2018-05-11

Minimum Hellinger Distance Estimation of ARCH/GARCH Models

Chen, Liang

UNIVERSITY OF CALGARY

Minimum Hellinger Distance Estimation of ARCH/GARCH Models

by

Liang Chen

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

May, 2018

Abstract

In this thesis, we proposed a minimum Hellinger distance estimator (MHDE) and a minimum profile Hellinger distance estimator (MPHDE) for estimating the parameters in the ARCH and GARCH models depending on whether the innovation distribution is specified or not. The asymptotic properties of MHDE and MPHDE were examined through graphs as the theoretical investigation of them are more involved and needs further study in the future research. Moreover, we demonstrated the finite-sample performance of both MHDE and MPHDE through simulation studies and compared them with the well-established methods including maximum likelihood estimation (MLE), Gaussian Quasi-MLE (GQMLE) and Non-Gaussian Quasi-MLE (NGQMLE). Our numerical results showed that MHDE and MPHDE have better performance in terms of bias, MSE and coverage probability (CP) when the data are contaminated, which testified to the robustness of MHD-type estimators.

Acknowledgements

This thesis would have not been accomplished without the valuable assistance and support from many people, who have contributed in the preparation and completion of this thesis.

First of all, I would like to take this opportunity to express my sincere gratitude to my supervisor Dr. Jingjing Wu. Dr. Wu guided me throughout my two-year study in the master's program in statistics at University of Calgary. She spent significant amount of time helping me to gain a comprehensive understanding of this research both theoretically and practically. As my research progressing, Dr. Wu provided me with a great number of valuable suggestions and guidances, which led to significant improvement of this thesis. Without her encouragement, commitment and patience, I would not be able to make it. Moreover, she kindly provided me the opportunity to give a talk about my research at the First International Conference on Econometrics and Statistics held in HongKong on June 2017. I can never thank Dr. Wu enough for what she has done for my academic accomplishment achieved in the past two years.

I would also like to give my great appreciation to other members of my committee, Dr. Xuewen Lu and Dr. Hua Shen. I am very grateful that they sacrificed their precious time to read and offer valuable advices to my thesis.

Special thanks go to the Department of Mathematics and Statistics. The department provided me with great learning environments and financial supports so that I can better concentrate on my study and research. Generous helps on my study, research and career development were offered by professors, staff and schoolmates in the department. I especially thank Dr. Ware for his support of the Mitacs project, in which I gained direct exposure to the pratice of energy market risk management. It enabled me to translate the knowledge and techniques that I have learned in school into real-world applications, which contributed to my professional development and competitiveness in job market.

Many thanks go to the primary references, such as Beran [4], Fan et al. [11] and Hall and Yao [14]. Their outstanding works greatly inspired me and guided my thesis.

Last but not the least, I am deeply indebted to my father and mother for their constant and unconditional support for me. Their encouragements and expectations are the motivation behind my hard work.

Table of Contents

List of Tables

List of Figures and Illustrations

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
ARCH	autoregressive conditional heteroscedasticity
GARCH	generalized autoregressive conditional heteroscedasticity
MHDE	minimum Hellinger distance estimation/estimator/estimate
MPHDE	minimum profile Hellinger distance estimation/estimator/estimate
SMHDE	simulated minimum Hellinger distance estimation/estimator/estimate
MLE	maximum likelihood estimation/estimator/estimate
GQMLE	Guassian quasi-maximum likelihood estimation/estimator/estimate
NGQMLE	non-Guassian quasi-maximum likelihood estimation/estimator/estimate
DCT	Dominated Convergence Theorem
p.d.f.	probability distribution function
i.i.d.	independent and identically distributed
r.v.	random variable
$a.e.$	almost everywhere
CP	Coverage Probability
$\|\cdot\|$	L^2-norm
$\xrightarrow{\mathscr{P}}$	converge in probability
$\xrightarrow{a.e.}$	converge almost everywhere

Chapter 1

INTRODUCTION

In this chapter, we give some background introduction for this research work. In Section 1.1, we introduce the ARCH and GARCH models that will be considered throughout the thesis. In Sections 1.2 and 1.3 we review the minimum Hellinger distance estimation (MHDE) and minimum profile Hellinger distance estimation (MPHDE) respectively, which are the methods that will be used to estimate the ARCH/GARCH models.

1.1. ARCH/GARCH Models

Volatility has played a vital role in financial time series analysis and risk management. It is an important indicator of market risk in that higher volatility generally indicates higher risk involved and as a result companies need to adjust their trading strategies based on the volatility forecasting [11]. Therefore, generating reliable volatility forecasting has long been the main concern of financial institutions. What complicates the volatility analysis is that volatility tends to cluster together, i.e. large/small changes tend to be followed by large/small changes. This phenomenon suggests that volatility is autocorrelated and changing over time. To deal with this issue, Engle [10] has proposed autoregressive conditional heteroscedasticity (ARCH) to model volatility dynamics as a function of past squared returns: for $i = 0, 1, \cdots, n$,

$$\begin{aligned} X_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \beta_0 + \sum_{i=1}^{p} \beta_i X_{t-i}^2, \end{aligned} \tag{1.1}$$

where X_t is observable and is generally interpreted as returns of an asset at time t, σ_t^2 describes the conditional variance of the current returns, and the innovation error ε_t is unobservable but independent of X_{t-j} for $j = 1, \cdots, t$. This model is commonly denoted as

ARCH(p) with p the degree of dependence.

In the ARCH(p) model (1.1), β_0 is the lower bound of the conditional and unconditional variance of X_t's, while β_i, $i = 1, \dots p$, measures how fast the impact of a shock in returns at time $t-i$ on volatility at time t fades over time. In other words, any large (in absolute value) shock in ε_t will be associated with a persistently large (conditional) variance in the returns X_t. The larger the $\sum_{i=1}^{p} \beta_i$, the longer the persistence of this effect. Theoretically, we need $\beta_0 \geq 0$, $0 \leq \beta_i < 1$ for $i = 1, \cdots, p$ and $\sum_{i=1}^{p} \beta_i < 1$ to ensure the stability of ARCH(p) process.

The ε_t's are i.i.d. random shocks featuring a white noise process with mean zero and variance one. This enables σ_t's to bear the interpretation of conditional volatility of X_t given all the information about X up to $t-1$, denoted as $\mathscr{F}_{t-1}$. To see this, notice that $E(\varepsilon_t^2|\mathscr{F}_{t-1}) = E(\varepsilon_t^2) = 1$ and $E(X_t|\mathscr{F}_{t-1}) = E(\sigma_t\varepsilon_t|\mathscr{F}_{t-1}) = \sigma_t E(\varepsilon_t|\mathscr{F}_{t-1}) = 0$, thus the variance of X_t conditional on the past history $X_{t-1}, X_{t-2}, \dots, X_0$ is

$$E(X_t^2|\mathscr{F}_{t-1}) \;=\; E(\sigma_t^2\varepsilon_t^2|\mathscr{F}_{t-1}) \;=\; \sigma_t^2 E(\varepsilon_t^2|\mathscr{F}_{t-1}) \;=\; \sigma_t^2.$$

Moreover, since $E(\varepsilon_t\varepsilon_{t-j}) = 0$ and ε_t and X_{t-j} are independent for $j \neq 0$, it also follows that

$$E(X_t X_{t-j}) \;=\; E[(\sigma_t\varepsilon_t)(\sigma_{t-j}\varepsilon_{t-j})] \;=\; E(\sigma_t\sigma_{t-j})E(\varepsilon_t\varepsilon_{t-j}) \;=\; 0.$$

This implies that X_t is serially uncorrelated but not independent. However, the squared returns are generally serially correlated and thus not independent as well.

The introduction of ARCH(p) model has led to the proliferation of related volatility models. Among numerous generalizations of ARCH model, the following GARCH(p,q) model proposed by Bollerslev [5] has been most widely used:

$$\begin{aligned} X_t &= \sigma_t\varepsilon_t, \\ \sigma_t^2 &= \beta_0 + \sum_{i=1}^{p} \beta_i X_{t-i}^2 + \sum_{j=1}^{q} \alpha_j \sigma_{t-j}^2. \end{aligned} \tag{1.2}$$

In the GARCH(p,q) model, the current conditional variance, σ_t^2, depends on the weighted averages of not only past squared returns but also historical (conditional) variances. When it comes to the interpretation of the model parameters, β_0 is the lower bound of the conditional and unconditional variance of X_t's. The β_i measures the extent to which a shock in returns at time $t-i$ affects the volatility at time t. The α_j measures the effect of volatility shock at time $t-j$ on the volatility at time t. In addition, $\sum_{i=1}^{p}\beta_i + \sum_{j=1}^{q}\alpha_j$ measures the rate at which these effects dies over time. The larger the $\sum_{i=1}^{p}\beta_i + \sum_{j=1}^{q}\alpha_j$, the longer the persistence of these effects. To ensure the stability of the GARCH(p,q) process, we need the constraints $\beta_0 \geq 0$, $0 \leq \beta_i < 1$ for $i = 1,\ldots,p$, $0 \leq \alpha_j < 1$ for $j = 1,\ldots,q$ and $0 < \sum_{i=1}^{p}\beta_i + \sum_{j=1}^{q}\alpha_j < 1$.

One benefit of using GARCH model is that a high-order ARCH model can be approximated by a more parsimonious GARCH representation that is easier to be identified and estimated [9]. Therefore, GARCH(1,1) model is usually used in volatility analysis and forecasting.

Since the introduction of ARCH and GARCH models, the estimation of model parameters has been a popular research topic in statistics and economics. The most common method is the Guassian Quasi-maximum likelihood estimation (GQMLE), which is based on the assumption that the distribution of innovation ε_t is normal. This in turn implies that the returns has a conditional normal distribution. Notwithstanding that a wealth of empirical evidence (see, for example, McFarland et al. [24] and Baillie and Bollerslev [2]) has documented the substantial excess kurtosis of the innovation distribution, it has been shown in Weiss [33], Lee and Hansen [19] and Lumsdaine [23] that GQMLE is consistent and asymptotically normally distributed even if the true innovation distribution is far from normal, provided that the innovation distribution has a finite fourth moment. This desirable property has made GQMLE a widely diffused approach to the estimation of ARCH/GARCH models. Nevertheless, deviation of true innovation density from Guassian may increase the variance of GQMLE considerably and thereby fail to reach the efficiency of MLE by a wide margin, reflecting the cost of not knowing the true innovation distribution.

To account for the excess kurtosis of innovation distribution, the quasi-MLE (QMLE) assuming heavy-tailed distributions, such as the Student's t distribution, is sometimes used; see, for example, Bollerslev [6] and Nelson [26]. The drawback of this method is obvious as it may result in inconsistent estimates of model parameters if the specified distribution of innovation is different from the true innovation distribution. As a consequence, several methods have been introduced in an attempt to obtain consistent and asymptotically efficient estimators of the model parameters that requires minimal knowledge about the innovation distribution. For example, Storti [31] has presented a distribution-free approach that relies on minimizing the weighted distance between estimated and sample auto-covariance of squared returns, in which the proposed estimator competes favourably with GQMLE although the robustness properties of the estimator was not investigated. Moreover, Fan et al. [11] has proposed a non-Gaussian quasi-maximum likelihood estimator (NGQMLE) featuring a three-step procedure. To be specific, they use non-Gaussian likelihood functions that includes a scale parameter to correct the inconsistency of using non-Gaussian likelihood function when it is not compatible with the true distribution. The resulting estimator is more efficient than the GQMLE, particularly when the innovation error has heavy tails. In addition, Andrews [1] has proposed a rank-based estimator by minimizing a rank-based residual dispersion function, which has also been shown to be robust against density misspecification and more efficient than GQMLE.

Although various techniques have been employed to estimate the ARCH/GARCH model parameters in the literature, few have utilized minimum Hellinger distance (MHD) method except for Kadjo et al. [16]. Specifically, they proposed the simulated MHD (SMHD) method to the estimation of model parameters but with the assumption that the innovation distribution is known so that the returns (as a function of unknown parameters) can be simulated. The literature on even the class of all minimum distance estimations is surprisingly sparse. To our best knowledge, there are only a few works considering minimum distance estimation for ARCH/GARCH models. Specifically, Baillie and Chung [3] and Storti [31] have constructed weighted squared distance estimators for GARCH$(1,1)$ model based on autocorrelation functions. For general GARCH(p,q) models, Galbraith et al. [12] has developed a squared distance estimator based on the estimated ARCH parameters. However, MHD-type estimators have been shown to be consistent, asymtotically normal and more importantly, robust to data contamination by [4] for the i.i.d. data. Therefore, we should give serious consideration to it when it comes to the estimation of ARCH/GARCH model parameters. As a result, I will investigate the estimation of ARCH/GARCH models using MHD-type estimators.

1.2. Minimum Hellinger Distance Estimation (MHDE)

MHDE was first proposed by Beran [4] for parametric models. Beran [4] has shown that MHDE has good properties such as consistency, asymptotic normality and asymptotic variance achieving Cramér-Rao lower bound. In fact, Lindsay [21] has proved that MLE and MHDE are members of a large class of efficient estimators with various second-order efficiency properties. On the other hand, Beran [4] has shown that MHDE also possesses excellent robustness properties for parametric models. More specifically, the MHDE is resistant against both outliers and model misspecification. With both asymptotic efficiency and robustness, MHDEs are receiving increasing attention in practice and constitute a desirable class of estimators.

Consider the class of parametric models $\{f_\theta : \theta \in \Theta\}$, where the parameter space Θ is a subset of $\mathbb{R}^p$ with $p \in \mathbb{N}$. Following Beran [4], the MHDE of the unknown θ is defined as the value in Θ that minimizes the Hellinger distance between the parametric model f_θ and its nonparametric density estimator. Mathematically, the MHDE $\hat{\theta}$ is defined as

$$\hat{\theta} \;=\; \arg\inf_{\theta\in\Theta} \left\| f_\theta^{1/2} - \hat{f}^{1/2} \right\| \tag{1.3}$$

where $\|\cdot\|$ denotes the L^2-norm and $\hat{f}$ is a nonparametric density estimator of f_θ, such as the kernel density estimator.

The literature on MHDE has been dominated by its investigation under parametric models. For example, Simpson [29] has examined MHDE for discrete data models, while Yang [47] and Ying [48] have studied MHDE for censored data. Sriram and Vidyashankar [30] and Woo and Sriram [35, 36] have investigated MHDE for branching processes and mixture complexity of a finite mixture model, respectively. MHDE for finite mixture models and their variants were studied in Woodward et al. [37], Cutler and na [8], Karlis and Xekalaki [17], Lu et al. [22] and Xiang et al. [45]. Takada [32], N'drin and Hili [25] and Prause et al. [27] have studied MHDE in stochastic volatility model, one-dimensional diffusion process and bivariate time series, respectively. In addition, Karunamuni and Wu [18] has proposed a one-step MHDE to overcome computational drawbacks of MHDE.

On the other hand, MHDE for semiparametric models hasn't been fully investigated until recently. Suppose we observe independent and identically distributed (i.i.d.) random variables (r.v.s) $X_1,\ldots,X_n$ with density $g_0 = f_{\theta,\eta}$ being a member of the general semiparametric model

$$\mathscr{F} = \{f_{\theta,\eta} : \theta \in \Theta, \eta \in \mathscr{H}\}, \tag{1.4}$$

where Θ is a compact subset of $\mathbb{R}^p$ and $\mathscr{H}$ is an arbitrary set of infinite dimension. In general, θ is the parameter of interest with η being the nuisance parameter. To make estimating θ meaningful, assume $\mathscr{F}$ is identifiable in the sense that, if the Hellinger distance between f_{θ_1,η_1} and f_{θ_2,η_2} is 0, i.e. $\|f^{1/2}_{\theta_1,\eta_1} - f^{1/2}_{\theta_2,\eta_2}\| = 0$, then $\theta_1 = \theta_2$ and $\eta_1 = \eta_2$. Wu and Karunamuni [38, 39] have extended the MHDE from parametric models to (1.4), the semiparametric model of the general form, and the resulting MHDE has been proved to retain the efficiency and robustness properties under regularity conditions. Wu and Karunamuni [41] has examined the hypothesis testing based on the MHDE for model (1.4). For semiparametric models of specific form, Wu et al. [43] has studied MHDE in a two-sample semiparametric model with the two population probability distribution functions (p.d.f.) linked by an exponential ‘tilt’. Zhu et al. [49] has investigated MHDE under the same model but for survival data with cure rate, while Chen and Wu [7] has applied this model to the classification of leukemia patients. Wu and Zhou [42] has applied and studied MHDE for a semiparametric two-component location-shifted mixture model.

1.3. Minimum Profile Hellinger Distance Estimation (MPHDE)

For model (1.4), the works mentioned in Section 1.2 focuses on the case when an estimator $\hat{\eta}$ of η, based on either the same data or other resources, is available. If an estimator of η is not available, Wu and Karunamuni [40] has for the first time introduced the minimum profile Hellinger distance estimation (MPHDE). The MPHDE is obtained by first profiling out the infinite-dimensional nuisance parameter η and then minimizing the profiled Hellinger distance. More specifically, let $\hat{f}$ denote an appropriate nonparametric estimator of $f_{\theta,\eta}$ based on the sample $X_1,\ldots,X_n$. For any fixed $\theta \in \Theta$, we define the profiled nuisance parameter by

$$\eta(\theta) \;=\; \arg\inf_{\eta \in \mathscr{H}} \left\| f_{\theta,\eta}^{1/2} - \hat{f}^{1/2} \right\|, \quad \theta \in \Theta.$$

Now the MPHDE of θ is defined as

$$\hat{\theta} \;=\; \arg\inf_{\theta \in \Theta} \left\| f_{\theta,\eta(\theta)}^{1/2} - \hat{f}^{1/2} \right\|. \tag{1.5}$$

It has been proved by Wu and Karunamuni [40] that the MPHDE $\hat{\theta}$ defined in (1.5) is consistent, asymptotically normal, efficient, and adaptive (provided that the specific semiparametric model under consideration is adaptive). Furthermore, Wu and Karunamuni [40] has also shown that the MPHDE retains good robustness properties against outliers and model misspecification. Wu et al. [44] and Xiang et al. [46] have developed the MPHDE for, respectively, a semiparametric two-component location-shifted mixture model and a semiparametric two-component mixture model with the first component known up to some unknown parameters and the second component being an unspecified distribution with an unknown location parameter.

In this thesis, we investigate the MHD-type estimations for ARCH/GARCH models. Specifically, we focus on the estimation of the ARCH(1) and GARCH(1,1) models. However, we consider both cases of known and unknown innovation distribution, which corresponds to a parametric model and a semiparametric model, respectively. We propose a MHDE when the innovation distribution is known and a MPHDE when it is unknown. Note that the proposed MHDE and MPHDE are not direct applications of those defined for general parametric and semiparametric models in the sense that the latter are based on i.i.d. r.v.s while the data under our consideration are time series data and thus are not independent. We would like to study whether the resulting MHD-type estimators for ARCH/GARCH models still possess good efficiency and robustness properties.

The remainder of the thesis is orgainized as follows. Chapter 2 proposes and studies MHDE and MPHDE for ARCH/GARCH models when the innovation distribution is known and unknown, respectively. The MHDE and MPHDE are compared with the commonly used estimators in the current literature including MLE, GQMLE and NGQMLE. In Chapter 3, finite-sample performance, including both efficiency and robustness, of the proposed MHDE and MPHDE are examined through Monte Carlo simulation. A real data is analyzed in Chapter 4 using the proposed MPHDE. Finally, concluding remarks and discussion of future work are presented in Chapter 5.

Chapter 2

METHODOLOGIES

This chapter discusses the methods of estimating ARCH/GARCH models with particular emphasis on ARCH(1) and GARCH$(1,1)$ models. We consider both cases when the innovation distribution is known and unknown in Section 2.1 and Section 2.2, respectively. In Section 2.1, when the innovation distribution is known, we construct a MHDE for the unknown parameters and demonstrate its asymptotic properties using graphs. For the purpose of comparison in Chapter 3, we also review the commonly used MLE. In Section 2.2, when the innovation distribution is unknown, we propose a MPHDE of the unknown parameters. For the purpose of comparison in the next chapter, we also review some other competing estimators used in the literature, including the highly diffused GQMLE and the NGQMLE. Simulation studies demonstrating the finite-sample performance of MHDE as well as MPHDE are deferred to Chapter 3.

2.1. Innovation Distribution is Known

As a special case of GARCH(p,q) given in (1.2), the GARCH$(1,1)$ model is represented with different parameter notations as follows

$$\begin{aligned} X_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \beta_0 + \beta_1 X_{t-1}^2 + \beta_2 \sigma_{t-1}^2, \quad t = 0, 1, \cdots, n, \end{aligned} \tag{2.1}$$

where X_t's are observed serially uncorrelated but non-i.i.d. returns and ε_t's are unobservable i.i.d. random shocks following the p.d.f. f_ε with mean zero and variance one. In this section, we consider the relatively simpler case that the innovation distribution f_ε is known. Note that in (2.1), the GARCH$(1,1)$ is a parametric model when the innovation distribution is known. Since the ARCH(1) model is a special case of the more complicated GARCH$(1,1)$ model when $\beta_2 = 0$, we illustrate the construction of MHDE under GARCH$(1,1)$ model. In Section 2.1.1 we propose and construct the MHDE for GARCH(1,1) parameters. Section 2.1.2 is devoted to demonstrating the properties of the proposed MHDE through graphs. For comparison purposes, we review in Section 2.1.3 the commonly used MLE.

2.1.1 Construction of MHDE

MLE is known to be the most efficient estimator under the regularity conditions, i.e. it achieves the Cramér-Rao lower bound. However, it is very sensitive to and thus distorted by outliers and data contamination. As shown in many studies, financial time series generally contain extreme values, especially during financial crisis when outliers are generated and contaminate the prevailing distribution of returns. To account for this fact and to overcome the potential problems caused by non-robustness of MLE, in this section we propose a MHDE for the models under our consideration. We first give the construction of the MHDE and then study the asymptotic properties of the resulting estimator.

Recall that the original definition of MHDE given in (1.3) for parametric models is based on i.i.d. samples. Since the X_t's in (2.1) are non-i.i.d. r.v.s, it is not appropriate to construct the MHDE based on the distribution of X_t. One may argue that we can give a MHDE based on the joint distribution of the non-i.i.d. X_t's, but the problems are: first, it is tedious and difficult to derive the joint distribution of the X_t's in terms of model parameters; second, the thus resulting estimator may not possess the efficiency and robustness properties of the originally defined MHDE given in (1.3) for i.i.d. samples. Therefore, under model (2.1), we need to modify the original MHDE to accommodate the difficulties invoked by the nature of non-i.i.d samples.

Let $\beta = (\beta_0, \beta_1, \beta_2)^\top$ denote the true model parameter vector. We assume that the parameter space Θ is a compact subset of $\mathbb{R}^3$, as β_1, β_2 are bounded between 0 and 1 and β_0

is a finite constant. Note that

$$\varepsilon_t = \frac{X_t}{\sqrt{\beta_0 + \beta_1 X_{t-1}^2 + \beta_2 \sigma_{t-1}^2}}, \quad t = 1, \cdots, n. \tag{2.2}$$

are i.i.d. random shocks that are unobserved. Thus when β is known, the r.v.s defined in (2.2) are i.i.d. and follow a known p.d.f. f_ε. However, since β is actually unknown, we let $b = (b_0, b_1, b_2)^\top$ denote a reasonable estimate of β and define

$$v_{b,t} := \frac{X_t}{\sqrt{b_0 + b_1 X_{t-1}^2 + b_2 \sigma_{t-1}^2}} = \frac{\sqrt{\beta_0 + \beta_1 X_{t-1}^2 + \beta_2 \sigma_{t-1}^2}}{\sqrt{b_0 + b_1 X_{t-1}^2 + b_2 \sigma_{t-1}^2}} \varepsilon_t. \tag{2.3}$$

Then the $\{v_{b,t} : t = 0, 1, \ldots, n\}$ are expected to be i.i.d. and follow a p.d.f. close to f_ε. Note that $v_{\beta,t} = \varepsilon_t$ and the p.d.f. of $v_{\beta,t}$'s are exactly f_ε. Following this idea, we can construct a nonparametric density estimator of f_ε based on $v_{b,t}$. Throughout this thesis, we always use kernel nonparametric function estimation. Specifically, we use the kernel density estimator $\hat{f}(b;\cdot)$ of f_ε defined by

$$\hat{f}(b;x) = \frac{1}{nh_n} \sum_{t=1}^{n} K\left(\frac{x - v_{b,t}}{h_n}\right), \tag{2.4}$$

where K is a kernel function (non-negative p.d.f.) and h_n is a sequence of bandwidths such that $h_n > 0$, $h_n \to 0$ and and $nh_n \to 0$ as $n \to \infty$. Kernel functions are usually symmetric and some commonly used ones include triangle, quartic, Epanechnikov and Gaussian kernels. It is noteworthy that all kernels are asymptotically equivalent and different choices of kernel won't influence the asymptotic properties of the estimators. Now the MHDE $\hat{\beta}_{MHD}$ of β can be defined as

$$\begin{aligned} \hat{\beta}_{MHD} &= \arg\inf_{b\in\Theta} \left\| \hat{f}^{1/2}(b;x) - f_\varepsilon^{1/2}(x) \right\| \\ &= \arg\inf_{b\in\Theta} \left(\int \hat{f}(b;x)dx + \int f_\varepsilon(x)dx - 2\int \hat{f}^{1/2}(b;x) f_\varepsilon^{1/2}(x)dx \right)^{1/2} \\ &= \arg\inf_{b\in\Theta} \left(2 - 2\int \hat{f}^{1/2}(b;x) f_\varepsilon^{1/2}(x)dx \right)^{1/2} \\ &= \arg\sup_{b\in\Theta} \int \hat{f}^{1/2}(b;x) f_\varepsilon^{1/2}(x)dx \end{aligned} \tag{2.5}$$

The last equality holds since $\hat{f}(b;\cdot)$ and f_ε are p.d.f.s, i.e. $\int \hat{f}(b;x)dx = 1$ and $\int f_\varepsilon(x)dx = 1$. Also note that the 'arg inf' and 'arg sup' in (2.5) can be safely replaced with 'arg min' and 'arg max' respectively simply due to the fact that the parameter space Θ is compact and thus the infimum and supremum are achievable within Θ.

By the definition (2.5), the MHDE $\hat{\beta}_{MHD}$ of β is the value of b which minimizes the Hellinger distance between the assumed innovation distribution f_ε and its kernel nonparametric estimator under the model (2.1), or equivalently the value that makes the distribution of $\nu_{b,t}$'s as close to f_ε as possible.

Remark 2.1. The nonparametric kernel density estimator $\hat{f}(b;\cdot)$ given in (2.4) contains the parameter b to be estimated while the underlying true density function is free of parameters. This is different from the original definition of MHDE given in (1.3).

Remark 2.2. In economics, we are not only interested in the estimation of the parameter β but also the innovation distribution f_ε. If f_ε is unknown, then $\hat{f}(\hat{\beta}_{MHD};\cdot)$ can serve as a reasonable empirical estimate of the innovation distribution based on the observed data X_t's. This aspect is further pursued in Section 2.2.

2.1.2 Graphical demonstration of MHDE

In this part, we use graphs to demonstrate visually the mechanism of the MHDE defined in (2.5). For demonstration purposes, we consider the ARCH(1) model given in (1.1) with the true innovation distribution f_ε being t_4 and the true parameter values $\beta = (1, 0.8)^\top$.

We first examine whether the proposed MHDE in (2.5) is reasonable or not. For a single sample of size $n = 1000$ and different values of $b = (b_0, b_1)^\top$, we calculate $\|b - \beta\|$ and the Hellinger distance $\|\hat{f}^{1/2}(b;x) - f_\varepsilon^{1/2}(x)\|$ and then plot them on x-axis and y-axis, respectively, in Figure 2.1. We take $b_0 = 0.6 + 0.04k$ and $b_1 = 0.6 + 0.02k$ with $k = 0, 1, \cdots, 20$. Note that the true β value is achieved when $k = 10$. There are two curves drawn in the figure, the blue one is for $0 \le k \le 10$ and the red one is for $10 \le k \le 20$. As can be seen from it, the L^2 distance between $f_\varepsilon^{1/2}(x)$ and $\hat{f}^{1/2}(b;x)$ is an increasing function of the L^2 distance between b and β. As a consequence, the minimizer $\hat{\beta}_{MHD}$ given in (2.5) is expected to be fairly close to the true value β since minimizing (2.5) is equivalent to minimizing the L^2 distance between b and β.

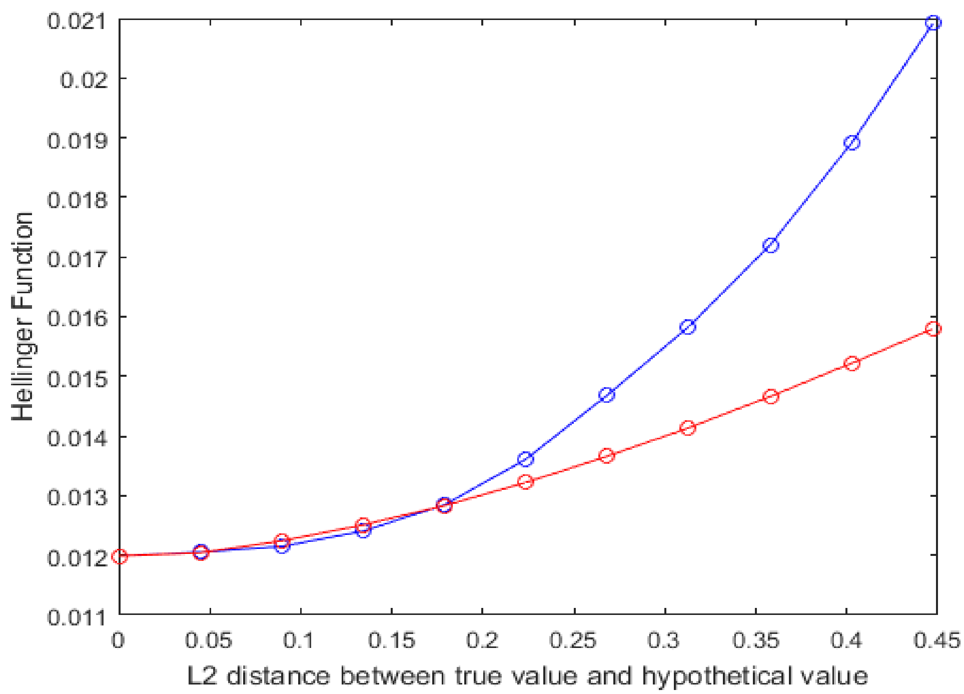


Figure 2.1: The Hellinger distance $\|\hat{f}^{1/2}(b;x) - f_\varepsilon^{1/2}(x)\|$ as a function of $\|b-\beta\|$.

As we were not able to derive theoretically the asymptotic distribution of the MHDE, we use QQ-plot to examine its distribution. In addition to considering the true innovation distribution being t_4, we also examine the contaminated distribution $0.95t_4 + 0.05\chi^2_{(2)}$ for comparison purposes. To obtain the QQ-plot presented in Figure 2.2, we run 500 simulations each with a sample size of $n = 3000$. From Figure 2.2 we can see that the large-sample distribution of the MHDE doesn't appear to be normal when there is no contamination. Rather, the distribution exhibits heavier tails than normal distribution. In contrast, the MLE appears to be normally distributed when there is no contamination, which is a proved property of MLE under regularity conditions. Nevertheless, when there is contamination, the distribution of MLE deviates from normal while that of MHDE appears to be less heavy-tailed.

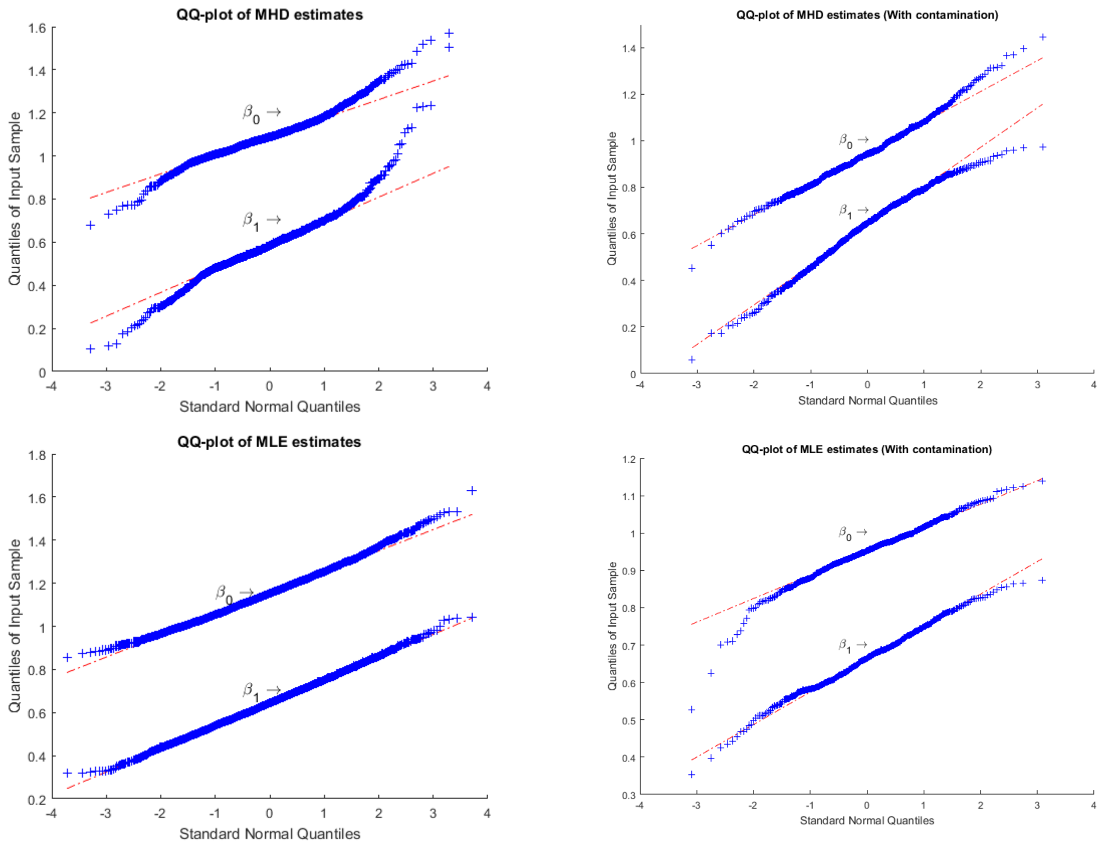


Figure 2.2: QQ-plots of the MHDE and the MLE.

Next we examine the estimated innovation density $\hat{f}^{1/2}(\hat{\beta}_{MHD};x)$, which is based on the kernel density estimator in (2.4). For a single sample of size $n = 50, 200, 500, 1000, 3000, 5000$, we first estimate the MHDE, then plug it into (2.3) to calculate the estimated residuals, and finally plug the estimated residuals into (2.4) to obtain the estimated innovation density $\hat{f}^{1/2}(\hat{\beta}_{MHD};x)$. The results are presented in Figure 2.3. Note that when using (2.3) we should set $b_2 = 0$ as we are considering ARCH(1) model. We can see from the figure that the estimated innovation distribution of ε_t's is getting increasingly closer to the true distribution f_ε as the sample size increases, which suggests that the MHDE of β is getting closer and closer to the true parameter values. The justification of this can be found in the Appendix.

This testifies to the consistency of MHDE to a certain degree. Figure 2.3 also indicates that the convergence rate of MHDE to the true parameter values might be relatively slow and the sample size needs to be at least 500 in order to obtain accurate parameter estimates. But this requirement of relatively large sample size is usually not a problem in financial time series as daily or even hourly data are usually available.

2.1.3 Review of MLE

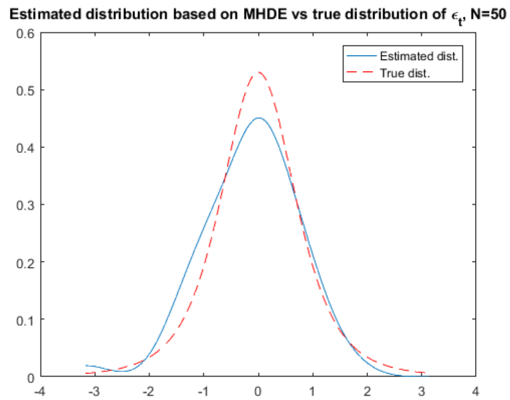
MLE is the most commonly used method in estimating ARCH/GARCH model parameters due to its simplicity. More importantly, it is consistent and the most efficient estimator if the distribution is correctly specified. In general, to estimate the parameters using maximum likelihood, we form a likelihood function, which is essentially a joint p.d.f.. But instead of treating it as a function of the data given the set of parameters, i.e. $f(x_1, x_2, \ldots, x_n|\beta)$, we think of the likelihood function as a function of the parameters given the data, $L(\beta|x_1, x_2, \ldots, x_n)$, and maximize the likelihood function with respect to the parameters.

In the ARCH/GARCH model framework, suppose ε_t has a known p.d.f. $f_\varepsilon(x)$. Since $X_t = \sigma_t \varepsilon_t$, by simple calculation of the p.d.f. of transformed r.v.s, the conditional p.d.f. of X_t, given all the information $\mathscr{F}_{t-1}$ up to $t-1$, is then

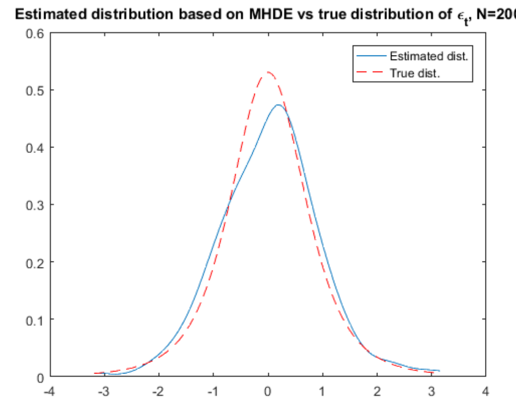
$$f_{X_t|\beta,\mathscr{F}_{t-1}}(x) \;=\; \frac{1}{\sigma_t} f_\varepsilon\left(\frac{x}{\sigma_t}\right),$$

where $\mathscr{F}_{t-1}$ is essentially the information about the returns up to $t-1$ and β is the model parameters to be estimated that is contained in σ_t.

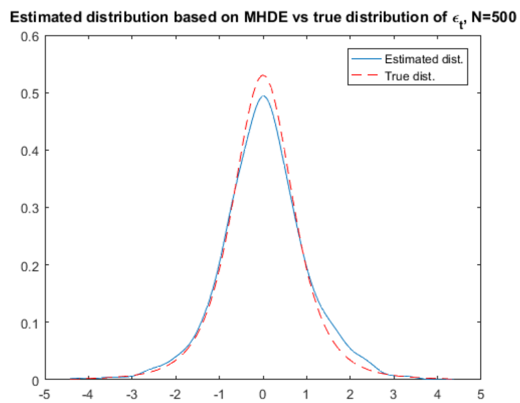
If the returns X_t's were independent of each other, we could write the joint density function of $X_1, X_2, \ldots, X_n$ or equivalently the likelihood function as the product of the marginal densities. However, the returns X_t are clearly not independent. In spite of this, we can still



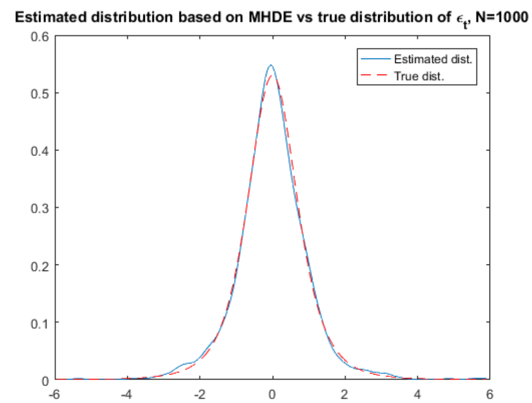
(a) n=50



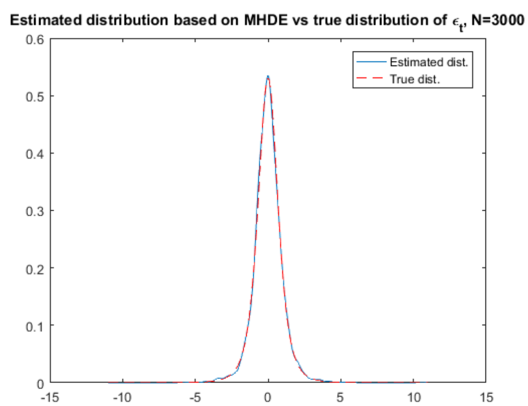
(b) n=200



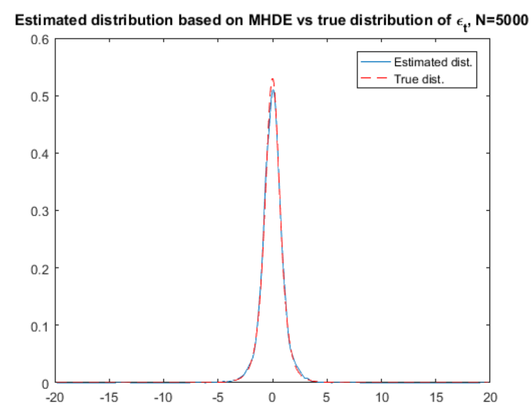
(c) n=500



(d) n=1000



(e) n=3000



(f) n=5000

Figure 2.3: Estimated innovation distribution $\hat{f}^{1/2}(\hat{\beta}_{MHD};x)$ v.s. the true innovation distribution.

write the joint p.d.f. as the product of conditional ones, i.e.

$$\begin{aligned}
f_{X_1,\ldots,X_n|\beta}(x_1,\ldots,x_n) &= f_{X_n|\beta,X_1,\ldots,X_{n-1}}(x_n) f_{X_1,\ldots,X_{n-1}|\beta}(x_1,\ldots,x_{n-1}) \\
&= f_{X_n|\beta,X_1,\ldots,X_{n-1}}(x_n) f_{X_{n-1}|\beta,X_1,\ldots,X_{n-2}}(x_{n-1}) f_{X_1,\ldots,X_{n-2}|\beta}(x_1,\ldots,x_{n-2}) \\
&\vdots \\
&= f_{X_n|\beta,X_1,\ldots,X_{n-1}}(x_n) f_{X_{n-1}|\beta,X_1,\ldots,X_{n-2}}(x_{n-1}) \cdots f_{X_2|\beta,X_1}(x_2) f_{X_1|\beta}(x_1).
\end{aligned}$$

Then with $f_{X_1|\beta,\mathscr{F}_0} = f_{X_1|\beta}$, the likelihood function is

$$L(\beta|x_1,x_2,\ldots,x_n) \;=\; \prod_{t=1}^{n} f_{X_t|\beta,\mathscr{F}_{t-1}}(x_t) \;=\; \prod_{t=1}^{n} \frac{1}{\sigma_t} f_\varepsilon\left(\frac{x_t}{\sigma_t}\right).$$

It is usually more convenient to work with the log-likelihood function as follows

$$l(\beta|x_1,x_2,\ldots,x_n) \;=\; \log L(\beta|x_1,x_2,\ldots,x_n) \;=\; \sum_{t=1}^{n} \log f_\varepsilon\left(\frac{x_t}{\sigma_t}\right) - \sum_{t=1}^{n} \log \sigma_t.$$

As an example, consider the GARCH$(1,1)$ model given in (2.1) and assume ε_t follows a Student's t distribution with variance normalized to one. If we use f_d to denote the t distribution with degrees of freedom d, then its variance is $\frac{d}{d-2}$ and the p.d.f. of ε_t is then $\sqrt{\frac{d}{d-2}} f_d\left(\sqrt{\frac{d}{d-2}}x\right)$. Now the log-likelihood function given $X_1,X_2,\ldots,X_n$ is

$$\begin{aligned}
l(\beta|x_1,x_2,\ldots,x_n) &= \sum_{t=1}^{n} \log \sqrt{\frac{d}{d-2}} f_d\left(\frac{x_t}{\sigma_t}\sqrt{\frac{d}{d-2}}\right) - \sum_{t=1}^{n} \log \sigma_t \\
&= \sum_{t=1}^{n} \log f_d\left(\frac{x_t}{\sigma_t}\sqrt{\frac{d}{d-2}}\right) + \sum_{t=1}^{n} \log \sqrt{\frac{d}{d-2}} - \sum_{t=1}^{n} \log \sigma_t.
\end{aligned}$$

The properties of MLE on ARCH/GARCH model has been studied extensively in literature; see, for example, Gourieroux et al. [13], Weiss [34] and Lee and Hansen [20] among many others.

2.2. Innovation Distribution is Unknown

In reality, it is very likely that the innovation distribution is unknown. Thus, estimation methods that require minimal distribution assumptions are highly sought-after. In this section, we will explore the scenario where the innovation distribution is unknown. Note that when the innovation distribution is unknown, the ARCH and GARCH models are semi-parametric models since the models involves both a finite dimensional unknown parameters and infinite dimensional unknown functions. In Section 2.2.1, we propose and construct a MPHDE of the parameters in ARCH/GARCH models. In Section 2.2.2, we discuss the properties of the proposed MPHDE using graphs. For comparison purposes, in Section 2.2.3 we review the commonly used methods, which are GQMLE and NGQMLE.

2.2.1 Construction of MPHDE

In this section, we generalize the MHDE defined in Section 2.1.1 to accommodate the case of unknown innovation distribution. Throughout this section, we assume that the innovation distribution f_ε in (2.1) is symmetric but otherwise unspecified. The symmetry assumption here is essential as it allows us to derive the MPHDE without too much difficulties. We will follow the idea of the MPHDE first introduced by Wu and Karunamuni [40] to construct the MPHDE for the GARCH$(1,1)$ model given in (2.1) by first profiling out the unknown innovation density f_ε. With the profiled innovation density, we then minimize over the parameter space to obtain the MPHDE of β.

For the kernel density estimator $\hat{f}(b;x)$ given in (2.4), note that

$$\hat{f}^{\frac{1}{2}}(b;x) \;=\; \frac{1}{2}\left(\left[\hat{f}^{\frac{1}{2}}(b;x)+\hat{f}^{\frac{1}{2}}(b;-x)\right]+\left[\hat{f}^{\frac{1}{2}}(b;x)-\hat{f}^{\frac{1}{2}}(b;-x)\right]\right)$$

and that the first term on the right-hand side of the above expression is an even function of x

while the second is an odd function of x. Let g be any even p.d.f., then

$$
\begin{aligned}
& \left\| \hat{f}^{\frac{1}{2}}(b;x) - g^{\frac{1}{2}}(x) \right\|^2 \\
= & \int \left[\hat{f}^{\frac{1}{2}}(b;x) - g^{\frac{1}{2}}(x) \right]^2 dx \\
= & \int \hat{f}(b;x)dx + \int g(x)dx - 2\int \hat{f}^{\frac{1}{2}}(b;x) g^{\frac{1}{2}}(x)dx \\
= & 1 + 1 - 2\int \frac{1}{2}\left[\hat{f}^{\frac{1}{2}}(b;x) + \hat{f}^{\frac{1}{2}}(b;-x) \right] g^{\frac{1}{2}}(x)dx - 2\int \frac{1}{2}\left[\hat{f}^{\frac{1}{2}}(b;x) - \hat{f}^{\frac{1}{2}}(b;-x) \right] g^{\frac{1}{2}}(x)dx \\
= & 2 - \int \left[\hat{f}^{\frac{1}{2}}(b;x) + \hat{f}^{\frac{1}{2}}(b;-x) \right] g^{\frac{1}{2}}(x)dx.
\end{aligned}
\tag{2.6}
$$

The last equality of (2.6) holds since the product of an even function and an odd function is an odd function which integrates to zero on the real line. When we try to minimize the Hellinger distance on the left hand side of (2.6), we only need to maximize the second term on the right hand side. By the Cauchy-Schwarz inequality, this quantity is maximized when

$$
\frac{g^{\frac{1}{2}}(x)}{\|g^{\frac{1}{2}}(x)\|} = \frac{\hat{f}^{\frac{1}{2}}(b;x) + \hat{f}^{\frac{1}{2}}(b;-x)}{\|\hat{f}^{\frac{1}{2}}(b;x) + \hat{f}^{\frac{1}{2}}(b;-x)\|},
$$

that is $g^{\frac{1}{2}}(x) = [\hat{f}^{\frac{1}{2}}(b;x) + \hat{f}^{\frac{1}{2}}(b;-x)]/\|\hat{f}^{\frac{1}{2}}(b;x) + \hat{f}^{\frac{1}{2}}(b;-x)\|_2$ since $\|g^{\frac{1}{2}}(x)\|^2 = \int g(x)dx = 1$. As a result, for all even p.d.f. g, $\|\hat{f}^{\frac{1}{2}}(b;x) - g^{\frac{1}{2}}(x)\|$ is minimized at

$$
g(x) = \frac{\left[\hat{f}^{\frac{1}{2}}(b;x) + \hat{f}^{\frac{1}{2}}(b;-x) \right]^2}{\left\| \hat{f}^{\frac{1}{2}}(b;x) + \hat{f}^{\frac{1}{2}}(b;-x) \right\|^2}.
$$

Plug the above value of $g(x)$ into (2.6), we obtain the profile Hellinger distance function

$$
H(b) = 2 - \left\| \hat{f}^{\frac{1}{2}}(b;x) + \hat{f}^{\frac{1}{2}}(b;-x) \right\|.
$$

Now we only need to minimize $H(b)$ over $b \in \Theta$, or equivalently maximize $\|\hat{f}^{\frac{1}{2}}(b;x) + \hat{f}^{\frac{1}{2}}(b;-x)\|$ over $b \in \Theta$. The MPHDE of β is thus defined as

$$
\hat{\beta}_{MPHD} = \underset{b\in\Theta}{\arg\max} \left\| \hat{f}^{\frac{1}{2}}(b;x) + \hat{f}^{\frac{1}{2}}(b;-x) \right\|. \tag{2.7}
$$

Wu and Karunamuni [40] has proved that the MPHDE is in general efficient and asymptotically normally distributed. However, the definition of MPHDE here is a bit different from the general one proposed in Wu and Karunamuni [40] in the sense that the latter has non-parametric density estimation and unknown parameters separated while the former has them combined in the same function due to the non-i.i.d. data.

2.2.2 Graphical demonstration of MPHDE

Since we were not able to study theoretically the asymptotic properties of the proposed MPHDE in (2.7) due to the complexity of the ARCH/GARCH model, we use graphs to demonstrate its properties in this part. We consider the same setting as that in Section 2.1.2, i.e. ARCH(1) with f_ε being t_4 and $\beta = (1, 0.8)^\top$. Figure 2.4, 2.5 and 2.6 are constructed in the same way as for Figure 2.1, 2.2 and 2.3 respectively, but for the MPHDE defined in (2.7) instead of the MHDE defined in (2.5). We observe similar phenomena in Figure 2.4, 2.5 and 2.6 to those in Figure 2.1, 2.2 and 2.3 respectively.

Figure 2.4 displays the objective function $\|\hat{f}^{\frac{1}{2}}(b;x) + \hat{f}^{\frac{1}{2}}(b;-x)\|$ given in (2.7) as a function of the L^2 distance between b and the true values β for a single sample size of $n = 1000$. The blue curve is for $0 \le k \le 10$ while the red curve is for $10 \le k \le 20$. From Figure 2.4 we can see that the objective function is a decreasing function of the L^2 distance between b and β, and in turn the profile Hellinger distance function is an increasing function of the L^2 distance between b and β. Theoretically the objective function can achieve the maximum value of 2 and Figure 2.4 does show a maximum close to 2. This indicates that maximizing the objective function in (2.7) is equivalent to minimizing the L^2 distance between b and β. Thus, the maximizer $\hat{\beta}_{MPHD}$ given in (2.7) will be very close to the true value β intuitively. In other words, the MPHDE is expected to be consistent even though we are not able to prove it theoretically.

Further, we use the QQ-plot presented in Figure 2.5 to examine the distribution of the MPHDE. From Figure 2.5 we can see that the large-sample ($n = 3000$) distribution of the MPHDE doesn't appear to be normal when there is no contamination, and instead the distri-

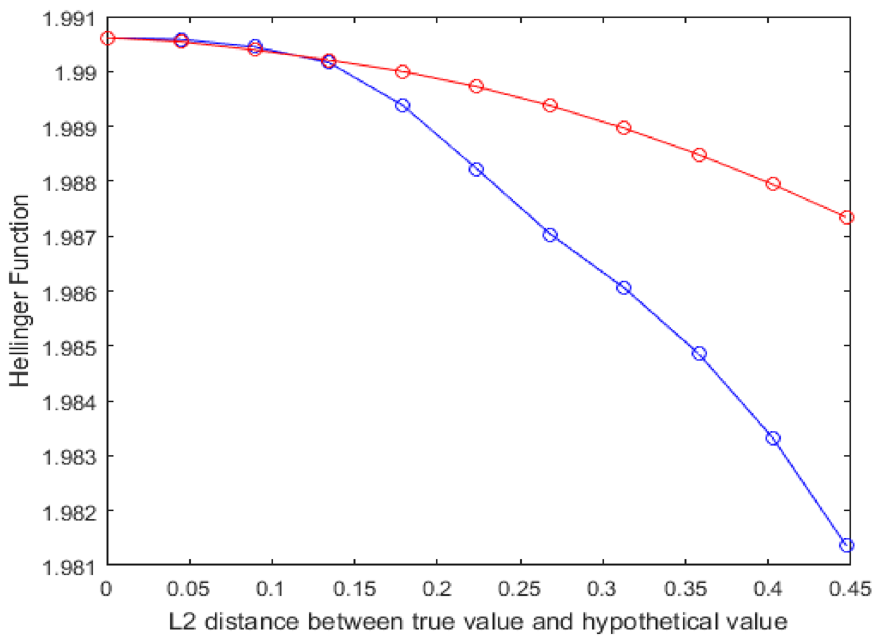


Figure 2.4: The Hellinger function $\|\hat{f}^{\frac{1}{2}}(b;x)+\hat{f}^{\frac{1}{2}}(b;-x)\|$ as a function of $\|b-\beta\|$.

bution exhibits heavier tails than normal distribution, espeically on the right tail. In contrast, the GQMLE and NGQMLE appear to be normally distributed, which is a proved property in literature of GQMLE and NGQMLE. However, in the case of contamination, the large-sample distribution of MPHDE appears to have less pronounced heavy tails. Whereas, the distribution of GQMLE and NGQMLE now have heavy tails and deviate from normal. The details of the GQMLE and NGQMLE are deferred to Section 2.2.3.

Figure 2.6 examines the estimated innovation function $\hat{f}^{1/2}(\hat{\beta}_{MPHD};x)$, which is also based on the kernel density esitmator in (2.4). We can see from the figure that the estimated innovation density of ε_t's is getting closer and closer to the true distribution f_ε as the sample size increases, which indicates that the MPHDE of β is getting increasingly closer to the true parameter values. This gives evidence to the consistency of MPHDE to some extent. Figure 2.6 also indicates that the convergence rate of the MPHDE to the true parameter value might be relatively slow as well and the sample size needs to be at least 500 in order to obtain an

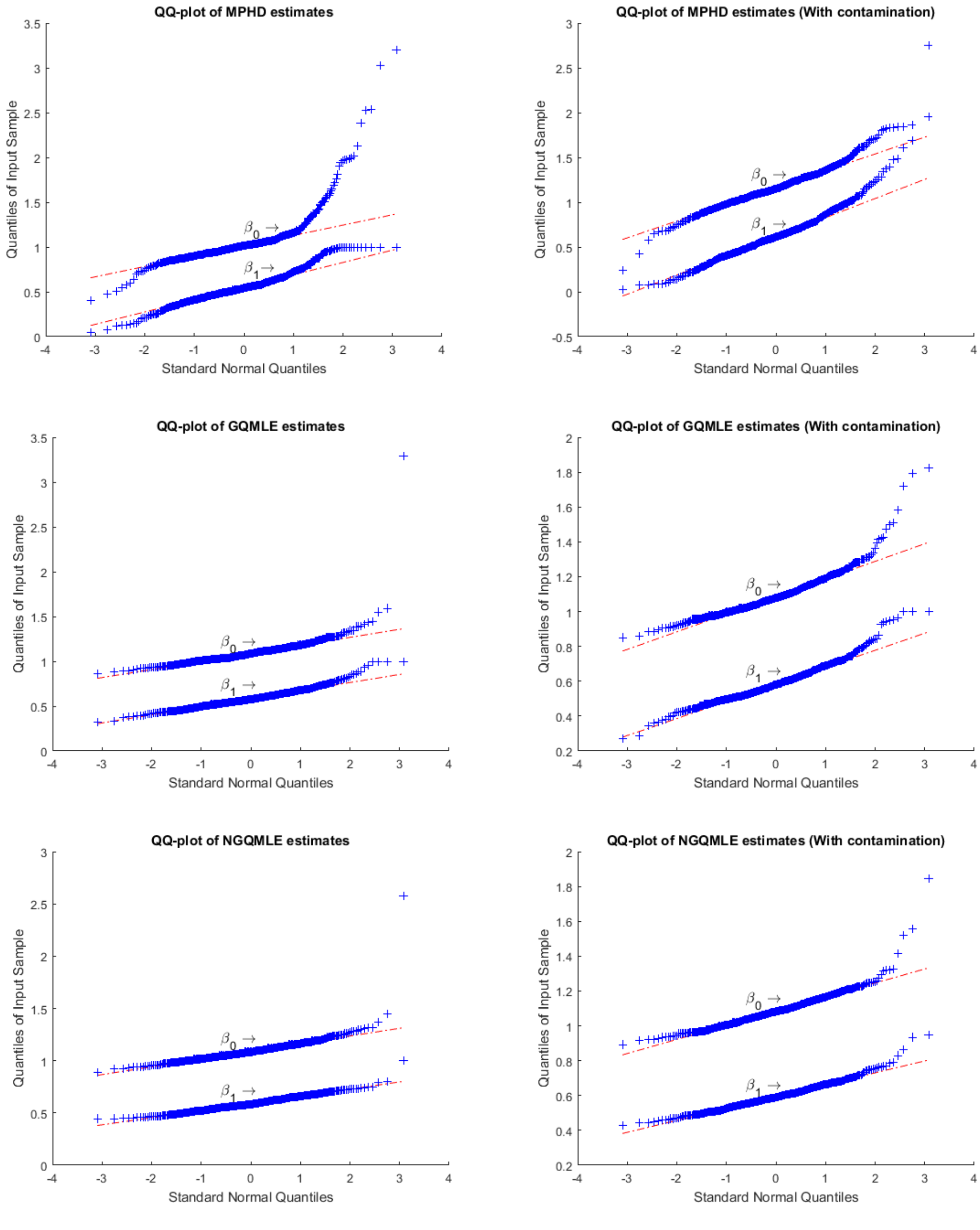


Figure 2.5: QQ-plots of the MPHDE, GQMLE and NGQMLE.

accurate MPHDE.

2.2.3 Review of GQMLE and NGQMLE

Due to its simplicity, GQMLE is the most popular method for estimating ARCH/GARCH models when the innovation distribution is unknown. GQMLE simply use the Gaussian function as the innovation distribution regardless of the true innovation distribution. GQMLE has been shown to be consistent provided that the innovation has a finite fourth moment. Despite its simplicity, GQMLE may suffer from considerable efficiency loss if the true innovation distribution is far from normal. To improve the efficiency, Fan et al. [11] suggested using the heavy-tailed likelihood function with a scale parameter η_f, which serves to correct the inconsistency of using non-Gaussian likelihood functions. Specifically, they constructed the likelihood function based on the following three-step procedure. In the first step, the GQMLE is calculated to obtain a reasonable estimate of the innovation ε_t, which will in turn be used to estimate the scale parameter η_f. Here the GQMLE is given by

$$\begin{aligned}\hat{\beta}_G &= \arg\max_{b\in\Theta} \frac{1}{n}\sum_{t=1}^{n} l_1(b;x_t) \\ &= \arg\max_{b\in\Theta} \frac{1}{n}\sum_{t=1}^{n}\left(-\log(\sigma_t) - \frac{x_t^2}{2\sigma_t^2}\right)\end{aligned} \tag{2.8}$$

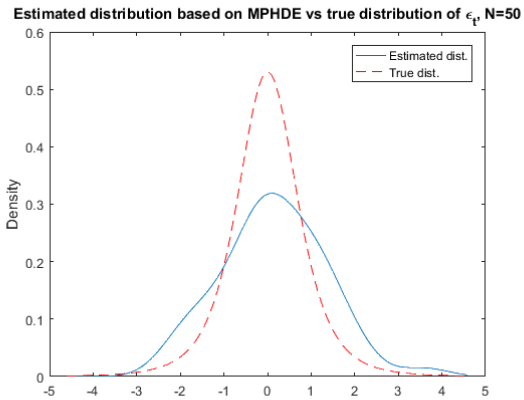
where l_1 is the likelihood function based on normal distribution. In the second step, an estimate of η_f, denoted by $\hat{\eta}_f$, is obtained by performing the following maximization with estimated residuals from the first step:

$$\begin{aligned}\hat{\eta}_f &= \arg\max_{\eta} \frac{1}{n}\sum_{t=1}^{n} l_2\left(\eta,\hat{\beta}_G;x_t\right) \\ &= \arg\max_{\eta} \frac{1}{n}\sum_{t=1}^{n}\left[-\log(\eta)+\log f\left(\frac{\tilde{\varepsilon}_t}{\eta}\right)\right],\end{aligned}$$

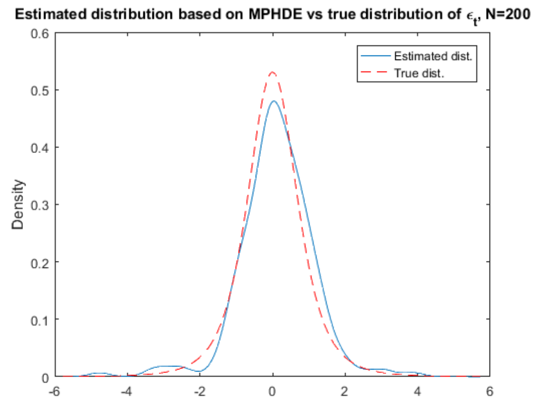
where $\tilde{\varepsilon}_t = x_t/\tilde{\sigma}_t$ are the estimated residuals based on the GQMLE $\hat{\beta}_G$ calculated in the first step and $\tilde{\sigma}_t$ is the σ_t value in (2.1) with β replaced by $\hat{\beta}_G$. Finally, the NGQMLE $\hat{\beta}_{NG}$ is defined as the maximizer of the non-Gaussian quasi-likelihood with plug-in $\hat{\eta}_f$, i.e.

$$\begin{aligned}\hat{\beta}_{NG} &= \underset{b\in\Theta}{\arg\max}\frac{1}{n}\sum_{t=1}^{n} l_3\left(\hat{\eta}_f, b; x_t\right) \\ &= \underset{b\in\Theta}{\arg\max}\frac{1}{n}\sum_{t=1}^{n}\left[-\log(\hat{\eta}_f\sigma_t)+\log f\left(\frac{x_t}{\hat{\eta}_f\sigma_t}\right)\right],\end{aligned} \tag{2.9}$$

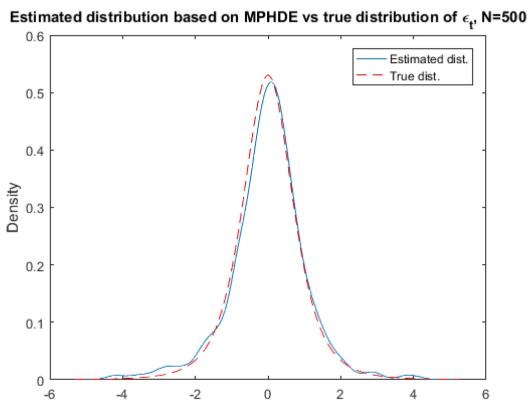
where f is a heavy-tailed likelihood function (we use t_4 in our simulation studies). Fan et al. [11] showed that NGQMLE generally outperforms GQMLE in terms of efficiency, particularly when the innovation distribution is heavy-tailed. However, the authors didn't investigate the robustness properties of NGQMLE. We will show in the next chapter that both GQMLE and NGQMLE are quite sensitive to outlier contamination, which is the common deficiency of MLE-type estimators.



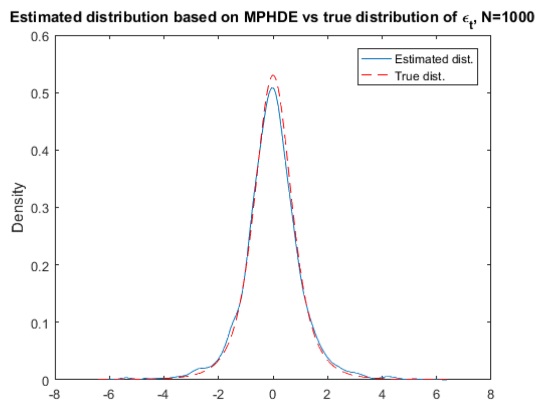
(a) $n = 50$



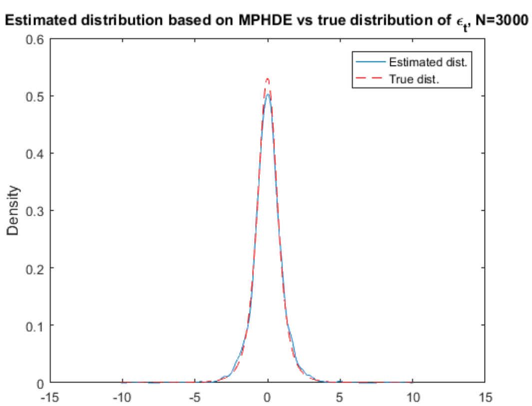
(b) $n = 200$



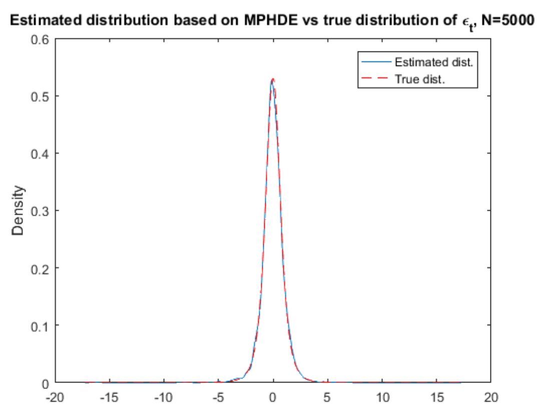
(c) $n = 500$



(d) $n = 1000$



(e) $n = 3000$



(f) $n = 5000$

Figure 2.6: Estimated innovation distribution $\hat{f}^{1/2}(\hat{\beta}_{MPHD};x)$ v.s. the true innovation distribution.

Chapter 3

SIMULATION STUDIES

In this chapter, we illustrate the finite-sample performance of the proposed MHDE and MPHDE, particularly for ARCH(1) and GARCH$(1,1)$ models, through Monte Carlo simulation. To check the robustness properties, we examine the performance of both MHDE and MPHDE when data is contaminated and not contaminated. Specifically, for ARCH(1) model, we consider the following four scenarios: (i) the innovation distribution f_ε is known and data is not contaminated (easiest case); (ii) f_ε is known and data is contaminated; (iii) f_ε is unknown and data is not contaminated; (iv) f_ε is unknown and data is contaminated (hardest case). For GARCH(1,1) model, we only consider the latter two harder cases as it is more relevant to the real data analysis. The proposed MHDE is compared with MLE When f_ε is known, while the proposed MPHDE is compared with GQMLE and NGQMLE when f_ε is unknown.

A variety of the innovation distributions are considered in our simulation studies. Specifically, when data is not contaminated, we consider Student's t distribution with various degrees of freedom and generalized normal distribution with different shape parameters for f_ε. All these distributions are standardized to have mean zero and variance one so as to align with the assumption of ARCH/GARCH models. In the case of data contamination, we consider generalized normal distribution with contaminating data from uniform or chi-square distribution.

The generalized normal distribution, also known as the exponential power distribution, is a parametric family of symmetric distributions that are governed by three parameters, namely the location parameter μ, the scale parameter α and the shape parameter d. Throughout the thesis, we use gg_d to denote the generalized normal p.d.f. with shape parameter d.

This family includes the normal distribution when $d = 2$ (with mean μ and variance $\frac{\alpha^2}{2}$) and the Laplace distribution when $d = 1$. As $d \to \infty$, the p.d.f. gg_d converges pointwise to a uniform density on $(\mu - d, \mu + d)$. This family allows for tails that are either heavier than normal (when $d < 2$) or lighter than normal (when $d > 2$).

For each type of innovation distribution, we run $N = 500$ simulations each with sample size ranging among $250, 500$ and 1000. All the simulations are done in MATLAB and we use the build-in function 'fmincon' to solve the associated optimization problems.

Kernel density estimation is used in the construction of both MHDE and MPHDE. In our simulation, we use the compact-supported Epanechnikov kernel function

$$K(x) \;=\; \frac{3}{4}(1-x^2)I_{[-1,1]}(x).$$

Moreover, data-driven bandwidths are employed to reduce the finite-sample bias and MSE of both MHDE and MPHDE. Specifically we use $h_n = S_n n^{-1/3}$ with S_n the robust scale estimator

$$S_n \;=\; 1.1926 \operatorname*{med}_i \left[\operatorname*{med}_j |\hat{\varepsilon}_i - \hat{\varepsilon}_j| \right]$$

proposed by [28], where

$$\hat{\varepsilon}_i \;=\; \frac{X_i}{\hat{\sigma}_i} \;=\; \frac{X_i}{\sqrt{\hat{\beta}_0 + \hat{\beta}_1 X_{i-1}^2 + \hat{\beta}_2 \hat{\sigma}_{i-1}^2}}$$

is the estimated residuals based on $\hat{\beta}_{GQMLE}$.

The performance of estimators is assessed by the estimated bias and MSE given by

$$\widehat{Bias} \;=\; \frac{1}{N}\sum_{i=1}^{N}(\hat{\beta}^i - \beta),$$

$$\widehat{MSE} \;=\; \frac{1}{N}\sum_{i=1}^{N}(\hat{\beta}^i - \beta)^2,$$

where $\hat{\beta}^i$ is the estimate $\hat{\beta}$ in the i^{th} simulation repetition. In addition to the bias and MSE of the estimators, we also report the coverage probability (CP) that is calculated using bootstrap. Specifically, for a generated single sample of a particular size n, we first calculate the estimate $\hat{\beta}$ of the model parameters. Then 100 bootstrapping samples are sampled with replacement from the original single sample. With the 100 bootstrapping samples, we can calculate 100 (bootstrapping) estimates $\hat{\beta}_{bs}$ and their sample standard deviation $SD(\hat{\beta}_{bs})$. Finally the level $1-\alpha$ confidence interval of β is obtained by $\hat{\beta} \pm z_{\alpha/2} SD(\hat{\beta}_{bs})$, where $z_{\alpha/2}$ is the upper $\alpha/2$-quantile of standard normal distribution. We use $\alpha = 5\%$ in our simulation. The coverage probability is simply the number of times that the true β value falls into the above constructed bootstrapping confidence interval divided by $N = 500$. We look at the performance of the estimators for ARCH(1) and GARCH(1) models in Section 3.1 and Section 3.2 respectively.

3.1. ARCH(1)

Consider ARCH(1) model given in (2.1) with $\beta_2 = 0$. Note that β_0 is the lower bound of the conditional and unconditional variance of X_t's, and β_1 measures the rate at which a shock in returns X_t today feeds into next period's volatility dies over time. In other words, any large (in absolute value) shock in ε_t will be associated with a persistently large (conditional) variance in the returns X_t. The larger the β_1, the longer the persistence of this effect. Theoretically, we need $\beta_0 > 0$ and $0 < \beta_1 < 1$ to ensure the stability of ARCH(1) model. As a result, we impose the constraints $\beta_0 > 0$ and $0 \le \beta_1 \le 1$ in our simulation studies, and we take the true parameter values to be $\beta = (\beta_0, \beta_1)^\top = (1, 0.7)^\top$.

Case 1: known innovation distribution

Under ARCH(1) model with $\beta = (1, 0.7)^\top$ and known innovation distribution, we compare the performance of the proposed MHDE given in (2.5) with the MLE by examining their biases, MSE and CP. The results are presented in Tables 3.1 and 3.2 for β_0 and β_1 respectively. Note that for t distribution, the smaller the degrees of freedom, the heavier the tail of the distribution. For generalized normal distribution gg_d, the smaller the shape parameter d, the heavier the tail of the distribution ($d < 2$ heavy-tailed; $d > 2$ light-tailed).

From Tables 3.1 and 3.2 we can see that, no matter which innovation distribution and which sample size is considered, MLE always has smaller MSE than those of MHDE. This is not surprising as MLE is the most efficient estimator under regularity conditions. MLE also has smaller bias in most cases than the MHDE. Moreover, MLE and MHDE have close performance in terms of CP, with both close to the nominal level 95% although the latter has a little bit lower CP in some cases. In summary, the MLE is more efficient and thus preferred over the MHDE when the data is not contaminated.

From Tables 3.1 and 3.2 we also observe that the bias and MSE of both the MLE and MHDE tend to increase and the CP tends to decrease, as the kurtosis of innovation distribution grows (i.e. the innovation distribution has heavier tail). This is probably due to the increase of the probability of extreme values as the tail gets heavier. Also as expected, when sample size n increases, the performance of both MLE and MHDE improves.

Table 3.1: $\hat{\beta}_{0,MLE}$ and $\hat{\beta}_{0,MHD}$ for ARCH(1) with known innovation distribution.

		$\hat{\beta}_{0,MLE}$		$\hat{\beta}_{0,MHD}$	
f_ε	n	Bias (MSE)	CP	Bias (MSE)	CP
t_{20}	250	0.023 (0.069)	0.92	0.027 (0.076)	0.91
	500	0.009 (0.026)	0.93	0.017 (0.030)	0.93
	1000	0.006 (0.006)	0.94	0.014 (0.008)	0.94
t_9	250	0.029 (0.071)	0.91	0.027 (0.082)	0.91
	500	0.021 (0.027)	0.90	0.025 (0.033)	0.90
	1000	0.009 (0.010)	0.93	0.015 (0.013)	0.92
t_6	250	0.024 (0.080)	0.92	0.028 (0.089)	0.92
	500	0.019 (0.028)	0.93	0.028 (0.033)	0.93
	1000	-0.015 (0.012)	0.94	-0.019 (0.014)	0.93
t_4	250	0.028 (0.089)	0.93	0.033 (0.098)	0.92
	500	0.021 (0.034)	0.93	0.023 (0.040)	0.92
	1000	0.021 (0.015)	0.90	0.019 (0.019)	0.90
t_3	250	0.035 (0.096)	0.89	0.039 (0.108)	0.88
	500	0.020 (0.040)	0.91	0.024 (0.047)	0.92
	1000	0.017 (0.017)	0.92	0.016 (0.020)	0.91
gg_4	250	0.007 (0.036)	0.95	0.010 (0.042)	0.96
	500	0.005 (0.010)	0.94	0.007 (0.014)	0.94
	1000	0.005 (0.004)	0.95	0.006 (0.006)	0.94
gg_2	250	0.011 (0.058)	0.94	0.014 (0.068)	0.94
	500	0.010 (0.019)	0.93	0.012 (0.022)	0.93
	1000	0.007 (0.006)	0.95	0.010 (0.008)	0.92
gg_1	250	0.016 (0.086)	0.96	-0.019 (0.095)	0.97
	500	0.018 (0.029)	0.90	-0.016 (0.034)	0.91
	1000	0.009 (0.013)	0.92	-0.013 (0.016)	0.91
$gg_{0.5}$	250	-0.034 (0.123)	0.90	-0.042 (0.131)	0.88
	500	-0.024 (0.040)	0.92	-0.028 (0.048)	0.91
	1000	-0.015 (0.019)	0.92	-0.015 (0.021)	0.92

Table 3.2: $\hat{\beta}_{1,MLE}$ and $\hat{\beta}_{1,MHD}$ for ARCH(1) with known innovation distribution.

		$\hat{\beta}_{1,MLE}$		$\hat{\beta}_{1,MHD}$	
f_ε	n	Bias (MSE)	CP	Bias (MSE)	CP
t_{20}	250	-0.013 (0.046)	0.92	-0.010 (0.056)	0.91
	500	-0.007 (0.017)	0.94	-0.008 (0.020)	0.93
	1000	-0.005 (0.004)	0.94	-0.005 (0.006)	0.93
t_9	250	-0.019 (0.059)	0.92	-0.015 (0.065)	0.91
	500	-0.014 (0.022)	0.93	-0.012 (0.026)	0.92
	1000	-0.015 (0.007)	0.93	-0.016 (0.009)	0.93
t_6	250	-0.025 (0.071)	0.94	-0.023 (0.077)	0.96
	500	-0.014 (0.024)	0.93	-0.015 (0.028)	0.93
	1000	-0.016 (0.010)	0.92	-0.015 (0.011)	0.92
t_4	250	-0.024 (0.081)	0.91	-0.025 (0.089)	0.91
	500	-0.018 (0.031)	0.92	-0.022 (0.035)	0.91
	1000	-0.006 (0.013)	0.93	-0.017 (0.015)	0.92
t_3	250	-0.025 (0.099)	0.89	-0.037 (0.111)	0.87
	500	0.022 (0.038)	0.91	-0.028 (0.041)	0.89
	1000	-0.024 (0.016)	0.91	-0.020 (0.018)	0.90
gg_4	250	-0.013 (0.023)	0.93	-0.015 (0.030)	0.92
	500	-0.006 (0.008)	0.94	-0.009 (0.011)	0.93
	1000	-0.006 (0.003)	0.95	-0.006 (0.004)	0.96
gg_2	250	-0.022 (0.040)	0.92	-0.018 (0.048)	0.92
	500	-0.008 (0.017)	0.93	-0.011 (0.019)	0.93
	1000	-0.003 (0.004)	0.95	0.009 (0.005)	0.94
gg_1	250	-0.023 (0.082)	0.93	0.027 (0.089)	0.92
	500	-0.015 (0.023)	0.92	0.019 (0.027)	0.92
	1000	-0.006 (0.009)	0.93	0.010 (0.011)	0.92
$gg_{0.5}$	250	-0.036 (0.112)	0.90	-0.033 (0.118)	0.88
	500	-0.019 (0.038)	0.91	-0.029 (0.042)	0.89
	1000	-0.012 (0.015)	0.91	-0.017 (0.018)	0.90

Case 2: known innovation distribution and contaminated

In this part, we still consider ARCH(1) model with known innovation distribution but now the innovation distribution (further the data) is contaminated. The purpose of this case is to examine the robustness properties of the proposed MHDE. We consider two types of contamination. The first one assumes that a certain percentage of the innovation distribution is contaminated by another distribution, i.e.

$$\varepsilon_t \sim (1-\delta)f_\varepsilon + \delta h \tag{3.1}$$

for some small contamination rate $\delta > 0$ and some contaminating distribution h. The second type of contamination assumes that the distribution of a small time period $\Delta t = \lfloor \delta n \rfloor$ (small portion of consecutive observations) is completely different from the assumed innovation distribution while the rest of the observations are from the assumed innovation distribution. In another word,

$$\varepsilon_t \sim \begin{cases} f_\varepsilon, & \text{if } 0 \le t \le t_1 \text{ or } t_1 + \Delta t \le t \le n, \\ h, & \text{if } t_1 < t < t_1 + \Delta t. \end{cases} \tag{3.2}$$

where $t_1 = 0.3n$. This kind of contamination is thought to be a better depiction of the asset returns in the event of financial crisis when the distribution of returns is generally different from those at normal times. Nevertheless, we will examine both types of contamination and the contamination rate $\delta = 5\%$ in both type of contamination. The contaminating distribution is chosen to be either uniform distribution, such as $U[-0.5, 0]$, or chi-square distribution, such as $\chi^2_{(2)}$. We present the results for the generalized normal distribution, ranging from light-tailed gg_4 to heavy-tailed $gg_{0.5}$, in Table 3.3 and 3.4 for the first type of contamination (3.1). Table 3.5 and 3.6 report the results for the second type of contamination (3.2). The results for t distributions considered in Case 1 are overall quite similar and thus omitted.

From Table 3.3 and 3.4 we can see that, for contamination structure (3.1), the biases and MSEs of both MLE and MHDE are significantly increased while the CPs of them are decreased, in comparison with the corresponding uncontaminated ones. This is expected as extra noises are introduced in the data, which skews the innovation distribution and invalidates the model assumptions of zero mean and unit variance of the innovation to some degree. Despite this, both the (absolute values of) biases and MSEs of MLE and MHDE decrease as sample size increases. More importantly, MHDE now outperforms MLE in terms of bias, MSE and CP in most cases. Specifically, the (absolute values of) bias and MSE of MHDE are remarkably smaller than those of MLE while the CPs of MHDE is slightly better than those of MLE. This testifies to the robustness of MHDE that MLE is generally lacking.

Under the second type of contamination (3.2), we observe from Table 3.5 and 3.6 similar phenomena as those under the first type of contamination (3.1), in which MHDE again has a competitive edge over MLE. Even though the MHDE is more robust than MLE, its performance is still severely undermined. In this sense, the proposed MPHDE given in (2.7) is expected to be more robust than the MHDE since the former doesn't use the assumed innovation function but instead estimates it nonparametrically.

When the two types of contamination are compared, i.e. Table 3.3 and 3.4 compared with Table 3.5 and 3.6, the second type of contamination affects both the MLE and the MHDE more severely than the first type in the sense of having much larger bias, MSE and much smaller CP. This is quite understandable as under (3.1) the observation follows a distribution that is very close to the assumed one at any time, while under (3.2) the data is from a completely different distribution during a certain time period.

Table 3.3: $\hat{\beta}_{0,MLE}$ and $\hat{\beta}_{0,MHD}$ for ARCH(1) with known innovation distribution and contamination (3.1).

		$\hat{\beta}_{0,MLE}$		$\hat{\beta}_{0,MHD}$	
Contamination (3.1)	n	Bias (MSE)	CP	Bias (MSE)	CP
$0.95gg_4 + 0.05U[-0.5, 0]$	250	-0.054 (0.056)	0.88	-0.044 (0.059)	0.90
	500	-0.052 (0.030)	0.87	-0.041 (0.027)	0.89
	1000	-0.055 (0.014)	0.87	-0.034 (0.012)	0.90
$0.95gg_4 + 0.05\chi^2_{(2)}$	250	-0.058 (0.061)	0.87	-0.047 (0.061)	0.88
	500	-0.064 (0.028)	0.86	-0.045 (0.024)	0.87
	1000	-0.053 (0.013)	0.87	-0.031 (0.011)	0.87
$0.95gg_2 + 0.05U[-0.5, 0]$	250	-0.068 (0.086)	0.85	-0.050 (0.089)	0.88
	500	-0.062 (0.040)	0.86	-0.047 (0.035)	0.86
	1000	-0.059 (0.017)	0.85	-0.039 (0.014)	0.86
$0.95gg_2 + 0.05\chi^2_{(2)}$	250	-0.055 (0.086)	0.86	-0.048 (0.083)	0.87
	500	-0.058 (0.036)	0.85	-0.033 (0.033)	0.87
	1000	-0.053 (0.016)	0.86	-0.031 (0.013)	0.88
$0.95gg_1 + 0.05U[-0.5, 0]$	250	-0.074 (0.119)	0.83	0.066 (0.122)	0.86
	500	-0.058 (0.054)	0.84	-0.056 (0.049)	0.86
	1000	-0.053 (0.026)	0.83	-0.044 (0.023)	0.86
$0.95gg_1 + 0.05\chi^2_{(2)}$	250	-0.075 (0.123)	0.84	-0.060 (0.117)	0.86
	500	-0.072 (0.052)	0.84	-0.051 (0.049)	0.87
	1000	-0.058 (0.025)	0.85	-0.053 (0.023)	0.88
$0.95gg_{0.5} + 0.05U[-0.5, 0]$	250	-0.102 (0.165)	0.81	-0.061 (0.161)	0.84
	500	-0.092 (0.075)	0.80	-0.056 (0.068)	0.83
	1000	-0.100 (0.042)	0.81	-0.052 (0.036)	0.84
$0.95gg_{0.5} + 0.05\chi^2_{(2)}$	250	0.096 (0.169)	0.81	0.072 (0.161)	0.84
	500	0.093 (0.071)	0.80	0.060 (0.067)	0.83
	1000	0.091 (0.038)	0.81	0.054 (0.033)	0.84

Table 3.4: $\hat{\beta}_{1,MLE}$ and $\hat{\beta}_{1,MHD}$ for ARCH(1) for ARCH(1) with known innovation distribution and contamination (3.1).

Contamination (3.1)	n	$\hat{\beta}_{1,MLE}$ Bias (MSE)	CP	$\hat{\beta}_{1,MHD}$ Bias (MSE)	CP
$0.95gg_4 + 0.05U[-0.5,0]$	250	-0.046 (0.047)	0.89	0.034 (0.050)	0.90
	500	-0.044 (0.025)	0.88	0.037 (0.022)	0.89
	1000	-0.047 (0.012)	0.89	0.028 (0.009)	0.90
$0.95gg_4 + 0.05\chi^2_{(2)}$	250	-0.051 (0.050)	0.88	0.039 (0.052)	0.91
	500	-0.054 (0.023)	0.87	0.037 (0.020)	0.89
	1000	-0.045 (0.011)	0.86	0.026 (0.009)	0.89
$0.95gg_2 + 0.05U[-0.5,0]$	250	-0.049 (0.073)	0.90	0.042 (0.077)	0.91
	500	-0.053 (0.033)	0.88	0.039 (0.031)	0.89
	1000	-0.059 (0.014)	0.89	0.033 (0.011)	0.90
$0.95gg_2 + 0.05\chi^2_{(2)}$	250	-0.045 (0.072)	0.90	0.040 (0.072)	0.90
	500	-0.047 (0.030)	0.88	0.027 (0.028)	0.89
	1000	-0.049 (0.013)	0.89	0.026 (0.011)	0.89
$0.95gg_1 + 0.05U[-0.5,0]$	250	-0.063 (0.098)	0.84	-0.055 (0.103)	0.86
	500	-0.049 (0.044)	0.85	0.046 (0.042)	0.87
	1000	-0.045 (0.022)	0.85	0.037 (0.018)	0.87
$0.95gg_1 + 0.05\chi^2_{(2)}$	250	-0.064 (0.095)	0.85	0.050 (0.100)	0.87
	500	-0.058 (0.045)	0.84	0.042 (0.042)	0.86
	1000	-0.053 (0.021)	0.85	0.044 (0.018)	0.87
$0.95gg_{0.5} + 0.05U[-0.5,0]$	250	-0.077 (0.152)	0.82	0.051 (0.149)	0.84
	500	-0.075 (0.068)	0.81	0.046 (0.062)	0.84
	1000	-0.069 (0.035)	0.82	0.043 (0.031)	0.85
$0.95gg_{0.5} + 0.05\chi^2_{(2)}$	250	0.082 (0.154)	0.81	-0.060 (0.148)	0.83
	500	0.077 (0.065)	0.82	-0.050 (0.061)	0.84
	1000	0.067 (0.034)	0.82	-0.049 (0.029)	0.85

Table 3.5: $\hat{\beta}_{0,MLE}$ and $\hat{\beta}_{0,MHD}$ for ARCH(1) with known innovation distribution and contamination (3.2).

		$\hat{\beta}_{0,MLE}$		$\hat{\beta}_{0,MHD}$	
Contamination (3.2)	n	Bias (MSE)	CP	Bias (MSE)	CP
$f_\varepsilon = gg_4$	250	-0.085 (0.209)	0.76	-0.068 (0.173)	0.78
$h = U[-0.5, 0]$	500	-0.079 (0.153)	0.75	-0.061 (0.123)	0.76
	1000	-0.084 (0.113)	0.76	-0.059 (0.096)	0.77
$f_\varepsilon = gg_4$	250	-0.080 (0.197)	0.76	-0.069 (0.168)	0.77
$h(x-1.5) = -\chi^2_{(2)}$	500	-0.074 (0.146)	0.75	-0.062 (0.114)	0.76
	1000	-0.068 (0.101)	0.76	-0.057 (0.089)	0.78
$f_{\varepsilon,} = gg_2$	250	0.086 (0.225)	0.76	0.070 (0.189)	0.77
$h = U[-0.5, 0]$	500	0.079 (0.149)	0.75	0.062 (0.123)	0.78
	1000	0.074 (0.108)	0.75	0.057 (0.100)	0.78
$f_\varepsilon = gg_2$	250	0.087 (0.222)	0.75	0.069 (0.184)	0.77
$h(x-1.5) = -\chi^2_{(2)}$	500	0.080 (0.143)	0.75	0.057 (0.122)	0.77
	1000	0.079 (0.101)	0.76	0.060 (0.095)	0.78
$f_\varepsilon = gg_1$	250	0.099 (0.316)	0.74	0.086 (0.260)	0.76
$h(x) = U[-0.5, 0]$	500	0.094 (0.194)	0.74	0.077 (0.167)	0.77
	1000	0.086 (0.132)	0.75	0.071 (0.101)	0.78
$f_\varepsilon = gg_1$	250	0.094 (0.308)	0.74	0.084 (0.253)	0.76
$h(x-1.5) = -\chi^2_{(2)}$	500	0.092 (0.187)	0.73	0.080 (0.169)	0.76
	1000	0.083 (0.125)	0.74	0.065 (0.099)	0.77
$f_\varepsilon = gg_{0.5}$	250	0.120 (0.392)	0.72	0.102 (0.325)	0.75
$h = U[-0.5, 0]$	500	0.114 (0.295)	0.71	0.089 (0.236)	0.74
	1000	0.106 (0.170)	0.72	0.093 (0.138)	0.74
$f_\varepsilon = gg_{0.5}$	250	0.121 (0.377)	0.73	0.096 (0.311)	0.75
$h(x-1.5) = -\chi^2_{(2)}$	500	0.109 (0.286)	0.72	0.098 (0.227)	0.74
	1000	0.113 (0.160)	0.72	0.084 (0.132)	0.75

Table 3.6: $\hat{\beta}_{1,MLE}$ and $\hat{\beta}_{1,MHD}$ for ARCH(1) with known innovation distribution and contamination (3.2).

		$\hat{\beta}_{1,MLE}$		$\hat{\beta}_{1,MHD}$	
Contamination (3.2)	n	Bias (MSE)	CP	Bias (MSE)	CP
$f_\varepsilon = gg_4$	250	-0.075 (0.173)	0.76	-0.057 (0.135)	0.78
$h = U[-0.5, 0]$	500	-0.074 (0.127)	0.76	-0.051 (0.094)	0.77
	1000	-0.070 (0.094)	0.77	-0.050 (0.065)	0.79
$f_\varepsilon = gg_4$	250	-0.070 (0.164)	0.76	-0.058 (0.131)	0.77
$h(x-1.5) = -\chi^2_{(2)}$	500	-0.065 (0.121)	0.75	-0.052 (0.086)	0.76
	1000	-0.060 (0.084)	0.76	-0.048 (0.054)	0.77
$f_\varepsilon = gg_2$	250	0.076 (0.187)	0.76	0.059 (0.149)	0.77
$h = U[-0.5, 0]$	500	0.070 (0.124)	0.75	0.052 (0.102)	0.76
	1000	0.065 (0.090)	0.75	0.048 (0.068)	0.76
$f_\varepsilon = gg_2$	250	0.077 (0.184)	0.76	0.058 (0.144)	0.77
$h(x-1.5) = -\chi^2_{(2)}$	500	0.070 (0.119)	0.75	0.050 (0.101)	0.76
	1000	0.070 (0.084)	0.76	0.048 (0.063)	0.77
$f_\varepsilon = gg_1$	250	0.087 (0.242)	0.73	0.072 (0.183)	0.75
$h = U[-0.5, 0]$	500	0.083 (0.161)	0.72	0.065 (0.118)	0.75
	1000	0.076 (0.110)	0.73	0.060 (0.076)	0.76
$f_\varepsilon = gg_1$	250	0.083 (0.246)	0.73	0.071 (0.185)	0.75
$h(x-1.5) = -\chi^2_{(2)}$	500	0.081 (0.155)	0.73	0.067 (0.110)	0.76
	1000	0.073 (0.104)	0.74	0.055 (0.071)	0.76
$f_\varepsilon = gg_{0.5}$	250	0.106 (0.298)	0.71	0.086 (0.228)	0.74
$h = U[-0.5, 0]$	500	0.100 (0.231)	0.71	0.075 (0.163)	0.73
	1000	0.093 (0.141)	0.72	0.078 (0.098)	0.75
$f_\varepsilon = gg_{0.5}$	250	0.106 (0.288)	0.71	0.082 (0.220)	0.74
$h(x-1.5) = -\chi^2_{(2)}$	500	0.096 (0.237)	0.70	0.077 (0.172)	0.73
	1000	0.099 (0.129)	0.71	0.071 (0.097)	0.74

Case 3: unknown innovation distribution

When the innovation distribution is unspecified, which is a more realistic case, we compare the performance of the proposed MPHDE in (2.7) with both the GQMLE given in (2.8) and the NGQMLE given in (2.9) proposed by Fan et al. [11]. We take t_4 as the quasi-likelihood function used in the construction of NGQMLE. We use the same underlying innovation distributions as those in Case 1 and the simulation results are reported in Table 3.7 and 3.8.

From Table 3.7 and 3.8 we can see that when the innovation distribution is normal (gg_2), close to normal (t_{20}) or light-tailed (gg_4), the GQMLE outperforms both MPHDE and NGQMLE in terms of bias, MSE and CP by a small margin. And for all other cases, NGQMLE and MPHDE performs better than GQMLE. The performance of GQMLE deteriorates as the tails of innovation distribution grow heavier, particularly in $gg_{0.5}$ and t_3. This is expected since the GQMLE assumes the innovation distribution is normal and it is most efficient when the true innovation is normal and the performance deteriorates when the innovation distribution deviates from normal. Comparing MPHDE with NGQMLE, we find that NGQMLE boasts smaller bias and MSE in most cases while their CPs are very close to each other. Thus, the NGQMLE is preferred over MPHDE when we know that the data is clean.

Table 3.7: $\hat{\beta}_{0,MPHD}$, $\hat{\beta}_{0,G}$ and $\hat{\beta}_{0,NG}$ for ARCH(1) with unknown innovation distribution.

		$\hat{\beta}_{0,MPHD}$		$\hat{\beta}_{0,G}$		$\hat{\beta}_{0,NG}$	
f_ε	n	Bias (MSE)	CP	Bias (MSE)	CP	Bias (MSE)	CP
t_{20}	250	0.029 (0.083)	0.91	0.030 (0.087)	0.92	0.031 (0.081)	0.91
	500	-0.021 (0.038)	0.90	0.016 (0.035)	0.91	0.020 (0.036)	0.90
	1000	-0.016 (0.010)	0.91	0.012 (0.009)	0.93	0.011 (0.009)	0.92
t_9	250	0.037 (0.093)	0.89	0.042 (0.096)	0.89	0.035 (0.089)	0.90
	500	-0.027 (0.040)	0.89	0.030 (0.041)	0.88	-0.029 (0.039)	0.89
	1000	-0.022 (0.016)	0.90	0.015 (0.017)	0.89	0.014 (0.015)	0.90
t_6	250	0.036 (0.101)	0.90	0.038 (0.119)	0.89	-0.038 (0.094)	0.90
	500	0.032 (0.040)	0.88	0.036 (0.047)	0.87	-0.034 (0.037)	0.89
	1000	-0.023 (0.017)	0.89	0.022 (0.019)	0.88	-0.021 (0.016)	0.89
t_4	250	0.034 (0.110)	0.90	0.042 (0.144)	0.89	-0.036 (0.103)	0.90
	500	-0.030 (0.048)	0.89	0.032 (0.062)	0.88	-0.023 (0.043)	0.90
	1000	-0.024 (0.025)	0.88	-0.026 (0.031)	0.87	-0.018 (0.023)	0.89
t_3	250	0.058 (0.121)	0.88	0.069 (0.162)	0.87	-0.060 (0.119)	0.88
	500	0.044 (0.056)	0.88	0.054 (0.079)	0.86	-0.041 (0.053)	0.88
	1000	-0.034 (0.027)	0.89	-0.042 (0.039)	0.88	-0.025 (0.025)	0.89
gg_4	250	0.019 (0.050)	0.92	0.015 (0.043)	0.93	0.017 (0.047)	0.92
	500	0.018 (0.021)	0.91	0.008 (0.016)	0.92	0.008 (0.019)	0.91
	1000	-0.014 (0.008)	0.92	0.007 (0.006)	0.94	0.010 (0.008)	0.93
gg_2	250	0.022 (0.066)	0.92	0.014 (0.061)	0.94	0.022 (0.064)	0.92
	500	-0.020 (0.025)	0.91	0.012 (0.020)	0.92	0.013 (0.023)	0.91
	1000	-0.018 (0.010)	0.92	0.008 (0.007)	0.93	0.010 (0.009)	0.92
gg_1	250	0.035 (0.101)	0.91	0.038 (0.106)	0.90	0.028 (0.097)	0.91
	500	-0.018 (0.040)	0.90	0.020 (0.044)	0.89	0.014 (0.038)	0.90
	1000	-0.016 (0.021)	0.91	-0.016 (0.027)	0.90	-0.014 (0.021)	0.91
$gg_{0.5}$	250	0.051 (0.133)	0.88	0.062 (0.151)	0.87	-0.048 (0.128)	0.89
	500	0.033 (0.056)	0.89	0.043 (0.072)	0.87	-0.028 (0.054)	0.89
	1000	0.022 (0.027)	0.90	0.042 (0.035)	0.88	-0.023 (0.025)	0.90

Table 3.8: $\hat{\beta}_{1,MPHD}$, $\hat{\beta}_{1,G}$ and $\hat{\beta}_{1,NG}$ for ARCH(1) with unknown innovation distribution.

		$\hat{\beta}_{1,MPHD}$		$\hat{\beta}_{1,G}$		$\hat{\beta}_{1,NG}$	
f_ε	n	Bias (MSE)	CP	Bias (MSE)	CP	Bias (MSE)	CP
t_{20}	250	-0.036 (0.058)	0.90	-0.029 (0.062)	0.91	-0.033 (0.064)	0.90
	500	-0.025 (0.029)	0.92	-0.010 (0.026)	0.93	-0.013 (0.027)	0.92
	1000	-0.016 (0.011)	0.92	-0.005 (0.011)	0.93	-0.006 (0.011)	0.93
t_9	250	-0.024 (0.072)	0.90	-0.021 (0.066)	0.89	-0.030 (0.066)	0.91
	500	-0.018 (0.032)	0.91	-0.024 (0.032)	0.90	-0.021 (0.030)	0.91
	1000	-0.016 (0.018)	0.92	-0.022 (0.020)	0.91	-0.017 (0.017)	0.92
t_6	250	-0.042 (0.083)	0.91	-0.044 (0.087)	0.89	-0.040 (0.082)	0.91
	500	-0.034 (0.035)	0.90	-0.026 (0.038)	0.89	-0.023 (0.033)	0.90
	1000	-0.029 (0.020)	0.91	-0.023 (0.023)	0.90	-0.024 (0.019)	0.91
t_4	250	-0.048 (0.092)	0.90	-0.052 (0.110)	0.88	-0.042 (0.088)	0.91
	500	-0.036 (0.042)	0.89	-0.053 (0.050)	0.88	-0.030 (0.040)	0.90
	1000	-0.023 (0.024)	0.91	-0.030 (0.029)	0.91	-0.017 (0.022)	0.91
t_3	250	-0.051 (0.125)	0.88	-0.068 (0.149)	0.87	-0.047 (0.118)	0.88
	500	-0.038 (0.050)	0.89	-0.058 (0.068)	0.87	-0.024 (0.047)	0.89
	1000	-0.028 (0.027)	0.90	-0.049 (0.037)	0.88	-0.027 (0.026)	0.90
gg_4	250	-0.026 (0.041)	0.92	-0.016 (0.034)	0.93	-0.021 (0.039)	0.92
	500	-0.017 (0.018)	0.91	-0.015 (0.016)	0.92	-0.014 (0.017)	0.91
	1000	-0.010 (0.010)	0.93	-0.011 (0.008)	0.94	-0.012 (0.010)	0.93
gg_2	250	-0.028 (0.050)	0.92	-0.033 (0.041)	0.92	-0.039 (0.044)	0.91
	500	-0.022 (0.023)	0.92	-0.013 (0.019)	0.93	-0.018 (0.021)	0.91
	1000	-0.012 (0.010)	0.93	-0.006 (0.006)	0.94	-0.010 (0.010)	0.93
gg_1	250	-0.037 (0.084)	0.91	-0.037 (0.091)	0.90	-0.038 (0.080)	0.91
	500	-0.025 (0.032)	0.91	-0.028 (0.034)	0.90	-0.029 (0.031)	0.92
	1000	-0.018 (0.015)	0.92	-0.015 (0.019)	0.91	0.014 (0.013)	0.92
$gg_{0.5}$	250	-0.049 (0.125)	0.89	-0.074 (0.146)	0.87	-0.042 (0.119)	0.89
	500	-0.037 (0.048)	0.89	-0.069 (0.064)	0.88	-0.033 (0.044)	0.89
	1000	-0.026 (0.022)	0.90	-0.042 (0.033)	0.88	-0.022 (0.019)	0.90

Case 4: unknown innovation distribution and contaminated

This case is parallel to the Case 2 but with unknown innovation distribution. So as in Case 2, we consider both types of contaminations, (3.1) and (3.2), in order to assess the robustness of MPHDE as compared to those of GQMLE and NGQMLE. We also consider the same contaminating distributions and contamination rate as those in Tables 3.3-3.6. The results are presented in Table 3.9 and 3.10 for the first type of contamination (3.1) while Table 3.11 and 3.12 report the second type (3.2).

From Table 3.9 and 3.10 we can see that, for contamination structure (3.1), the biases and MSEs of all the three estimators are significantly increased while the CPs of them are decreased, in comparison with the corresponding uncontaminated ones. Despite this, as expected, both the (absolute values of) biases and MSEs of MLE and MHDE decrease as sample size increases. When the three estimators are compared with each other, GQMLE generally has the worst performance while MPHDE has the best performance among the three estimators in terms of bias, MSE and CP, which gives evidence of the robustness of MPHDE in the event of contamination.

Under the second type of contamination (3.2), we observe from Table 3.11 and 3.12 similar phenomena as those under the first type of contamination (3.1), in which MPHDE performs better than both GQMLE and NGQMLE. Specifically, MPHDE is around 40% more efficient than GQMLE and 20% more effficient than NGQMLE on average. MPHDE also has a clear advantage over GQMLE and NGQMLE in terms of bias and CP.

When the two types of contamination are compared, i.e. Table 3.9 and 3.10 compared with Table 3.11 and 3.12, the second type of contamination affects the performance of MPHDE, GQMLE and NGQMLE more severely than the first type in the sense of having much larger bias, MSE and much smaller CP. This is consistent with our observations in Case 2 when the innovation distribution is known.

When the MHDE and MPHDE are compared with each other, MPHDE performs better when it comes to bias, MSE and CP under the second type of contamination while MHDE performs better under the case of no contamination and the first type of contamination. This observation is quite reasonable. Specifically, when there is no contamination, the MHDE makes use of the correctly specified innovation distribution and thus is more efficient than the MPHDE. Under the second type of contamination which is more severe than the first type, the MHDE makes use of the wrong specified innovation distribution and thus produce worse results than those of MPHDE which estimates the innovation distribution nonparametrically based on the data. On the other hand, under the first type of contamination, MHDE still benefits from the specified innovation distribution that is close to the contaminated one.

Table 3.9: $\hat{\beta}_{0,MPHD}$, $\hat{\beta}_{0,G}$ and $\hat{\beta}_{0,NG}$ for ARCH(1) with unknown innovation distribution and contamination (3.1).

		$\hat{\beta}_{0,MPHD}$		$\hat{\beta}_{0,G}$		$\hat{\beta}_{0,NG}$	
Contamination (3.1)	n	Bias (MSE)	CP	Bias (MSE)	CP	Bias (MSE)	CP
$0.95gg_4+$	250	-0.053 (0.068)	0.87	-0.070 (0.072)	0.87	-0.064 (0.066)	0.87
$0.05U[-0.5,0]$	500	-0.047 (0.034)	0.87	-0.060 (0.040)	0.86	-0.055 (0.037)	0.86
	1000	-0.043 (0.020)	0.87	-0.053 (0.026)	0.86	-0.050 (0.025)	0.86
$0.95gg_4+$	250	-0.051 (0.066)	0.88	-0.070 (0.072)	0.88	-0.062 (0.064)	0.88
$0.05\chi^2_{(2)}$	500	-0.046 (0.029)	0.87	-0.064 (0.033)	0.86	-0.057 (0.031)	0.86
	1000	-0.042 (0.017)	0.87	-0.057 (0.023)	0.87	-0.049 (0.020)	0.86
$0.95gg_2+$	250	-0.059 (0.095)	0.86	-0.070 (0.098)	0.86	-0.069 (0.094)	0.86
$0.05U[-0.5,0]$	500	-0.050 (0.038)	0.87	-0.065 (0.042)	0.85	-0.056 (0.040)	0.86
	1000	-0.054 (0.016)	0.87	-0.059 (0.019)	0.86	-0.058 (0.019)	0.86
$0.95gg_2+$	250	-0.058 (0.093)	0.86	-0.066 (0.100)	0.86	-0.062 (0.093)	0.86
$0.05\chi^2_{(2)}$	500	-0.046 (0.036)	0.87	-0.053 (0.040)	0.86	-0.054 (0.040)	0.86
	1000	-0.042 (0.016)	0.88	-0.055 (0.018)	0.87	-0.059 (0.017)	0.87
$0.95gg_1+$	250	-0.070 (0.133)	0.85	-0.090 (0.142)	0.82	-0.086 (0.136)	0.84
$0.05U[-0.5,0]$	500	-0.072 (0.059)	0.85	-0.080 (0.068)	0.82	-0.075 (0.064)	0.83
	1000	-0.055 (0.029)	0.86	-0.074 (0.039)	0.84	-0.072 (0.034)	0.85
$0.95gg_1+$	250	-0.067 (0.131)	0.85	-0.087 (0.139)	0.82	-0.083 (0.135)	0.83
$0.05\chi^2_{(2)}$	500	-0.064 (0.055)	0.84	-0.079 (0.063)	0.81	-0.078 (0.060)	0.83
	1000	-0.056 (0.030)	0.85	-0.074 (0.037)	0.82	-0.070 (0.033)	0.83
$0.95gg_{0.5}+$	250	-0.072 (0.182)	0.83	-0.101 (0.195)	0.80	-0.086 (0.185)	0.81
$0.05U[-0.5,0]$	500	-0.066 (0.078)	0.82	-0.092 (0.094)	0.78	-0.081 (0.085)	0.79
	1000	-0.062 (0.045)	0.84	-0.095 (0.058)	0.79	-0.078 (0.051)	0.80
$0.95gg_{0.5}+$	250	-0.073 (0.179)	0.84	-0.096 (0.191)	0.81	-0.088 (0.187)	0.82
$0.05\chi^2_{(2)}$	500	-0.068 (0.075)	0.83	-0.092 (0.090)	0.80	-0.079 (0.083)	0.81
	1000	-0.064 (0.044)	0.83	-0.086 (0.056)	0.79	-0.075 (0.051)	0.80

Table 3.10: $\hat{\beta}_{1,MPHD}$, $\hat{\beta}_{1,G}$ and $\hat{\beta}_{1,NG}$ for ARCH(1) with unknown innovation distribution and contamination (3.1).

		$\hat{\beta}_{1,MPHD}$		$\hat{\beta}_{1,G}$		$\hat{\beta}_{1,NG}$	
Contamination (3.1)	n	Bias (MSE)	CP	Bias (MSE)	CP	Bias (MSE)	CP
$0.95gg_4+$	250	-0.058 (0.059)	0.88	-0.070 (0.062)	0.87	-0.064 (0.059)	0.88
$0.05U[-0.5,0]$	500	-0.054 (0.032)	0.88	-0.064 (0.038)	0.86	-0.059 (0.036)	0.87
	1000	-0.046 (0.017)	0.89	-0.061 (0.022)	0.86	-0.055 (0.020)	0.88
$0.95gg_4+$	250	-0.056 (0.058)	0.90	-0.069 (0.061)	0.87	-0.056 (0.056)	0.90
$0.05\chi^2_{(2)}$	500	-0.048 (0.030)	0.89	-0.060 (0.036)	0.86	-0.055 (0.036)	0.88
	1000	-0.047 (0.017)	0.89	-0.058 (0.021)	0.87	-0.050 (0.020)	0.87
$0.95gg_2+$	250	-0.051 (0.077)	0.88	-0.069 (0.077)	0.89	-0.056 (0.080)	0.88
$0.05U[-0.5,0]$	500	-0.050 (0.034)	0.89	-0.060 (0.037)	0.88	-0.053 (0.037)	0.88
	1000	-0.041 (0.016)	0.89	-0.062 (0.018)	0.88	-0.049 (0.020)	0.88
$0.95gg_2+$	250	-0.054 (0.076)	0.90	-0.072 (0.074)	0.90	-0.060 (0.079)	0.90
$0.05\chi^2_{(2)}$	500	-0.050 (0.030)	0.89	-0.060 (0.034)	0.88	-0.054 (0.034)	0.88
	1000	-0.048 (0.014)	0.90	-0.057 (0.016)	0.88	-0.057 (0.017)	0.89
$0.95gg_1+$	250	-0.065 (0.111)	0.85	-0.073 (0.128)	0.82	-0.068 (0.109)	0.83
$0.05U[-0.5,0]$	500	-0.058 (0.055)	0.84	-0.068 (0.065)	0.80	-0.064 (0.059)	0.82
	1000	-0.049 (0.027)	0.85	-0.059 (0.032)	0.81	-0.056 (0.030)	0.84
$0.95gg_1+$	250	-0.059 (0.110)	0.87	-0.068 (0.124)	0.85	-0.061 (0.115)	0.87
$0.05\chi^2_{(2)}$	500	-0.052 (0.052)	0.86	-0.064 (0.062)	0.83	-0.059 (0.057)	0.85
	1000	-0.049 (0.026)	0.85	-0.060 (0.033)	0.83	-0.055 (0.031)	0.84
$0.95gg_{0.5}+$	250	-0.071 (0.165)	0.82	-0.084 (0.184)	0.80	-0.076 (0.177)	0.81
$0.05U[-0.5,0]$	500	-0.064 (0.076)	0.82	-0.075 (0.092)	0.78	-0.069 (0.080)	0.80
	1000	-0.055 (0.042)	0.83	-0.072 (0.051)	0.79	-0.064 (0.046)	0.80
$0.95gg_{0.5}+$	250	-0.069 (0.161)	0.83	-0.075 (0.178)	0.80	-0.074 (0.177)	0.82
$0.05\chi^2_{(2)}$	500	-0.064 (0.072)	0.82	-0.073 (0.088)	0.79	-0.059 (0.075)	0.80
	1000	-0.051 (0.039)	0.83	-0.068 (0.056)	0.79	-0.063 (0.043)	0.81

Table 3.11: $\hat{\beta}_{0,MPHD}$, $\hat{\beta}_{0,G}$ and $\hat{\beta}_{0,NG}$ for ARCH(1) with unknown innovation distribution and contamination (3.2).

		$\hat{\beta}_{0,MPHD}$		$\hat{\beta}_{0,G}$		$\hat{\beta}_{0,NG}$	
Contamination (3.2)	n	Bias (MSE)	CP	Bias (MSE)	CP	Bias (MSE)	CP
$f_\varepsilon = gg_4$	250	-0.064 (0.156)	0.79	-0.093 (0.202)	0.76	-0.075 (0.180)	0.76
$h = U[-0.5, 0]$	500	-0.059 (0.101)	0.77	-0.086 (0.148)	0.75	-0.070 (0.130)	0.75
	1000	-0.053 (0.080)	0.78	-0.083 (0.116)	0.76	-0.066 (0.087)	0.76
$f_\varepsilon = gg_4$	250	-0.060 (0.145)	0.78	-0.091 (0.199)	0.76	-0.071 (0.170)	0.77
$h(x-1.5) = -\chi^2_{(2)}$	500	-0.050 (0.099)	0.77	-0.084 (0.140)	0.76	-0.076 (0.116)	0.76
	1000	-0.053 (0.075)	0.79	-0.076 (0.104)	0.76	-0.065 (0.089)	0.77
$f_\varepsilon = gg_2$	250	0.065 (0.172)	0.77	0.087 (0.223)	0.76	0.080 (0.205)	0.76
$h = U[-0.5, 0]$	500	0.056 (0.117)	0.76	0.082 (0.146)	0.74	0.074 (0.131)	0.74
	1000	0.054 (0.078)	0.76	0.078 (0.110)	0.75	0.064 (0.094)	0.75
$f_\varepsilon = gg_2$	250	0.061 (0.174)	0.77	0.085 (0.222)	0.75	0.080 (0.194)	0.75
$h(x-1.5) = -\chi^2_{(2)}$	500	0.052 (0.105)	0.77	0.086 (0.145)	0.75	0.071 (0.127)	0.76
	1000	0.051 (0.076)	0.78	0.075 (0.100)	0.76	0.066 (0.088)	0.76
$f_\varepsilon = gg_1$	250	0.076 (0.242)	0.76	0.109 (0.329)	0.73	0.086 (0.279)	0.74
$h = U[-0.5, 0]$	500	0.068 (0.151)	0.76	0.106 (0.205)	0.72	0.078 (0.171)	0.74
	1000	0.063 (0.094)	0.76	0.093 (0.139)	0.73	0.072 (0.115)	0.75
$f_\varepsilon = gg_1$	250	0.072 (0.238)	0.76	0.105 (0.304)	0.73	0.089 (0.263)	0.74
$h(x-1.5) = -\chi^2_{(2)}$	500	0.066 (0.148)	0.77	0.096 (0.181)	0.73	0.077 (0.162)	0.75
	1000	0.060 (0.083)	0.77	0.085 (0.130)	0.73	0.068 (0.128)	0.75
$f_\varepsilon = gg_{0.5}$	250	0.088 (0.284)	0.74	0.129 (0.367)	0.70	0.101 (0.309)	0.72
$h = U[-0.5, 0]$	500	0.080 (0.196)	0.74	0.118 (0.265)	0.69	0.097 (0.246)	0.71
	1000	0.073 (0.114)	0.73	0.109 (0.185)	0.70	0.089 (0.152)	0.72
$f_{\varepsilon,1} = gg_{0.5}$	250	0.084 (0.266)	0.75	0.121 (0.353)	0.71	-0.095 (0.289)	0.72
$h(x-1.5) = -\chi^2_{(2)}$	500	0.077 (0.196)	0.73	0.112 (0.260)	0.70	-0.088 (0.239)	0.71
	1000	0.070 (0.104)	0.74	0.105 (0.174)	0.71	-0.081 (0.151)	0.72

Table 3.12: $\hat{\beta}_{1,MPHD}$, $\hat{\beta}_{1,G}$ and $\hat{\beta}_{1,NG}$ for ARCH(1) with unknown innovation distribution and contamination (3.2).

		$\hat{\beta}_{1,MPHD}$		$\hat{\beta}_{1,G}$		$\hat{\beta}_{1,NG}$	
Contamination (3.2)	n	Bias (MSE)	CP	Bias (MSE)	CP	Bias (MSE)	CP
$f_\varepsilon = gg_4$	250	0.055 (0.120)	0.80	0.081 (0.184)	0.76	0.069 (0.141)	0.77
$h = U[-0.5, 0]$	500	0.051 (0.087)	0.79	0.077 (0.137)	0.77	0.064 (0.115)	0.77
	1000	0.046 (0.060)	0.79	0.066 (0.105)	0.78	0.060 (0.078)	0.78
$f_\varepsilon = gg_4$	250	-0.058 (0.127)	0.78	-0.077 (0.164)	0.76	-0.072 (0.148)	0.76
$h(x-1.5) = -\chi^2_{(2)}$	500	-0.052 (0.080)	0.77	0.079 (0.126)	0.75	0.075 (0.108)	0.76
	1000	-0.041 (0.050)	0.79	0.064 (0.082)	0.76	0.062 (0.065)	0.76
$f_\varepsilon = gg_2$	250	0.051 (0.133)	0.78	0.078 (0.186)	0.76	0.079 (0.162)	0.76
$h = U[-0.5, 0]$	500	0.053 (0.100)	0.78	0.074 (0.120)	0.76	0.068 (0.113)	0.76
	1000	0.046 (0.062)	0.79	0.061 (0.093)	0.77	0.066 (0.084)	0.77
$f_\varepsilon = gg_2$	250	0.057 (0.135)	0.79	0.077 (0.186)	0.77	0.079 (0.169)	0.77
$h(x-1.5) = -\chi^2_{(2)}$	500	0.047 (0.095)	0.78	0.075 (0.117)	0.75	0.071 (0.106)	0.76
	1000	0.040 (0.057)	0.79	0.071 (0.084)	0.76	0.072 (0.073)	0.76
$f_\varepsilon = gg_1$	250	-0.070 (0.163)	0.76	-0.097 (0.252)	0.72	-0.091 (0.210)	0.74
$h = U[-0.5, 0]$	500	-0.059 (0.103)	0.76	-0.087 (0.175)	0.72	-0.079 (0.145)	0.73
	1000	-0.057 (0.067)	0.77	-0.079 (0.120)	0.73	-0.068 (0.096)	0.75
$f_\varepsilon = gg_1$	250	-0.071 (0.163)	0.76	-0.091 (0.246)	0.73	-0.064 (0.193)	0.75
$h(x-1.5) = -\chi^2_{(2)}$	500	-0.066 (0.091)	0.77	-0.083 (0.158)	0.73	-0.074 (0.119)	0.74
	1000	-0.058 (0.061)	0.77	-0.076 (0.118)	0.74	-0.064 (0.086)	0.75
$f_\varepsilon = gg_{0.5}$	250	-0.080 (0.207)	0.74	-0.108 (0.293)	0.70	-0.094 (0.240)	0.72
$h = U[-0.5, 0]$	500	-0.072 (0.142)	0.73	-0.097 (0.222)	0.70	-0.086 (0.193)	0.71
	1000	-0.067 (0.082)	0.74	-0.088 (0.133)	0.71	-0.081 (0.118)	0.72
$f_\varepsilon = gg_{0.5}$	250	-0.075 (0.197)	0.75	-0.103 (0.280)	0.71	-0.086 (0.540)	0.73
$h(x-1.5) = -\chi^2_{(2)}$	500	-0.074 (0.150)	0.74	-0.095 (0.205)	0.70	-0.080 (0.184)	0.73
	1000	-0.062 (0.083)	0.74	-0.086 (0.132)	0.71	-0.072 (0.115)	0.72

3.2. GARCH$(1,1)$

Now we assess and compare the finite-sample performance of the proposed estimator MPHDE with GQMLE and NGQMLE under the GARCH$(1,1)$ model. We can similarly consider the same Cases 1-4 under GARCH$(1,1)$ model, but since some results are quite similar, we only report the more general Case 3 and Case 4 under the second type of contamination (3.2).

For the GARCH$(1,1)$ model given in (2.1), β_0 is the lower bound of the conditional and unconditional variance of X_t's, β_1 measures the extent to which a shock in returns X_t today feeds into next period's volatility while β_2 measures the effect of volatility shock today on the next period's volatility. In addition, $\beta_1+\beta_2$ measures the rate at which these effects dies over time. The larger the $\beta_1+\beta_2$, the longer the persistence of these effects. To ensure the stability of the GARCH process, we impose the constraints $\beta_i > 0$, $i = 0,1,2$ and $0 < \beta_1 + \beta_2 < 1$ in the simulation. In our simulation studies, we take the true parameter values $\beta = (\beta_0,\beta_1,\beta_2)^\top = (0.5,0.3,0.6)^\top$. Note that in GARCH$(1,1)$, σ_0 is also an unknown parameter. Following the convention in the existing literature, we use the sample unconditional standard deviation as an estimate for σ_0 to reduce dimensionality, i.e. $\sigma_0 = std(X_t)$.

Case 5: unknown innovation distribution

When the innovation distribution is unknown and data are not contaminated, the results are presented in Table 3.13, 3.14 and 3.15 for β_0, β_1 and β_2 respectively. As in Case 3, we take t_4 as the quasi-likelihood function used in the construction of NGQMLE. The performance of MPHDE relative to GQMLE and NGQMLE is similar to those in the ARCH(1) model. Specifically, When the innovation distribution is normal (gg_2), close to normal (t_{20}) or light-tailed (gg_4), GQMLE delivers the best performance among the three followed by NGQMLE. On the other hand, when the innovation distribution has heavy tails, NGQMLE generally performs the best followed by MPHDE.

Case 6: unknown innovation distribution and contaminated

When the innovation distribution is unknown and data are subject to the second type of contamination (3.2), the results are presented in Table 3.16, 3.17 and 3.18 for β_0, β_1 and β_2 respectively. Similar to the simulation studies of the ARCH(1) model in Case 4, MPHDE again tops both GQMLE and NGQMLE in terms of bias, MSE and CP by a considerable margin, which justifies its robustness to data contamination.

Table 3.13: $\hat{\beta}_{0,MPHD}$, $\hat{\beta}_{0,G}$ and $\hat{\beta}_{0,NG}$ for GARCH$(1,1)$ with unknown innovation distribution.

		$\hat{\beta}_{0,MPHD}$		$\hat{\beta}_{0,G}$		$\hat{\beta}_{0,NG}$	
f_ε	n	Bias (MSE)	CP	Bias (MSE)	CP	Bias (MSE)	CP
t_{20}	250	-0.028 (0.088)	0.90	0.024 (0.086)	0.91	0.025 (0.087)	0.91
	500	-0.019 (0.034)	0.90	0.014 (0.032)	0.91	0.017 (0.035)	0.90
	1000	-0.015 (0.013)	0.91	0.012 (0.012)	0.92	0.014 (0.012)	0.91
t_9	250	-0.034 (0.090)	0.90	0.033 (0.088)	0.89	0.031 (0.084)	0.90
	500	-0.028 (0.036)	0.89	0.031 (0.039)	0.88	0.029 (0.034)	0.89
	1000	-0.023 (0.016)	0.88	0.020 (0.020)	0.89	0.017 (0.014)	0.89
t_6	250	0.034 (0.100)	0.90	0.035 (0.117)	0.88	-0.030 (0.090)	0.90
	500	0.023 (0.040)	0.89	0.026 (0.046)	0.87	-0.018 (0.037)	0.89
	1000	-0.019 (0.018)	0.90	0.019 (0.024)	0.88	-0.018 (0.016)	0.90
t_4	250	0.033 (0.116)	0.90	0.048 (0.145)	0.88	-0.037 (0.106)	0.90
	500	-0.028 (0.045)	0.89	0.038 (0.057)	0.87	-0.029 (0.043)	0.89
	1000	-0.024 (0.022)	0.89	-0.031 (0.030)	0.87	-0.025 (0.020)	0.90
t_3	250	0.046 (0.126)	0.88	0.061 (0.153)	0.86	-0.050 (0.118)	0.88
	500	0.039 (0.058)	0.88	0.051 (0.074)	0.86	-0.043 (0.052)	0.89
	1000	-0.026 (0.025)	0.89	-0.044 (0.033)	0.87	-0.033 (0.024)	0.89
gg_4	250	0.021 (0.045)	0.92	0.018 (0.041)	0.93	0.019 (0.044)	0.92
	500	0.023 (0.023)	0.91	0.012 (0.019)	0.92	0.016 (0.022)	0.91
	1000	-0.016 (0.010)	0.91	0.010 (0.009)	0.93	0.012 (0.010)	0.92
gg_2	250	0.027 (0.062)	0.92	0.020 (0.057)	0.93	0.025 (0.060)	0.92
	500	-0.016 (0.029)	0.91	0.013 (0.026)	0.92	0.015 (0.030)	0.91
	1000	-0.013 (0.014)	0.92	0.010 (0.013)	0.93	0.012 (0.014)	0.91
gg_1	250	-0.035 (0.098)	0.92	0.038 (0.106)	0.91	0.030 (0.091)	0.92
	500	-0.028 (0.038)	0.91	0.035 (0.044)	0.89	0.025 (0.036)	0.91
	1000	-0.017 (0.020)	0.91	0.022 (0.025)	0.89	0.017 (0.019)	0.90
$gg_{0.5}$	250	0.045 (0.113)	0.89	0.055 (0.134)	0.87	-0.039 (0.109)	0.89
	500	0.027 (0.051)	0.88	0.042 (0.059)	0.86	-0.025 (0.049)	0.88
	1000	0.024 (0.026)	0.89	0.032 (0.033)	0.87	-0.022 (0.024)	0.89

Table 3.14: $\hat{\beta}_{1,MPHD}$, $\hat{\beta}_{1,G}$ and $\hat{\beta}_{1,NG}$ for GARCH$(1,1)$ with unknown innovation distribution.

		$\hat{\beta}_{1,MPHD}$		$\hat{\beta}_{1,G}$		$\hat{\beta}_{1,NG}$	
f_ε	n	Bias (MSE)	CP	Bias (MSE)	CP	Bias (MSE)	CP
t_{20}	250	-0.031 (0.046)	0.90	-0.026 (0.043)	0.91	-0.028 (0.044)	0.90
	500	-0.021 (0.021)	0.91	-0.018 (0.022)	0.92	-0.016 (0.020)	0.91
	1000	-0.009 (0.008)	0.92	-0.008 (0.006)	0.92	-0.011 (0.007)	0.92
t_9	250	-0.027 (0.053)	0.91	-0.025 (0.051)	0.90	-0.020 (0.049)	0.91
	500	-0.017 (0.024)	0.90	-0.020 (0.026)	0.90	-0.015 (0.022)	0.90
	1000	-0.013 (0.013)	0.91	-0.015 (0.015)	0.91	-0.012 (0.011)	0.91
t_6	250	-0.030 (0.060)	0.91	-0.031 (0.064)	0.90	-0.027 (0.057)	0.91
	500	-0.019 (0.031)	0.90	-0.020 (0.037)	0.89	-0.016 (0.027)	0.90
	1000	-0.016 (0.017)	0.91	-0.018 (0.020)	0.90	-0.014 (0.016)	0.91
t_4	250	-0.035 (0.078)	0.90	-0.041 (0.088)	0.89	-0.031 (0.070)	0.91
	500	-0.029 (0.037)	0.89	-0.037 (0.042)	0.88	-0.024 (0.034)	0.90
	1000	-0.019 (0.022)	0.91	-0.024 (0.028)	0.89	-0.016 (0.020)	0.91
t_3	250	-0.037 (0.110)	0.88	-0.047 (0.133)	0.87	-0.033 (0.096)	0.88
	500	-0.030 (0.055)	0.88	-0.042 (0.069)	0.87	-0.028 (0.051)	0.88
	1000	-0.023 (0.029)	0.89	-0.034 (0.035)	0.88	-0.020 (0.027)	0.89
gg_4	250	-0.021 (0.033)	0.92	-0.017 (0.029)	0.93	-0.018 (0.032)	0.92
	500	-0.012 (0.019)	0.91	-0.009 (0.017)	0.92	-0.009 (0.019)	0.92
	1000	-0.007 (0.010)	0.92	-0.005 (0.009)	0.93	-0.006 (0.012)	0.92
gg_2	250	-0.020 (0.039)	0.92	-0.018 (0.034)	0.92	-0.019 (0.038)	0.92
	500	-0.016 (0.025)	0.91	-0.015 (0.021)	0.92	-0.019 (0.024)	0.91
	1000	-0.009 (0.015)	0.92	-0.007 (0.012)	0.92	-0.009 (0.015)	0.92
gg_1	250	-0.030 (0.070)	0.91	-0.034 (0.075)	0.90	-0.029 (0.067)	0.91
	500	-0.018 (0.030)	0.91	-0.022 (0.034)	0.90	-0.016 (0.028)	0.92
	1000	-0.011 (0.014)	0.92	-0.015 (0.016)	0.91	0.009 (0.012)	0.92
$gg_{0.5}$	250	-0.035 (0.119)	0.90	-0.058 (0.142)	0.88	-0.033 (0.111)	0.90
	500	-0.030 (0.043)	0.89	-0.050 (0.058)	0.87	-0.026 (0.039)	0.89
	1000	-0.019 (0.018)	0.90	-0.031 (0.026)	0.89	-0.017 (0.016)	0.90

Table 3.15: $\hat{\beta}_{2,MPHD}$, $\hat{\beta}_{2,G}$ and $\hat{\beta}_{2,NG}$ for GARCH$(1,1)$ with unknown innovation distribution.

		$\hat{\beta}_{2,MPHD}$		$\hat{\beta}_{2,G}$		$\hat{\beta}_{2,NG}$	
f_ε	n	Bias (MSE)	CP	Bias (MSE)	CP	Bias (MSE)	CP
t_{20}	250	-0.037 (0.066)	0.90	-0.031 (0.061)	0.91	-0.034 (0.061)	0.90
	500	-0.026 (0.032)	0.91	-0.018 (0.031)	0.92	-0.022 (0.030)	0.91
	1000	-0.013 (0.013)	0.92	-0.011 (0.012)	0.92	-0.015 (0.011)	0.92
t_9	250	-0.030 (0.067)	0.90	-0.031 (0.065)	0.90	-0.026 (0.062)	0.91
	500	-0.021 (0.034)	0.91	-0.022 (0.036)	0.90	-0.018 (0.032)	0.91
	1000	-0.018 (0.017)	0.92	-0.018 (0.017)	0.91	-0.016 (0.015)	0.92
t_6	250	-0.036 (0.084)	0.91	-0.044 (0.086)	0.90	-0.038 (0.079)	0.91
	500	-0.028 (0.039)	0.90	-0.029 (0.040)	0.89	-0.026 (0.036)	0.90
	1000	-0.016 (0.024)	0.91	-0.017 (0.027)	0.91	-0.016 (0.022)	0.91
t_4	250	-0.046 (0.102)	0.90	-0.050 (0.109)	0.90	-0.040 (0.097)	0.91
	500	-0.035 (0.049)	0.89	-0.048 (0.055)	0.88	-0.032 (0.044)	0.89
	1000	-0.024 (0.022)	0.91	-0.030 (0.026)	0.89	-0.020 (0.020)	0.91
t_3	250	-0.049 (0.126)	0.88	-0.062 (0.156)	0.87	-0.043 (0.121)	0.88
	500	-0.037 (0.051)	0.89	-0.049 (0.068)	0.88	-0.031 (0.048)	0.89
	1000	-0.028 (0.025)	0.90	-0.042 (0.032)	0.88	-0.026 (0.023)	0.90
gg_4	250	-0.027 (0.049)	0.92	-0.022 (0.042)	0.93	-0.024 (0.045)	0.93
	500	-0.013 (0.021)	0.91	-0.010 (0.019)	0.92	-0.015 (0.020)	0.91
	1000	-0.011 (0.013)	0.92	-0.009 (0.011)	0.92	-0.010 (0.013)	0.92
gg_2	250	-0.029 (0.054)	0.92	-0.026 (0.051)	0.92	-0.029 (0.054)	0.92
	500	-0.021 (0.030)	0.91	-0.019 (0.024)	0.91	-0.023 (0.026)	0.91
	1000	-0.011 (0.016)	0.92	-0.010 (0.013)	0.93	-0.011 (0.015)	0.92
gg_1	250	-0.038 (0.077)	0.91	-0.041 (0.083)	0.90	-0.037 (0.071)	0.91
	500	-0.023 (0.033)	0.90	-0.026 (0.040)	0.89	-0.023 (0.030)	0.90
	1000	-0.014 (0.016)	0.91	-0.017 (0.019)	0.90	0.012 (0.014)	0.91
$gg_{0.5}$	250	-0.048 (0.127)	0.89	-0.070 (0.150)	0.87	-0.043 (0.118)	0.89
	500	-0.035 (0.045)	0.88	-0.061 (0.062)	0.87	-0.033 (0.048)	0.89
	1000	-0.029 (0.023)	0.89	-0.040 (0.032)	0.88	-0.024 (0.021)	0.90

Table 3.16: $\hat{\beta}_{0,MPHD}$, $\hat{\beta}_{0,G}$ and $\hat{\beta}_{0,NG}$ for GARCH$(1,1)$ with unknown innovation distribution and contamination (3.2).

		$\hat{\beta}_{0,MPHD}$		$\hat{\beta}_{0,G}$		$\hat{\beta}_{0,NG}$	
Contamination (3.2)	n	Bias (MSE)	CP	Bias (MSE)	CP	Bias (MSE)	CP
$f_\varepsilon = gg_4$	250	-0.060 (0.138)	0.78	-0.087 (0.181)	0.75	-0.079 (0.151)	0.76
$h = U[-0.5, 0]$	500	-0.052 (0.097)	0.77	-0.084 (0.137)	0.74	-0.072 (0.111)	0.74
	1000	-0.050 (0.063)	0.77	-0.083 (0.096)	0.75	-0.073 (0.082)	0.76
$f_\varepsilon = gg_4$	250	-0.053 (0.134)	0.78	0.085 (0.174)	0.76	0.072 (0.146)	0.77
$h(x-1.5) = -\chi^2_{(2)}$	500	-0.050 (0.093)	0.77	0.079 (0.131)	0.75	0.072 (0.109)	0.76
	1000	-0.051 (0.060)	0.79	0.073 (0.091)	0.76	0.067 (0.075)	0.76
$f_\varepsilon = gg_2$	250	-0.064 (0.157)	0.77	-0.083 (0.195)	0.75	-0.072 (0.172)	0.75
$h = U[-0.5, 0]$	500	-0.058 (0.112)	0.76	-0.085 (0.149)	0.73	-0.068 (0.120)	0.74
	1000	-0.055 (0.073)	0.76	-0.077 (0.112)	0.74	-0.069 (0.088)	0.75
$f_\varepsilon = gg_2$	250	-0.060 (0.152)	0.78	0.077 (0.193)	0.75	0.069 (0.170)	0.76
$h(x-1.5) = -\chi^2_{(2)}$	500	-0.057 (0.089)	0.77	0.080 (0.128)	0.74	0.067 (0.106)	0.76
	1000	-0.052 (0.062)	0.78	0.075 (0.090)	0.75	0.064 (0.074)	0.76
$f_\varepsilon = gg_1$	250	-0.060 (0.213)	0.76	-0.092 (0.269)	0.73	-0.066 (0.236)	0.74
$h = U[-0.5, 0]$	500	-0.055 (0.141)	0.76	-0.088 (0.181)	0.73	-0.060 (0.157)	0.75
	1000	-0.058 (0.083)	0.77	-0.084 (0.130)	0.74	-0.072 (0.103)	0.75
$f_\varepsilon = gg_1$	250	0.061 (0.211)	0.76	0.090 (0.254)	0.73	0.069 (0.221)	0.75
$h(x-1.5) = -\chi^2_{(2)}$	500	0.064 (0.134)	0.77	0.086 (0.167)	0.74	0.067 (0.146)	0.75
	1000	0.053 (0.081)	0.77	0.082 (0.122)	0.74	0.062 (0.100)	0.75
$f_\varepsilon = gg_{0.5}$	250	0.070 (0.252)	0.74	-0.101 (0.295)	0.71	-0.085 (0.271)	0.72
$h = U[-0.5, 0]$	500	0.064 (0.179)	0.74	-0.098 (0.233)	0.70	-0.084 (0.203)	0.72
	1000	0.060 (0.099)	0.75	-0.093 (0.132)	0.71	-0.081 (0.117)	0.73
$f_\varepsilon = gg_{0.5}$	250	0.067 (0.248)	0.75	0.097 (0.290)	0.72	-0.087 (0.268)	0.73
$h(x-1.5) = -\chi^2_{(2)}$	500	0.063 (0.169)	0.74	0.095 (0.221)	0.71	-0.080 (0.196)	0.73
	1000	0.060 (0.090)	0.76	0.095 (0.131)	0.72	-0.081 (0.117)	0.74

Table 3.17: $\hat{\beta}_{1,MPHD}$, $\hat{\beta}_{1,G}$ and $\hat{\beta}_{1,NG}$ for GARCH$(1,1)$ with unknown innovation distribution and contamination (3.2).

Contamination (3.2)	n	$\hat{\beta}_{1,MPHD}$ Bias (MSE)	CP	$\hat{\beta}_{1,G}$ Bias (MSE)	CP	$\hat{\beta}_{1,NG}$ Bias (MSE)	CP
$f_\varepsilon = gg_4$	250	0.052 (0.096)	0.80	0.064 (0.135)	0.77	0.054 (0.107)	0.78
$h = U[-0.5,0]$	500	0.048 (0.070)	0.79	0.063 (0.094)	0.76	0.053 (0.083)	0.78
	1000	0.042 (0.049)	0.81	0.058 (0.066)	0.77	0.050 (0.061)	0.78
$f_\varepsilon = gg_4$	250	0.050 (0.094)	0.79	0.062 (0.124)	0.77	0.053 (0.105)	0.78
$h(x-1.5) = -\chi^2_{(2)}$	500	0.049 (0.067)	0.78	0.062 (0.089)	0.75	0.051 (0.072)	0.76
	1000	0.040 (0.044)	0.79	0.053 (0.063)	0.76	0.045 (0.052)	0.77
$f_\varepsilon = gg_2$	250	0.052 (0.112)	0.78	0.068 (0.153)	0.76	0.062 (0.122)	0.76
$h = U[-0.5,0]$	500	0.048 (0.076)	0.78	0.058 (0.104)	0.75	0.052 (0.086)	0.76
	1000	0.043 (0.049)	0.79	0.058 (0.073)	0.77	0.049 (0.066)	0.77
$f_\varepsilon = gg_2$	250	-0.050 (0.107)	0.78	-0.066 (0.147)	0.76	-0.061 (0.118)	0.76
$h(x-1.5) = -\chi^2_{(2)}$	500	-0.048 (0.066)	0.78	-0.058 (0.103)	0.75	-0.053 (0.080)	0.76
	1000	-0.042 (0.044)	0.79	-0.053 (0.067)	0.76	-0.048 (0.058)	0.76
$f_\varepsilon = gg_1$	250	0.057 (0.148)	0.76	0.075 (0.192)	0.73	0.067 (0.162)	0.74
$h = U[-0.5,0]$	500	0.054 (0.088)	0.76	0.067 (0.118)	0.73	0.060 (0.108)	0.73
	1000	0.044 (0.056)	0.77	0.060 (0.081)	0.73	0.053 (0.070)	0.75
$f_\varepsilon = gg_1$	250	-0.058 (0.139)	0.76	-0.073 (0.189)	0.73	-0.066 (0.153)	0.75
$h(x-1.5) = -\chi^2_{(2)}$	500	-0.048 (0.085)	0.77	0.067 (0.119)	0.73	0.058 (0.102)	0.74
	1000	-0.043 (0.057)	0.77	0.061 (0.075)	0.74	0.054 (0.066)	0.75
$f_\varepsilon = gg_{0.5}$	250	0.062 (0.168)	0.74	0.077 (0.218)	0.71	0.065 (0.184)	0.73
$h = U[-0.5,0]$	500	0.056 (0.121)	0.74	0.073 (0.171)	0.70	0.064 (0.133)	0.72
	1000	0.050 (0.070)	0.75	0.067 (0.108)	0.72	0.062 (0.085)	0.73
$f_\varepsilon = gg_{0.5}$	250	-0.061 (0.160)	0.75	-0.076 (0.207)	0.71	0.068 (0.182)	0.73
$h(x-1.5) = -\chi^2_{(2)}$	500	-0.055 (0.105)	0.74	-0.071 (0.165)	0.70	0.063 (0.135)	0.73
	1000	-0.048 (0.069)	0.76	0.062 (0.106)	0.71	0.055 (0.087)	0.73

Table 3.18: $\hat{\beta}_{2,MPHD}$, $\hat{\beta}_{2,G}$ and $\hat{\beta}_{2,NG}$ for GARCH$(1,1)$ with unknown innovation distribution and contamination (3.2).

		$\hat{\beta}_{2,MPHD}$		$\hat{\beta}_{2,G}$		$\hat{\beta}_{2,NG}$	
Contamination (3.2)	n	Bias (MSE)	CP	Bias (MSE)	CP	Bias (MSE)	CP
$f_\varepsilon = gg_4$	250	0.059 (0.132)	0.79	0.080 (0.187)	0.76	0.068 (0.147)	0.77
$h = U[-0.5,0]$	500	0.055 (0.096)	0.79	0.079 (0.137)	0.77	0.067 (0.115)	0.77
	1000	0.048 (0.067)	0.80	0.073 (0.096)	0.78	0.063 (0.085)	0.78
$f_\varepsilon = gg_4$	250	0.057 (0.129)	0.80	0.078 (0.181)	0.78	0.067 (0.145)	0.78
$h(x-1.5) = -\chi^2_{(2)}$	500	0.056 (0.093)	0.78	0.077 (0.129)	0.77	0.064 (0.099)	0.77
	1000	0.046 (0.061)	0.79	0.067 (0.092)	0.77	0.057 (0.072)	0.78
$f_\varepsilon = gg_2$	250	0.060 (0.154)	0.78	0.085 (0.213)	0.76	0.078 (0.169)	0.76
$h = U[-0.5,0]$	500	0.055 (0.105)	0.79	0.073 (0.151)	0.76	0.066 (0.118)	0.77
	1000	0.049 (0.068)	0.79	0.073 (0.106)	0.77	0.062 (0.091)	0.77
$f_\varepsilon = gg_2$	250	-0.057 (0.147)	0.79	-0.082 (0.204)	0.77	-0.076 (0.163)	0.78
$h(x-1.5) = -\chi^2_{(2)}$	500	-0.055 (0.091)	0.78	-0.073 (0.150)	0.75	-0.067 (0.110)	0.76
	1000	-0.048 (0.070)	0.79	-0.067 (0.108)	0.76	-0.061 (0.089)	0.77
$f_\varepsilon = gg_1$	250	0.066 (0.171)	0.77	0.095 (0.253)	0.73	0.084 (0.199)	0.74
$h = U[-0.5,0]$	500	0.062 (0.092)	0.76	0.084 (0.144)	0.72	0.075 (0.129)	0.73
	1000	0.051 (0.065)	0.77	0.075 (0.105)	0.73	0.067 (0.084)	0.74
$f_\varepsilon = gg_1$	250	-0.067 (0.168)	0.77	-0.092 (0.247)	0.73	-0.082 (0.187)	0.75
$h(x-1.5) = -\chi^2_{(2)}$	500	-0.055 (0.088)	0.77	0.084 (0.146)	0.73	0.073 (0.111)	0.74
	1000	-0.050 (0.065)	0.78	0.076 (0.090)	0.74	0.068 (0.078)	0.75
$f_\varepsilon = gg_{0.5}$	250	0.071 (0.193)	0.75	0.097 (0.244)	0.71	0.081 (0.211)	0.73
$h = U[-0.5,0]$	500	0.064 (0.147)	0.73	0.092 (0.207)	0.70	0.080 (0.163)	0.72
	1000	0.057 (0.092)	0.74	0.084 (0.141)	0.71	0.078 (0.112)	0.72
$f_\varepsilon = gg_{0.5}$	250	-0.070 (0.185)	0.75	-0.096 (0.233)	0.71	0.085 (0.188)	0.72
$h(x-1.5) = -\chi^2_{(2)}$	500	-0.063 (0.139)	0.74	-0.089 (0.198)	0.71	0.079 (0.166)	0.73
	1000	-0.055 (0.087)	0.75	0.078 (0.138)	0.72	0.069 (0.104)	0.73

Chapter 4

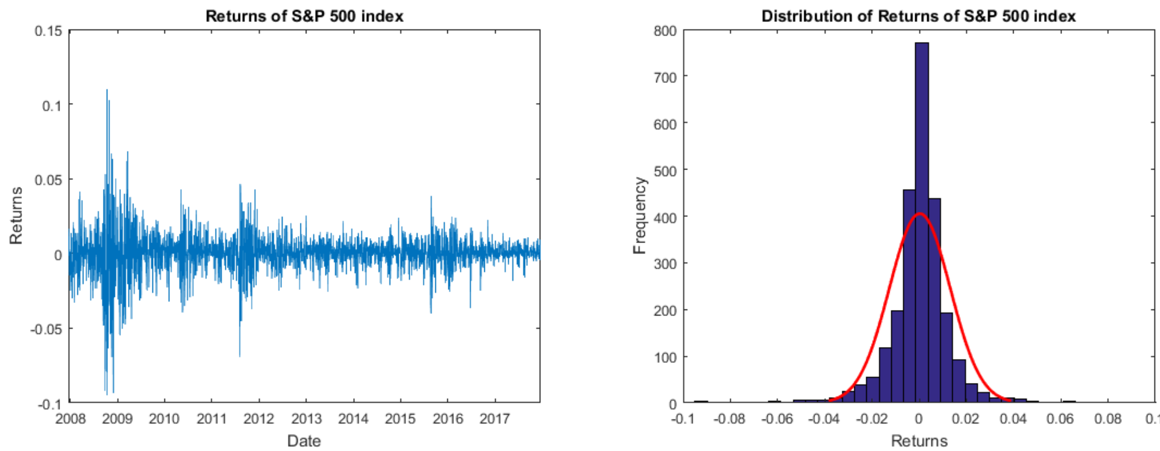
REAL DATA ANALYSIS

In this chapter, we demonstrate the implementation of the proposed estimators through real data analysis. Specifically, we use GARCH$(1,1)$ model to fit the daily (log) returns of S&P 500 index collected from December 18, 2007 to December 18, 2017, which contains 2531 trading days in total. The daily (log) returns X_t is defined as

$$X_t \;=\; \log P_t - \log P_{t-1} = \log\left(1 + \left(\frac{P_t}{P_{t-1}} - 1\right)\right) \approx \frac{P_t - P_{t-1}}{P_{t-1}}, \qquad (4.1)$$

where P_t is the index at the end of trading day t. The approximation is based on $\log(1+x) \approx x$ when x is small. Therefore, the log returns is essentially the percentage change in the asset price. Moreover, the log returns is preferred over the change in price $P_t - P_{t-1}$ because the former is a better indicator of the performance of an asset. For example, consider two assets in which \$10 and \$1000 are invested by investors A and B respectively. Suppose both assets are down \$10 the next day, then the two investors seem to be equally worse off in terms of change in price. However, investor A has lost all of his money while investor B just lost 1% of his initial investment.

Figure 4.1a displays the returns of S&P 500 index over this time period. It is clear from Figure 4.1a that the daily returns exhibit volatility clustering where the returns appear to be more volatile from the year of 2008 to the mid of the year 2009 when the global financial crisis occurred, suggesting the need to use ARCH/GARCH model. We also observe from Figure 4.1a that the returns fluctuate around zero horizontal line which indicates that the unconditional mean of the daily returns is around zero. In Figure 4.1b, we present the histogram of the daily returns, along with the normal curve in red. From it we can see again that the innovation distribution is roughly symmetric about zero with quite a heavy tail. These observations justifies our use of GARCH$(1,1)$ model to fit the data. We use the more general GARCH(1,1) model instead of ARCH model since high-order ARCH model can be approximated by GARCH(1,1) model.



(a) Time series plot

(b) Histogram

Figure 4.1: Returns of S&P 500 index

Since we don't know the exact innovation distribution, we use the MPHDE, GQMLE and NGQMLE to estimate the semiparametric GARCH$(1,1)$ model. The estimates along with their standard errors are reported in Table 4.1. The standard errors are calculated using bootstrap with 500 bootstrapping samples.

From Table 4.1 we can see that β_0 seems to be statistically insignificant while β_1 and β_2 appear to be significant, regardless of estimation method used. Moreover, the MPHDE has the smallest bootstrapping standard errors while the GQMLE has the largest. This might be explained by the fact that the data is 'contaminated' during the global financial crisis in 2008 and 2009. When comparing the parameter estimates from these three methods, we find that MPHDE and NGQMLE are close to each other while the GQMLE is a bit different. From Table 4.1 we can also see that the estimated GARCH process is stable as $\hat{\beta}_1 + \hat{\beta}_2 < 1$ based on either method. Being stable here means that the conditional variance of returns does not grow without bounds over time and the unconditional variance of returns is finite. This is consistent with our visual inspection of Figure 4.1a. On the other hand, the sum $\hat{\beta}_1 + \hat{\beta}_2$ is very close to 1, regardless of estimation method used. This indicates that the shock in the financial market in the current period has quite a persistent effect on the volatility of S&P 500 index in the future period. When $\hat{\beta}_1$ and $\hat{\beta}_2$ are compared, since $\hat{\beta}_2$ is much bigger than $\hat{\beta}_1$, we can firmly conclude that past volatility has a much greater impact than past returns on current volatility. Overall, the analysis results are consistent with those in existing literature (see, for example, Huang [15]).

Table 4.1: Estimates and their standard errors (in parenthesis) of MPHDE, GQMLE and NGQMLE.

	MPHDE	GQMLE	NGQMLE
β_0	1.165×10^{-5}	1.254×10^{-5}	1.283×10^{-5}
	(1.791×10^{-5})	(2.079×10^{-5})	(1.934×10^{-5})
β_1	0.083	0.153	0.093
	(0.027)	(0.075)	(0.031)
β_2	0.889	0.823	0.894
	(0.082)	(0.124)	(0.096)

Chapter 5

SUMMARY AND DISCUSSION

This thesis focuses on the estimation of ARCH(1) and GARCH(1,1) models, for which the MHDE and MPHDE are proposed depending on whether the innovation distribution is known or unknown. In the real-world application, GARCH$(1,1)$ model is usually a reasonable model to consider as it can approximate the high-order ARCH process [9].

Chapter 1 reviews the ARCH/GARCH models and the MHDE and MPHDE. In Chapter 2, we propose and construct the MHDE and the MPHDE for the ARCH/GARCH models. In Section 2.1, when the innovation distribution is known, we construct a MHDE for the unknown parameters. This MHDE is different from the original general definition of MHDE in that the nonparametric p.d.f. estimation and the unknown parameters are mixed together, which in turn increases the difficulty of investigating the asymptotic properties of the resulting estimator. As a consequence, we are not able to prove the consistency of MHDE mathematically. However, we examine the asymptotic properties of MHDE using graphs. Specifically, we show graphically that MHDE can recover accurately the true innovation distribution when the sample size is relatively large. This suggests that MHDE is expected to converge to the true parameter values. Moreover, MHDE does not appear to be asymptotically normally distributed while MLE does when there is no contamination. However, when data is contaminated, the distribution of MLE deviates from normal and that of MHDE becomes less heavy-tailed. In Section 2.2, when the innovation distribution is unknown, we propose the MPHDE by first profiling out the unknown innovation distribution that is assumed to be symmetric. Even though we are not able to prove its consistency from the theoretical point of view, we show graphically that the profile Hellinger distance is an increasing function of the L^2-distance between the parameters and their true value, which implies that the proposed MPHDE should be fairly close to the true parameter values. Graphical demonstration shows that the MPHDE is also capable of recovering accurately the unknown innovation distribution when sample size is relatively large. In addition, the large-sample distribution of GQMLE and NGQMLE no longer appear to be normal and that of MPHDE exhibits less heavy tails when there is contamination.

Chapter 3 examines the finite-sample performance of the proposed MHDE and MPHDE through Monte Carlo simulation studies. The results under ARCH(1) and GARCH$(1,1)$ models are presented in Sections 3.1 and 3.2 respectively. In Section 3.1 for ARCH(1) model, we consider the following four cases: (i) the innovation distribution is known and data is not contaminated; (ii) the innovation distribution is known and data is contaminated; (iii) the innovation distribution is unknown and data is not contaminated; (iv) the innovation distribution is unknown and data is contaminated. We compare the performance of the MHDE with that of the MLE under Cases (i) and (ii). And the performance of MPHDE is compared to that of GQMLE and the NGQMLE under Cases (iii) and (iv). We consider two types of contamination. The first type contaminates the innovation distribution at each time point with 5% contamination rate, while the second type assumes a completely different innovation distribution 5% of the time. The simulation results show that MLE generally has better performance in terms of bias, MSE and CP when the innovation distribution is known and uncontaminated (Case i) while MHDE has an advantage in the first kind of contamination. Although MHDE also outperforms MLE in the second kind of contamination, they both perform poorly. When the innovation distribution is unknown and without contamination (Case iii), NGQMLE generally has the best performance followed by MPHDE if the true innovation distribution is heavy-tailed. However, GQMLE has the best performance followed by NGQMLE if the true innovation distribution is normal, close to normal or light-tailed. On the other hand, when the innovation distribution is known and with contamination (Case iv), MPHDE has the best performance followed by NGQMLE. In Section 3.2 which considers GARCH$(1,1)$ model, we have similar observations and conclusions to those in Section 3.1. In summary, MPHDE is generally recommended in the real-world applications where innovation distribution is unknown due to its robustness to data contamination and the fact that with data contamination GQMLE and NGQMLE no longer enjoy normal distribution, even for large sample.

In Chapter 4, we analyze a real data, which is the daily (log) returns of S&P 500 index from December 18, 2007 to December 18, 2017. We fit a GARCH$(1,1)$ model to the data using MPHDE, GQMLE and NGQMLE, respectively. The three methods give similar estimates, even though the MPHDE and the NGQMLE are close to each other while the GQMLE is a bit different. All the three methods give consistent estimation results compared with those in existing literature in which β_0 is close to zero, β_1 is close to 0.1 and β_2 is close to 0.9.

In my future research, there are two problems that needs to be addressed. First, the consistency of the MHDE and MPHDE needs to be proved rigorously in order to further justify its use in ARCH/GARCH models. Second, even though the MPHDE has been derived under the symmetric assumption of the innovation distribution, we hope that a MPHDE that is free of any distribution assumption can be developed in the future so that it can accommodate asymmetric (actually any) innovation distributions.

Bibliography

[1] Beth Andrews. Rank-based estimation for GARCH processes. *Econometric Theory*, 28:1037–1064, 2012.

[2] Richard T. Baillie and Tim Bollerslev. Intra-day and inter-market volatility in foreign exchange rates. *Review of Economic Studies*, 58:565–585, 1991.

[3] Richard T. Baillie and Huimin Chung. Estimation of GARCH models from the autocorrelations of the squares of a process. *Journal of Time Series Analysis*, 22(6):631–650, 2001.

[4] Rudolf Beran. Minimum Hellinger distance estimates for parametric models. *Annals of Statistics*, 5:445–463, 1977.

[5] Tim Bollerslev. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[6] Tim Bollerslev. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69:542–547, 1987.

[7] Guoqiang Chen and Jingjing Wu. Molecular classification of acute leukemia. In *Some Recent Advances in Mathematics and Statistics*, pages 60–74. World Scientific, 2013.

[8] Adele Cutler and Olga I Cordero-Bra na. Minimum hellinger distance estimation for finite mixture models. *Journal of the American Statistical Association*, 91(436):1716–1723, 1996.

[9] Walter Enders. *Applied Econometric Time Series*. Wiley, fourth edition, 2015.

[10] Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50:987–1008, 1982.

[11] Jianqing Fan, Lei Qi, and Dacheng Xiu. Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *Journal of Business Economic Statistics*, 32:178–191, 2014.

[12] John W Galbraith, Victoria Zinde-Walsh, and Jingmei Zhu. Garch model estimation using estimated quadratic variation. *Econometric reviews*, 34(6-10):1172–1192, 2015.

[13] Christian Gourieroux, Alain Monfort, and Alain Trognon. Pseudo maximum likelihood methods: Theory. *Econometrica: Journal of the Econometric Society*, pages 681–700, 1984.

[14] Peter Hall and Qiwei Yao. Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica*, 71:285–317, 2003.

[15] Kun Huang. Modeling volatility of S&P 500 index daily returns: a comparison between model based forecasts and implied volatility. *Finance and Statistics, Hanken School of Economics*, 2011.

[16] Roger Kadjo, Ouagnina Hill, and Aubin N'dri. Minimum hellinger distance estimation of a univariate GARCH process. *Journal of Mathematics Research*, 9:80–94, 2017.

[17] Dimitris Karlis and Evdokia Xekalaki. Minimum hellinger distance estimation for poisson mixtures. *Computational Statistics & Data Analysis*, 29(1):81–103, 1998.

[18] Rohana J Karunamuni and Jingjing Wu. One-step minimum hellinger distance estimation. *Computational Statistics & Data Analysis*, 55(12):3148–3164, 2011.

[19] Sang-Won Lee and Bruce E. Hansen. Asymptotic theory for the GARCH(1, 1) quasi-maximum likelihood estimator. *Econometric theory*, 10:29–52, 1994.

[20] Sang-Won Lee and Bruce E Hansen. Asymptotic theory for the garch (1, 1) quasi-maximum likelihood estimator. *Econometric theory*, 10(1):29–52, 1994.

[21] Bruce G. Lindsay. Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Annals of Statistics*, 22:1081–1114, 1994.

[22] Zudi Lu, Yer Van Hui, and Andy H. Lee. Minimum hellinger distance estimation for finite mixtures of poisson regression models and its applications. *Biometrics*, 59(4): 1016–1026, 2003.

[23] Robin L. Lumsdaine. Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH (1,1) and covariance stationary GARCH (1,1) models. *Econometrica: Journal of the Econometric Society*, 64:575–596, 1996.

[24] James W. McFarland, R. Richardson Pettit, and Sam K. Sung. The distribution of foreign exchange price changes: Trading day effects and risk measurement. *The Journal of Finance*, 37:693–715, 1982.

[25] Julien Apala N'drin and Ouagnina Hili. Parameter estimation of one-dimensional diffusion process by minimum hellinger distance method. *Random Operators and Stochastic Equations*, 21(4):403–424, 2013.

[26] Daniel B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59:347–370, 1991.

[27] Annabel Prause, Ansgar Steland, and Mohammed Abujarad. Minimum hellinger distance estimation for bivariate samples and time series with applications to nonlinear regression and copula-based models. *Metrika*, 79(4):425–455, 2016.

[28] Peter J. Rousseeuw and Christophe Croux. Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88:1273–1283, 1993.

[29] Douglas G. Simpson. Minimum hellinger distance estimation for the analysis of count data. *Journal of the American Statistical Association*, 82(399):802–807, 1987.

[30] TN Sriram and AN Vidyashankar. Minimum hellinger distance estimation for supercritical galton–watson processes. *Statistics & probability letters*, 50(4):331–342, 2000.

[31] Giuseppe Storti. Minimum distance estimation of GARCH(1, 1) models. *Computational statistics & data analysis*, 51(3):1803–1821, 2006.

[32] Teruko Takada. Simulated minimum hellinger distance estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, 53(6):2390–2403, 2009.

[33] Andrew A. Weiss. Asymptotic theory for ARCH models: estimation and testing. *Econometric theory*, 2:107–131, 1986.

[34] Andrew A Weiss. Asymptotic theory for arch models: estimation and testing. *Econometric theory*, 2(1):107–131, 1986.

[35] Mi-Ja Woo and TN Sriram. Robust estimation of mixture complexity. *Journal of the American Statistical Association*, 101(476):1475–1486, 2006.

[36] Mi-Ja Woo and TN Sriram. Robust estimation of mixture complexity for count data. *Computational Statistics & Data Analysis*, 51(9):4379–4392, 2007.

[37] Wayne A Woodward, Paul Whitney, and Paul W Eslinger. Minimum hellinger distance estimation of mixture proportions. *Journal of Statistical Planning and Inference*, 48 (3):303–319, 1995.

[38] Jingjing Wu and Rohana J Karunamuni. On minimum hellinger distance estimation. *Canadian Journal of Statistics*, 37(4):514–533, 2009.

[39] Jingjing Wu and Rohana J Karunamuni. Efficient hellinger distance estimates for semiparametric models. *Journal of Multivariate Analysis*, 107:1–23, 2012.

[40] Jingjing Wu and Rohana J Karunamuni. Profile hellinger distance estimation. *Statistics*, 49(4):711–740, 2015.

[41] Jingjing Wu and Rohana J Karunamuni. Efficient and robust tests for semiparametric models. *Annals of the Institute of Statistical Mathematics*, pages 1–28, 2017.

[42] Jingjing Wu and Xiaofan Zhou. *Minimum Hellinger distance estimation for a semi-parametric location-shifted mixture model*. Technical Report, 2018.

[43] Jingjing Wu, Rohana Karunamuni, and Biao Zhang. Minimum hellinger distance estimation in a two-sample semiparametric model. *Journal of Multivariate Analysis*, 101 (5):1102–1122, 2010.

[44] Jingjing Wu, Weixin Yao, and Sijia Xiang. Computation of an efficient and robust estimator in a semiparametric mixture model. *Journal of Statistical Computation and Simulation*, 87(11):2128–2137, 2017.

[45] Liming Xiang, Kelvin KW Yau, Yer Van Hui, and Andy H Lee. Minimum hellinger distance estimation for k-component poisson mixture with random effects. *Biometrics*, 64(2):508–518, 2008.

[46] Sijia Xiang, Weixin Yao, and Jingjing Wu. Minimum profile hellinger distance estimation for a semiparametric mixture model. *Canadian Journal of Statistics*, 42(2): 246–267, 2014.

[47] Song Yang. Minimum hellinger distance estimation of parameter in the random censorship model. *The Annals of Statistics*, 19(2):579–602, 1991.

[48] Zhiliang Ying. Minimum hellinger-type distance estimation for censored data. *The Annals of Statistics*, 20(3):1361–1390, 1992.

[49] Yayuan Zhu, Jingjing Wu, and Xuewen Lu. Minimum hellinger distance estimation for a two-sample semiparametric cure rate model with censored survival data. *Computational Statistics*, 28(6):2495–2518, 2013.