol{\theta}_0) \geq \chi^{-1}_{(4S)}(1-\alpha)$ $(\alpha = 0.1, 0.05, 0.01)$ based on 1,000 replications. In this case the data is generated from an ARMA(1,1) process with parameter vector $(\phi_0, \phi_1, \theta_1, \sigma_Z^2)$ = (0.0082, 0.9904, -0.9505, 1) and a slightly perturbed parameter of $\phi_1 = 0.9854$ is tested with the other parameters fixed at the true value.

Comparing the results in Table 4.4 and that in Table 6.2, we observe that the tests based on three- or four-scale EEs has a higher power against the small deviation in the AR parameter than the corresponding tests based on two-scale and at the same levels of aggregation.

TABLE 6.2: Empirical power of the multiscale tests (in %) when the observations are generated from the ARMA(1,1) model with parameter vector $(\phi_0, \phi_1, \theta_1, \sigma_Z^2)$ = (0.0082, 0.9904, -0.9505, 1). Testing parameter $(0.0082, 0.9854, -0.9505, 1)$.

$m_2 - m_S$	$T =$	1,500		$m_2 - m_S$	$T =$	3,000	
	10%	5%	1%		10%	5%	1%
2, 5	34.5	23.2	11.1	2, 5	71.7	63.3	44.1
2, 10	55.8	44.8	25.0	2, 10	84.8	78.5	58.7
5, 10	59.4	49.3	27.7	5, 10	88.5	83.3	67.1
2, 5, 10	52.9	42.7	26.8	2, 5, 10	71.6	63.1	44.2
$m_2 - m_S$	$T =$	6,000		$m_2 - m_S$	$T =$	12,000	
	10%	5%	1%		10%	5%	1%
2, 5	97.6	95.5	85.3	2, 5	100	100	100
2, 10	99.6	99.4	96.1	2, 10	100	100	99.1
5, 10	99.7	99.5	97.0	5, 10	100	100	100
2, 5, 10	99.2	96.7	88.8	2, 5, 10	100	100	98.4

6.3.2 Testing a model

In this part, we conduct simulation studies to investigate the empirical size and power properties of our proposed tests for testing a model. To investigate the empirical size property, we generate the data from an ARMA(1,1) model and test the model specified through the two-scale estimating equations (6.1) optimized under the constraints of temporal aggregation relations. To investigate the empirical power property, we generate data from a particular ARMA(2,2) model.

Size of the test

Table 6.3 shows the percentage of $W_B(\tilde{\boldsymbol{\theta}}) \geq \chi^{-1}_{(4(S-1))}(1-\alpha)$ $(\alpha = 0.1, 0.05, 0.01)$ based on 1,000 repetitions for the ARMA(1,1) process with parameter vector $(\phi_0, \phi_1, \theta_1, \sigma_Z^2)$ = (0.0082, 0.9904, -0.9505, 1).

TABLE 6.3: Empirical size of the multiscale tests when the observations are from the ARMA(1,1) model with parameter vector $(\phi_0, \phi_1, \theta_1, \sigma_Z^2)$ = (0.0082, 0.9904, -0.9505, 1).

$m_2 - m_S$	$T =$	1,500		$m_2 - m_S$	$T =$	3,000	
	10%	5%	1%		10%	5%	1%
2, 5	4.3	3.2	1.3	2, 5	6.2	4.8	2.7
2, 10	5.2	4.4	1.4	2, 10	5.4	3.1	2.2
5, 10	7.4	5.4	2.1	5, 10	5.1	3.3	1.8
2, 5, 10	4.7	3.9	1.3	2, 5, 10	5.3	3.2	1.3
$m_2 - m_S$	$T =$	6,000		$m_2 - m_S$	$T =$	12,000	
	10%	5%	1%		10%	5%	1%
2, 5	4.1	2.3	0.8	2, 5	3.5	1.5	0.2
2, 10	5.1	3.3	1.0	2, 10	6.2	3.4	1.1
5, 10	6.1	3.5	0.9	5, 10	8.9	4.6	2.1
2, 5, 10	5.2	3.6	0.9	2, 5, 10	7.6	4.3	0.7

As one can see from the empirical sizes in Table 6.3, the multiscale test generally has conservative empirical sizes in this case.

Power of the test against multiscale-type higher order data generating mechanism

To demonstrate the empirical power of the test, we test against a ARMA(2,2) process with a choice of parameter values motivated by the ARMA representation of two-component GARCH models, which has a multiscale volatility interpretation.

Table 6.4 shows the percentage of $W_B(\tilde{\boldsymbol{\theta}}) \geq \chi^{-1}_{(4)}(1-\alpha)$ ($\alpha = 0.1, 0.05, 0.01$) based on 1,000 replications. In this case data is generated from an ARMA(2,2) process with parameter vector $(\phi_0, \phi_1, \phi_2, \theta_1, \theta_2)$ = (0.00008, 1.9279, - 0.9280, -1.8644, 0.8652).

Comparing the results in Table 6.4 and that in Table 4.6, we observe that the tests based on three- or four-scale EEs has a higher power against the small deviation in the AR parameter than the corresponding tests based on two-scale and at the same levels of aggregation.

TABLE 6.4: Empirical power the multiscale tests (in %) when the observations are generated from the ARMA(2,2) model with parameter vector $(\phi_0, \phi_1, \phi_2, \theta_1, \theta_2)$ = (0.00008, 1.9279, - 0.9280, -1.8644, 0.8652). Testing the ARMA(1,1) model.

$m_2 - m_S$	$T =$	1,500		$m_2 - m_S$	$T =$	3,000	
	10%	5%	1%		10%	5%	1%
2, 5	35.5	32.8	30.8	2, 5	54.5	51.2	47.7
2, 10	31.2	26.3	14.9	2, 10	49.0	38.3	24.5
5, 10	48.6	42.5	38.1	5, 10	67.4	53.0	31.1
2, 5, 10	59.1	56.8	53.0	2, 5, 10	70.6	64.9	53.5
$m_2 - m_S$	$T =$	6,000		$m_2 - m_S$	$T =$	12,000	
	10%	5%	1%		10%	5%	1%
2, 5	75.6	68.5	60.3	2, 5	84.1	76.0	60.4
2, 10	75.2	65.1	46.9	2, 10	97.8	94.7	87.6
5, 10	89.0	82.8	64.1	5, 10	100	98.5	91.8
2, 5, 10	83.9	73.5	60.2	2, 5, 10	98.1	95.2	88.3

6.4 Simulation Study for GARCH Process

In this section, we conduct simulation studies to investigate the finite sample properties of the S-scale tests of GARCH models.

Corresponding to the design of the simulation studies for the two-scale tests, we conduct two types of tests. The first type tests a model for a given set of parameters, and the second type tests the model itself. For each type of tests, we study their empirical size properties and demonstrate the empirical power properties against particular alternatives motivated by practical situations.

As for the alternatives in the first type of tests, we consider data generated from the true model but with slightly different parameters. For the second type of tests, we generate data from some higher order models, which mimics the practical situation of an under-specified model due to the existence of multiscale phenomenon as observed in financial volatility time series.

For every data generating process, we simulate from a strong model with i.i.d. normal innovations. The HF data corresponds to $m_1 = 1$, i.e. we use all of the simulated data. The LF m_2 is chosen at various values.

6.4.1 Testing a model with a particular set of parameters

Size of test

Table 6.5 shows the percentage of $W_B(\boldsymbol{\theta}_0) \geq \chi^{-1}_{(4S)}(1-\alpha)$ $(\alpha = 0.1, 0.05, 0.01)$ based on 1,000 repetitions for the GARCH(1,1) process with parameter vector $(\alpha_0, \alpha_1, \beta_1) =$ (0.0082, 0.9904, 0.9505).

As one can see from the empirical sizes in Table 6.5, the test has empirical sizes in broad agreement with the nominal sizes as sample size increases.

Power of the test against small deviations in the parameters

To demonstrate the empirical power of the test, we test against a small deviation from the true parameter.

Table 6.6 shows the percentage of $W_B(\boldsymbol{\theta}_0) \geq \chi^{-1}_{(4S)}(1-\alpha)$ $(\alpha = 0.1, 0.05, 0.01)$ based on 1,000 replications. In this case the data is generated from a GARCH(1,1)

TABLE 6.5: Empirical size of the multiscale tests when the observations are generated from the GARCH(1,1) model with parameter vector $(\alpha_0, \alpha_1, \beta_1) =$ (0.0082, 0.9904, 0.9505).

$m_2 - m_S$	$T =$	1,500		$m_2 - m_S$	$T =$	3,000	
	10%	5%	1%		10%	5%	1%
2, 5	8.3	4.8	2.0	2, 5	6.7	4.6	1.8
2, 10	7.4	5.4	2.4	2, 10	8.6	4.8	2.0
5, 10	7.8	4.8	2.4	5, 10	11.2	7.5	1.6
2, 5, 10	8.1	5.5	2.3	2, 5, 10	10.3	6.8	1.7
$m_2 - m_S$	$T =$	6,000		$m_2 - m_S$	$T =$	12,000	
	10%	5%	1%		10%	5%	1%
2, 5	7.8	4.5	1.4	2, 5	8.2	4.4	1.2
2, 10	8.6	5.2	1.8	2, 10	9.3	5.2	1.2
5, 10	9.8	5.1	1.9	5, 10	10.8	4.6	1.0
2, 5, 10	8.2	5.2	1.6	2, 5, 10	9.1	4.8	1.3

process with parameter vector $(\alpha_0, \alpha_1, \beta_1) = (0.0082, 0.9904, 0.9505)$ and a slightly perturbed parameter value of $\alpha_1 = 0.0349$ is tested with the other parameters fixed at the true value.

Comparing the results in Table 6.6 and that in Table 5.2, we observe that the tests based on three- or four-scale EEs has a higher power against the small deviation in the parameter than the corresponding tests based on two-scale and at the same level of aggregation.

TABLE 6.6: Empirical power of the multiscale tests (in %) when the observations are generated from the GARCH(1,1) model with parameter vector $(\alpha_0, \alpha_1, \beta_1) = (0.0082, 0.9904, 0.9505)$.
Testing parameter $(0.0082, 0.0349, 0.9505)$.

$m_2 - m_S$	$T =$	1,500		$m_2 - m_S$	$T =$	3,000	
	10%	5%	1%		10%	5%	1%
2, 5	35.2	25.3	10.4	2, 5	80.5	70.6	47.5
2, 10	36.1	29.1	13.6	2, 10	89.6	84.5	70.8
5, 10	42.6	32.6	17.3	5, 10	93.2	89.1	79.2
2, 5, 10	38.0	29.4	15.6	2, 5, 10	87.2	83.0	69.4
$m_2 - m_S$	$T =$	6,000		$m_2 - m_S$	$T =$	12,000	
	10%	5%	1%		10%	5%	1%
2, 5	99.3	98.8	96.5	2, 5	100	100	100
2, 10	99.8	99.6	98.7	2, 10	100	100	100
5, 10	99.6	99.5	99.6	5, 10	100	100	100
2, 5, 10	99.4	99.0	99.5	2, 5, 10	100	100	100

6.4.2 Testing a model

In this part, we conduct simulation studies to investigate the empirical size and power of our proposed tests when testing a model. To investigate the empirical size property, we generate the data from a GARCH(1,1) model and test the model specified through the two-scale estimating equations (4.7) optimized under the constraints of temporal aggregation relations. To investigate the empirical power property, we generate data from a particular GARCH(2,2) model.

Size of test

Table 6.7 shows the percentage of $W_B(\tilde{\boldsymbol{\theta}}) \geq \chi^{-1}_{(4(S-1))}(1-\alpha)$ $(\alpha = 0.1, 0.05, 0.01)$ based on 1,000 repetitions for the GARCH(1,1) process with parameter vector $(\alpha_0, \alpha_1, \beta_1)$ = (0.0082, 0.9904, 0.9505).

TABLE 6.7: Empirical size of the multiscale tests when the observations are generated from the GARCH(1,1) model with parameter vector $(\alpha_0, \alpha_1, \beta_1)$ = (0.0082, 0.9904, 0.9505).

$m_2 - m_S$	$T =$	1,500		$m_2 - m_S$	$T =$	3,000	
	10%	5%	1%		10%	5%	1%
5, 30	6.4	3.8	2.0	5, 30	6.0	4.1	2.0
10, 30	7.0	5.0	2.0	10, 30	4.2	3.8	0.9
5, 10, 60	21.0	19.0	15.8	5, 10, 60	6.2	4.3	1.2
$m_2 - m_S$	$T =$	6,000		$m_2 - m_S$	$T =$	12,000	
	10%	5%	1%		10%	5%	1%
5, 30	6.8	3.9	0.8	5, 30	7.2	4.3	0.8
10, 30	6.9	4.1	0.7	10, 30	7.9	4.5	0.8
5, 10, 60	7.2	4.2	0.8	5, 10, 60	8.2	4.8	1.1

As one can see from the empirical sizes in Table 6.7, the multiscale test generally has a smaller number of rejections of the null hypothesis than the nominal values except in the case of $m_2 - m_S = 5, 10, 60$ and $T = 1,500$. The reason of the large size in this case is likely due to the high level of aggregation involved, which is 60, together with the relatively small sample size. In this case, violation of EEs associated with scale 60 could occur more often then the nominal size of the test would indicate.

Power of the test against multiscale-type higher order data generating mechanism

To demonstrate the empirical power of the test, we generate data from a particular GARCH(2,2) model corresponding to a two-component GARCH model estimated with real data, which has a multiscale volatility interpretation.

Table 6.8 shows the percentage of $W_B(\tilde{\boldsymbol{\theta}}) \geq \chi^{-1}_{(4(S-1))}(1-\alpha)$ $(\alpha = 0.1, 0.05, 0.01)$ based on 1,000 replications. In this case data is generated from GARCH(2,2) process with parameters vector $(\alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2)$ = (7e-5, 0.0635, -0.0628, 1.8644, -0.8652) and a GARCH(1,1) model is tested.

Table 6.8: Empirical power of the multiscale tests (in %) when the observations are generated from the GARCH(2,2) model with parameters vector $(\alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2)$ = (7e-5, 0.0635, -0.0628, 1.8644, -0.8652). Testing the GARCH(1,1) model.

$m_2 - m_S$	$T =$	1,500		$m_2 - m_S$	$T =$	3,000	
	10%	5%	1%		10%	5%	1%
5, 30	18.7	16.6	10.2	5, 30	25.4	21.2	16.4
10, 30	26.2	23.1	19.8	10, 30	33.4	29.8	25.6
5, 10, 60	100	100	100	5, 10, 60	100	100	100
$m_2 - m_S$	$T =$	6,000		$m_2 - m_S$	$T =$	12,000	
	10%	5%	1%		10%	5%	1%
5, 30	31.2	23.4	12.0	5, 30	61.4	53.6	36.6
10, 30	29.8	22.8	12.0	10, 30	57.4	48.4	32.8
5, 10, 60	100	100	100	5, 10, 60	100	100	100

Comparing the results in Table 6.8 and that in Table 5.4, we observe that the tests based on three- or four-scale EEs has a higher power against the small deviation in the parameter than the corresponding tests based on two-scale and at the same level of aggregation.

6.5 Empirical Study

In this section, we apply the multiscale test to the Dow Jones Industrial Average (DJIA) index return sample used in Engle and Patton (2001)[22], which forms a comparison with the empirical studies in Section 5.4.

6.5.1 Multiscale testing of QMLE estimates

Table 6.9 summarizes the test results of multiscale tests for the DJIA sample data. First, we test the GARCH(1,1) model with a three-scale test based on daily, 2-day, and 3-day returns to seek confirmation of the findings of the two-scale tests that the GARCH(1,1) model as estimated in Engle and Patton (2001)[22] is inconsistent with the sample data at 2-day and 3-day scales. The test result shows that the GARCH(1,1) model is rejected with high level of confidence, which seems to confirm the results of the related two-scale tests. However, as we are going to elaborate in the following subsection, a closer examination of the matrix condition number of the variance-covariance matrix of the estimation equations involved in this three-scale test raises alarm about potential problems of numerical stability in computing the value of the test statistics.

Next, the estimated GARCH(1,1) model is inconsistent with the sample data at scales of 60-day (or higher, as we have tested but not reported here for the sake of space).

Finally, for the intermediates scales, from 5-day to 30-day, the estimated GARCH(1,1) is not rejected by the multiscale tests of consistency.

TABLE 6.9: Empirical likelihood testing results of GARCH(1,1) model with parameter vector $(\alpha_0, \alpha_1, \beta_1) = (0.0082,\ 0.0399,\ 0.9505)$ and $\hat{\sigma}_\nu^2 = 4.34$ on the DJIA return sample as in Engle and Patton (2001)[22].

Three-Scale Test			
	10%	5%	1%
Test statistics $\chi^2(12)$			
Critical values	18.55	21.03	26.22
Test statistics value (average):			
$m_2 = 2, m_3 = 3$	53.30		
Test Conclusion	R	R	R
$m_2 = 2, m_3 = 5$	14.36		
Test Conclusion	A	A	A
$m_2 = 2, m_3 = 10$	9.61		
Test Conclusion	A	A	A
$m_2 = 3, m_3 = 5$	20.80		
Test Conclusion	R	A	A
$m_2 = 3, m_3 = 10$	8.740		
Test Conclusion	A	A	A
$m_2 = 5, m_3 = 10$	10.58		
Test Conclusion	A	A	A
$m_2 = 5, m_3 = 20$	9.50		
Test Conclusion	A	A	A
$m_2 = 5, m_3 = 30$	9.98		
Test Conclusion	A	A	A
$m_2 = 5, m_3 = 60$	32.73		
Test Conclusion	R	R	R
$m_2 = 10, m_3 = 20$	10.03		
Test Conclusion	A	A	A
$m_2 = 10, m_3 = 60$	43.46		
Test Conclusion	R	R	R
$m_2 = 20, m_3 = 60$	49.23		
Test Conclusion	R	R	R
$m_2 = 30, m_3 = 60$	61.44		
Test Conclusion	R	R	R
Four-Scale Test			
	10%	5%	1%
Test statistics $\chi^2(16)$			
Critical values	23.54	26.30	32.00
Test statistics value (average):			
$m_2 = 5, m_3 = 10, m_4 = 20$	12.25		
Test Conclusion	A	A	A
$m_2 = 5, m_3 = 10, m_4 = 60$	52.76		
Test Conclusion	R	R	R
$m_2 = 5, m_3 = 10, m_4 = 120$	27.60		
Test Conclusion	R	R	A
$m_2 = 10, m_3 = 20, m_4 = 30$	12.32		
Test Conclusion	A	A	A
$m_2 = 10, m_3 = 20, m_4 = 60$	54.11		
Test Conclusion	R	R	R

A: accept; R: reject.

(Test statistic values are averaged over a range of choices of block length.)

6.5.2 Multiscale testing of the GARCH(1,1) model

In this section, we have conducted multiscale tests of the GARCH(1,1) model using the DJIA sample data and form comparison with the results of the two-scale tests as presented in Section 5.4.

Table 6.10 shows the results of multiscale testing of the GARCH(1,1) model. We observe that the conclusions of multiscale tests of the GARCH(1,1) model for the DJIA return sample generally follows similar patterns of the two-scale tests of the GARCH(1,1) model as well as the multiscale tests with a particular set of QMLE. Some notable differences are summarized below.

Firstly, the test of the GARCH(1,1) model itself shows that it is not rejected by the three-scale test based on daily, 2-day, and 3-day data whereas the two-scale test based on daily and 3-day data rejected the GARCH(1,1) model, which seems to be counterintuitive. (See Table 5.6.) A calculation of the matrix condition numbers (under the 2-norm) of the variance-covariance matrix associated with the three-scale tests based on daily, 2-day, and 3-day, in both of the cases of testing the model with QMLE and the model itself, yields, respectively, the values of 845 and 680, which indicate that the inversion of the these matrices tend to be numerically unstable. In contrast, the values of the condition numbers of the corresponding matrices associated with the tests based on more separated scales are usually less than 50 and generally less than 100 and the numerical inversion of those matrices are much more stable. Since the inverse of the variance-covariance matrix associated with the estimating equations used to construct the test plays a key role in the computation of the test statistics, we suspect the result of the three-scale test based on daily, 2-day and 3-day data to be numerically unreliable. We suggest that the users of the multiscale test shall not use scales too close to each other.

Secondly, comparing the test results corresponding to the set of scales $m_2 = 10$, $m_3 = 20$, and $m_4 = 60$, we observe that certain sets of parameter values could lead to the model not being rejected at 60-day scale at confidence levels of 5% and 1%. However, the GARCH(1,1) model is clearly rejected at scale 60.

TABLE 6.10: Log profile empirical likelihood statistics values for testing the GARCH(1,1) model using sample from HF and various LFs on the DJIA return sample as in Engle and Patton (2001)[22].

Three-Scale Test			
	10%	5%	1%
Test statistics $\chi^2(8)$			
Critical values	13.36	15.51	20.09
Test statistics value (average):			
$m_2 = 2, m_3 = 3$	7.24		
Test Conclusion	A	A	A
$m_2 = 2, m_3 = 5$	4.22		
Test Conclusion	A	A	A
$m_2 = 2, m_3 = 10$	4.55		
Test Conclusion	A	A	A
$m_2 = 3, m_3 = 5$	3.70		
Test Conclusion	A	A	A
$m_2 = 3, m_3 = 10$	4.32		
Test Conclusion	A	A	A
$m_2 = 5, m_3 = 10$	4.89		
Test Conclusion	A	A	A
$m_2 = 5, m_3 = 20$	6.18		
Test Conclusion	A	A	A
$m_2 = 5, m_3 = 30$	5.78		
Test Conclusion	A	A	A
$m_2 = 5, m_3 = 60$	15.54		
Test Conclusion	R	R	A
$m_2 = 10, m_3 = 20$	6.72		
Test Conclusion	A	A	A
$m_2 = 10, m_3 = 60$	18.85		
Test Conclusion	R	R	A
$m_2 = 20, m_3 = 60$	17.06		
Test Conclusion	R	R	A
$m_2 = 30, m_3 = 60$	17.95		
Test Conclusion	R	R	A
Four-Scale Test			
	10%	5%	1%
Test statistics $\chi^2(12)$			
Critical values	18.55	21.03	26.22
Test statistics value (average):			
$m_2 = 5, m_3 = 10, m_4 = 20$	8.90		
Test Conclusion	A	A	A
$m_2 = 5, m_3 = 10, m_4 = 60$	25.84		
Test Conclusion	R	R	R
$m_2 = 10, m_3 = 20, m_4 = 30$	7.59		
Test Conclusion	A	A	A
$m_2 = 10, m_3 = 20, m_4 = 60$	20.98		
Test Conclusion	R	A	A
$m_2 = 10, m_3 = 60, m_4 = 120$	61.26		
Test Conclusion	R	R	R

A: accept; R: reject.

(Test statistic values are averaged over a range of choices of block length.)

6.6 Applications of the Multiscale Testing Framework to Estimation

In this section, we apply the empirical likelihood based approach to the multiscale testing framework to estimate the model parameters.

As we have seen in the multiscale tests of a model, our proposed testing framework results in a set of optimized parameters corresponding to the maximal value of the corresponding empirical likelihood function. This set of parameters is called the *maximum empirical likelihood estimate (MELE)*. See, for example, Qin and Lawless (1994)[65] and Owen (2001)[62]. In particular, Qin and Lawless (1994)[65], Corollary 2, showed that the MELE based on a given set of r estimating equations (4.2) for estimating the s-dimensional parameter $\boldsymbol{\theta}$ is fully efficient in the sense that it has the same asymptotic variance as the optimal estimator obtained from the class of s estimating equations that are linear combinations of the r estimating equations (4.2).

In addition to the general properties of the MELE known in the literature, our simulation study suggests that there are two additional advantages of the MELE resulting from our proposed two-scale inference framework. The first one is that the HF model parameters can be estimated by using only the LF sample. This can be useful when one only has access to the LF sample but wants to make predictions at the HF. The second one is that the MELE obtained by using the multi-frequency sample can reduce the bias in estimating the parameters, especially when the the estimation errors are examined at LF scales which is key to the multiple step forecasts. We illustrate these two benefits with simulation studies.

Specifically, we generate the data from a strong GARCH(1,1) model with a parameter vector $(\alpha_0, \alpha_1, \beta_1) = (0.0082, 0.9904, 0.9505)$ and standard normal innovations. We consider the situation where one only has access to samples already subject to some known level of aggregation and we focus on the estimation of the persistence parameter $\alpha_1 + \beta_1$. We conduct multiscale maximum empirical likelihood estimation by using the data sampled at the highest *available* frequency (i.e. m_1) combined with samples at some lower frequencies. For comparison, we conduct single scale MELE and quasi-maximum likelihood estimation based on normal innovations using highest available frequency sample.

In the maximum empirical likelihood estimation, we parameterize the model at the data generating scale, which is assumed to be known. As in the multiscale testing framework, we derive estimating equations at the scales where the corresponding samples are used for multiscale estimation. As we assume that one knows the data generating process but only has access to aggregated sample, we parameterize the estimating equations at all scales using the parameters in the data generating process. For example, when conducting multiscale estimation using samples at the scales $m_1 = 5$ and $m_2 = 10$, the corresponding estimating equations based on the ARMA(1,1) representation of the GARCH(1,1) at the two scales have their AR(1) parameters parameterized as $(\alpha_1 + \beta_1)^5$ and $(\alpha_1 + \beta_1)^{10}$, respectively. The other model parameters are similarly specified. This is different from the multiscale testing situation where we parameterize the LF parameters as independent parameters. In the current situation, we have an over-identified system of estimation equations.

Table. 6.11 below summarizes the estimation results based on various data aggregation levels (i.e. various levels of aggregation of the highest frequency available sample) and an generating data of length 600 times of the level of aggregation of the highest frequency available sample. For example, when the highest frequency available sample is assumed to be aggregated at the level $m_1 = 5$, we generate data of length $3,000$. And when the highest frequency available sample is assumed to be aggregated at the level $m_1 = 10$, we simulated data of length $6,000$. The summary statistics for the estimates are based on 1,000 replications. The QMLE estimates are obtained by using the built-in function in **Matlab 2014** for estimating GARCH models specified with a normal innovation distribution.

At any level of aggregation assumed for the highest frequency available sample, the parameters of the data generating model are naturally obtained through the MELE due to the parameterization we use. For the QMLE, the directly observed estimates correspond to the model at the aggregation level m_1. As we can see from the temporal aggregation results in Chapter 2, not every parameter in a HF GARCH(1,1) model can be uniquely recovered from the aggregated LF GARCH(1,1) model. This illustrates the first advantage of using the multiscale MELE estimation since the model parameters at the data generating level can be naturally estimated, which may not be possible if we had used QMLE.

The persistence parameter $(\alpha_1 + \beta_1)$ in a GARCH(1,1) model can be converted from a LF to a HF through a simple power relation. Using this relation, we convert the estimated persistence parameters among various scales and compare the corresponding percentage errors with respect to their true values. As can be observed from Table. 6.11, the medians of the point estimates based on the 1,000 replications obtained through the multiscale MELEs generally have smaller biases than the corresponding medians of the single scale MELE and the means and medians obtained through QMLE. The advantages of multiscale MELE are especially large at larger scales. Admittedly, the MELEs generally have larger standard deviations than the QMLEs, and this could result in quite large deviations of the means of the MELE point estimates based on the 1,000 replications from the corresponding true value. In comparison, the median of the MELE point estimates are less affected by the standard deviations and this is why we choose to use the median instead of the mean for the MELEs. However, we would like to emphasize the advantage brought by using multiple frequency samples in reducing the biases in estimating the persistence parameter evaluated at a range of scales, which will be beneficial to the forecasting using the GARCH(1,1) model at the corresponding horizons.

Table 6.11: Comparison of the estimation of the persistence parameter $\alpha_1+\beta_1$ in a GARCH(1,1) model.

Scale	1	5	10	20	60	120
True value of $\alpha_1+\beta_1$	0.9904	0.9529	0.9080	0.8245	0.5606	0.3145
MELE (median)						
$m_1 = 5$	0.9871	0.9371	0.8782	0.7713	0.4588	0.2105
(percentage error)	(0.3%)	(1.7%)	(3.3%)	(6.5%)	(18.1%)	(33.0%)
$m_1 = 5, m_2 = 10$	0.9885	0.9438	0.8908	0.7935	0.4996	0.2496
(percentage error)	(0.2%)	(1.0%)	(1.9%)	(3.8%)	(10.9%)	(20.6%)
$m_1 = 5, m_2 = 20$	0.9892	0.9472	0.8971	0.8048	0.5213	0.2717
(percentage error)	(0.1%)	(0.6%)	(1.2%)	(2.4%)	(7.0%)	(13.5%)
$m_1 = 5, m_2 = 30$	0.9908	0.9548	0.9117	0.8312	0.5743	0.3299
(percentage error)	(0.0%)	(0.2%)	(0.4%)	(0.8%)	(2.5%)	(5.0%)
$m_1 = 5, m_2 = 60$	0.9884	0.9433	0.8899	0.7919	0.4966	0.2466
(percentage error)	(0.2%)	(1.0%)	(2.0%)	(4.0%)	(11.4%)	(21.5%)
QMLE estimates						
QMLE (mean)*	0.9920	0.9606	0.9228	0.8516	0.6176	0.3814
(percentage error)	(0.2%)	(0.8%)	(1.6%)	(3.3%)	(10.2%)	(21.4%)
QMLE (median)	0.9935	0.9679	0.9369	0.8777	0.6762	0.4572
(percentage error)	(0.3%)	(1.6%)	(3.2%)	(6.4%)	(20.6%)	(45.5%)
MELE (median)						
$m_1 = 10$	0.9886	0.9443	0.8917	0.7951	0.5026	0.2526
(percentage error)	(0.2%)	(0.9%)	(1.8%)	(3.6%)	(10.3%)	(19.6%)
$m_1 = 10, m_2 = 20$	0.9902	0.9520	0.9062	0.8212	0.5538	0.3067
(percentage error)	(0.0%)	(0.1%)	(0.2%)	(0.4%)	(1.2%)	(2.4%)
$m_1 = 10, m_2 = 30$	0.9908	0.9548	0.9117	0.8312	0.5743	0.3299
(percentage error)	(0.0%)	(0.2%)	(0.4%)	(0.8%)	(2.5%)	(5.0%)
$m_1 = 10, m_2 = 60$	0.9903	0.9524	0.9071	0.8229	0.5572	0.3105
(percentage error)	(0.0%)	(0.1%)	(0.1%)	(0.2%)	(0.6%)	(1.2%)
QMLE estimates						
QMLE (mean)	0.9934	0.9674	0.9359	0.8760	0.6721	0.4518
(percentage error)	(0.3%)	(1.5%)	(3.1%)	(6.2%)	(19.9%)	(43.8%)
QMLE (median)	0.9939	0.9699	0.9406	0.8848	0.6927	0.4799
(percentage error)	(0.4%)	(1.8%)	(3.6%)	(7.3%)	(23.6%)	(52.7%)
MELE (median)						
$m_1 = 20$	0.9894	0.9481	0.8989	0.8080	0.5276	0.2784
(percentage error)	(0.1%)	(0.5%)	(1.0%)	(2.0%)	(5.9%)	(11.4%)
$m_1 = 20, m_2 = 30$	0.9909	0.9553	0.9126	0.8329	0.5778	0.3339
(percentage error)	(0.1%)	(0.3%)	(0.5%)	(1.0%)	(3.1%)	(6.2%)
$m_1 = 20, m_2 = 60$	0.9901	0.9515	0.9053	0.8196	0.5505	0.3030
(percentage error)	(0.0%)	(0.2%)	(0.3%)	(0.6%)	(1.8%)	(3.6%)
QMLE estimates						
QMLE (mean)	0.9940	0.9704	0.9416	0.8866	0.6969	0.4857
(percentage error)	(0.4%)	(1.8%)	(3.7%)	(7.5%)	(24.3%)	(54.6%)
QMLE (median)	0.9942	0.9713	0.9435	0.8902	0.7054	0.4976
(percentage error)	(0.4%)	(1.9%)	(3.9%)	(8.0%)	(25.8%)	(58.3%)

* The QMLEs corresponding to the scales smaller than m_1 are obtained by taking the $(1/m_1)^{th}$ power of the QMLEs estimated with the aggregated data.

6.7 Section Conclusion

In this section, we have presented the multiscale tests of consistency for the ARMA and GARCH processes. The theoretical development of the multiscale tests extends straightforwardly the corresponding theory for the two-scale tests. However, cautions need to be exercised in choosing the scales to be used in the multiscale tests in order to avoid a potentially numerical instability in computing the value of the test statistics resulting from inverting matrices with large condition numbers

Overall, simulation studies show that the multiscale tests are able to significantly improve powers over the corresponding two-scale tests.

Empirical study on a sample of Dow Jones Industrial Average (DJIA) index return sample corroborate results from the two-scale tests that although a GARCH(1,1) model is accepted at the high frequency data, it may not be adequate for data at lower frequencies.

Chapter 7

Conclusions and Future Research Questions

In this chapter, we conclude the works completed in this thesis and outline several future research questions. Each section corresponds to one future research question.

7.1 Conclusion of the Works Completed in This Thesis

In this thesis, we have proposed a novel statistical inference method for testing whether an ARMA or GARCH type model is consistent with data at multiple time scales. Our proposed method uses functional relations derived from temporal aggregation of time series models and is based on the framework of empirical likelihood which are implemented through a set of estimating equations. Simulation studies demonstrated that our proposed tests have good empirical size property and high power against some particular alternatives which are motivated by empirical studies on financial asset returns data.

A particular focus of our study is on the modeling and testing of the GARCH models for financial asset return volatility. Some practical issues related to modeling multiscale type volatility dynamics with the GARCH models are discussed. In particular, we have highlighted the usage of component GARCH models as alternative parameterizations for the GARCH models with general parameter constraints in order to better capture multiscale dynamics in financial asset return volatility.

7.2 Some Future Research Questions

7.2.1 Two open problems with block empirical likelihood inference related to this thesis

Since block empirical likelihood inference plays a fundamental role in the multiscale inference procedure developed in this thesis, it is worth pointing out two important open problems with the BEL approach.

The first problem concerns how to deal with dependency in the data in the EL method. As far as blocking techniques are concerned, it may be useful to develop methods which automatically determine various block sizes. Another direction for exploration, as already mentioned in Chapter 4, it to consider alternative techniques to the blocking methods. For example, kernel smoothing and expansive block empirical likelihood are among the alternative class of methods to the blocking technique. Relevant reference in these directions include Smith(2011)[68] and Nordman, et.al. (2013)[58].

The second important problem concerning the empirical likelihood approach is the so-called "convex hull" problem (CHP). The CHP is a common practical problem that, with realized data, the constraint set in (A.9) and

$$\mathcal{R}(\boldsymbol{\theta}) = \underset{w_i}{sup}\left\{\prod_{i=1}^{n} nw_i | w_i > 0, \sum_{i=1}^{n} w_i = 1, \sum_{i=1}^{n} w_i g_i = 0\right\},$$

may be empty at possibly many points of $\boldsymbol{\theta}$), even all the points in $\boldsymbol{\Theta}$. The conventional approach in the literature, as mentioned in Chen, et.al. (2008)[8], is to set the value of the profile log empirical likelihood function to be infinity, which is equivalent to saying that the convex hull of the n real vectors $g_1, \cdots, g_n$ does not contain the origin. As a remedy to the CHP, the so called "adjusted empirical likelihood" (AEL) approach had been developed by Chen, et.al. (2008)[8] and Liu and Chen (2010)[45] in the context of i.i.d. data. However, whether we can extend the AEL to the case of weakly dependent data remains an open research question.

7.2.2 Aggregation and multiscale inference of multivariate GARCH models

It is both natural and important to consider temporal aggregation and scale-consistency of multivariate GARCH models. For example, practical problems like portfolio selection requires modeling the joint dynamics of returns on several assets. Hafner(2008)[31] studied the temporal aggregation of multivariate GARCH(1,1) processes and showed that the class of multivariate weak GARCH(1,1) processes in the general vector specification is closed under temporal aggregation. Relation between coefficients in the high frequency process and aggregated low frequency process, similar to the Drost-Nijman formula as given in our Section 2.3.3, was established. The techniques used to derive temporal aggregation relation draws heavily on the techniques for studying temporal aggregation of vector ARMA(VARMA) processes. Therefore, temporal aggregation relation for multivariate GARCH(2,2) and thus multivariate component GARCH models can be derived, at least in principle.

In an empirical study, Hafner and Rombouts(2007)[32] estimated a bivariate GARCH(1,1) model with daily DJIA index return and NASDAQ index return. They found that the estimates using daily data are inconsistent with estimates obtained from using weekly and biweekly data. We translate their parameter estimates into volatility and co-volatility half-life and it essentially points to the same scale-inconsistency problem as reported in Engle and Patton(2001)[22]. Therefore, it is empirically interesting to investigate whether some bivariate version of the two component GARCH model could solve the scale inconsistency problem of the bivariate GARCH(1,1) model for modeling DJIA/NASDAQ variance and covariance.

7.2.3 Aggregation and multiscale inference of variations of the standard GARCH models and applications [1]

In this thesis, we only considered standard GARCH models of Engle (1982)[20] and Bollerslev (1986)[5] with normal innovations. There have been many variations of the standard GARCH model proposed in the literature over the years. Bollerslev

[1]We thank the external examiner, Professor Lars Stentoft, for emphasizing this research direction.

(2008)[6] provides a summary of them. In particular, asymmetric GARCH models and GARCH models with heavy-tailed innovations are important extensions of the standard GARCH models, especially for financial applications. While a complete understanding of the probabilistic properties, such as stationarity and mixing conditions, of all those variations of GARCH models may be difficult to guarantee, we plan to conduct more extensive studies to examine the properties of our proposed inference framework using GARCH models with heavy-tailed and skewed innovation distributions.

In particular, it would be interesting to study applications of GARCH models in financial risk management, such as Value-at-Risk (VaR) estimation. See, for example, McNeil and Frey (2000)[52] and McNeil, Frey, and Embrechts (2005)[53]. A particular problem involving multi-scale characterization of financial asset return process is the problem of "scaling of VaR". This problem arises as, in some situations, one may want to scale a the model estimated with daily return data, to several-day scales in order to obtain estimates of VaR measure over various horizons. Kaufmann (2004)[39] investigated scaling rules based on GARCH(1,1) model. It is of interest to investigate the scaling rules under higher order GARCH models following the line of Kaufmann (2004)[39].

7.2.4 Scale-consistency of continuous-time Processes

The problem of scale-consistency also naturally concerns continuous-time diffusion models which are widely used in the pricing of financial derivatives. Drost and Werker (1996)[17] and Meddahi and Renault (2004)[54] showed that the the weak ARMA structure is also the underlying dependency structure of some commonly used continuous-time volatility models. For example, the famous Heston model of Heston (1993)[35] has an underlying weak AR(1) structure. When estimating the continuous-time models, the choice of sample frequency may be arbitrary. And the estimated models are often used on time scales different from the scale where the sample data are from. Therefore, testing the scale-consistency of continuous-time models is important for practical purposes just as it is for the ARMA and GARCH models.

Drost and Werker (1996)[17] considered some continuous-time models which has a weak GARCH structure at all discrete time scale and derived the functional relation between the parameters of the continuous-time models and that of the weak

GARCH representation. Meddahi and Renault (2004)[54] proposed the square-root stochastic autoregressive volatility (SR-SARV) representation which nests many commonly used continuous-time volatility models. Using the SR-SARV representation, one can derive the weak ARMA representation of the continuous-time models and the corresponding functional relations between the model parameters at different time scales. Chen, et.al. (2008)[9] considered the problem of model specification test for continuous-time diffusion models using an empirical likelihood approach. Therefore, it is an interesting and feasible direction to explore the testing of scale-consistency of continuous-time processes commonly studied in Finance.

Another related stream of research focuses on studying the continuous-time limits of discrete-time GARCH processes and the related problems of statistical inference and financial applications. Nelson(1990)[56] lists a set of conditions to guarantee weak convergence of a discrete time Markov chain, defined by a system of stochastic difference equations, towards a diffusion. This approach requires convergence, as the interval between observations shrinks to zero, of a number of conditional moments to well defined limits at an appropriate rate. In the context of GARCH-type models, Nelson (1990)[56] shows convergence results for a series of GARCH specifications. This approach is later exploited by Duan(1997)[18] to derive a diffusion limit of an Augmented GARCH model and by Alexander and Lazar(2005)[1] to derive a diffusion limit of a weak to derive a diffusion limit of a weak GARCH process. When a discrete time model is cast as a diffusion approximation, inference on the parameters of a diffusion model can be conducted through parameter estimates of a discrete time GARCH-type model. This suggests that, instead of direct estimation of the diffusion parameters, we can infer the diffusion parameters by means of a tractable likelihood function of an approximating discrete time multivariate GARCH process. This approach is known as quasi-approximated maximum likelihood (QAML), and has been used in studies such as Stentoft(2011)[69]. However, it is potentially difficult to show consistency of the QAML estimator even if the discrete time approximation is closed under temporal aggregation, as pointed out by Drost and Werker(1996)[17]. For the univariate GARCH model, Wang(2002)[72] proves that the statistical experiments resulting from the estimation of the diffusion model and its approximating discrete time model are not equivalent. This suggests that the QAML estimator are unlikely to be consistent in both the univariate and multivariate GARCH setting.

References

[1] Alexander, C. and E. Lazar (2005, November). On The Continuous Limit of GARCH . *ICMA Centre Dicussion Papers in Finance 2005-13*.

[2] Augustyniak, M., L. Bauwens, and A. Dufays (2018, June). A New Approach to Volatility Modeling: The Factorial Hidden Markov Volatility Model. *Journal of Business and Economic Statistics*.

[3] Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996, September). Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics 74*(1), 3–30.

[4] Barndorff-Nielsen, O. E. and N. Shephard (2001). On The Correlation Structure for the Generalized Autoregressive Conditional Heteroskedastic Process. *Journal of the Royal Statistical Society: Series B 63*(2), 167–241.

[5] Bollerslev, T. (1986, April). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics 50*(3), 307–327.

[6] Bollerslev, T. (2008). Glossary to ARCH (GARCH). *CREATES Research Paper 2008-49*.

[7] Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods*. Springer Series in Statistics. Springer-Vrelag.

[8] Chen, J., A. M. Variyath, and B. Abraham (2008, June). Adjusted Empirical Likelihood and Its Properties. *Journal of Computational and Graphical Statistics 17*(2), 426–443.

[9] Chen, S. X., J. Gao, and C. Y. Tang (2008). A test for model specification of diffusion processes. *The Annals of Statistics 36*(1), 167–198.

[10] Chernov, M., A. Gallant, E. Ghysels, and G. Tauchen (2003, September). Modelling and forecasting time series sampled at different frequencies. *Journal of Econometrics 116*(1), 225–257.

[11] Christoffersen, P., K. Jacobs, C. Ornthanalai, and Y. Wang (2008, December). Option valuation with long-run and short-run volatility components. *Journal of Financial Economics 90*(3), 272–297.

[12] Dias, D. A. and C. R. Marques (2010, January). Using mean reversion as a measure of persistence. *Economic Modelling 27*(1), 262–273.

[13] Ding, Z. (2016, June). Volatility Modeling Using GARCH: Theory and Practice. *Available at SSRN: http://dx.doi.org/10.2139/ssrn.2359212*.

[14] Ding, Z. and C. Granger (1996, July). Modeling Volatility Persistence of Speculative Returns: A New Approach. *Journal of Econometrics 73*(1), 185–215.

[15] Drgulescu, A. A. and V. M. Yakovenko (2002). Probability distribution of returns in the Heston model with stochastic volatility. *Quantitative Finance 2*(1), 443–453.

[16] Drost, F. and T. Nijman (1993, July). Temporal Aggregation of GARCH Processes. *Econometrica 61*(4), 909–927.

[17] Drost, F. and B. Werker (1996, September). Closing the GARCH Gap: Continuous Time GARCH Modeling. *Journal of Econometrics 74*(1), 31–57.

[18] Duan, J.-c. (1997, July). Augmented GARCH (p,q) process and its diffusion limit. *Journal of Econometrics* (1), 97–127.

[19] Eberlein, E. and F. Ozkan (2003). Time consistency of Lvy models. *Quantitative Finance 3*, 40–50.

[20] Engle, R. F. (1982, July). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica 50*(4), 987–1007.

[21] Engle, R. F. and G. G. J. Lee (1999). A Long-Run and Short-Run Component Model of Stock Return Volatility. *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, 475–497.

[22] Engle, R. F. and A. J. Patton (2001). What Good Is A Volatility Model? *Quantitative Finance 1*(2), 237–245.

[23] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications, Volume 1.* Wiley Series in Probability and Statistics. Wiley.

[24] Fouque, J.-P., G. Papanicolaou, R. Sircar, and K. Solna (2011). *Multiscale Stochastic Volatility for Equity, Interest Rate, and Credit Derivatives.* Cambridge University Press.

[25] Francq, C. and J.-M. Zakoan (1998, May). Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference 68*(1), 145–165.

[26] Francq, C. and J.-M. Zakoan (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications.* Wiley.

[27] Ghashghaie, S., W. Breymann, J. Peinke, P. Talkner, and Y. Dodge (1996, June). Turbulent Cascades In Foreign Exchange Markets. *Nature 381*, 767–770.

[28] Godambe, V. P. (1960). An Optimum Property of Regular Maximum Likelihood Estimation. *The Annals of Mathematical Statistics 31*(4), 1208–1211.

[29] Godambe, V. P. (1991). *Estimating Functions.* Oxford University Press.

[30] Godambe, V. P. and C. C. Heyde (1987). Quasi-Likelihood and Optimal Estimation. *International Statistical Review 55*(3), 231–244.

[31] Hafner, C. M. (2008, January). Temporal Aggregation of Multivariate GARCH Processes. *Journal of Econometrics 142*(1), 467–483.

[32] Hafner, C. M. and J. V. K. Rombouts (2007, August). Estimation of Temporally Aggregated Multivariate GARCH Models. *Journal of Statistical Computation and Simulation 77*(8), 629–650.

[33] Hansen, P. R. and A. Lunde (2014, February). Estimating the Persistence and the Autocorrelation Function of a Time Series that is Measured with Error. *Econometric Theory 30*(1), 60–93.

[34] He, C. and T. Terasvirta (1999, December). Fourth Moment Structure of the GARCH(p, q) Process. *Econometric Theory 15*(6), 824–846.

[35] Heston, S. L. (1993, April). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies 6*(2), 327–343.

[36] Hyndman, R. J. (1996, May). Computing and graphing highest density regions. *The American Statistician 50*(2), 120–126.

[37] Ibragimov, I. and Y. V. Linnik (1971). *Independent and Stationary Sequence of Random Variables.* Wolters-Noordhoff, Groningen.

[38] Karanasos, M. (1999, May). The Second Moment and the Autocovariance Function of the Squared Errors of the GARCH model. *Journal of Econometrics 90*(1), 63–76.

[39] Kaufmann, R. (2004). Long-Term Risk Management. *ETH Zurich, PhD Thesis*.

[40] Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *The Annals of Statistics 25*(5), 2084–2102.

[41] Kitamura, Y. (2006, June). Empirical Likelihood Methods in Econometrics: Theory and Practice. *Cowles Foundation Discussion Paper No. 1569*.

[42] Kunsch, H. R. (1989). The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics 17*(3), 1217–1241.

[43] LeBaron, B. (2001). Stochastic volatility as a simple generator of apparent financial power laws and long memory. *Quantitative Finance 1*, 621–631.

[44] Ling, S. and M. McAleer (2002, January). Stationarity and the existence of moments of a family of GARCH processes. *Journal of Econometrics 106*(2), 109–117.

[45] Liu, Y. and J. Chen (2010). Adjusted empirical likelihood with high-order precision. *The Annals of Statistics 38*(3), 1341–1362.

[46] Lo, A. and C. MacKinlay (1988, January). Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test. *The Review of Financial Studies 1*(1), 41–66.

[47] Mandelbrot, B., A. Fisher, and L. Calvet (1997). A Multifractal Model of Asset Returns. *Cowles Foundation Discussion Paper No. 1164*.

[48] Mandelbrot, B. B. (1963, October). The Variation of Certain Speculative Prices. *The Journal of Business 36*(4), 394–419.

[49] Mantegna, R. and E. Stanley (1995, July). Temporal Aggregation of Volatility Models. *Nature* (376), 46–49.

[50] Mantegna, R. and E. Stanley (1996, October). Turbulence and financial markets. *Nature* (383), 587–588.

[51] McLeish, D. and C. Small (1988). *The Theory and Applications of Statistical Inference Functions*. Lecture Notes in Statistics. Springer.

[52] McNeil, A. J. and R. Frey (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance 7*(3-4), 271–300.

[53] McNeil, A. J., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

[54] Meddahi, N. and E. Renault (2004, April). Temporal aggregation of volatility models. *Journal of Econometrics 119*(2), 355–379.

[55] Muller, U. A., M. M.Dacorogna, R. D.Dave, R. B.Olsen, O. V.Pictet, and J. E. Weizscker (1997, June). Volatilities of different time resolutions Analyzing the dynamics of market components. *Journal of Empirical Finance 4*(2-3), 213–239.

[56] Nelson, D. B. (1990, July/August). ARCH models as diffusion approximations. *Journal of Econometrics 45*(1-2), 7–38.

[57] Nelson, D. B. and C. Q. Cao (1992, April). Inequality Constraints in the Univariate GARCH Model. *Journal of Business ' Economic Statistics 10*(2), 229–235.

[58] Nordman, D. J., H. Bunzel, and S. N. Lahiri (2013). A nonstandard empirical likelihood for time series. *The Annals of Statistics 41*(26), 3050–3073.

[59] Nordman, D. J. and S. N.Lahiri (2014, December). A review of empirical likelihood methods for time series. *Journal of Statistical Planning and Inference 155*, 1–18.

[60] Ohanissian, A., J. R. Russell, and R. S. Tsay (2008, APril). True or Spurious Long Memory? A New Test. *Journal of Business and Economic Statistics 26*(2), 161–175.

[61] Owen, A. (1990). Empirical Likelihood Ratio Confidence Regions. *The Annals of Statistics 18*(1), 90–120.

[62] Owen, A. B. (2001). *Empirical Likelihood*. Chapman and Hall/CRC Monographs on Statistics and Applied Probability. Chapman and Hall/CRC.

[63] Politis, D. N., J. P. Romano, and M. Wolf (1999). *Subsampling*. Springer Series in Statistics. Springer.

[64] Qin, J. (1992). Empirical Likelihood and Semiparametric Models. *PhD Thesis, University of Waterloo*.

[65] Qin, J. and J. Lawless (1994). Empirical Likelihood and General Estimating Equations. *The Annals of Statistics 22*(1), 300–325.

[66] Silvestrini, A. and D. Veredas (2008, July). Temporal Aggregation of Univariate and Multivariate Time Series Models: A Survey. *Journal of Economic Surveys 22*(3), 497–458.

[67] Small, C., J. Wang, and Z. Yan (2000). Eliminating Multiple Root Problems in Estimation. *Statistical Science 15*(4), 313–341.

[68] Smith, R. J. (2011, December). GEL Criteria for Moment Condition Models. *Econometric Theory 27*(6), 1192–1235.

[69] Stentoft, L. (2011, December). American option pricing with discrete and continuous time models: An empirical comparison. *Journal of Empirical Finance 18*(5), 880–902.

[70] Stock, J. H. (1991, December). Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. *Journal of Monetary Economics 3*(28), 435–459.

[71] Tsai, H. and K.-S. Chan (2008, June). A Note on Inequality Constraints in the GARCH Model. *Econometric Theory 24*(3), 823–828.

[72] Wang, Y. (2002). Asymptotic nonequivalence of GARCH models and diffusions. *The Annals of Statistics 30*(3), 30.

[73] Wei, W. W. (2005). *Time Series Analysis: Univariate and Multivariate Methods*. Pearson.

[74] Wirjanto, T. (1997). Estimating Functions and Over-Identified Models. *Lecture Notes – Monograph Series 32*, 239–257.

Appendix A

Appendix: Some Background Material for Empirical Likelihood Inference

We first introduce the empirical likelihood method for the inference of the mean of an i.i.d. (vector-valued) sample, and then we describe how to combine EL with general estimating equations in the context of weakly dependent data. We mainly follow Owen (2001)[62] for the exposition. The proofs of the results used in our exposition can be found in Owen(1990)[61], Qin and Lawless (1994)[65], and Kitamura (1997)[40].

A.1 EL for i.i.d. data and inference about the mean

The key idea behind inference based on the empirical likelihood approach can be defined through a nonparametric likelihood ratio function under a set of estimating equations. It results in using a parametric family that is a multinomial distribution over the observed data values.

Suppose that $X_1, ..., X_n \in \mathbb{R}^d$ are independent vector-valued random variables with a common distribution function (DF) F_0 with some $d \geq 1$. It is convenient to describe distributions by the probabilities that they attach to sets in the vector case. Therefore, we denote by $F(A) = Pr(X \in A)$ for $X \sim F$ and $A \subseteq \mathbb{R}^d$.

We let δ_x denote the distribution under which $X = x$ with probability 1. Thus $\delta_x(A) = 1_{x \in A}$. Let $x_1, \cdots, x_n$ be a realization of $X_1, \cdots, X_n$.

Definition The *empirical cumulative distribution function (ECDF)* of $X_1, ..., X_n$ is

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(x) \text{ for } x \in \mathbb{R}^d.$$

Definition Given $X_1, ..., X_n \in \mathbb{R}^d$, which are assumed to be independent with a common DF F_0, the *nonparametric likelihood function* of any DF F (not necessarily F_0) is

$$L(F) := \prod_{i=1}^{n} F(\{x_i\}).$$

where $F(\{x_i\})$ is the probability of obtaining the observation, or realization of $X_i, i = 1, \cdots n$ under F.

Theorem 3.1 of Owen (2001)[62] Let $X_1, ..., X_n \in \mathbb{R}^d$ be independent random variables with a common DF F_0. Let F_n be their EDF and let F be any DF. If $F \neq F_n$, then $L(F) < L(F_n)$.
Proof: See Owen (2001)[62] Theorem 3.1. □

We use a ratio of nonparametric likelihoods as a basis for hypothesis testing and confidence intervals. For a distribution F, we define

$$R(F) := \frac{L(F)}{L(F_n)}.$$

Denote by $p_i \geq 0$ the probability that the distribution F assigns to the realization of $x_i \in \mathbb{R}^d$, where $\sum_{i=1}^{n} p_i \leq 1$. Then $L(F) = \prod_{i=1}^{n} p_i$ and

$$R(F) = \frac{L(F)}{L(F_n)} = \prod_{i=1}^{n} np_i. \tag{A.1}$$

In practice, it is possible to have ties in the data, i.e. $x_i = x_j$ for some $i \neq j$. To deal with ties in a more convenient fashion, we may replace the p_j's with a set of observation specific weights $w_i \geq 0$, for $i = 1, ..., n$. The w_i's are chosen such that p_j is equal to the sum of w_i over all i with $x_i = x_j$. Such a distribution which puts weight w_i on observation x_i reproduces F. So we work with

$$R(F) = \prod_{i=1}^{n} nw_i, \tag{A.2}$$

where $w_i \geq 0$, $\sum_{i=1}^{n} w_i = 1$, and F puts probability $\sum_{j:x_i=x_j} w_j$ on X_i.

Assume that we are interested in making inference about the common mean μ of the random vectors $X_1, ..., X_n \in \mathbb{R}^d$, i.e. $E(X_i) = \mu$. Using the distributions with $\sum_{i=1}^{n} w_i = 1$, we write the profile empirical likelihood ratio function for the vector mean as

$$\mathcal{R}(\mu) = \sup_{w_i} \left\{ \prod_{i=1}^{n} n w_i | w_i \geq 0, \sum_{i=1}^{n} w_i = 1, \sum_{i=1}^{n} w_i X_i = \mu \right\}. \tag{A.3}$$

The replacement of $\sum_{i=1}^{n} w_i \leq 1$ by $\sum_{i=1}^{n} w_i = 1$ is justified by the following argument. If we have $1 - \sum_{i=1}^{n} w_i > 0$, then this probability can be reassigned to data points in such a way that the new distribution $\tilde{F}$ has the same mean as F but has $L(\tilde{F}) > L(F)$.

Confidence region for the mean is

$$C_{r_0,n} = \left\{ \sum_{i=1}^{n} w_i X_i | w_i \geq 0, \sum_{i=1}^{n} w_i = 1, \prod_{i=1}^{n} n w_i \geq r_0 \right\}, \tag{A.4}$$

for some threshold value r_0.

For testing the null hypothesis $H_0 : \mu = \mu_0$, we reject H_0 when $\mathcal{R}(\mu_0)$ is less then some threshold value r_0. The following empirical likelihood theorem serves as a basis for determining the threshold value r_0.

Empirical Likelihood Theorem (ELT) for Vector Mean (Owen 2001[62], Theorem 3.2). Let $X_1, ..., X_n$ be independent random vectors in $\mathbb{R}^d$ with a common distribution F_0 having mean μ_0 and a finite variance matrix V_0 of rank $b > 0$. Then $C_{r,n}$ is a convex set and $-2 \log \mathcal{R}(\mu_0)$ converges in distribution to a $\chi^2_{(b)}$ random variable as $n \to \infty$.

$$-2 \log \mathcal{R}(\mu_0) = n(\bar{X} - \mu_0)' S^{-1} (\bar{X} - \mu_0) \to \chi^2_b.$$

Proof: See Owen (2001)[62].

We denote $W(\mu) = -2 \log \mathcal{R}(\mu)$. Approximate α-level confidence regions for μ may be obtained as the set of points μ such that $W(\mu) \leq c_\alpha$, where c_α is defined such that $Pr(\chi^2_{(b)} \leq c_\alpha) = 1 - \alpha$.

Solving the EL problem

As described in Owen (2001)[62], section 2.1, a unique value for the right-hand side of (A.3) exists, provided that 0 is inside the convex hull of the points $x_1 - \mu, ..., x_n - \mu$. An explicit expression for $\mathcal{R}(\mu)$ can be derived by a Lagrange multiplier argument: the maximum of $\prod_{i=1}^n nw_i$ subject to the constraints $w_i > 0$, $\sum_{i=1}^n w_i = 1$ and $\sum_{i=1}^n w_i(x_i - \mu) = 0$ is attained when

$$w_i = w_i(\mu) = n^{-1}\{1 + \lambda'(x_i - \mu)\}^{-1}, \tag{A.5}$$

where $\lambda = \lambda(\mu)$ is a $d \times 1$ vector given as the solution to

$$\sum_{i=1}^n \{1 + \lambda'(x_i - \mu)\}^{-1}(x_i - \mu) = 0. \tag{A.6}$$

Thus,

$$\mathcal{R}(\mu) = \prod_{i=1}^n \{1 + \lambda'(x_i - \mu)\}^{-1}. \tag{A.7}$$

The profile log-empirical likelihood ratio is

$$\log \mathcal{R}(\mu) = \sum_{i=1}^n \log\{1 + \lambda'(x_i - \mu)\}. \tag{A.8}$$

A.2 EL with estimating equations

While the empirical likelihood inference method proposed by Owen (1990)[61] initially focused on the mean of random vectors, Qin and Lawless (1994)[65] combined empirical likelihood and estimating equations to form a very general inference framework.

Assume that there are r estimating equations as given in (4.2) and s parameters which are summarized in the vector of parameters $\boldsymbol{\theta}$. We consider cases where there are at least as many estimating equations as the number of parameters, i.e. $r \geq s$. Assume that there are n i.i.d. d-dimensional samples $X_1, \cdots, X_n$.

For the estimating equations (4.2), the empirical likelihood approach is based on the profile empirical likelihood ratio function:

$$\mathcal{R}(\boldsymbol{\theta}) = \sup_{w_i} \left\{ \prod_{i=1}^{n} n w_i | w_i > 0, \sum_{i=1}^{n} w_i = 1, \sum_{i=1}^{n} w_i g_i = 0 \right\}. \tag{A.9}$$

We may maximize $\log \mathcal{R}(\boldsymbol{\theta})$ to obtain an estimate $\tilde{\boldsymbol{\theta}}$ of the parameter $\boldsymbol{\theta}$, called the maximum empirical likelihood estimate (MELE), and empirical likelihood ratio statistics can be constructed, similar to the case for the vector mean, to test various hypotheses about parameter values and model specification.

The following assumptions are made by Qin and Lawless (1994)[65] in order to prove the asymptotic results:
i) $E[g(X, \boldsymbol{\theta}_0) g'(X, \boldsymbol{\theta}_0)]$ is positive definite,
ii) $\partial g(X, \boldsymbol{\theta})/\partial \boldsymbol{\theta}$ is continuous in a neighborhood of the true parameter $\boldsymbol{\theta}_0$,
iii) $||\partial g(X, \boldsymbol{\theta})/\partial \boldsymbol{\theta}||$ and $||g(X, \boldsymbol{\theta})||^3$ are bounded by some integrable function in this neighborhood,
iv) the rank of $E[\partial g(X, \boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}]$ is r,
v) $\partial^2 g(x, \boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$ is continuous in $\boldsymbol{\theta}$ in a neighborhood of the true value $\boldsymbol{\theta}_0$,
vi) $||\partial g(x, \boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'||$ can be bounded by some integrable function in the neighborhood of the true value $\boldsymbol{\theta}_0$.

Then, Theorem 1 of Qin and Lawless (1994)[65] showed that $\tilde{\boldsymbol{\theta}}$, $\tilde{\lambda}$, and $\tilde{F}_n(x)$ are asymptotically normally distributed. In addition, the MELE $\tilde{\boldsymbol{\theta}}$ is fully efficient in the sense that it has the same asymptotic variance as the optimal estimator obtained from the class of $r \times 1$ estimating equations that are linear combinations of $g(X, \boldsymbol{\theta})$.

Empirical likelihood ratio statistics can be constructed to test hypothesis about parameters and the model. The following theorems provide a basis for the testing problems that we are interested in.

Empirical Likelihood Theorem (ELT) for Testing a Model with a Particular Set of Parameters. The empirical likelihood ratio statistic for testing $H_0 : E[g(X, \boldsymbol{\theta}_0)] = 0$ is

$$W_1(\boldsymbol{\theta}_0) = -2 \log \mathcal{R}(\boldsymbol{\theta}_0). \tag{A.10}$$

Under the regularity conditions, $W_1(\boldsymbol{\theta}_0)$ converges to a χ^2_r random variable as $n \to \infty$ when H_0 is true.

Empirical Likelihood Theorem (ELT) for Testing a Model (Qin and Lawless (1994)[65], Corollary 4). The empirical likelihood ratio statistic for testing $H_0 : E[g(X, \boldsymbol{\theta})] = 0$ is

$$W_2(\tilde{\boldsymbol{\theta}}) = -2 \log \mathcal{R}(\tilde{\boldsymbol{\theta}}). \tag{A.11}$$

Under the regularity conditions, $W_2(\tilde{\boldsymbol{\theta}})$ converges to a χ^2_{r-s} random variable as $n \to \infty$ when H_0 is true.

A.3 EL for dependent data

While Qin and Lawless (1994)[65] considered i.i.d. data, there are many other situations where we have dependent data, such as in the cases of ARMA processes that we will deal with. A naive application of the empirical likelihood theorems of Qin and Lawless (1994)[65] in the case of dependent data will cause the theorems to fail because the covariance estimator for i.i.d. data is improper for dependent data. In this respect, Kitamura (1997)[40] proposed block empirical likelihood (BEL) method which applies to weakly dependent data under α-mixing assumption. We use the BEL approach to deal with temporal dependency in our case. For reviews of methods for empirical likelihood methods for dependent data, see Kitamura (2006)[41] and Nordman and Lahiri (2014)[59].

Particularly, Kitamura (1997)[40] considered a strong mixing type of dependency:
Definition (Strong Mixing): Let $\{X_t\}$ be a d-dimensional real-valued stationary stochastic process satisfying

$$\alpha_X(k) \to 0, \text{ when } k \to \infty, \tag{A.12}$$

where $\alpha_X(k) = \underset{A,B}{sup}|P(A \cap B) - P(A)P(B)|$, $A \in \mathcal{F}^0_{-\infty}$, $B \in \mathcal{F}^\infty_k$ and $\mathcal{F}^n_m = \sigma(X_t,\ m \leq t \leq n)$. And it is further assumed there exists some constant c such that

$$\sum_{k=1}^{\infty} \alpha_X(k)^{1-1/c} < \infty \tag{A.13}$$

Let $M \leq n$ be block length and some $L \leq M$ be the separation between consecutive blocks. Assume $M \to \infty$, $M = o(n^{1/2})$, $L = O(M)$ as $n \to \infty$, and $L \leq M$. Denote by B_i, $i \in \mathbb{N}$ a vector of M consecutive observations $(X_{(i-1)L+1}, \cdots, X_{(i-1)L+M})$. M is called the "window width" of the blocking scheme and L is the separation between block starting points. Then, each block of observation is mapped by the mapping $\phi_M(B_i)$ defined as

$$T_i(\boldsymbol{\theta}) = \phi_M(B_i, \boldsymbol{\theta}) = \sum_{j=1}^{M} g(X_{(i-1)L+j}, \boldsymbol{\theta})/M,$$

in which $i = 1, \cdots, Q$ where $Q = [(n-M)/L]+1$ is the total number of blocks and $[c]$ is the biggest integer smaller than c. Then, the empirical likelihood method is applied to the *blocked* observations T_i's. The profile empirical likelihood ratio function is

$$\mathcal{R}_B(\boldsymbol{\theta}) = \underset{w_i^B}{sup} \left\{ \prod_{i=1}^{Q} Q w_i^B | w_i^B \geq 0, \sum_{i=1}^{Q} w_i^B = 1, \sum_{i=1}^{Q} w_i^B T_i = 0 \right\}. \tag{A.14}$$

Using the Lagrange multiplier argument, $\mathcal{R}_B(\boldsymbol{\theta})$ can be written as

$$\mathcal{R}_B(\boldsymbol{\theta}) = \prod_{i=1}^{Q} \{1 + \lambda' T_i\}^{-1}. \tag{A.15}$$

Notice that in the block empirical likelihood method, w_i^B and λ depend on the blocking parameters M and L.

The key to the success of the blocking EL method is that $\sum T_i(\boldsymbol{\theta}) T_i(\boldsymbol{\theta})'$ constitutes a consistent estimator of the variance of $g(X_i, \boldsymbol{\theta})$, whereas $\sum g_i(\boldsymbol{\theta}) g_i(\boldsymbol{\theta})'$ does not as the latter ignores the dependency in the $g(X_i, \boldsymbol{\theta})$'s. Ignorance of the dependency in the $g(X_i, \boldsymbol{\theta})$'s would cause the empirical likelihood ratio statistics fail to converge to a χ^2 limit.

Kitamura (1997)[40] proved empirical likelihood theorems for the following block version of the log-empirical likelihood ratio statistics with weakly dependent observations, which are counterparts to the empirical likelihood theorems for estimating equations with i.i.d. observations.

Denote by $\Gamma(z, \delta)$ an open sphere with center z and radius δ and $||\cdot||$ the Eclidean norm. The following regularity conditions are assumed:
i) The parameter space Θ is compact;
ii) $\boldsymbol{\theta}_0$ is the unique root of $Eg(X_t, \boldsymbol{\theta}_0) = 0$;
iii) For sufficiently small $\delta > 0$ and $\eta > 0$, $E \sup_{\theta^* \in \Gamma(\boldsymbol{\theta}, \delta)} ||g(X_t, \theta^*)||^{2(1+\eta)} < \infty$ for all $\boldsymbol{\theta} \in \Theta$;
iv) If a sequence $\boldsymbol{\theta}_j$, $j = 1, 2, \cdots$ converges to some $\boldsymbol{\theta} \in \Theta$ as $j \to \infty$, $g(x, \boldsymbol{\theta})$ for all x except perhaps on a null set, which may vary with $\boldsymbol{\theta}$;

v) $\boldsymbol{\theta}_0$ is an interior point of Θ and $g(x, \boldsymbol{\theta})$ is twice continuously differentiable at $\boldsymbol{\theta}_0$;
vi) $Var[n^{-1/2} \sum_{i=1}^{n} g(X_i, \boldsymbol{\theta}_0)] \to S > 0$ as $n \to \infty$;
vii) $E||g(x, \boldsymbol{\theta}_0)||^2 < \infty$ for $c > 1$ defined in (A.13), $E \sup_{\theta^* \in \Gamma(\boldsymbol{\theta}_0, \delta)} ||g(X_t, \theta^*)||^{2+\epsilon} < K$, $M = o(n^{1/2 - 1/(2+\epsilon)})$ for some $\epsilon > 0$, $E \sup_{\theta^* \in \Gamma(\boldsymbol{\theta}_0, \delta)} ||\partial g(X_t, \theta^*)/\partial \theta'||^2 < K$ and

$E \sup_{\theta^* \in \Gamma(\boldsymbol{\theta}_0, \delta)} ||\partial^2 g_j(X_t, \theta^*)/\partial \theta \partial \theta'||^2 < K$ for all $j = 1, \cdots, r$ where $K < \infty$;
viii) $E \partial g(X_t, \boldsymbol{\theta}_0)/\partial \theta'$ is of full column rank.

Then, we have
Block Empirical Likelihood Theorem for Testing a Model with a Particular Set of Parameter. The empirical likelihood ratio statistic for testing $H_0 : E[g(X, \boldsymbol{\theta}_0)] = 0$ is

$$W_{B1}(\boldsymbol{\theta}_0) = -2A_n^{-1} \log \mathcal{R}_B(\boldsymbol{\theta}_0), \tag{A.16}$$

where $A_n = QM/n$. Under the regularity conditions, $W_{B1}(\boldsymbol{\theta}_0)$ converges to a χ_r^2 random variable as $n \to \infty$ when H_0 is true.

Block Empirical Likelihood Theorem for Testing a Model. The empirical likelihood ratio statistic for testing $H_0 : E[g(X, \boldsymbol{\theta})] = 0$ is

$$W_{B2}(\tilde{\boldsymbol{\theta}}) = -2A_n^{-1} \log \mathcal{R}_B(\tilde{\boldsymbol{\theta}}). \tag{A.17}$$

Under the regularity conditions, $W_{B2}(\tilde{\boldsymbol{\theta}})$ converges to a χ_{r-s}^2 random variable as $n \to \infty$ when H_0 is true.