THE UNIVERSITY OF THE SOUTH PACIFIC
LIBRARY
Author Statement of Accessibility

Name of Candidate : Dinesh Krishna Rao

Degree : MSc in Statistics

~~Department~~/School : Computing Information and Mathematical Sciences.

Institution/University : U.S.P.

Thesis Title : Asymptotic Distribution of Cramer-von Mises Statistics for ARCH residual Empirical Processes.

Date of completion of requirements for award :

1. This thesis may be consulted in the Library without the author's permission — Yes/No

2. This thesis may be cited without the author's permission providing it is suitably acknowledged. — Yes/No

3. This thesis may be photocopied in whole without the author's written permission — Yes/No

4. This thesis may be photocopied in proportion without the author's written permission — Yes/No
   Part that may be copied:

   Under 10% ---------- 40-60% ---------- ✓

   10-20% ---------- 60-80% ----------

   20-40% ---------- Over 80% ----------

5. I authorise the University to produce a microfilm or microfiche copy for retention and use in the Library according to rules 1-4 above (for security and preservation purposes mainly) — Yes/No

6. I authorise the Library to retain a copy of this thesis in e-format for archival and preservation purposes. — Yes/No

7. After a period of 5 years from the date of publication, the USP Library may issue the thesis in whole or in part, in photostat or microfilm or e-format or other copying medium, without first seeking the author's written permission. — Yes/No

8. I authorise the University to make this thesis available on the Internet for access by authorised users. — Yes/No

Signed:

Date: 06/02/06

Contact Address

C/o School of Computing, Information and Mathematical Sciences

Permanent Address

16, Lemaki st, Vatuwaqa,
Suva.

2001

THE UNIVERSITY OF THE SOUTH PACIFIC
LIBRARY

DIGITAL THESES PROJECT

Author Statement of Accessibility- Part 2- Permission for Internet Access

Name of Candidate : Dinesh Krishna Rao

Degree : MSc in statistics

Department/School : Computing, Information and Mathematical Sc.

Institution/University : U.S.P

Thesis Title : Asymptotic Distribution of Cramer-von Mises Statistics for ARCH residual Empirical Processes

Date of completion of requirements for award :

1. I authorise the University to make this thesis available on the Internet for access by USP authorised users. Yes/No

2. I authorise the University to make this thesis available on the Internet under the International digital theses project Yes/No

Signed:

Date: 06/02/06

Contact Address
c/o School of Computing Information and Mathematical Sciences

e-mail : rao-di@usp.ac.fj

Permanent Address
16, Lemaki, st, Vatuwaqa, Suva.

e-mail: rao-di@usp.ac.fj

July 2004

EF

Asymptotic Distribution of Cramér-von Mises Statistics for ARCH Residual Empirical Processes

Declaration of Originality

I Dinesh Krishna Rao, hereby declare that this thesis, entitled "Asymptotic Distribution of Cramér-von Mises Statistics for ARCH Residual Empirical Processes" is original to the best of my knowledge; citations and references have been acknowledged and the main result formulated has not been previously submitted for a university degree either in whole or in part elsewhere.

Dinesh Krishna Rao

January, 2006.

Contents

Notations

R	: set of all real numbers
T	: transpose
R^m	: set of all real $m \times 1$ vectors
a.e.	: almost everywhere
a.s.	: almost surely
i.i.d.	: independent and identically distributed
r.v.	: random variable
d.f.	: distribution function
$o_p(1)$	: convergence in probability to 0
$O_p(1)$	: bounded in probability
CLS	: conditional least squares
min	: minimum, minimize
sup	: supremum
e	: exponential
$\in$	: belongs to

$\subset$	: is a set of
Θ	: parameter space
d	: differential
$\|\mathsf{a}\|$	: norm of a vector a
$\|a\|$	: absolute value of a scalar a
$\mathbb{I}(A)$	: indicator function of an event A
$\phi'(\nu)$	: partial derivative of $\phi(\nu)$
$\phi''(\nu)$	: second partial derivative of $\phi(\nu)$
$\mathcal{F}$	: σ-field
$E(X)$	: expectation of a r.v. X
$E(X\|F)$	: conditional expectation of r.v. X with respect to F
$Var(X)$	: variance of a r.v. X
$Var(X\|F)$	: conditional variance of r.v. X with respect to F
$Cov(X,Y)$	: covariance of X and Y
$X_n \xrightarrow{a.s.} X$	: X_n converges almost surely to X
$X_n \xrightarrow{p} X$	: X_n converges in probability to X
$X_n \xrightarrow{d} X$	: X_n converges in distribution to X
$N(\mu,\sigma^2)$	: normal distribution with mean μ and variance σ^2
$\Phi(x)$	: standard normal probability distribution function
$\{X_t\}$	: a time series process

$\{\widehat{T}_N\}$	: Cramér-von Mises statistics
AR	: autoregressive process
MA	: moving average process
ARMA	: autoregressive moving average process
TAR	: Threshold autoregressive process
ARCH(p)	: autoregressive conditional heteroskedastic process of order p
GARCH	: generalized ARCH process

Abstract

In this thesis, the limiting Gaussian distribution of a class of Cramér-von Mises statistics $\{\widehat{T}_N\}$ for two-sample problem pertaining to empirical processes of the squared residuals from two independent samples of ARCH processes is elucidated. A distinctive feature is that, unlike the residuals of ARMA processes, the asymptotics of $\{\widehat{T}_N\}$ depend on those of ARCH volatility estimators. Based on the asymptotics of $\{\widehat{T}_N\}$, we numerically assess the relative asymptotic efficiency and ARCH volatility effect for some ARCH residual distributions. Moreover, a measure of robustness for $\{\widehat{T}_N\}$ is introduced. Then this aspect of $\{\widehat{T}_N\}$ based on such residual distributions is illustrated numerically. In contrast with the i.i.d. or ARMA settings, these studies illuminate some interesting features of ARCH residuals.

Key phrases

ARMA process; ARCH process; squared residuals; empirical process; two-sample Cramér-von Mises statistic; asymptotic normality; asymptotic relative efficiency; ARCH volatility effect; robustness.

Preface

In the i.i.d. settings, two-sample problem is one of the important statistical problems. For this problem, the study of the asymptotic properties based on the celebrated Cramér-von Mises statistics is fundamental and an essential part of nonparametric statistics. Many researchers have contributed to their development, and numerous theorems have been formulated in many testing problems. Most of the techniques employed in one-sample case have very close counterparts in the two-sample situation.

For a two-sample Cramér-von Mises statistic in the i.i.d. settings, Anderson (1962) derived the exact distribution and he compared its limiting distribution with the exact one, and found that a good approximation to the exact distribution for moderate sample sizes. He also reported that the accuracy of his approximation is better than that of the two-sample Kolmogrov-Smirnov statistics studied by Hodges (1957). An excellent account of Cramér-von Mises tests is given in Durbin (1973) and we refer the reader to this reference for details and further references.

The main object of this thesis is to elucidate the asymptotic theory of the two-sample Cramér-von Mises statistics $\{\widehat{T}_N\}$ for ARCH residual empirical processes based on the techniques of Chernoff and Savage (1958) and Horváth *et al.* (2001). Since the asymptotics of the residual empirical processes are different from those for the usual ARMA case, the limiting distribution of $\{\widehat{T}_N\}$ is greatly different from that of ARMA case (of course i.i.d. case). More concretely, the thesis is organized as follows.

Chapter 1 provides the introduction and summary of the thesis. It briefly discusses some basic and important results, which will help to better understand the main result formulated in this thesis.

Chapter 2 gives the setting of $\{\widehat{T}_N\}$ pertaining to empirical processes based on the squared residuals from two independent samples of ARCH(p) processes and establishes its limiting Gaussian distribution.

This result, in Chapter 3 facilitates the study of asymptotic performance of $\{\widehat{T}_N\}$, like the relative asymptotic efficiency and ARCH volatility effect for some ARCH residual distributions. Moreover, we introduce a robustness measure for $\{\widehat{T}_N\}$ by means of the influence function. Then this aspect of $\{\widehat{T}_N\}$ based on such residual distributions is illustrated by simulations.

Chapter 4 gives the proof of our theorem formulated in Chapter 2.

Finally, Chapter 5 provides the concluding remarks and gives a brief outline of the related research that can be carried out in future.

Chapter 1

Introduction and Summary

In this chapter, we provide the introduction and summary to the thesis. In particular, the preliminary concepts, CLS estimation, ARCH processes, background and related work, and motivation. These aspects facilitates the understanding of the main result (Theorem 2) given in Chapter 2.

1.1 Preliminary Concepts

1.1.1 Time Series Analysis

A time series $\{X_t\}$ is a sequence of values of a variable at equally spaced time interval t. Statisticians usually view a time series as a realization from a stochastic process. One distinguishing feature in time series is that the records are usually dependent. Due to different applications, the data may be collected hourly, daily weekly, monthly, or yearly, and so on. The objectives of time series analysis are diverse, depending on the background of applications. The main objectives of time series analysis are to understand the underlying dynamics and structure that produced the observed data, forecast future events, and control future events via intervention. Time series analysis is used for many applications such as economic forecasting, sales forecasting, budgetary analysis, stock market analysis, yield projections, process and quality control, inventory studies, workload projections, utility studies, census studies and so forth.

A common assumption in many time series techniques is that the data is stationary. A stationary process has the property that the mean, variance and autocorrelation structure do not change overtime. Stationarity can be defined in precise mathematical terms, but for our purpose we mean a flat looking series, without trend, constant variance over time, a constant autocorrelation structure over time and no periodic fluctuations (seasonality).

A Discrete-time series is one in which the set t of times at which observations are made at fixed time intervals, e.g., $t = 1, \ldots, m$. A Continuous-time series is obtained when observations are recorded continuously over some time interval, e.g., $t = [0, 1]$.

Linear Time Series Models

The most popular class of linear time series models is the autoregressive moving average (ARMA) models, which includes the autoregressive (AR) and moving average (MA) models as special cases. The ARMA model is the most commonly used to model linear dynamic structures to depict linear relationships among lagged variables, and to serve as vehicles for linear forecasting.

Nonlinear Time Series Models

The long-lasting popularity of ARMA models convincingly justifies the usefulness of linear models for analyzing time series data. Nevertheless, in view of the fact that any statistical model is an approximation to the real world, a linear model is merely a first step in representing an unknown dynamic relationship in terms of a mathematical formula. The truth is that the world is nonlinear! Therefore, it is not surprising that there exists an abundance of empirical evidence indicating the limitation of the linear ARMA family, when applied to the field of financial and monetary economics. To model a number of nonlinear features such as dependence beyond linear correlation, we need to appeal to nonlinear models. Several typical examples of nonlinear models are ARCH, Threshold autoregressive (TAR), generalized autoregressive conditional heteroskedastic (GARCH) and exponential ARCH models.

1.1.2 Convergence and Bounded in Probability

Concepts of relative magnitude or order of magnitude are useful in investigating limiting behavior of r.v.'s. We first define the concepts of order as used in real analysis. Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers and $\{b_n\}_{n=1}^{\infty}$ be a sequence of positive real numbers (see e.g., Sen and Singer (1993)).

Convergence in Probability to Zero

Definition 1. *We say that a_n converges in probability to zero, written $a_n = o_p(1)$ or $a_n \xrightarrow{p} 0$, if for every $\epsilon > 0$,*

$$P(|a_n| > \epsilon) \to 0 \quad as \quad n \to \infty.$$

Definition 2. *We say that $a_n = o(b_n)$ as $n \to \infty$ if*

$$a_n/b_n \to 0.$$

When a_n and b_n tend to infinity, this states that a_n tends to infinity more slower than b_n; when both tends to 0, it states that a_n tends to 0 much faster than b_n.

Note: $o(a_n)$ denotes any quantity tending to 0 faster than a_n.

Bounded in Probability

Definition 3. *A sequence $\{a_n\}$ is bounded in probability (or tight), written $a_n = O_p(1)$ if for every $\epsilon > 0$, there exists $\delta(\epsilon) \in (0, \infty)$ such that*

$$P(|a_n| > \delta(\epsilon)) < \epsilon \quad \textit{for all } n.$$

Note:

1. $a_n = O(b_n)$ means that a_n is of order smaller than or equal to that of b_n.

2. $a_n = O(b_n)$ if $|a_n/b_n|$ is bounded.

Related Properties:

(i) a_n converges in probability to a, $a \in R$, written $a_n \xrightarrow{p} a$, if and only if $a_n - a = o_p(1)$.

(ii) $a_n = o_p(b_n)$ if and only if ${b_n}^{-1} a_n = o_p(1)$.

(iii) $a_n = O_p(b_n)$ if and only if ${b_n}^{-1} a_n = O_p(1)$.

1.1.3 Taylor Expansion in Probability

Let $\{X_n\}$ be a sequence of random variables such that $X_n = a + O_p(1)$, where $a \in R$. If g is continuous at a then $g(X_n) = g(a) + o_p(1)$. If we strengthen the assumptions on g to include the existence of derivatives, then it is possible to derive probabilistic analogues of the Taylor expansions of non-random functions about a given point a. Now, $X_n = a + O_p(r_n)$, where $0 < r_n \to 0$ as $n \to \infty$. If g is a function with s derivatives at a, then,

$$G(X_n) = \sum_{j=0}^{s} \frac{g^{(j)}(a)}{j!}(X_n - a)^j + o_p(r_n^s),$$

where $g^{(j)}$ is the j^{th} derivative of g and $g^{(0)} = g$.

1.1.4 Empirical Distribution Function

Let X be a real-valued r.v. with d.f. $F(= \{F(x) : x \in R\})$. Consider a sample of m i.i.d. r.v.'s $\{X_1, X_2, ...X_m\}$ drawn from the d.f. F. Write $F_m(x) = m^{-1} \sum_{i=1}^{m} \mathbb{I}_{\{X_i \le x\}}$. Then $mF_m(x)$ is the number of $X_i's$, $1 \le i \le m$ that are $\le x$. The quantity $F_m(x)$ is called the sample or empirical distribution function. We note that $0 \le F_m(x) \le 1$ for all x, and moreover, that F_m is right continuous, nondecreasing, and $F_m(-\infty) = 0$ and $F_m(\infty) = 1$.

If $X_{(1)}, X_{(2)}, ..., X_{(m)}$ is the ordered statistic for $X_1, X_2, ..., X_m$, then

$$F_m(x) = \begin{cases} 0, & \text{if } x < X_{(1)} \\ \frac{k}{m}, & \text{if } X_{(k)} \le x < X_{(k+1)} \\ 1, & \text{if } x \ge X_{(m)} \end{cases} \qquad (k = 1, 2, ..., m-1).$$

The r.v. $F_m(x)$ has the probability function

$$P\left[F_m(x) = \frac{j}{m}\right] = \binom{m}{j} [F(x)]^j [1 - F(x)]^{m-j}, \ j = 0, 1, ..., m,$$

with

$$\begin{aligned} E(F_m(x)) &= m^{-1} \sum_{i=1}^{m} E\{\mathbb{I}_{\{X_i \le X\}}\} = m^{-1} \sum_{i=1}^{m} P\{X_i \le x\} = F(x), \\ Var(F_m(x)) &= \frac{F(x)[1 - F(x)]}{m}, \end{aligned}$$

and

$$E\{[F_m(x) - F(x)]\}\{[F_m(y) - F(y)]\} = \frac{F(x)\{1 - F(y)\}}{m}, \quad x \le y.$$

It is also known that

$$F_m(x) \xrightarrow{p} F(x) \quad \text{as } m \to \infty,$$

$$\frac{\sqrt{m}[F_m(x) - F(x)]}{\sqrt{F(x)[1 - F(x)]}} \xrightarrow{d} \mathcal{N}(0, 1) \quad \text{as } m \to \infty,$$

(see e.g., Sen and Singer (1993)).

1.2 Conditional Least Squares Estimation

In this section, we state Tjøstheim's theorem (1986) which was essentially obtained by reformulating and extending the arguments of Klimko and Nelson (1978) to nonlinear time series.

Let $\{X_t; t = 0, \pm 1, ...\}$ be a strictly stationary and ergodic process taking values in R^p and defined on the probability space $(\Omega, \mathcal{F}, P)$. Here, $\{X_t\}$ is possibly a strictly stationary ergodic nonlinear time series. In addition, suppose that $E\{\|X_t\|^2\} < \infty$ so that $\{X_t\}$ is second order stationary, where $\|.\|$ denotes the Euclidean norm. We assume that observations $(X_1, ..., X_n)$ are available. The probability distribution of $(X_1, ..., X_n)$ is specified by unknown parameter $\theta = (\theta_1, ..., \theta_q)^T \in \Theta \subset R^q$. Its true value is denoted by θ^0. Then consider a general real-valued penalty function $\mathcal{Q}_n(\theta) = \mathcal{Q}_n(X_1, ..., X_n; \theta)$ depending on the observations and $\theta \in \Theta$.

Let us now specify the penalty function. Let $\mathcal{F}_t(m)$ be the σ-field generated by $\{X_s;\, t - m \le s \le t\}$, where m is an appropriate integer. If $\{X_t\}$ is a nonlinear autoregressive model of order k, we can take $m = k$. Let $\tilde{X}_{t|t-1}(\theta) = E_\theta\{X_t | \mathcal{F}_{t-1}(m)\}$ be an optimal one-step least squares predictor of X_t based on $X_{t-1}, ..., X_{t-m}$. Then the penalty function becomes

$$\mathcal{Q}_n(\theta) = \sum_{t=m+1}^{n} \{X_t - \tilde{X}_{t|t-1}(\theta)\}^T \{X_t - \tilde{X}_{t|t-1}(\theta)\}.$$

The conditional least squares (CLS) estimator $\hat{\theta}_n^{(CL)}$ of θ is defined by

$$\hat{\theta}_n^{(CL)} = \arg\min_{\theta \in \Theta} \mathcal{Q}_n(\theta).$$

Hence, we have the following theorem.

Theorem 1. *(Tjøstheim, 1986). Suppose that $\{X_t\}$ is a p-dimensional strictly stationary process with $E\{\|X_t\|^2\} < \infty$ and that $\tilde{X}_{t|t-1}(\theta) = E_\theta\{X_t | \mathcal{F}_{t-1}(m)\}$ is almost surely three times continuously differentiable with respect to θ in an open* *set Θ containing θ^0. Moreover, suppose that the following conditions hold:*

(i)

$$E\left\{\left\|\frac{\partial}{\partial\theta_i}\tilde{X}_{t|t-1}(\theta^0)\right\|^2\right\} < \infty \quad \text{and} \quad E\left\{\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{X}_{t|t-1}(\theta^0)\right\|^2\right\} < \infty$$

for $i, j = 1, \ldots, p$.

(ii) The vectors $\partial\tilde{X}_{t|t-1}(\theta^0)/\partial\theta_i$, $i = 1, \ldots, p$, are linearly independent in the sense that if $c_1, \ldots, c_p$, are arbitrary real numbers such that

$$E\left\{\left\|\sum_{i=1}^{p} c_i \frac{\partial}{\partial\theta_i}\tilde{X}_{t|t-1}(\theta^0)\right\|^2\right\} = 0,$$

then $c_1 = \cdots = c_p = 0$.

(iii) For $\theta \in \Theta$, there exist functions $G_{t-1}^{ijk}(X_1, \ldots, X_{t-1})$ and $H_t^{ijk}(X_1, \ldots, X_t)$ such that

$$\left|\frac{\partial}{\partial\theta_i}\tilde{X}_{t|t-1}^T(\theta)\frac{\partial^2}{\partial\theta_j\partial\theta_k}\tilde{X}_{t|t-1}(\theta)\right| \leq G_{t-1}^{ijk}, \quad E(G_{t-1}^{ijk}) < \infty,$$

$$\left|\{X_t - \tilde{X}_{t|t-1}(\theta)\}^T\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}\tilde{X}_{t|t-1}(\theta)\right| \leq H_t^{ijk}, \quad E(H_t^{ijk}) < \infty,$$

for $i, j, k = 1, \ldots, p$.

(iv)

$$\begin{aligned} R &= E\left\{\frac{\partial}{\partial\theta}\tilde{X}_{t|t-1}^T(\theta^0)\{X_t - \tilde{X}_{t|t-1}(\theta^0)\}\right. \\ &\quad \left.\times\{X_t - \tilde{X}_{t|t-1}(\theta^0)\}^T\frac{\partial}{\partial\theta}\tilde{X}_{t|t-1}(\theta^0)\right\} < \infty. \end{aligned}$$

Then there exists a sequence of estimators $\hat{\theta}_n^{(CL)}$ such that $\hat{\theta}_n^{(CL)} \xrightarrow{a.s} \theta^0$ as $n \to \infty$, and for any $\epsilon > 0$, there exists an event $\mathcal{E}$ with $P(\mathcal{E}) > 1 - \epsilon$ and an n_0 such that on $\mathcal{E}$, for $n > n_0$, $(\partial/\partial\theta)\mathcal{Q}_n(\hat{\theta}_n^{(CL)}) = 0$, and $\mathcal{Q}_n$ attains a relative minimum at

$\hat{\theta}_n^{(CL)}$. *Moreover, if there exists a positive integer* m *satisfying* $E_\theta\{X_t|\mathcal{F}_t(m)\} = E_\theta(X_t|\mathcal{F}_t)$, *where* $\mathcal{F}_t$ *is the* σ*-field generated by* $\{X_s, s \le t\}$, *then as* $n \to \infty$,

$$\sqrt{n}(\hat{\theta}_n^{(CL)} - \theta^0) \xrightarrow{d} \mathcal{N}(0, U^{-1}RU^{-1}),$$

where

$$U = E\left\{\frac{\partial}{\partial\theta}\tilde{X}_{t|t-1}^T(\theta^0)\frac{\partial}{\partial\theta}\tilde{X}_{t|t-1}(\theta^0)\right\}.$$

The conditional least squares (CLS) estimation approach provides a unified treatment of estimation problems for widely used classes of nonlinear time series.

1.3 ARCH Process

Models that make use of recent available information will be able to forecast better than other models that do not take into account this information. This is one of the reasons why these models benefit particularly from focussing on establishing the difference between conditional and unconditional moments. Conventional econometric models do not allow for a conditional variance whose values depend on the past information, so volatility clustering is not a phenomenon that can be understood with the aid of these traditional models.

Analysis of financial data has received a considerable amount of attention in the literature during the past two decades. Several models have been suggested to capture special features of financial data and most of these models have the property that the conditional variance depends on the past. One of the well known and most heavily used examples is the class of ARCH(p) processes, introduced by Engle (1982) to model the volatility of the UK inflation data. Since then, ARCH related processes have become perhaps the most popular and extensively studied financial econometric models (Engle (1995), Tsay (2002), Chandra and

Taniguchi (2003)). An ARCH(p) process is characterized by the equations

$$X_t = \begin{cases} \sigma_t(\theta)\varepsilon_t, \ \sigma_t^2(\theta) = \theta^0 + \sum\limits_{i=1}^{p} \theta^i X_{t-i}^2, & t = 1, \ldots, m, \\ 0, & t = -p+1, \ldots, 0. \end{cases} \tag{1.1}$$

where $\{\varepsilon_t\}$ is a sequence of i.i.d.(0,1) random variables with fourth-order cumulant κ_4, $\theta = (\theta^0, \theta^1, ..., \theta^p)^T \in \Theta \subset R^{p+1}$ is an unknown parameter vector satisfying $\theta^0 > 0$, $\theta^i \geq 0$, $i = 1, \ldots, p$, and ε_t is independent of $X_s, s < t$.

Since traditional time series models assume a constant one-period forecast variance, the ARCH model was introduced to overcome this implausible assumption. The process $\{X_t\}$ is serially uncorrelated with zero mean and nonconstant variance conditional on the past values.

It became clear that ARCH models could efficiently and quite easily represent the typical empirical findings in financial time series, e.g the conditional heteroskedasticity. Financial time series present nonlinear dynamic characteristics and the ARCH models offer a more adaptive framework for this type of problem. In particular after the collapse of the Bretton Woods system and the implementation of flexible exchange rates in the seventies ARCH models are increasingly used by researchers and practitioners. The financial data is known to have fat tailed distributions and volatility clustering. It has been shown that realizations from ARCH type models can exhibit this behavior so that it is of interest to consider the implications if financial data follow ARCH models. As the name suggests, the model has the following properties:

- Autoregression- uses previous estimates of volatility to calculate subsequent future values. Hence volatility values are closely related.

- Heteroskedasticity- the probability distributions of the volatility varies with the current value.

Existing literature assumes as a minimal requirement of $\{X_t\}$ to be ergodic or stationary so that the laws of large numbers can be applied. Moreover the generic assumption for asymptotic normality is that the squared error process has finite variance. From the view point of statistical theory, ARCH models may be considered as a specific nonlinear time series models which allow for an exhaustive study of the underlying dynamics. The literature on the subject is so vast that we restrict ourselves to directing the reader to fairly comprehensive reviews by Bollerslev *et al.* (1992) and Shepard (1996). A detail treatment of ARCH models at a textbook level is also given by Gouriéroux (1997).

Simulated ARCH Graphs

Let us consider the ARCH(1) process defined by the equations

$$X_t = \begin{cases} \sigma_t(\theta)\varepsilon_t, \quad \sigma_t^2(\theta) = \theta^0 + \theta^1 X_{t-1}^2 & \text{for} \quad t = 1, \ldots, m, \\ 0 & \text{for} \quad t \leq 0, \end{cases}$$

where $\{\varepsilon_t\}$ is a sequence of i.i.d. $(0,1)$ r.v., $\theta = (\theta^0, \theta^1)^T, \theta^0 > 0, 0 \leq \theta^1 < 1$, and ε_t is independent of $X_s, s < t$.

For values $\theta^0 = 0.2, \theta^1 = 0; 0.8$ and $n = 100$, the graphs are plotted in Figures 1 and 2. It is apparent from these graphs the effect on the appearance of the time series $\{X_t\}$ of varying the parameter θ^1.

Figure 1 displays white noise ($\theta^1 = 0$). A series with no autocorrelation looks choppy and patternless to the eye; the value of the observation gives no information about the value of the next observation. In Figure 2, the series seems smoother, with observations above or below the mean often appearing in clusters of modest duration.

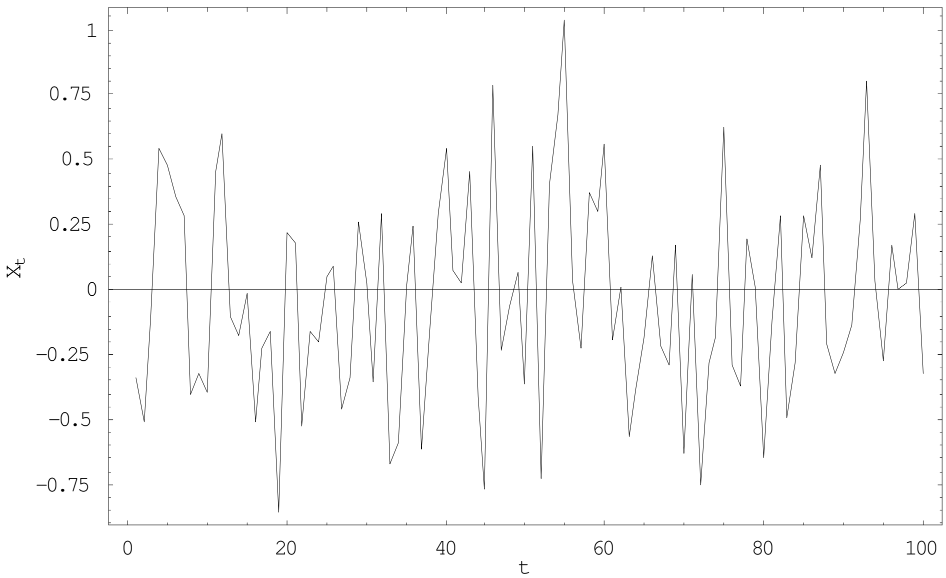

Figure 1: $\theta^0 = 0.2, \theta^1 = 0, n = 100$

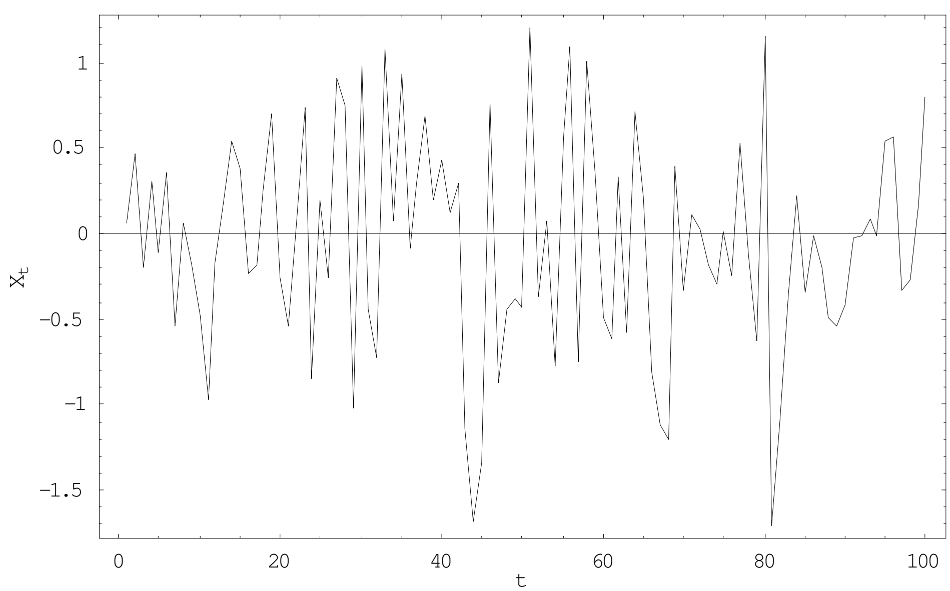

Figure 2: $\theta^0 = 0.2, \theta^1 = 0.8, n = 100$

1.4 Background and Related Work

One of the first ARCH papers was on "ARCH with Estimates of the Variance of UK Inflation" by Engle, published in *Econometrica* in 1982. Traditional econometric models assume a constant one-period forecast variance. To overcome this implausible assumption, this new class of stochastic processes called ARCH processes was used to estimate the means and variances of the UK inflation data. While some aspects of this paper have been bypassed by subsequent research, it remains an excellent introduction.

Since then, variations, extensions, and applications of this model have been breathtaking and intimidating. Many papers have appeared in many different places and have been applied to many different settings. The implementation of this model is relatively simple and from practical a point of view, it's well known how to identify, estimate and test this model.

In the estimation of ARCH model, Tjøstheim (1986) proposed a conditional least squares (CLS) estimator, and discussed its asymptotics. As another method, Godambe (1985) developed the theory of estimating function and introduced the concept of asymptotically optimal estimating function (see e.g., Chandra (2001a)).

For an ARCH(p) process, Horváth *et al.* (2001) derived the limiting distribution of the empirical process based on the squared residuals which is considered of fundamental importance for statistical analysis. Then they showed that, unlike the residuals of ARMA models, these residuals do not behave in this context like asymptotically independent random variables, and the asymptotic distribution involves a term depending on estimators of the volatility parameters of the process. Also Lee and Taniguchi (2005) proved the local asymptotic normality for ARCH(∞) models, and discussed the residual empirical process for an ARCH(p) model with stochastic mean.

Further, Giraitis *et al.* (2000) discussed a class of ARCH(∞) models, which includes that of ARCH(p) models as a special case, and established sufficient conditions for the existence of a stationary solution and gave its explicit representation.

For a two-sample Cramér-von Mises statistic in the i.i.d. settings, Anderson (1962) derived the exact distribution and he compared its limiting distribution with the exact one, and found that a good approximation to the exact distribution for moderate sample sizes. He also reported that the accuracy of his approximation is better than that of the two-sample Kolmogrov-Smirnov statistics studied by Hodges (1957).

More recently, Chandra and Taniguchi (2003) elucidated the asymptotics of the rank order statistics for ARCH residual empirical processes. Most of the techniques employed in the preceding paper have very close counterparts in this thesis.

1.5 Motivation

As volatility ebbs and flows in financial market and as more and more volatility and correlation- dependent securities are priced and traded, the demand for good models, processes and forecasts pushes the research forward.

Thus this study motivates us to consider two independent samples from ARCH(p) processes $\{X_t\}$ (a target process) and $\{Y_t\}$. The corresponding squared innovation processes are, say, $\{U_{\mathrm{x},t}\}$ and $\{U_{\mathrm{y},t}\}$ with possibly non-Gaussian distributions F and G. To know the possible differences between these distributions, an appropriate nonparametric technique is employed based on a class of Cramér-von Mises statistics $\{\widehat{T}_N\}$. Such statistics serves as a basis for the comparison in terms of tests of goodness of fit.

Chapter 2

Two-sample Cramér-von Mises statistics and main result

In this chapter we study a class of Cramér-von Mises statistics (see e.g., Durbin (1973, p.44)) for two-sample problem pertaining to empirical processes based on the squared residuals from two independent samples of ARCH processes.

2.1 Two Independent ARCH Processes

A class of ARCH(p) processes is characterized by the equations

$$X_t = \begin{cases} \sigma_t(\theta_{\rm x})\varepsilon_t, \ \sigma_t^2(\theta_{\rm x}) = \theta_{\rm x}^0 + \sum_{i=1}^{p_{\rm x}} \theta_{\rm x}^i X_{t-i}^2, & t = 1, \ldots, m, \\ 0, & t = -p_{\rm x}+1, \ldots, 0. \end{cases} \tag{2.1}$$

For this model, we impose the following condtions.

Assumption 1.

(i) $\{\varepsilon_t\}$ is a sequence of i.i.d.(0,1) random variables with fourth-order cumulant $\kappa_4^{\rm x}$.

(ii) $\theta_{\rm x} = (\theta_{\rm x}^0, \theta_{\rm x}^1, ..., \theta_{\rm x}^{p_{\rm x}})^T \in \Theta_{\rm x} \subset R^{p_{\rm x}+1}$ is an unknown parameter vector satisfying $\theta_{\rm x}^0 > 0$, $\theta_{\rm x}^i \geq 0$, $i = 1, \dots, p_{\rm x} - 1$, $\theta_{\rm x}^{p_{\rm x}} > 0$.

(iii) $\theta_{\rm x}^1 + \cdots + \theta_{\rm x}^{p_{\rm x}} < 1$ for stationarity (see Milhøj (1985)).

(iv) ε_t is independent of $X_s, s < t$.

(v) $F(x)$ is the distribution function of ε_t^2 and its density $f(x) = F'(x)$ is continuous on $(0, \infty)$.

Another class of ARCH(p) processes, independent of $\{X_t\}$, is defined similarly by the equations

$$Y_t = \begin{cases} \sigma_t(\theta_{\rm y})\xi_t, \ \sigma_t^2(\theta_{\rm y}) = \theta_{\rm y}^0 + \sum_{i=1}^{p_{\rm y}} \theta_{\rm y}^i Y_{t-i}^2, & t = 1, \dots, n, \\ 0, & t = -p_{\rm y} + 1, \dots, 0, \end{cases} \tag{2.2}$$

satisfying the following conditions.

Assumption 2.

(i) $\{\xi_t\}$ is a sequence of i.i.d.(0,1) random variables with fourth-order cumulant $\kappa_4^{\rm y}$.

(ii) $\theta_{\rm y} = (\theta_{\rm y}^0, \theta_{\rm y}^1, ..., \theta_{\rm y}^{p_{\rm y}})^T \in \Theta_{\rm y} \subset R^{p_{\rm y}+1}$, $\theta_{\rm y}^0 > 0$, $\theta_{\rm y}^i \geq 0$, $i = 1, \dots, p_{\rm y} - 1$, $\theta_{\rm y}^{p_{\rm y}} > 0$, are unknown parameters.

(iii) $\theta_{\rm y}^1 + \cdots + \theta_{\rm y}^{p_{\rm y}} < 1$ for stationarity.

(iv) ξ_t is independent of $Y_s, s < t$.

(v) $G(x)$ is the distribution function of ξ_t^2 and $g(x) = G'(x)$ is continuous on $(0, \infty)$.

2.2 Two-Sample Problem and ARCH Estimation

In the following, we are concerned with the two-sample problem of testing

$$H_0 : F(x) = G(x) \text{ for all } x \quad \text{against} \quad H_A : F(x) \neq G(x) \text{ for some } x. \tag{2.3}$$

We first consider the estimation of θ_{x} and θ_{y}. Write $\zeta_{\mathrm{x},t} = (\varepsilon_t^2 - 1)\theta_{\mathrm{x}}^T W_{\mathrm{x},t-1}$ and $Z_{\mathrm{x},t} = X_t^2$, $W_{\mathrm{x},t} = (1, Z_{\mathrm{x},t}, \ldots, Z_{\mathrm{x},t-p_{\mathrm{x}}+1})^T$. Then the autoregressive representation is given by

$$Z_{\mathrm{x},t} = \theta_{\mathrm{x}}^T W_{\mathrm{x},t-1} + \zeta_{\mathrm{x},t}, \quad 1 \leq t \leq m,$$

and analogously for (2.2),

$$Z_{\mathrm{y},t} = \theta_{\mathrm{y}}^T W_{\mathrm{y},t-1} + \zeta_{\mathrm{y},t}, \quad 1 \leq t \leq n,$$

where $\zeta_{\mathrm{y},t} = (\xi_t^2 - 1)\theta_{\mathrm{y}}^T W_{\mathrm{y},t-1}$ and $Z_{\mathrm{y},t} = Y_t^2$, $W_{\mathrm{y},t} = (1, Z_{\mathrm{y},t}, \ldots, Z_{\mathrm{y},t-p_{\mathrm{y}}+1})^T$. Note that $\zeta_{\mathrm{x},t}$ and $\zeta_{\mathrm{y},t}$ are the martingale difference since

$$E(\zeta_{\mathrm{x},t} | \mathcal{F}_{t-1}^{\mathrm{x}}) = E(\zeta_{\mathrm{y},t} | \mathcal{F}_{t-1}^{\mathrm{y}}) = 0,$$

where $\mathcal{F}_t^{\mathrm{x}} = \sigma\{Z_{\mathrm{x},t}, Z_{\mathrm{x},t-1}, ...\}$ and $\mathcal{F}_t^{\mathrm{y}} = \sigma\{Z_{\mathrm{y},t}, Z_{\mathrm{y},t-1}, ...\}$. Suppose that observed stretches $Z_{\mathrm{x},1}, ..., Z_{\mathrm{x},m}$ and $Z_{\mathrm{y},1}, ..., Z_{\mathrm{y},n}$ from $\{Z_{\mathrm{x},t}\}$ and $\{Z_{\mathrm{y},t}\}$, respectively, are available. Thus from Theorem 1, the corresponding conditional least squares estimators of θ_{x} and θ_{y} are given by

$$\hat{\theta}_{\mathrm{x},m} = (\hat{\theta}_{\mathrm{x},m}^0, \ldots, \hat{\theta}_{\mathrm{x},m}^{p_{\mathrm{x}}})^T = \arg\min_{\theta_{\mathrm{x}}} \mathcal{Q}_m(\theta_{\mathrm{x}})$$

and

$$\hat{\theta}_{\mathrm{y},n} = (\hat{\theta}_{\mathrm{y},n}^0, \ldots, \hat{\theta}_{\mathrm{y},n}^{p_{\mathrm{y}}})^T = \arg\min_{\theta_{\mathrm{y}}} \mathcal{Q}_n(\theta_{\mathrm{y}}),$$

where

$$\mathcal{Q}_m(\theta_{\mathrm{x}}) = \sum_{t=1}^{m} (Z_{\mathrm{x},t} - \theta_{\mathrm{x}}^T W_{\mathrm{x},t-1})^2 \quad \text{and} \quad \mathcal{Q}_n(\theta_{\mathrm{y}}) = \sum_{t=1}^{n} (Z_{\mathrm{y},t} - \theta_{\mathrm{y}}^T W_{\mathrm{y},t-1})^2.$$

Here, we assume that $\hat{\theta}_{\mathrm{x},m}$ and $\hat{\theta}_{\mathrm{y},n}$ are asymptotically consistent and normal with rate $m^{-1/2}$ and $n^{-1/2}$, respectively, i.e.,

$$m^{1/2}\|\hat{\theta}_{\mathrm{x},m} - \theta_{\mathrm{x}}\| = O_p(1) \quad \text{and} \quad n^{1/2}\|\hat{\theta}_{\mathrm{y},n} - \theta_{\mathrm{y}}\| = O_p(1). \tag{2.4}$$

where $\|.\|$ denotes the Euclidean norm. For validity of (2.4), (Tjøstheim (1986), pp.254-256) gave a set of sufficient conditions. Conditions (2.4) are also satisfied by the pseudo-maximum likelihood and conditional likelihood estimators (see e.g., Gouriéroux (1997)).

The corresponding empirical squared residuals are given by

$$\hat{\varepsilon}_t^2 = X_t^2/\sigma_t^2(\hat{\theta}_{\mathrm{x},m}), \quad 1 \le t \le m \quad \text{and} \quad \hat{\xi}_t^2 = Y_t^2/\sigma_t^2(\hat{\theta}_{\mathrm{y},n}), \quad 1 \le t \le n, \tag{2.5}$$

where

$$\sigma_t^2(\hat{\theta}_{\mathrm{x},m}) = \hat{\theta}_{\mathrm{x},m}^0 + \sum_{i=1}^{p_{\mathrm{x}}} \hat{\theta}_{\mathrm{x},m}^i X_{t-i}^2 \quad \text{and} \quad \sigma_t^2(\hat{\theta}_{\mathrm{y},n}) = \hat{\theta}_{\mathrm{y},n}^0 + \sum_{i=1}^{p_{\mathrm{y}}} \hat{\theta}_{\mathrm{y},n}^i Y_{t-i}^2.$$

2.3 Derivation of $\widehat{H}_N(x)$

For the testing problem (2.3), we begin by describing our approach in line with Chernoff and Savage (1958). Put $N = m+n$ and $\lambda_N = m/N$. For (2.5), the sizes m and n are assumed to be such that the inequalities

$$0 < \lambda_0 \le \lambda_N \le 1 - \lambda_0 < 1$$

hold for some fixed $\lambda_0 \le \frac{1}{2}$. Then the combined distribution function is defined by

$$H_N(x) = \lambda_N F(x) + (1 - \lambda_N)G(x),$$

where $0 < H_N < 1$. In the same way, if $\hat{F}_m(x)$ and $\hat{G}_n(x)$ denote the empirical distribution functions of $\{\hat{\varepsilon}_t^2\}$ and $\{\hat{\xi}_t^2\}$, the corresponding empirical distribution function is

$$\widehat{H}_N(x) = \lambda_N \hat{F}_m(x) + (1 - \lambda_N)\hat{G}_n(x). \tag{2.6}$$

Write

$$\begin{aligned}\hat{B}_m(x) &= m^{1/2}(\hat{F}_m(x) - F(x)) = m^{-1/2}\sum_{t=1}^{m}(\mathbb{I}(\hat{\varepsilon}_t^2 \le x) - F(x)) \quad \text{and} \\ \hat{B}_n(x) &= n^{1/2}(\hat{G}_n(x) - G(x)) = n^{-1/2}\sum_{t=1}^{n}(\mathbb{I}(\hat{\xi}_t^2 \le x) - G(x)),\end{aligned}$$

where $\mathbb{I}(A)$ is the indicator function of the event A. Then from the result by Horváth *et al.* (2001) (see also Lee and Taniguchi (2005)), we observe that the quantity $\hat{B}_m(x)$ has the following representation,

$$\hat{B}_m(x) = m^{1/2}(F_m(x) - F(x)) + A_{\mathrm{x}} x f(x) + \text{ lower order terms,} \tag{2.7}$$

where

$$F_m(x) = \frac{1}{m}\sum_{t=1}^{m}\mathbb{I}(\varepsilon_t^2 \le x) \quad \text{and} \quad A_{\mathrm{x}} = \sum_{0\le i\le p_{\mathrm{x}}} m^{1/2}(\hat{\theta}_{\mathrm{x},m}^i - \theta_{\mathrm{x}}^i)\tau_{\mathrm{x},i} \tag{2.8}$$

with

$$\tau_{\mathrm{x},0} = E(1/\sigma_t^2(\theta_{\mathrm{x}})) \quad \text{and} \quad \tau_{\mathrm{x},i} = E(X_{t-i}^2/\sigma_t^2(\theta_{\mathrm{x}})), 1 \le i \le p_{\mathrm{x}}.$$

By analogy with (2.7), the corresponding representation of $\hat{B}_n(x)$ is given by

$$\hat{B}_n(x) = n^{1/2}(G_n(x) - G(x)) + A_{\mathrm{y}} x g(x) + \text{ lower order terms,} \tag{2.9}$$

where

$$G_n(x) = \frac{1}{n}\sum_{t=1}^{n}\mathbb{I}(\xi_t^2 \le x) \quad \text{and} \quad A_{\mathrm{y}} = \sum_{0\le i\le p_{\mathrm{y}}} n^{1/2}(\hat{\theta}_{\mathrm{y},n}^i - \theta_{\mathrm{y}}^i)\tau_{\mathrm{y},i} \tag{2.10}$$

with

$$\tau_{\mathrm{y},0} = E(1/\sigma_t^2(\theta_{\mathrm{y}})) \quad \text{and} \quad \tau_{\mathrm{y},i} = E(Y_{t-i}^2/\sigma_t^2(\theta_{\mathrm{y}})), 1 \le i \le p_{\mathrm{y}}.$$

Hence, from (2.7) and (2.9), the expression (2.6) becomes

$$\begin{aligned}\widehat{H}_N(x) &= \mathcal{H}_N(x) + m^{-1/2}\lambda_N A_{\mathrm{x}} x f(x) \\ &\quad + n^{-1/2}(1-\lambda_N) A_{\mathrm{y}} x g(x) + \text{ lower order terms,}\end{aligned} \tag{2.11}$$

where

$$\mathcal{H}_N(x) = \lambda_N F_m(x) + (1-\lambda_N) G_n(x)$$

with $0 < \mathcal{H}_N < 1$. The decomposition (2.11) is basic and will be used repeatedly in the sequel.

2.4 Cramér-von Mises Statistics $\{\widehat{T}_N\}$

For the testing problem (2.3), let us consider a class of Cramér-von Mises statistics of the form

$$\widehat{T}_N = \int (\hat{F}_m(x) - \hat{G}_n(x))^2 d\widehat{H}_N(x). \tag{2.12}$$

Note that (2.12) is constructed from the empirical residuals $\{\hat{\varepsilon}_t^2\}$ and $\{\hat{\xi}_t^2\}$. Likewise, if we construct it replacing $\{\hat{\varepsilon}_t^2\}$ and $\{\hat{\xi}_t^2\}$ by $\{\varepsilon_t^2\}$ and $\{\xi_t^2\}$, respectively, then it becomes the usual Cramér-von Mises statistic (see e.g., Durbin (1973, p.47))

$$T_N^D = \int (F_m(x) - G_n(x))^2 d\mathcal{H}_N(x).$$

This statistic was essentially proposed by Lehmann (1951) and studied by many researchers (Anderson (1962), Ahmad (1996)) who contributed to its development, and numerous theorems have been formulated in many testing problems. Noting that under $H_0 : F = G$, the quantity $\mathcal{H}_N$ converges to F almost surely, and we may conclude under certain regularity conditions that $(mn/N)T_N^D$ converges in law to

$$\int_0^1 Z^2(t)dt,$$

where $\{Z(t);\, 0 \le t \le 1\}$ is a Gaussian process with

$$E(Z(t)) = 0 \quad \text{and} \quad E(Z(s)Z(t)) = \min(s,t) - st, \quad 0 \le s,t \le 1.$$

2.5 Asymptotic Theory of $\{\widehat{T}_N\}$

The object of this section is to elucidate the asymptotics of (2.12). In what follows, K will denote a generic constant which does not depend on F, G, m, n and N.

We impose the following regularity conditions.

Assumption 3.

(A.1) $|(F-G)(x)| < K(H_N(x)(1-H_N(x)))^{1/2}$ for all $x > 0$ and $K > 0$.

(A.2) $xf(x)$, $xg(x)$, $xf'(x)$ and $xg'(x)$ are uniformly bounded continuous, and integrable functions on $(0, \infty)$.

(A.3) $|xf(x)| < KH_N(x)(1-H_N(x))$ and $|xg(x)| < KH_N(x)(1-H_N(x))$ for all $x > 0$ and $K > 0$.

(A.4) There exists $c > 0$ such that $F(x) \geq c\{xf(x)\}$ and $G(x) \geq c\{xg(x)\}$ for all $x > 0$.

Returning to the models $\{X_t\}$ and $\{Y_t\}$, we now impose a further condition on θ_{x} and θ_{y}, and the moment of ε_t and ξ_t. For this purpose, write

$$\mathcal{A}_{\mathrm{x},t} = \begin{pmatrix} \theta_{\mathrm{x}}^1 \varepsilon_t^2 & \cdots & \theta_{\mathrm{x}}^{p_{\mathrm{x}}-1} \varepsilon_t^2 & \theta_{\mathrm{x}}^{p_{\mathrm{x}}} \varepsilon_t^2 \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}$$

and

$$\mathcal{A}_{\mathrm{y},t} = \begin{pmatrix} \theta_{\mathrm{y}}^1 \xi_t^2 & \cdots & \theta_{\mathrm{y}}^{p_{\mathrm{y}}-1} \xi_t^2 & \theta_{\mathrm{y}}^{p_{\mathrm{y}}} \xi_t^2 \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Introduce the notation

$$\mathcal{A}_{\mathrm{x},t}^{\otimes s} = \mathcal{A}_{\mathrm{x},t} \overbrace{\otimes \cdots \otimes}^{s \text{ times}} \mathcal{A}_{\mathrm{x},t}$$

(e.g., Hannan (1970, p.518)), and define

$$\Sigma_{\mathrm{x},s} = E(\mathcal{A}_{\mathrm{x},t}^{\otimes s}) \quad \text{and} \quad \Sigma_{\mathrm{y},s} = E(\mathcal{A}_{\mathrm{y},t}^{\otimes s}),$$

where $\otimes$ denotes the tensor product.

Assumption 4.

(B.1) ε_t^2 and ξ_t^2 are nondegenerate random variables.

(B.2) $E|\varepsilon_t|^8 < \infty$ and $\|\Sigma_{\mathrm{x},3}\| < 1$, $E|\xi_t|^8 < \infty$ and $\|\Sigma_{\mathrm{y},3}\| < 1$,

where $\|\cdot\|$ is the spectral matrix norm. From this and the result by Chen and An (1998), it follows that $E(Z_{\mathrm{x},t}^4) < \infty$ and $E(Z_{\mathrm{y},t}^4) < \infty$. For the case when $p_{\mathrm{x}} = 1$, and $\{\varepsilon_t\}$ is Gaussian, we see that $\|\Sigma_{\mathrm{x},3}\| < 1$ implies $\theta_{\mathrm{x}}^1 < 15^{-\frac{1}{3}} \approx 0.4$.

In order to state the main result, we observe that the matrices

$$\mathcal{U}_{\mathrm{x}} = E(W_{\mathrm{x},t-1}W_{\mathrm{x},t-1}^T), \quad \mathcal{U}_{\mathrm{y}} = E(W_{\mathrm{y},t-1}W_{\mathrm{y},t-1}^T),$$

and

$$\mathcal{R}_{\mathrm{x}} = (\kappa_4^{\mathrm{x}} + 2)E(\sigma_t^4(\theta_{\mathrm{x}})W_{\mathrm{x},t-1}W_{\mathrm{x},t-1}^T)$$

and

$$\mathcal{R}_{\mathrm{y}} = (\kappa_4^{\mathrm{y}} + 2)E(\sigma_t^4(\theta_{\mathrm{y}})W_{\mathrm{y},t-1}W_{\mathrm{y},t-1}^T)$$

are positive definite. To justify R_{x} as an illustration, first observe that it is evidently nonnegative definite,

$$\alpha^T\mathcal{R}_{\mathrm{x}}\alpha = (\kappa_4^{\mathrm{x}} + 2)E(\alpha^T\sigma_t^2(\theta_{\mathrm{x}})W_{\mathrm{x},t-1})^2 \geq 0$$

for any $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_{p_{\mathrm{x}}})^T \in R^{p_{\mathrm{x}}+1}$. Moreover, if we suppose that $\mathcal{R}_{\mathrm{x}}$ is not positive definite, then there exists a vector $(\alpha_0, \alpha_1, \ldots, \alpha_{j_0})$ with $\alpha_{j_0} \neq 0$ ($j_0 \leq p_{\mathrm{x}}$) such that

$$\alpha_0 + \alpha_1 Z_{\mathrm{x},s-1} + \cdots + \alpha_{j_0} Z_{\mathrm{x},s-j_0} = 0 \quad \text{a.e.}.$$

Here, note that $\sigma_t^2(\theta_{\mathrm{x}}) > 0$ a.e., because of $\theta_{\mathrm{x}}^0 > 0$. In this case, we can write

$$Z_{\mathrm{x},s-j_0} = -\beta_0 - \beta_1 Z_{\mathrm{x},s-1} - \cdots - \beta_{j_0-1}Z_{\mathrm{x},s-j_0+1},$$

where $\beta_k = \alpha_k/\alpha_{j_0}$. Hence, substituting this into the last term of $\sigma_s^2(\theta_{\mathrm{x}})$ in (2.1) with setting $s - j_0 = t - p_{\mathrm{x}}$ reveals that the dimension of our ARCH(p_{x}) is reduced to be less than $p_{\rm x}$, leading to a contradiction.

Now recalling the definition of $\mathcal{Q}_m(\theta_{\rm x})$ and $\mathcal{Q}_n(\theta_{\rm y})$, we observe that

$$\begin{aligned}
\frac{\partial \mathcal{Q}_m}{\partial \theta_{\rm x}^0} &= -2\sum_{t=1}^{m}(\varepsilon_t^2-1)\sigma_t^2(\theta_{\rm x}) = -2\sum_{t=1}^{m}\phi_{\rm x}(\varepsilon_t^2)\sigma_t^2(\theta_{\rm x}),\\
\frac{\partial \mathcal{Q}_m}{\partial \theta_{\rm x}^i} &= -2\sum_{t=1}^{m}(\varepsilon_t^2-1)\sigma_t^2(\theta_{\rm x})Z_{{\rm x},t-i} = -2\sum_{t=1}^{m}\phi_{\rm x}(\varepsilon_t^2)\sigma_t^2(\theta_{\rm x})Z_{{\rm x},t-i}, \quad 1\le i\le p_{\rm x},
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \mathcal{Q}_n}{\partial \theta_{\rm y}^0} &= -2\sum_{t=1}^{n}(\xi_t^2-1)\sigma_t^2(\theta_{\rm y}) = -2\sum_{t=1}^{n}\phi_{\rm y}(\xi_t^2)\sigma_t^2(\theta_{\rm y}),\\
\frac{\partial \mathcal{Q}_n}{\partial \theta_{\rm y}^i} &= -2\sum_{t=1}^{n}(\xi_t^2-1)\sigma_t^2(\theta_{\rm y})Z_{{\rm y},t-i} = -2\sum_{t=1}^{n}\phi_{\rm y}(\xi_t^2)\sigma_t^2(\theta_{\rm y})Z_{{\rm y},t-i}, \quad 1\le i\le p_{\rm y},
\end{aligned}$$

where $\phi_\cdot(u) = u-1$. Write

$$\gamma_{{\rm x},t} = (\sigma_t^2(\theta_{\rm x}), \sigma_t^2(\theta_{\rm x})Z_{{\rm x},t-1}, \ldots, \sigma_t^2(\theta_{\rm x})Z_{{\rm x},t-p_{\rm x}})^T$$

and

$$\gamma_{{\rm y},t} = (\sigma_t^2(\theta_{\rm y}), \sigma_t^2(\theta_{\rm y})Z_{{\rm y},t-1}, \ldots, \sigma_t^2(\theta_{\rm y})Z_{{\rm y},t-p_{\rm y}})^T.$$

Then, under certain regularity conditions, it is seen that the corresponding ith element of $\hat\theta_{{\rm x},m}$ and $\hat\theta_{{\rm x},n}$ admits the stochastic expansions,

$$\begin{aligned}
\hat\theta_{{\rm x},m}^i - \theta_{\rm x}^i &= \frac{1}{m}\sum_{t=1}^{m}V_{{\rm x},t}^i\phi_{\rm x}(\varepsilon_t^2) + o_p(m^{-1/2}), \quad 0\le i\le p_{\rm x} \quad \text{and} \qquad (2.13)\\
\hat\theta_{{\rm y},n}^i - \theta_{\rm y}^i &= \frac{1}{n}\sum_{t=1}^{n}V_{{\rm y},t}^i\phi_{\rm y}(\xi_t^2) + o_p(n^{-1/2}), \quad 0\le i\le p_{\rm y},
\end{aligned}$$

where $V_{{\rm x},t}^i$ and $V_{{\rm y},t}^i$ are the ith elements of $\mathcal{U}_{\rm x}^{-1}\gamma_{\rm x,t}$ and $\mathcal{U}_{\rm y}^{-1}\gamma_{\rm y,t}$. Write

$$\delta_{{\rm x},i} = E(V_{{\rm x},t}^i), \quad 0\le i\le p_{\rm x}, \quad \delta_{{\rm y},i} = E(V_{{\rm y},t}^i), \quad 0\le i\le p_{\rm y},$$

and $\tau_{\rm x} = (\tau_{{\rm x},0},\ldots,\tau_{{\rm x},p_{\rm x}})^T$ and $\tau_{\rm y} = (\tau_{{\rm y},0},\ldots,\tau_{{\rm y},p_{\rm y}})^T$ (recall (2.8) and (2.10)). Then, under $H_A: F\ne G$, we have the following result, whose proof is given in Chapter 4.

Theorem 2. *Suppose that Assumptions* $1-4$ *hold and that, in addition,* $\hat{\theta}_{\mathrm{x},m}$ *and* $\hat{\theta}_{\mathrm{y},n}$ *are the conditional least squares estimators of* θ_{x} *and* θ_{y} *satisfying (2.4). Then*

$$N^{1/2}(\widehat{T}_N-\mu_N)/\sigma_N \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as } N\to\infty,$$

where

$$\mu_N=\int (F(x)-G(x))^2 dH_N(x) \quad \text{and} \quad \sigma_N^2=\sigma_{1N}^2+\sigma_{2N}^2+\sigma_{3N}^2+\zeta_N\neq 0 \text{ with}$$

$$\begin{aligned}
\sigma_{1N}^2 &= 8\{\lambda_N^{-1}\iint_{x<y}\mathcal{A}(x,y)dG(x)dG(y)+(1-\lambda_N)^{-1}\iint_{x<y}\mathcal{B}(x,y)dF(x)dF(y)\},\\
\sigma_{2N}^2 &= \omega_{\mathrm{x},N}^T\mathcal{U}_{\mathrm{x}}^{-1}\mathcal{R}_{\mathrm{x}}\mathcal{U}_{\mathrm{x}}^{-1}\omega_{\mathrm{x},N}, \quad \sigma_{3N}^2=\omega_{\mathrm{y},N}^T\mathcal{U}_{\mathrm{y}}^{-1}\mathcal{R}_{\mathrm{y}}\mathcal{U}_{\mathrm{y}}^{-1}\omega_{\mathrm{y},N}, \quad \text{and}\\
\zeta_N &= -8\{\lambda_N^{-1}\sum_{0\le i\le p_{\mathrm{x}}}\tau_{\mathrm{x},i}\delta_{\mathrm{x},i}\iint \psi_{\mathrm{x}}(x)\rho_f(x,y)dG(x)dG(y)\\
&\quad -(1-\lambda_N)^{-1}\sum_{0\le i\le p_{\mathrm{y}}}\tau_{\mathrm{y},i}\delta_{\mathrm{y},i}\iint \psi_{\mathrm{y}}(x)\rho_g(x,z)dF(x)dF(z)\},
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{A}(x,y) &= F(x)(F-G)(x)(1-F(y))(F-G)(y),\\
\mathcal{B}(x,y) &= G(x)(F-G)(x)(1-G(y))(F-G)(y),\\
\omega_{\mathrm{x},N} &= -2\lambda_N^{-1/2}\int xf(x)(F-G)(x)dG(x)\times\tau_{\mathrm{x}},\\
\omega_{\mathrm{y},N} &= -2(1-\lambda_N)^{-1/2}\int zg(z)(F-G)(z)dF(z)\times\tau_{\mathrm{y}},\\
\rho_f(x,y) &= yf(y)(F-G)(x)(F-G)(y),\\
\rho_g(x,z) &= zg(z)(F-G)(x)(F-G)(z),\\
\psi_{\mathrm{x}}(x) &= \int_0^x \phi_{\mathrm{x}}(u)f(u)du, \quad \psi_{\mathrm{y}}(x)=\int_0^x \phi_{\mathrm{y}}(u)g(u)du.
\end{aligned}$$

Remark 1. Observe that the terms (see Theorem 2) σ_{2N}^2, σ_{3N}^2 and ζ_N depend on the volatility estimators $\hat{\theta}_{\mathrm{x},m}$ and $\hat{\theta}_{\mathrm{y},n}$. Hence, the asymptotics of $\{\widehat{T}_N\}$ are greatly different in comparison with the independent, identically distributed or ARMA settings, this result illuminates some interesting features of ARCH residuals.

Remark 2. For $\{\widehat{T}_N\}$ to be practically feasible, it is necessary to replace σ_N^2

which depends on several unknown parameters and functions by a consistent estimator $\hat{\sigma}_N^2$. Observe that $\delta_{\mathrm{x},i}$, $\tau_{\mathrm{x},i}$, $\delta_{\mathrm{y},j}$, $\tau_{\mathrm{y},j}$; $0 \le i \le p_{\mathrm{x}}$, $0 \le i \le p_{\mathrm{y}}$, and $\psi_{\mathrm{x}}(x)$ and $\psi_{\mathrm{y}}(x)$ are expected values and can be consistently estimated by the corresponding averages. Note also that $\mathcal{U}_{\mathrm{x}}^{-1}\mathcal{R}_{\mathrm{x}}\mathcal{U}_{\mathrm{x}}^{-1}$ and $\mathcal{U}_{\mathrm{y}}^{-1}\mathcal{R}_{\mathrm{y}}\mathcal{U}_{\mathrm{y}}^{-1}$ are the asymptotic covariance matrices of $\sqrt{m}(\hat{\theta}_{\mathrm{x},m} - \theta_{\mathrm{x}})$ and $\sqrt{n}(\hat{\theta}_{\mathrm{y},n} - \theta_{\mathrm{y}})$, respectively, and their estimation is discussed in Gouriéroux (1997).

Chapter 3

Asymptotic Performance of $\{\widehat{T}_N\}$

The limiting distribution of $\{\widehat{T}_N\}$ given in the preceding chapter provides a useful guide to the reliability of asymptotic relative efficiency and ARCH volatility effect. Thus we may proceed to illustrate these aspects of $\{\widehat{T}_N\}$ numerically for some ARCH residual distributions. Moreover, a measure of robustness for $\{\widehat{T}_N\}$ is introduced by means of Hampel's influence function. Then this quantitative information based on such ARCH residual distributions is illustrated by simulations. The same study of $\{\widehat{T}_N\}$ is also demonstrated using the daily stock returns of AMOCO and IBM companies of New York Stock Exchange from February 2, 1984, to December 31, 1991.

3.1 Asymptotic Relative Efficiency

In this section, we consider the assessment of asymptotic relative efficiency of the statistics T_N^D and $\widehat{T}_N$ for some residual distributions in the i.i.d. and in our ARCH residual settings, respectively. The results help to highlight some interesting features of $\widehat{T}_N$ in comparison with T_N^D.

For simplicity, let us consider the ARCH(1) model

$$X_t = \begin{cases} \sigma_t(\theta_{\mathrm{x}})\varepsilon_t, \quad \sigma_t^2(\theta_{\mathrm{x}}) = \theta_{\mathrm{x}}^0 + \theta_{\mathrm{x}}^1 X_{t-1}^2 & \text{for} \quad t = 1, \ldots, m, \\ 0 & \text{for} \quad t \le 0, \end{cases}$$

where $\theta_{\mathrm{x}} = (\theta_{\mathrm{x}}^0, \theta_{\mathrm{x}}^1)^T$, $\theta_{\mathrm{x}}^0 > 0$, $0 \le \theta_{\mathrm{x}}^1 < 1$, $\{\varepsilon_t\}$ is a sequence of i.i.d.(0,1) random variables with fourth-order cumulant κ_4^{x}, and ε_t is independent of X_s, $s < t$.

Another ARCH(1) model, independent of $\{X_t\}$, is given by

$$Y_t = \begin{cases} \sigma_t(\theta_{\mathrm{y}})\xi_t, \quad \sigma_t^2(\theta_{\mathrm{y}}) = \theta_{\mathrm{y}}^0 + \theta_{\mathrm{y}}^1 Y_{t-1}^2 & \text{for} \quad t = 1, \ldots, n, \\ 0 & \text{for} \quad t \le 0, \end{cases}$$

where $\theta_{\mathrm{y}} = (\theta_{\mathrm{y}}^0, \theta_{\mathrm{y}}^1)^T$, $\theta_{\mathrm{y}}^0 > 0$, $0 \le \theta_{\mathrm{y}}^1 < 1$, $\{\xi_t\}$ is a sequence of i.i.d.(0,1) random variables with fourth-order cumulant κ_4^{y}, and ξ_t is independent of Y_s, $s < t$.

Recall that $F(x)$ and $G(x)$ are the distribution functions of ε_t^2 and ξ_t^2, respectively. The hypothesis of interest in the two-sample problem is that $H_0 : F(x) = G(x)$ for all $x > 0$. If one imposes conditions on the form of the common distribution together with the assumption that a difference between the distributions exist, it is only between means or between variances. The proposed test procedure may be sensitive to violations of those assumptions which are inherent in the construction of the test. In practice, other assumptions are often made about the form of the underlying distributions. One common assumption is called the location model.

Let us now consider the location problem in the case of $G(x) = F(x + \theta)$ for some parameter θ. Henceforth, it is assumed that F is arbitrary and has finite variance σ_F^2. The two-sample testing problem for location can be described as follows;

$$H_0 : \theta = 0 \quad \text{against} \quad H_A : \theta > 0.$$

In light of Theorem 2, we can readily see under $H_0 : \theta = 0$ that the distributions $F(x)$ and $G(x)$ coincide for all $x > 0$. Thus, it is instructive to apply this theorem under $H_A : \theta > 0$ since $F(x) \le G(x)$ for all $x > 0$. In such a case, we may take, for example, $H_A : \theta = 1$. Assuming that $m = n = N/2$, the mean becomes

$$\mu_F(\theta) = \frac{1}{2}\int (F(x) - F(x+\theta))^2 d[F(x) + F(x+\theta)]$$

and the variance under $H_A : \theta = 1$ is $\sigma_F^2 = \sigma_1^2(F) + \sigma_2^2(F) + \sigma_3^2(F) + \gamma(F)$, where

$$\begin{aligned}
\sigma_1^2(F) &= 16 \iint\limits_{x<y} \mathcal{A}^*(x,y)dxdy + 16 \iint\limits_{x<y} \mathcal{B}^*(x,y)dxdy, \\
\sigma_2^2(F) &= 8C_{\mathrm{x}}\left(\int xf(x)f(x+1)[F(x) - F(x+1)]dx\right)^2, \\
\sigma_3^2(F) &= 8C_{\mathrm{y}}\left(\int zf(z)f(z+1)[F(z) - F(z+1)]dz\right)^2, \\
\gamma(F) &= -16k_1 \iint \left[\int_0^x (u-1)f(u)du\right]\rho_f^*(x,y)dxdy \\
&\quad +16k_2 \iint \left[\int_0^x (u-1)f(u+1)du\right]\rho_f^{**}(x,z)dxdz
\end{aligned}$$

with

$$\begin{aligned}
\mathcal{A}^*(x,y) &= f(x+1)f(y+1)F(x) \\
&\quad \times[F(x) - F(x+1)][1 - F(y)][F(y) - F(y+1)], \\
\mathcal{B}^*(x,y) &= f(x)f(y)F(x+1) \\
&\quad \times[F(x) - F(x+1)][1 - F(y+1)][F(y) - F(y+1)], \\
C_{\mathrm{x}} &= \tilde{\tau}_{\mathrm{x}}^T \mathcal{U}_{\mathrm{x}}^{-1} R_{\mathrm{x}} \mathcal{U}_{\mathrm{x}}^{-1} \tilde{\tau}_{\mathrm{x}}, \quad C_{\mathrm{y}} = \tilde{\tau}_{\mathrm{y}}^T \mathcal{U}_{\mathrm{y}}^{-1} R_{\mathrm{y}} \mathcal{U}_{\mathrm{y}}^{-1} \tilde{\tau}_{\mathrm{y}}, \\
k_1 &= \tau_{\mathrm{x},0}\delta_{\mathrm{x},0} + \tau_{\mathrm{x},1}\delta_{\mathrm{x},1}, \quad k_2 = \tau_{\mathrm{y},0}\delta_{\mathrm{y},0} + \tau_{\mathrm{y},1}\delta_{\mathrm{y},1}, \\
\rho_f^*(x,y) &= f(x+1)[F(x) - F(x+1)]yf(y)f(y+1)[F(y) - F(y+1)], \\
\rho_f^{**}(x,z) &= f(x)[F(x) - F(x+1)]zf(z)f(z+1)[F(z) - F(z+1)].
\end{aligned}$$

where $\tilde{\tau}_{\mathrm{x}} = (\tau_{\mathrm{x},0}, \tau_{\mathrm{x},1})^T$ and $\tilde{\tau}_{\mathrm{y}} = (\tau_{\mathrm{y},0}, \tau_{\mathrm{y},1})^T$.

To begin with, let us state a set of Pitman regularity conditions which makes the computation of efficiency for two test sequences quite easy in the case of finite sample sizes. Suppose that $\widehat{T}_N$ is a test statistic based on the first N observations for testing $H_0 : \theta = \theta_0$ against $H_A : \theta > \theta_0$ with critical region $\widehat{T}_N \geq \lambda_{N,\alpha}$. Further, suppose

(i) $\lim_{N\to\infty} P_{\theta_0}(\widehat{T}_N \geq \lambda_{N,\alpha}) = \alpha$, where $0 < \alpha < 1$ is a given level;

(ii) there exist functions $\mu_N(\theta)$ and $\sigma_N(\theta)$ such that $N^{1/2}(\widehat{T}_N - \mu_N(\theta))/\sigma_N(\theta) \xrightarrow{d}$ $\mathcal{N}(0,1)$

uniformly in $\theta \in [\theta_0, \theta_0 + \epsilon]$, $\epsilon > 0$;

(iii) $\mu_N'(\theta_0) > 0$;

(iv) for a sequence $\{\theta_N = \theta_0 + N^{-1/2}\delta,\ \delta > 0\}$,

$$\lim_{N\to\infty}[\mu_N'(\theta_N)/\mu_N'(\theta_0)] = 1, \quad \lim_{N\to\infty}[\sigma_N(\theta_N)/\sigma_N(\theta_0)] = 1;$$

(v) $\lim_{N\to\infty}[\mu_N'(\theta_0)/\sigma_N(\theta_0)] = c > 0$.

For $\alpha \in (0,1)$, write $\lambda_\alpha = \Phi^{-1}(1-\alpha)$, where $\Phi(x) = \int_{-\infty}^{x}(2\pi)^{-1/2}e^{-t^2/2}dt$. Then the asymptotic power is given by $1 - \Phi(\lambda_\alpha - \delta c)$. The quantity c defined by (v) is called the efficacy of $\widehat{T}_N$. It is known that the asymptotic power, in addition to providing a measure of performance, also serves as a basis for the comparison of different tests.

Let $T^{(1)} = \{T_N^{(1)}\}$ and $T^{(2)} = \{T_N^{(2)}\}$ be test sequences with efficacies c_1 and c_2, respectively. Then the asymptotic relative efficiency (ARE) of $T^{(1)}$ relative to $T^{(2)}$ is given by $e(T^{(1)}, T^{(2)}) = c_1^2/c_2^2$.

In order to evaluate the ARE of $T_N^D = T_N^D(F)$ and $\widehat{T}_N = \widehat{T}_N(F)$, it is necessary to specify F. For this purpose, let us suppose that $\{\varepsilon_t\}$ is a sequence of i.i.d.(0,1) random variables with continuous symmetric distribution F^* and density f^*. Then

$$F(x) = P(\varepsilon_t^2 \leq x) = \begin{cases} 2F^*(\sqrt{x}) - 1, & x > 0, \\ 0, & x \leq 0. \end{cases} \tag{3.1}$$

We shall now compute (3.1) in the following particular choices of F^*.

(i) F^* (Normal):

$$F_{\mathbb{N}}^*(x) = \int_{-\infty}^{x}(2\pi)^{-1/2}e^{-t^2/2}dt, \quad f_{\mathbb{N}}^*(x) = (2\pi)^{-1/2}e^{-x^2/2}, \quad x \in \mathbb{R}.$$

In this case,

$$F_{\mathbb{N}}(x) = 2F^*_{\mathbb{N}}(\sqrt{x}) - 1, \quad f_{\mathbb{N}}(x) = \frac{1}{\sqrt{2\pi x}} e^{-x/2}, \quad x > 0.$$

(ii) F^* (Double exponential):

$$F^*_{\mathbb{DE}}(x) = \int_{-\infty}^{x} \frac{1}{2} e^{-|t|} dt = 1 - \frac{1}{2} e^{-x}, \quad f^*_{\mathbb{DE}}(x) = \frac{1}{2} e^{-|x|}, \quad x \in \mathbb{R}.$$

In this case,

$$F_{\mathbb{DE}}(x) = 1 - e^{-\sqrt{x}}, \quad f_{\mathbb{DE}}(x) = \frac{1}{2\sqrt{x}} e^{-\sqrt{x}}, \quad x > 0.$$

(iii) F^* (Logistic):

$$F^*_{\mathbb{L}}(x) = \frac{1}{1 + e^{-x}}, \quad f^*_{\mathbb{L}}(x) = \frac{e^{-x}}{(1 + e^{-x})^2}, \quad x \in \mathbb{R}.$$

In this case,

$$F_{\mathbb{L}}(x) = \frac{1 - e^{-\sqrt{x}}}{1 + e^{-\sqrt{x}}}, \quad f_{\mathbb{L}}(x) = \frac{e^{-\sqrt{x}}}{\sqrt{x}(1 + e^{-\sqrt{x}})^2}, \quad x > 0.$$

Recalling the definition of $\mu_F(\theta)$, σ_F^2 and $\sigma_1^2(F)$ and assuming that $\mu_F(\theta)$ is continuously differentiable with respect to θ at $H_A : \theta = 1$ under the integral sign, we have

$$\begin{aligned} \mu_F'(1) &= \frac{1}{2} \int (F(x) - F(x+1))^2 f'(x+1) dx \\ &\quad - \int f(x+1)(f(x) + f(x+1))(F(x) - F(x+1)) dx \end{aligned}$$

so that the ARE of T_N^D and $\widehat{T}_N$ between distributions F_1 and F_2 is

$$e(F_2, F_1) = c_{F_2}^2 / c_{F_1}^2, \tag{3.2}$$

where $c_F = \mu_F'(1)/\sigma(F)$, with $\sigma(F) = \sigma_1(F)$ and σ_F. In an attempt to evaluate (3.2), we need to approximate values of $\sigma_F^2 = \sigma_F^2(C_{\mathrm{x}}, C_{\mathrm{y}}, k_1, k_2)$ for various $m = n = N/2$ and parameters based on $F = F_{\mathbb{N}}$, $F_{\mathbb{DE}}$ and $F_{\mathbb{L}}$. Set $\theta^0 = \theta^0_{\mathrm{x}} = \theta^0_{\mathrm{y}}$ and $\theta^1 = \theta^1_{\mathrm{x}} = \theta^1_{\mathrm{y}}$. Then, for $\theta^0 = 1, \theta^1 = 0.1, 0.3$, and $m = n = N/2 = 100, 500$, we generate realizations of X_t and Y_t. Note that the above choice of parameter values satisfies necessary conditions. On the basis of conditional least squares estimators $\hat{\theta}^0_m$ and $\hat{\theta}^1_m$ of θ^0 and θ^1, respectively, the quantities $C_{\mathrm{x}}, C_{\mathrm{y}}, k_1$ and k_2 are estimated by the corresponding averages. In the actual computation of $\mu'_F(1), \sigma^2_1(F)$ and σ^2_F, we evaluate the integrals by a rectangular numerical integration with n terms. All the estimation results in the tables below are based on 100 replications. Table 1 provides these results.

Table 1. Approximate values of $e(\cdot,\cdot)$ for $T^D_N = T^D_N(F)$ and $\widehat{T}_N = \widehat{T}_N(F)$ based on $F = F_*$

ARE	$T^D_N(F)$	$\widehat{T}_N(F)$: $m = n = 100,\ \theta^0 = 1$		$m = n = 500,\ \theta^0 = 1$	
		$\theta^1 = 0.1$	$\theta^1 = 0.3$	$\theta^1 = 0.1$	$\theta^1 = 0.3$
$e(F_{\mathbb{N}}, F_{\mathbb{DE}})$	1.3331	1.3067	1.3048	1.3062	1.3049
$e(F_{\mathbb{N}}, F_{\mathbb{L}})$	0.7105	0.7237	0.7248	0.7240	0.7248
$e(F_{\mathbb{DE}}, F_{\mathbb{L}})$	0.5330	0.5538	0.5555	0.5543	0.5554

A closer examination of the ARE values in Table 1 reveals some distinctive characteristics. It is fairly clear that the values in Table 1 are stable with respect to the choice of parameters and distributions, and $m = n$. We also observe that the corresponding values for $T^D_N(F)$ differ from those for $\widehat{T}_N(F)$. These differences are due to the effect of the ARCH volatility estimators $\hat{\theta}^0_m$ and $\hat{\theta}^1_m$. In addition, it is seen that the case of $F = F_{\mathbb{L}}$ is more efficient than the other cases for all chosen values of $m = n$ and the parameters. However, the efficiency for $F = F_{\mathbb{L}}$ decreases as θ^1 or $m = n$ increases. We also observe that $\widehat{T}_N(F)$ for $F = F_{\mathbb{DE}}$ is a strong competitor to that for $F = F_{\mathbb{L}}$ when θ^1 becomes small. Another point worth noting is that $T^D_N(F)$ and $\widehat{T}_N(F)$ for $F = F_{\mathbb{DE}}$ outperform that for $F = F_{\mathbb{N}}$ in all cases. A striking feature of this study agrees that this testing problem is best in the case of heavy-tailed ARCH residual distributions.

3.2 ARCH Volatility Effect

In this section, we study a distinction of $\widehat{T}_N = \widehat{T}_N(F)$ and $T_N^D = T_N^D(F)$ in terms of their levels of test for two-sample location problem under $H_A : \theta = 1$ based on $F = F_{\mathbb{N}}$, $F_{\mathbb{DE}}$ and $F_{\mathbb{L}}$.

Suppose that $N^{1/2}(T_N^D - \mu_F)/\sigma_1(F) \xrightarrow{d} \mathcal{N}(0,1)$ holds. Then the test

$$N^{1/2}(T_N^D - \mu_F)/\sigma_1(F) \geq \lambda_\alpha$$

has nominal asymptotic level α as $N \to \infty$. We assume α to be less than 0.5 so that $\lambda_\alpha > 0$. For this λ_α, let

$$\tilde{\alpha}_N = P\{N^{1/2}(\widehat{T}_N - \mu_F)/\sigma_F \geq \lambda_\alpha\}.$$

Then $\tilde{\alpha} = \lim_{N\to\infty} \tilde{\alpha}_N$ exists and is given by $\tilde{\alpha} = 1 - \Phi(\lambda_\alpha \delta_F)$, where $\delta_F = \sigma_1(F)/\sigma_F$. Since $\sigma_F \geq \sigma_1(F)$, we have $\tilde{\alpha} \geq \alpha$.

To distinguish how much the actual $\tilde{\alpha}$ varies from the nominal α, we use the level $\alpha = 0.05$ for which $\lambda_{0.05} = 1.645$. Using the same values of σ_F and $\sigma_1(F)$ for $F = F_{\mathbb{N}}$, $F_{\mathbb{DE}}$ and $F_{\mathbb{L}}$, we provide the results in Table 2.

Table 2. Actual $\tilde{\alpha} = 1 - \Phi(\lambda_\alpha \delta_F)$, $\delta_F = \sigma_1(F)/\sigma_F$, $F = F_*$ when nominal level $\alpha = 0.05$ for which $\lambda_{0.05} = 1.645$

Distribution	$m = n = 100$, $\theta^0 = 1$		$m = n = 500$, $\theta^0 = 1$	
	$\theta^1 = 0.1$	$\theta^1 = 0.3$	$\theta^1 = 0.1$	$\theta^1 = 0.3$
$\delta_{F_{\mathbb{N}}}$	0.9804	0.9790	0.9799	0.9790
$\tilde{\alpha}_{F_{\mathbb{N}}}$	0.0534	0.0536	0.0535	0.0537
$\delta_{F_{\mathbb{DE}}}$	0.9902	0.9896	0.9900	0.9896
$\tilde{\alpha}_{F_{\mathbb{DE}}}$	0.0517	0.0518	0.0517	0.0518
$\delta_{F_{\mathbb{L}}}$	0.9714	0.9694	0.9694	0.9714
$\tilde{\alpha}_{F_{\mathbb{L}}}$	0.0550	0.0554	0.0554	0.0554

Table 2 shows that the values of $\tilde{\alpha}_*$ are differ from the nominal $\alpha = 0.05$ with respect to the choice of parameters and distributions, and $m = n$. It is also seen that these values tend to increase slightly as θ^1 or $m = n$ increases. Such an increase is due to the asymptotics of the ARCH volatility estimators $\hat{\theta}^0_m$ and $\hat{\theta}^1_m$. In addition, it shows the effect of skewness on the level. As is typically the case when $F = F_*$ is skewed to the right, $\tilde{\alpha}_* > \alpha$ for the lower-tail rejection region. It should be pointed out that, in general, the closeness of $\tilde{\alpha}_*$ to α depends not only on the parameters but also on other aspects of $F = F_*$. We can therefore say that the asymptotic level of $\widehat{T}_N$ is fairly different from that of T_N^D because of the ARCH specification effect.

3.3 Robustness Measures

Hampel's influence function IFH is a heuristic tool which provides rich quantitative robustness information. It measures the sensitivity of a statistic T to infinitesimal deviations from an underlying distribution F. In the following, we introduce some measures which indicate a robustness of $\widehat{T}_N$ given by (2.12).

It was shown in the proof of Theorem 2 (see (4.1)) that

$$\widehat{T}_N - \mu_N = U_N(F_m, G_n) + V_{1N}(\hat{\theta}_{\mathrm{x},m}; F, G) + V_{2N}(\hat{\theta}_{\mathrm{y},n}; F, G) + o_p(N^{-1/2}),$$

where

$$U_N(F_m, G_n) = 2\left\{\int s(x)d(G_n - G)(x) - \int s^*(x)d(F_m - F)(x)\right\},$$

$$V_{1N}(\hat{\theta}_{\mathrm{x},m}; F, G) = -2\tau_{\mathrm{x}}^T(\hat{\theta}_{\mathrm{x},m} - \theta_{\mathrm{x}}) \int xf(x)(F - G)(x)dG(x)$$

and

$$V_{2N}(\hat{\theta}_{\mathrm{y},n}; F, G) = -2\tau_{\mathrm{y}}^T(\hat{\theta}_{\mathrm{y},n} - \theta_{\mathrm{y}}) \int xg(x)(F - G)(x)dF(x).$$

Let us first study a robustness of $U_N(F_m, G_n)$. To simplify the presentation, assume that $m = n = N/2$. Then

$$U_N(F_m, G_m) = 2\left\{\int s(x)d(G_m - G)(x) - \int s^*(x)d(F_m - F)(x)\right\},$$

where

$$s(x) = \int_{x_0}^{x} (F - G)(y)dF(y) \quad \text{and} \quad s^*(x) = \int_{x_0}^{x} (F - G)(y)dG(y)$$

with $x_0 > 0$ determined somewhat arbitrarily. As a measure of its robustness, we can introduce the following influence function:

$$IFH(F, G) = \lim_{h \searrow 0} \frac{U_N\{(1-h)F + h\delta_a, (1-h)G + h\delta_b\}}{h},$$

where $h \in (0, 1)$ and δ_t is the probability distribution with point-mass one at t. Thus, we obtain

$$IFH(F, G) = 2\{s(b) - s^*(a) + \int s^*(x)dF(x) - \int s(x)dG(x)\},$$

Next, we discuss a robust property of $V_{1N}(\hat{\theta}_{\mathrm{x},m}; F, G)$. Let us now consider $\hat{\theta}_{\mathrm{x},m}$. Write $S_{\mathrm{x},t} = (X_t^2, \ldots, X_{t-p_{\mathrm{x}}+1}^2)^T$ and $W_{S_{\mathrm{x}},t} = (1, S_{\mathrm{x},t}^T)^T$, and let $S_{\mathrm{x},t}^{(1)}$ be the first component of $S_{\mathrm{x},t}$. Then we can write $\hat{\theta}_{S_{\mathrm{x}},m} = \widehat{\mathcal{U}}_{S_{\mathrm{x}}}^{-1}\widehat{\gamma}_{S_{\mathrm{x}}}$, where

$$\widehat{\gamma}_{S_{\mathrm{x}}} = \frac{1}{m}\sum_{t=2}^{m} S_{\mathrm{x},t}^{(1)} W_{S_{\mathrm{x}},t-1} \quad \text{and} \quad \widehat{\mathcal{U}}_{S_{\mathrm{x}}} = \frac{1}{m}\sum_{t=2}^{m} W_{S_{\mathrm{x}},t-1} W_{S_{\mathrm{x}},t-1}^T.$$

Since $\widehat{\gamma}_{S_{\mathrm{x}}}$ and $\widehat{\mathcal{U}}_{S_{\mathrm{x}}}$ are sample versions of

$$\gamma_{S_{\mathrm{x}}} = E(S_{\mathrm{x},t}^{(1)} W_{S_{\mathrm{x}},t-1}) \quad \text{and} \quad \mathcal{U}_{S_{\mathrm{x}}} = E(W_{S_{\mathrm{x}},t-1}, W_{S_{\mathrm{x}},t-1}^T),$$

respectively, the corresponding functional of $\hat{\theta}_{S_{\mathrm{x}},m}$ is $T_{S_{\mathrm{x}}} \equiv \mathcal{U}_{S_{\mathrm{x}}}^{-1}\gamma_{S_{\mathrm{x}}}$. Let us now consider the following contaminated process

$$S_{\mathrm{x},t}^h = (1-h)S_{\mathrm{x},t} + hK_{\mathrm{x},t} = S_{\mathrm{x},t} + hL_{\mathrm{x},t}, \quad \text{(say)}.$$

For $S_{\mathrm{x}}^h = \{S_{\mathrm{x},t}^h\}$, we can introduce an influence function

$$T'_{S_{\mathrm{x}}} = \lim_{h \searrow 0} \frac{T_{S_{\mathrm{x}}^h} - T_{S_{\mathrm{x}}}}{h}.$$

Noting the differential formula for matrix $dZ^{-1} = -Z^{-1}(dZ)Z^{-1}$, we obtain

$$\frac{d}{dh}\mathcal{U}_{S_{\mathrm{x}}^h}^{-1}\bigg|_{h=0} = -\mathcal{U}_{S_{\mathrm{x}}}^{-1}(\Delta_{\mathrm{x}} + \Delta_{\mathrm{x}}^T)\mathcal{U}_{S_{\mathrm{x}}}^{-1}, \Delta_{\mathrm{x}} = E\left[\begin{pmatrix} 0 \\ L_{\mathrm{x},t-1} \end{pmatrix} W_{S_{\mathrm{x}},t-1}^T\right].$$

Also,

$$\frac{d}{dh}\gamma_{S_{\mathrm{x}}^h}\bigg|_{h=0} = E(L_{\mathrm{x},t}^{(1)}W_{S_{\mathrm{x}},t-1}) + E\left[S_{\mathrm{x},t}^{(1)}\begin{pmatrix} 0 \\ L_{\mathrm{x},t-1} \end{pmatrix}\right] = \gamma'_{S_{\mathrm{x}}}, \quad \text{(say)},$$

where $L_{\mathrm{x},t}^{(1)}$ is the first component of $L_{\mathrm{x},t}$. Hence,

$$T'_{S_{\mathrm{x}}} = \mathcal{U}_{S_{\mathrm{x}}}^{-1}(\gamma'_{S_{\mathrm{x}}} - (\Delta_{\mathrm{x}} + \Delta_{\mathrm{x}}^T)T_{S_{\mathrm{x}}})$$

and similarly for $V_{2N}(\hat{\theta}_{\mathrm{y},n}; F, G)$,

$$T'_{S_{\mathrm{y}}} = \mathcal{U}_{S_{\mathrm{y}}}^{-1}(\gamma'_{S_{\mathrm{y}}} - (\Delta_{\mathrm{y}} + \Delta_{\mathrm{y}}^T)T_{S_{\mathrm{y}}}).$$

The quantities $IFH(F,G)$, $T'_{S_{\mathrm{x}}}$ and $T'_{S_{\mathrm{y}}}$ will facilitate the fundamental description of sensitiveness or insensitiveness of $\widehat{T}_N$.

Returning to the setup of Section 3.1, we describe the quantitative information for $U_N(F_m, G_m)$, $V_{1N}(\hat{\theta}_{\mathrm{x},m}; F, G)$ and $V_{2N}(\hat{\theta}_{\mathrm{y},m}; F, G)$ by computing $IFH(F,G) = I(F)$, $V_{1N}(T'_{S_{\mathrm{x}}}, F)$ and $V_{2N}(T'_{S_{\mathrm{y}}}, F)$, respectively. Using the same realizations of X_t and Y_t for $m = n = N/2 = 100, 500$ and $(\theta^0, \theta^1) = (1, 0.1), (1, 0.3)$, Tables 3 and 4 provide these results for $F = F_{\mathbb{N}}$, $F_{\mathbb{DE}}$ and $F_{\mathbb{L}}$.

Table 3. Approximate values of $I(F)$, $V_{1N}(\mathbf{T}'_{S_{\mathrm{x}}}, F)$ and $V_{2N}(\mathbf{T}'_{S_{\mathrm{y}}}, F)$ for $\widehat{T}_N = \widehat{T}_N(F)$ based on various $F = F_*$ and $m = n = 100$, $\theta^0 = 1$

Distribution	$I(F)$	$V_{1N}(\mathbf{T}'_{S_{\mathrm{x}}}, F)$		$V_{2N}(\mathbf{T}'_{S_{\mathrm{y}}}, F)$	
		$\theta^1 = 0.1$	$\theta^1 = 0.3$	$\theta^1 = 0.1$	$\theta^1 = 0.3$
$F_{\mathbb{N}}$	0.0350	0.0647	0.0193	0.0187	0.0096
$F_{\mathbb{DE}}$	0.0224	0.0364	0.0108	0.0105	0.0054
$F_{\mathbb{L}}$	0.0249	0.0517	0.0154	0.0149	0.0077

Table 4. Approximate values of $I(F)$, $V_{1N}(\mathbf{T}'_{S_x}, F)$ and $V_{2N}(\mathbf{T}'_{S_y}, F)$ for $\widehat{T}_N = \widehat{T}_N(F)$ based on various $F = F_*$ and $m = n = 500$, $\theta^0 = 1$

Distribution	$I(F)$	$V_{1N}(\mathbf{T}'_{S_x}, F)$		$V_{2N}(\mathbf{T}'_{S_y}, F)$	
		$\theta^1 = 0.1$	$\theta^1 = 0.3$	$\theta^1 = 0.1$	$\theta^1 = 0.3$
$F_{\mathbb{N}}$	0.0350	0.0390	0.0173	0.0263	0.0151
$F_{\mathbb{DE}}$	0.0224	0.0219	0.0098	0.0148	0.0085
$F_{\mathbb{L}}$	0.0249	0.0311	0.0138	0.0210	0.0121

An examination of the values in Tables 3 and 4 shows some interesting features about the sensitivity of $\widehat{T}_N = \widehat{T}_N(F)$ for $F = F_{\mathbb{N}}$. First it is apparent that the values are stable with respect to the choice of parameters, distributions and $m = n = N/2$. It is also interesting to note that when θ^1 increases, the values of $V_{1N}(\cdot, \cdot)$ and $V_{2N}(\cdot, \cdot)$ tend to decrease for each $m = n$. This behavior depends not only on the choice of parameters but also on other aspects of F_*. We summarize by saying that $\widehat{T}_N$ is robust in terms of goodness of fit for such heavy-tail ARCH residual distributions.

3.4 Real Data Analysis

To assess the usefulness of the asymptotic result obtained in chapter 2, the proposed two-sample testing problem for location is applied to real data sets. The data sets of interest are the daily stock return data points ($m = n = 2000$) of AMOCO and IBM companies of New York Stock Exchange from February 2, 1984, to December 31, 1991.

For the ARCH residual distributions $F = F_{\mathbb{N}}$, $F_{\mathbb{DE}}$ and $F_{\mathbb{L}}$, the asymptotic relative efficiency, the ARCH volatility effect and the measure of robustness of $\widehat{T}_N = \widehat{T}_N(F)$ are demonstrated numerically in Tables $5 - 7$, respectively. Note that the values in the tables are stable with respect to the choice of the distributions. These results provide enough evidence in support of the simulation results. We summarize by saying that the two-sample testing problem for location works well in the case of heavy-tailed ARCH residual distributions.

Table 5. Estimated values of $e(\cdot, \cdot)$ based on various $F = F_*$.

$e(F_{\mathbb{N}}, F_{\mathbb{DE}})$	$e(F_{\mathbb{N}}, F_{\mathbb{L}})$	$e(F_{\mathbb{DE}}, F_{\mathbb{L}})$
1.2320	0.7646	0.6206

Table 6. Actual $\tilde{\alpha} = 1 - \Phi(\lambda_\alpha \delta_F)$, $\delta_F = \sigma_1(F)/\sigma_F, F = F_*$ when nominal level $\alpha = 0.05$ for which $\lambda_{0.05} = 1.645$

Distribution	δ_{F_*}	$\tilde{\alpha}_{F_*}$
$F_{\mathbb{N}}$	0.9306	0.0629
$F_{\mathbb{DE}}$	0.9680	0.0557
$F_{\mathbb{L}}$	0.8970	0.0700

Table 7. Estimated values of $V_{1N}(T'_{S_x}, F)$ and $V_{2N}(T'_{S_y}, F)$ based on various $F = F_*$.

Distribution	$V_{1N}(T'_{S_x}, F)$	$V_{2N}(T'_{S_y}, F)$
$F_{\mathbb{N}}$	-0.0025	0.0023
$F_{\mathbb{DE}}$	-0.0014	0.0013
$F_{\mathbb{L}}$	-0.0020	0.0018

Chapter 4

Proof

In this chapter we give the proof of Theorem 2.

Write $\hat{F}_m = (\hat{F}_m - F) + F$, $\hat{G}_n = (\hat{G}_n - G) + G$ and $d\widehat{H}_N = d(\widehat{H}_N - H_N) + dH_N$. Then the statistics (2.12) after a little simplification becomes

$$\widehat{T}_N = \mu_N + B_{1N} + B_{2N} + C_{1N} + C_{2N} + C_{3N},$$

where

$$\begin{aligned}
\mu_N &= \int (F - G)^2 dH_N(x), \\
B_{1N} &= \int (F - G)^2 d(\widehat{H}_N - H_N)(x), \\
B_{2N} &= 2\int (F - G)((\hat{F}_m - F) - (\hat{G}_n - G))dH_N(x), \\
C_{1N} &= \int ((\hat{F}_m - F) - (\hat{G}_n - G))^2 dH_N(x), \\
C_{2N} &= \int ((\hat{F}_m - F) - (\hat{G}_n - G))^2 d(\widehat{H}_N - H_N)(x), \\
C_{3N} &= 2\int (F - G)((\hat{F}_m - F) - (\hat{G}_n - G))d(\widehat{H}_N - H_N)(x).
\end{aligned}$$

To establish the proof of this theorem, we proceed to show that:

(i) the term μ_N is finite,

(ii) $B_{1N} + B_{2N}$ has a limiting Gaussian distribution, and

(iii) the C_* terms are uniformly of higher order.

Let us first show the statement (i). From (A.1), it is seen that

$$\left|\int (F-G)^2 dH_N(x)\right| \leq K \int_0^1 H_N(1-H_N)dH_N(x) \leq K < \infty.$$

Next we show the statement (ii). From (2.11) and integrating B_{1N} by parts, we observe that

$$\begin{aligned}
B_{1N} &= -2\int (F-G)(\widehat{H}_N - H_N)d(F-G)(x) \\
&= -2\{\lambda_N \int (F-G)(F_m - F)d(F-G)(x) \\
&\quad +(1-\lambda_N)\int (F-G)(G_n - G)d(F+G)(x) \\
&\quad +m^{-1/2}\lambda_N A_{\mathrm{x}} \int xf(x)(F-G)(x)d(F-G)(x) \\
&\quad +n^{-1/2}(1-\lambda_N)A_{\mathrm{y}} \int xg(x)(F-G)(x)d(F-G)(x)\} \\
&\quad + \text{lower order terms.}
\end{aligned}$$

Then, from (2.7), (2.9) and (2.11), we obtain

$$\begin{aligned}
N^{1/2}(B_{1N}+B_{2N}) &= 2N^{1/2}\Big\{\int s(x)d(G_n - G)(x) \\
&\quad -\int s^*(x)d(F_m - F)(x) \\
&\quad -m^{-1/2}A_{\mathrm{x}}\int xf(x)(F-G)(x)dG(x) \\
&\quad -n^{-1/2}A_{\mathrm{y}}\int zg(z)(F-G)(z)dF(z)\Big\} \\
&\quad + \text{lower order terms} \\
&= a_N + b_N + c_N + d_N + \text{lower order terms, (say),} \quad (4.1)
\end{aligned}$$

where

$$s(x) = \int_{x_0}^{x} (F-G)(y)dF(y) \quad \text{and} \quad s^*(x) = \int_{x_0}^{x} (F-G)(y)dG(y)$$

with $x_0 > 0$ determined somewhat arbitrarily.

To compute the variance of (4.1), we shall first find a bound on the moments of $s(x)$ and $s^*(x)$. Using (A.1) and the fact that $dH_N \geq \lambda_0 dF$, we see that

$$E\{|s(x)|\}^{2+\delta} \leq K \int_0^1 (H_N(1-H_N))^{1+\frac{\delta}{2}} dH_N(x) \leq K < \infty,$$

and similarly, we can establish that

$$E\{|s^*(x)|\}^{2+\delta} < \infty, \quad 0 < \delta \leq 1.$$

We shall now find the variance of (4.1). Noting that a_N and b_N are mutually independent random variables, and using the result by Chernoff and Savage (1958, p.976), we obtain

$$\sigma_{1N}^2 = Var(a_N + b_N). \tag{4.2}$$

Similarly, we can compute the same for c_N and d_N by first observing Theorem 1 that

$$Var(m^{1/2}(\hat{\theta}_{\mathrm{x},m} - \theta_{\mathrm{x}})) = \mathcal{U}_{\mathrm{x}}^{-1}\mathcal{R}_{\mathrm{x}}\mathcal{U}_{\mathrm{x}}^{-1}$$

and

$$Var(n^{1/2}(\hat{\theta}_{\mathrm{y},n} - \theta_{\mathrm{y}})) = \mathcal{U}_{\mathrm{y}}^{-1}\mathcal{R}_{\mathrm{y}}\mathcal{U}_{\mathrm{y}}^{-1}.$$

Thus, recalling (2.8), (2.10) and (2.11), we get

$$\sigma_{2N}^2 = Var(c_N) \quad \text{and} \quad \sigma_{3N}^2 = Var(d_N). \tag{4.3}$$

We next compute the covariance terms. Since $\{X_t\}$ and $\{Y_t\}$ are independent, we have only to evaluate

$$K_{1N} = 2E(b_N c_N) \quad \text{and} \quad K_{2N} = 2E(a_N d_N).$$

From (3.1), we obtain

$$K_{1N} = -8\lambda_N^{-1} \iint E\{m^{1/2}(F_m - F)(x)A_{\mathrm{x}}\}\rho_f(x,y)dG(x)dG(y),$$

for which, it is necessary to find $E\{\cdot\}$. Using the result by Horváth *et al.* (2001), it follows from (2.8) and (2.13) that

$$E(m^{1/2}(F_m - F)(x)A_{\mathrm{x}}) = \psi_{\mathrm{x}}(x) \sum_{0 \le i \le p_{\mathrm{x}}} \tau_{\mathrm{x},i}\delta_{\mathrm{x},i},$$

where $\psi_{\mathrm{x}}(x)$ is defined in Theorem 2. Thus,

$$K_{1N} = -8\lambda_N^{-1} \sum_{0 \le i \le p_{\mathrm{x}}} \tau_{\mathrm{x},i}\delta_{\mathrm{x},i} \iint \psi_{\mathrm{x}}(x)\rho_f(x,y)dG(x)dG(y)$$

and similarly

$$K_{2N} = 8(1-\lambda_N)^{-1} \sum_{0 \le i \le p_{\mathrm{y}}} \tau_{\mathrm{y},i}\delta_{\mathrm{y},i} \iint \psi_{\mathrm{y}}(x)\rho_f(x,z)dF(x)dF(x).$$

Adding K_{1N} and K_{2N} produces ζ_N defined in Theorem 2.

Hence, using the term ζ_N, (4.2), (4.3), Theorem 1 and the central limit theorems given by Horváth *et al.* (2001), we may conclude that

$$N^{1/2}(B_{1N} + B_{2N})/\sigma_N \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as } N \to \infty.$$

We finally show the statement (iii). For this, we need the following elementary results (see Chernoff and Savage (1958, p.986)).

(E.1) $dH_N \ge \lambda_N dF \ge \lambda_0 dF$.

(E.2) $dH_N \ge (1-\lambda_N)dG \ge \lambda_0 dG$.

(E.3) $1 - F \le (1-H_N)/\lambda_N \le (1-H_N)/\lambda_0$.

(E.4) $1 - G \le (1-H_N)/(1-\lambda_N) \le (1-H_N)/\lambda_0$.

(E.5) $F(1-F) \le H_N(1-H_N)/\lambda_N^2 \le H_N(1-H_N)/\lambda_0^2$.

(E.6) $G(1-G) \le H_N(1-H_N)/\lambda_0^2$.

Let (α_N, β_N) be the interval S_{N_ϵ}, where

$$S_{N_\epsilon} = \{x : H_N(1-H_N) > \eta_\epsilon\lambda_0 N^{-1}\}. \tag{4.4}$$

Then η_ϵ can be chosen independently of F, G, and λ_N so that

$$P[\varepsilon_t^2 \in S_{N_\epsilon},\ t=1,\ldots,m, \quad \xi_t^2 \in S_{N_\epsilon},\ t=1,\ldots,n] \geq 1-\epsilon.$$

Let us first evaluate the random variable C_{1N}. Using (2.7) and (2.9), we obtain

$$\begin{aligned}
C_{1N} &= \int ((F_m - F) - (G_n - G))^2 dH_N(x) \\
&\quad +2m^{-1/2} A_{\mathrm{x}} \int x f(x)(F_m - F)(x) dH_N(x) \\
&\quad -2m^{-1/2} A_{\mathrm{x}} \int x f(x)(G_n - G)(x) dH_N(x) \\
&\quad -2n^{-1/2} A_{\mathrm{y}} \int x g(x)(F_m - F)(x) dH_N(x) \\
&\quad +2n^{-1/2} A_{\mathrm{y}} \int x g(x)(G_n - G)(x) dH_N(x) \\
&\quad +m^{-1} A_{\mathrm{x}}^2 \int x^2 f^2(x) dH_N(x) \\
&\quad +n^{-1} A_{\mathrm{y}}^2 \int x^2 g^2(x) dH_N(x) \\
&\quad -2m^{-1/2} n^{-1/2} A_{\mathrm{x}} A_{\mathrm{y}} \int (x f(x))(x g(x)) dH_N(x) \\
&\quad + \text{lower order terms} \\
&= \sum_{i=1}^{8} C_{1iN} + \text{lower order terms}, \quad \text{(say)}.
\end{aligned}$$

We first deal with C_{11N}. In what follows, we mean that all mathematical relations, e.g., $\leq$, $=$ etc. hold with probability $1-\epsilon$. Since $\{X_t\}$ and $\{Y_t\}$ are independent, it follows from (E.5), (E.6) and (4.4) that

$$\begin{aligned}
E(|C_{11N}|) &= \frac{1}{N} \int_{S_{N_\epsilon}} \left[\frac{1}{\lambda_N} F(1-F) + \frac{1}{1-\lambda_N} G(1-G) \right] dH_N(x) \\
&\leq \frac{K}{N} \int_{S_{N_\epsilon}} H_N(1-H_N) dH_N(x) \\
&= \frac{1}{N} O[(H_N(\beta_N)(1-H_N(\beta_N)))^2] = o(N^{-1}). \qquad (4.5)
\end{aligned}$$

Therefore, by the dominated convergence theorem, we have $C_{11N} = o_p(N^{-1/2})$. Next we turn to C_{12N}, for which, it suffices to show

$$\int_{S_{N\epsilon}} \left\{ \int_{x_0}^{x} yf(y)dH_N(y) \right\} d(F_m - F)(x) = o_p(1). \tag{4.6}$$

In view of (A.1)-(A.3) and (4.4), we see that (4.6) is dominated by

$$\begin{aligned}
&\int_{S_{N\epsilon}} \left\{ \int_{x_0}^{x} |yf(y)|dH_N(y) \right\} |d(F_m - F)(x)| \\
&\leq K \int_{S_{N\epsilon}} \left\{ \int_{x_0}^{x} H_N(y)(1 - H_N(y))dH_N(y) \right\} |d(F_m - F)(x)| \\
&\leq \int_{S_{N\epsilon}} O[(H_N(x)(1 - H_N(x)))^2] |d(F_m - F)(x)| \\
&= m^{-1/2} \int_{S_{N\epsilon}} O(N^{-2}) |d(m^{1/2}(F_m - F)(x))| \\
&= o_p(1) \quad \text{(e.g., Puri and Sen (1993), Theorem 2.11.6),} \qquad (4.7)
\end{aligned}$$

which, together with the fact $(m^{-1/2}|\mathcal{A}_\text{x}|) = O_p(m^{-1/2})$, implies $C_{12N} = o_p(N^{-1/2})$. The proof for $C_{13N} = C_{14N} = C_{15N} = o_p(N^{-1/2})$ is analogous to (4.7). Now we consider C_{16N}. Following the arguments of (4.5) and (4.7), it is seen that

$$\begin{aligned}
|C_{16N}| &\leq m^{-1}|\mathcal{A}_\text{x}|^2 \int_{S_{N\epsilon}} |xf(x)|^2 dH_N(x) \\
&\leq O_p(m^{-1}) \int_{S_{N\epsilon}} (H_N(1 - H_N))^2 dH_N(x) = o_p(N^{-1}), \qquad (4.8)
\end{aligned}$$

hence, we have $C_{16N} = o_p(N^{-1/2})$. To complete the assertion for C_{1N}, we can similarly show $C_{17N} = C_{18N} = o_p(N^{-1/2})$. Consequently, we have

$$C_{1N} = o_p(N^{-1/2}).$$

Next we deal with C_{2N}. Recalling (2.12), we obtain

$$\begin{aligned}
C_{2N} &= \int ((\hat{F}_m - F) - (\hat{G}_n - G))^2 d(\mathcal{H}_N - H_N)(x) \\
&\quad + m^{-1/2}\lambda_N \mathcal{A}_\text{x} \int ((\hat{F}_m - F) - (\hat{G}_n - G))^2 d(xf(x)) \\
&\quad + n^{-1/2}(1 - \lambda_N)\mathcal{A}_\text{y} \int ((\hat{F}_m - F) - (\hat{G}_n - G))^2 d(xg(x)) \\
&\quad + \text{lower order terms} \\
&= C_{21N} + C_{22N} + C_{23N} + \text{ lower order terms,} \quad \text{(say),}
\end{aligned}$$

where $(\mathcal{H}_N - H_N)(x) = \lambda_N(F_m - F)(x) + (1-\lambda_N)(G_n - G)(x)$. Let us first evaluate C_{21N}. By analogy with the first C term, we have

$$\begin{aligned}
C_{21N} &= \int((F_m - F) - (G_n - G))^2 d(\mathcal{H}_N - H_N)(x) \\
&\quad +2m^{-1/2}\mathcal{A}_{\mathrm{x}} \int xf(x)(F_m - F)(x)d(\mathcal{H}_N - H_N)(x) \\
&\quad -2m^{-1/2}\mathcal{A}_{\mathrm{x}} \int xf(x)(G_n - G)(x)d(\mathcal{H}_N - H_N)(x) \\
&\quad -2n^{-1/2}\mathcal{A}_{\mathrm{y}} \int xg(x)(F_m - F)(x)d(\mathcal{H}_N - H_N)(x) \\
&\quad +2n^{-1/2}\mathcal{A}_{\mathrm{y}} \int xg(x)(G_n - G)(x)d(\mathcal{H}_N - H_N)(x) \\
&\quad +m^{-1}\mathcal{A}_{\mathrm{x}}^2 \int x^2 f^2(x)d(\mathcal{H}_N - H_N)(x) \\
&\quad +n^{-1}\mathcal{A}_{\mathrm{y}}^2 \int x^2 g^2(x)d\mathcal{H}_N - H_N)(x) \\
&\quad -2m^{-1/2}n^{-1/2}\mathcal{A}_{\mathrm{x}}\mathcal{A}_{\mathrm{y}} \int (xf(x))(xg(x))d(\mathcal{H}_N - H_N)(x) \\
&\quad + \text{ lower order terms} \\
&= \sum_{i=1}^{8} C_{21iN} + \text{ lower order terms}, \quad \text{(say)}.
\end{aligned}$$

Let us first consider C_{211N}. Since $\{X_t\}$ and $\{Y_t\}$ are independent, we have only to evaluate

$$\begin{aligned}
E(|C_{211N}|) &= E\Big\{\lambda_N \int_{S_{N\epsilon}} (F_m - F)^2 d(F_m - F)(x) \\
&\quad +(1-\lambda_N)\int_{S_{N\epsilon}} (G_n - G)^2 d(G_n - G)(x)\Big\}.
\end{aligned}$$

From the result by Chernoff and Savage (1958, p.990) and (4.5), it follows that

$$\begin{aligned}
E(|C_{211N}|) &= \frac{1}{N^2}\Big[\frac{1}{\lambda_N}\int_{S_{N\epsilon}} (1-F)(1-2F)dF(x) \\
&\quad +\frac{1}{1-\lambda_N}\int_{S_{N\epsilon}} (1-G)(1-2G)dG(x)\Big] \\
&\le \frac{K}{N^2}\int_{S_{N\epsilon}} dH_N(x) = o(N^{-1}),
\end{aligned}$$

which implies $C_{211N} = o_p(N^{-1/2})$. Next we consider C_{212N}, which on integrating by parts gives

$$C_{212N} = m^{-1/2}\mathcal{A}_{\mathrm{x}}\{-\lambda_N C^*_{212N} + 2(1-\lambda_N)C^{**}_{212N}\},$$

where

$$\begin{aligned} C^*_{212N} &= \int_{S_{N_\epsilon}} (F_m - F)^2 d(xf(x)), \\ C^{**}_{212N} &= \int_{S_{N_\epsilon}} xf(x)(F_m - F)(x)d(G_n - G)(x). \end{aligned}$$

Let us first deal with C^*_{212N}. From (A.2), (A.4) and (4.5), it follows that

$$\begin{aligned} E(|C^*_{212N}|) &\leq \frac{1}{cN\lambda_N}\int_{S_{N_\epsilon}} F(1-F)dF(x) \\ &\leq \frac{K}{N}\int_{S_{N_\epsilon}} H_N(1-H_N)dH_N(x) = o(N^{-1}). \end{aligned} \tag{4.9}$$

Next we turn to C^{**}_{212N}. Since $\{X_t\}$ and $\{Y_t\}$ are independent, we have

$$E(C^{**}_{212N}) = E[E(C^{**}_{212N}|\xi_1^2,\dots,\xi_n^2)] = 0, \quad E[(C^{**}_{212N})^2|\xi_1^2,\dots,\xi_n^2] = C^{***}_{212N},$$

$$\begin{aligned} C^{***}_{212N} &= \frac{2}{m}\iint\limits_{\substack{x,y\varepsilon S_{N_\epsilon}\\ x<y}} xyf(x)f(y)F(x)(1-F(y))d((G_n-G)(x)(G_n-G)(y)), \\ E(|C^{***}_{212N}|) &\leq \frac{2}{mn}\iint\limits_{\substack{x,y\varepsilon S_{N_\epsilon}\\ x<y}} |xyf(x)f(y)|F(x)(1-F(y))dG(x)dG(y) \\ &\leq \frac{K}{N^2}\iint\limits_{\substack{x,y\varepsilon S_{N_\epsilon}\\ x<y}} |xyf(x)f(y)|H_N(x)(1-H_N(y))dH_N(x)dH_N(y) \\ &\leq \frac{K}{N^2}\iint\limits_{0<x<y<1} x^2(1-x)y(1-y)^2dxdy = o(N^{-1}). \end{aligned} \tag{4.10}$$

Thus, using the dominated convergence theorem, $(m^{-1/2}|\mathcal{A}_{\mathrm{x}}|) = O_p(m^{-1/2})$, (4.9) and (4.10), we have $C_{212N} = o_p(N^{-1})$. The proof for $C_{213N} = C_{214N} = C_{215N} = o_p(N^{-1})$ can be handled similar to C_{212N}. Now we turn to evaluate C_{216N}, where

$$C_{216N} = m^{-1}\mathcal{A}_{\mathrm{x}}^2\{\lambda_N C^*_{216N} + (1-\lambda_N)C^{**}_{216N}\} \tag{4.11}$$

with

$$\begin{aligned}
C^*_{216N} &= \int_{S_{N_\epsilon}} (xf(x))^2 d(F_m - F)(x), \\
C^{**}_{216N} &= \int_{S_{N_\epsilon}} (xf(x))^2 d(G_n - G)(x).
\end{aligned}$$

Following the arguments of (4.7), we can easily show $C^*_{216N} = C^{**}_{216N} = o_p(1)$, which, together with $(m^{-1}|\mathcal{A}_{\mathrm{x}}|^2) = O_p(m^{-1})$, implies $C_{216N} = o_p(N^{-1})$. Similarly, we can prove $C_{217N} = C_{218N} = o_p(N^{-1})$. Hence, we have $C_{21N} = o_p(N^{-1/2})$.

Next we consider C_{22N}. In the same way as for C_{1N}, we obtain

$$\begin{aligned}
C_{22N} &= m^{-1/2}\lambda_N \mathcal{A}_{\mathrm{x}} \int ((F_m - F) - (G_n - G))^2 d(xf(x)) \\
&\quad + 2m^{-1}\lambda_N \mathcal{A}_{\mathrm{x}}^2 \int xf(x)(F_m - F)(x) d(xf(x)) \\
&\quad - 2m^{-1}\lambda_N \mathcal{A}_{\mathrm{x}}^2 \int xf(x)(G_n - G)(x) d(xf(x)) \\
&\quad - 2m^{-1/2}n^{-1/2}\lambda_N \mathcal{A}_{\mathrm{x}}\mathcal{A}_{\mathrm{y}} \int xg(x)(F_m - F)(x) d(xf(x)) \\
&\quad + 2m^{-1/2}n^{-1/2}\lambda_N \mathcal{A}_{\mathrm{x}}\mathcal{A}_{\mathrm{y}} \int xg(x)(G_n - G)(x) d(xf(x)) \\
&\quad + m^{-3/2}\lambda_N \mathcal{A}_{\mathrm{x}}^3 \int x^2 f^2(x) d(xf(x)) \\
&\quad + m^{-1/2}n^{-1}\lambda_N \mathcal{A}_{\mathrm{x}}\mathcal{A}_{\mathrm{y}}^2 \int x^2 g^2(x) d(xf(x)) \\
&\quad - 2m^{-1}n^{-1/2}\lambda_N \mathcal{A}_{\mathrm{x}}^2\mathcal{A}_{\mathrm{y}} \int (xf(x))(xg(x)) d(xf(x)) \\
&\quad + \text{lower order terms} \\
&= \sum_{i=1}^{8} C_{22iN} + \text{lower order terms}, \quad \text{(say)}.
\end{aligned}$$

Let us first consider $C_{221N} = m^{-1/2}\lambda_N \mathcal{A}_{\mathrm{x}} C^*_{221N}$, where

$$C^*_{221N} = \int ((F_m - F) - (G_n - G))^2 d(xf(x)).$$

Recalling (A.2), (A.4) and (4.7), we obtain

$$\begin{aligned}
E(|C^*_{221N}|) &\le \frac{1}{cN} \int_{S_{N_\epsilon}} \left[\frac{F(1-F)}{\lambda_N} + \frac{G(1-G)}{1-\lambda_N}\right] dF(x) \\
&\le \frac{K}{N} \int_{S_{N_\epsilon}} H_N(1-H_N) dH_N(x) = o(N^{-1}).
\end{aligned}$$

Therefore, by the dominated convergence theorem and $(m^{-1/2}|\mathcal{A}_{\mathrm{x}}|) = O_p(m^{-1/2})$, we have $C_{221N} = o_p(N^{-1})$. Next we evaluate C_{222N}, which on integrating by parts produces $C_{222N} = -m^{-1}\lambda_N \mathcal{A}_{\mathrm{x}}^2 C^*_{222N}$, where

$$C^*_{222N} = \int_{S_{N_\epsilon}} (xf(x))^2 d(F_m - F)(x). \tag{4.12}$$

Using the result of (4.11), it follows that $C^*_{222N} = o_p(1)$, which implies $C_{222N} = o_p(N^{-1})$. Similarly we can prove $C_{223N} = o_p(N^{-1})$. We now turn to evaluate C_{224N}. Now using (A.2)-(A.4) and (4.7), it is easy to show

$$\int_{S_{N_\epsilon}} \left\{ \int_{x_0}^{x} yg(y) d(yf(y)) \right\} d(F_m - F)(x) = o_p(1),$$

which, combined with $(m^{-1/2}n^{-1/2}|\mathcal{A}_{\mathrm{x}}||\mathcal{A}_{\mathrm{y}}|) = O_p(m^{-1/2}n^{-1/2})$, entails $C_{224N} = o_p(N^{-1})$. Similarly, we can prove $C_{225N} = o_p(N^{-1})$. Next, we consider C_{226N}. In view of (A.2)-(A.4) and (4.8), we see that

$$|C_{226N}| \leq O_p(m^{-3/2}) \int_{S_{N_\epsilon}} (H_N(1 - H_N))^2 dF(x) = o_p(N^{-1}).$$

Thus, $C_{226N} = o_p(N^{-1/2})$. Similarly, we can prove $C_{227N} = C_{228N} = o_p(N^{-1/2})$. Hence, we have $C_{22N} = o_p(N^{-1/2})$. The proof of $C_{23N} = o_p(N^{-1/2})$ follows precisely on the same lines as that of C_{22N}. Consequently, we have

$$C_{2N} = o_p(N^{-1/2}).$$

Finally we evaluate C_{3N}. By analogy with the second C term, we obtain

$$\begin{aligned}
C_{3N} &= 2\int (F-G)(F_m-F)d(\mathcal{H}_N-H_N)(x) \\
&\quad -2\int (F-G)(G_n-G)d(\mathcal{H}_N-H_N)(x) \\
&\quad +2m^{-1/2}\mathcal{A}_{\mathrm{x}}\int xf(x)(F-G)(x)d(\mathcal{H}_N-H_N)(x) \\
&\quad -2n^{-1/2}\mathcal{A}_{\mathrm{y}}\int xg(x)(F-G)(x)d(\mathcal{H}_N-H_N)(x) \\
&\quad +2m^{-1/2}\lambda_N\mathcal{A}_{\mathrm{x}}\int (F-G)(F_m-F)d(xf(x)) \\
&\quad -2m^{-1/2}\lambda_N\mathcal{A}_{\mathrm{x}}\int (F-G)(G_n-G)d(xf(x)) \\
&\quad +2n^{-1/2}(1-\lambda_N)\mathcal{A}_{\mathrm{y}}\int (F-G)(F_m-F)d(xg(x)) \\
&\quad -2n^{-1/2}(1-\lambda_N)\mathcal{A}_{\mathrm{y}}\int (F-G)(G_n-G)d(xg(x)) \\
&\quad +2m^{-1}\lambda_N\mathcal{A}_{\mathrm{x}}^2\int xf(x)(F-G)(x)d(xf(x)) \\
&\quad -2n^{-1}(1-\lambda_N)\mathcal{A}_{\mathrm{y}}^2\int xg(x)(F-G)(x)d(xg(x)) \\
&\quad -2m^{-1/2}n^{-1/2}\lambda_N\mathcal{A}_{\mathrm{x}}\mathcal{A}_{\mathrm{y}}\int xg(x)(F-G)(x)d(xf(x)) \\
&\quad +2m^{-1/2}n^{-1/2}(1-\lambda_N)\mathcal{A}_{\mathrm{x}}\mathcal{A}_{\mathrm{y}}\int xf(x)(F-G)(x)d(xg(x)) \\
&\quad + \text{lower order terms} \\
&= \sum_{i=1}^{12} C_{3iN} + \text{lower order terms}, \quad \text{(say)}.
\end{aligned}$$

Let us first consider C_{31N}, which on integrating by parts yields

$$C_{31N} = \lambda_N(C_{31N}^{*} - C_{31N}^{**}) + 2(1-\lambda_N)C_{31N}^{***},$$

where

$$\begin{aligned}
C_{31N}^{*} &= \int_{S_{N_\epsilon}} (F_m-F)^2 dG(x), \\
C_{31N}^{**} &= \int_{S_{N_\epsilon}} (F_m-F)^2 dF(x), \\
C_{31N}^{***} &= \int_{S_{N_\epsilon}} (F-G)(F_m-F)d(G_n-G)(x).
\end{aligned}$$

Recalling (4.10) and (4.11), we can easily show $C_{31N} = o_p(N^{-1/2})$, and analogously $C_{32N} = o_p(N^{-1/2})$. Next we turn to evaluate

$$C_{33N} = 2m^{-1/2}\mathcal{A}_{\mathrm{x}}\{\lambda_N C^*_{33N} + (1-\lambda_N)C^{**}_{33N}\},$$

where

$$\begin{aligned} C^*_{33N} &= \int_{S_{N_\epsilon}} xf(x)(F-G)(x)d(F_m - F)(x), \\ C^{**}_{33N} &= \int_{S_{N_\epsilon}} xf(x)(F-G)(x)d(G_n - G)(x). \end{aligned}$$

Using (A.1)-(A.3) and (4.12), we can show $C^*_{33N} = C^{**}_{33N} = o_p(1)$, which, combined with $(m^{-1/2}|\mathcal{A}_{\mathrm{x}}|) = O_p(m^{-1/2})$, implies $C_{33N} = o_p(N^{-1/2})$. The proof for $C_{34N} = o_p(N^{-1/2})$ can be handled similarly. Now we turn to C_{35N}, for which, it suffices to show

$$\int_{S_{N_\epsilon}} \left\{ \int_{x_0}^{x} (F-G)(y)d(yf(y)) \right\} d(F_m - F)(x) = o_p(1). \tag{4.13}$$

In view of (A.1), (A.2), (A.4) and (4.7), it is seen that (4.13) is dominated by

$$\begin{aligned} & \int_{S_{N_\epsilon}} O[(H_N(x)(1-H_N(x)))^{3/2}]|d(F_m - F)(x)| \\ &= m^{-1/2} \int_{S_{N_\epsilon}} O(N^{-3/2})|d(m^{1/2}(F_m - F)(x))| = o_p(1). \end{aligned}$$

Therefore, $C_{35N} = o_p(N^{-1/2})$. Similarly, we can show $C_{36N} = C_{37N} = C_{38N} = o_p(N^{-1/2})$. Now consider C_{39N}. From (A.1)-(A.4) and (4.8), we obtain

$$|C_{39N}| \leq O_p(m^{-1}) \int_{S_{N_\epsilon}} (H_N(1-H_N))^{3/2} dF(x) = o_p(N^{-1}),$$

hence, $C_{39N} = o_p(N^{-1/2})$. Similarly we can show $C_{310N} = C_{311N} = C_{312N} = o_p(N^{-1/2})$. Consequently, we have

$$C_{3N} = o_p(N^{-1/2}).$$

This completes the proof of the theorem. ■

Chapter 5

Conclusion

This chapter provides the concluding remarks to the thesis. It also gives a brief overview of the related research that can be carried out in future by reformulating the results obtained in this thesis. Moreover it also discusses some implication and application aspects of the results.

In this thesis, we have derived the limiting Gaussian distribution of the two-sample Cramér-von Mises Statistics $\{\widehat{T}_N\}$ for ARCH residual empirical processes based on the techniques of Chernoff and Savage (1958) and Horváth *et al.* (2001). More concretely, we concluded that $N^{1/2}(\widehat{T}_N - \mu_N)/\sigma_N \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as } N \to \infty$, that is $\widehat{T}_N$ is normally distributed with mean μ_N and variance σ_N^2.

Under the null hypothesis $H_0 : F = G$, we observe $\sigma_N^2 = 0$, which indicates that we do not have a normal limit. However, $N\widehat{T}_N$ has a non-normal limit and this is the analogue result given by Anderson (1962).

It may be noted that the above results can easily be reformulated to the case of the one-sample as well as $c(\geq 2)$-sample problem and that the same result is true for GARCH processes as well, using the result by Berkes and Harváth (2003).

Finally the results obtained are widely used to study the asymptotic power and power efficiency of a class of two-sample tests. Thus the study motivates us to consider two independent samples from ARCH(p) processes as stated in section 1.5. For instance, let $\{X_t\}$ be a data set for the stock market in Australia and let $\{Y_t\}$ be another data set for the stock market in New Zealand with possibly non-Gaussian distributions F and G. In order to highlight the possible differences between these two sets of data, a nonparametric technique is used based on the two-sample Cramér-von Mises statistics. These statistics serve as a basis for the comparison in terms of tests of goodness of fit.

Acknowledgements

I would like to sincerely thank those who have assisted me in the production of this thesis.

My profound thanks to my supervisor, **Dr. Subhash Ajay Chandra**, lecturer in Statistics of the Mathematics and Computing Science department, University of the South Pacific, for introducing me to time series analysis. I deeply express my sincere appreciation for his copious valuable suggestions, motivation and counselling in every phase of the arduous but meritorious task and proper guidance till completion.

Special thanks to the head of Mathematics and Computing Science department, Dr. Jito Vanualailai for rendering services from the available facilities, including computer, photocopying and printing. Moreover, he also gave me an opportunity to present a part of my thesis at the Napier Conference, New Zealand, 2005.

I also extend my thanks to Mr. Bibhya Nand Sharma who generously supported, guided and encouraged me in the course of my study.

My sincere thanks goes to University of the South Pacific for providing a graduate assistant scholarship.

Finally, I am greatly indebted to my parents, Mr. and Mrs. Rao, and my two brothers for their moral support and motivation. I dedicate this piece of work to my wife (Shalvina).

Bibliography

Ahmad, I. A. (1996). *Modification of some goodness of fit statistics II: two-sample and symmetry testing*. Sankhya: The Indian J. Stat. **58**, Series A, Pt. 3, 464-472.

Anderson, T. W. (1962). *On the distribution of the two-sample Cramér-von Mises criterion*. Annals of Mathematical Statistics **33**, 1148-1159.

Anderson, T. W. (1977). *The Statistical Analysis of Time Series*. Wiley, New York.

Bollerslev, T. (1986). *Generalized Autoregressive Conditional Heteroskedasticity*. Journal of Econometrics **31**, 307-327.

Bollerslev, T., Chou, R. Y. and Kroner, K.F. (1992). *ARCH modelling in finance: a review of the thoery and empirical evidence*. J. Econometrics **52**, 5-59.

Bollerslev, T., Engle, R. and Nelson (1993). *"ARCH models"*. In handbook of Econometrics, Volume IV, Amsterdam, North Holland.

Brockwell, P.J. and Davis, R.A.(1994). *Time series: Data Analysis and Theory*. expanded ed. Holden-Day, San Francisco.

Chan, N.H. (2002). *Time Series: Application to finance*. Wiley, New York.

Chandra, S.A. and Taniguchi, M. (2001a). *Estimating functions for nonlinear time series models*. Ann. Inst. Statist. Math. **53**, 125-141.

Chandra, S.A. and Taniguchi, M. (2003). *Asymptotics of rank order statistics for ARCH residual empirical processes.* Stochastic Processes and their Applications **104**, 301-324.

Chen, M. and An, H.Z.(1998). *A note on the stationarity of the existence of moments of the GARCH model.* Statist. Sinica **8**, 505-510.

Chernoff. H. and Salvage, I. R. (1958). *Asymptotic normality efficiency of certain nonparametric test statistics.* Ann. Math. Statist. **29**, 972-994.

Dunis, C. and Zhou, B. (1998). *Nonlinear Modelling of High Frequency Financial Time series.* Wiley, New York.

Durbin, J. (1973). *Distribution Theory for Tests based on the Sample Distribution Function.* Regional Conference Series in Applied Mathematics 9, SIAM, Philadelphia, Pennsylvania.

Engle, R.F. (1982). *Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation.* Econometrica **50**, 987-1007.

Engle, R. (1995). *ARCH Selected Readings.* Oxford University Press, New York.

Fan, J. and Yao, Q. (2002). *Nonlinear Time Series: Nonparametric and Parametric methods.* Springer, New York.

Fuller, W. A. (1996). *Introduction to Statistical Time Series, 2nd ed..* Wiley, New York.

Godambe, V. P. (1985). *The foundations of finite sample estimation in stochastic processes.* Biometrika, **72**, 419-428.

Giraitis, L., Kokoszka, P. and Leipus, R. (2000). *Stationary ARCH models: dependence structure and central limit theorem.* Econometric theory **16**, 3-22.

Gouriéroux, C. (1997). *ARCH Models and Financial Applications.* Springer, New York.

Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering.* Springer, New York.

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

Hannan , E. J. (1970). *Multiple Time Series Analysis.* Wiley, New York.

Harvey , A., Ruiz, E. and Sheperd, N (1991a). *Modelling Volatility: Some alternatives to ARCH* . Discussion paper, Athens University.

Hill , D. L. and Rao, P.V. (1977). *Test of symmetry based on Cramér-von Mises statistics.* Biometrika **64**, 489-494.

Hodges, J. L. (1957). *The significance probability of the Smirnov two-sample test).* Arkiv för Matematik. **3**, 469-486.

Horváth, L., Kokoszka, P. and Teyssiere, G. (2001). *Empirical process of the squared residuals of the ARCH sequence.* Ann. Statis. **29**, 445-469.

Jan, R.M. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley, New York.

Klimko, L. A. and Nelson, P. I. (1978). *On conditional least squares estimation for stochastic processes.* Ann. Statist. **6**, 629-642.

Kurtz, T. G. and Protter, P. (1991). *Weak limit theorems for stochastic integrals and stochastic differential equations.* Annals of Mathematical Statistics **19**, 1035-1070.

Lee, S. and Taniguchi, M. (2005). *Asymptotic theory for ARCH-SM models: LAN and residual empirical processes.* Stat. Sinica **15**, 215-234.

Lehmann, E. L. (1951). *Consistency and unbiasedness of certain nonparametric tests*. Annals of Mathematical Statistics **22**, 165-179.

Lehmann, E. L. (1999). *Elements of Large Sample Theory*. Springer, New York.

Milhøj, A. (1985). *The moment structure of ARCH processes*. Scand. J. Statist. **12**, 281-292.

Puri, M.L. and Sen, P.K (1993). *Nonparametric Methods in Multivariate Analysis.* Wiley, New York.

Rosenblatt, M. (1952). *Limit theorems associated with variants of the von Mises statistic*. Annals of Mathematical Statistics **23**, 617-623.

Sen, P.K. and Singer, J.M (1993). *Large Sample Methods in Statistics: An Introduction with Applications*. Chapman & Hall, London.

Shephard, N. (1996). *Statistical aspects of ARCH and stochastic volatility: Time series models in econometrics, Finance and other fields*. Chapman & Hall, London.

Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Springer, New York.

Tjøstheim, D. (1986). *Estimation in nonlinear time series models*. Stoch. Proc. Appl. **21**, 251-273.

Tsay, R. S (2002). *Analysis of Financial Time Series*. Wiley, New York.

Weiss, A. A. (1986). *Asymptotic Theory for ARCH Models: Estimation and Testing*. Econometric theory **2**, 107-131-22.

Zivot, E. and Wang, J. (2003). *Modelling Financial Time series with S-Plus*. Springer, New York.