Asymptotic Distribution of Cramér-von Mises Statistics for ARCH Processes

S. Ajay Chandra*, Dinesh K. Rao
School of Computing, Information and Mathematical Sciences
The University of the South Pacific
Suva, Fiji

Sangyeol Lee
College of Statistics
Seoul National University
Korea

Abstract. This paper elucidates 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. 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 and it is then illustrated numerically based on such residual distributions. 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. In contrast with the independent, identically distributed or ARMA settings, these studies illuminate some interesting features of ARCH residuals.

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

1 Introduction

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). Since then, ARCH related processes have become perhaps the most popular and extensively studied financial econometric models (Engle (1995), Tsay (2002), Francq and Zakoian (2004), Chandra and Taniguchi (2005)). For a class of ARCH(∞) processes, which includes that of ARCH(p) processes as a special case, established sufficient conditions for the existence of a stationary solution and gave its explicit representation.

For time series data, residuals must be considered, and these residuals necessarily depend on parameter estimates, and inference based on these residuals, especially model goodness-of-fit tests, is a basic tool in the statistical analysis (see Brockwell and Davis (1994)). For an ARCH(p) process, Horváth *et al.* (2001) derived the limiting distribution

*chandra_sa@usp.ac.fj

of the empirical process based on the squared residuals. 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 local asymptotic normality for ARCH(∞) processes, and discussed the residual empirical process for an ARCH(p) process with stochastic mean.

In the i.i.d. settings, the 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. It is well known that these results are widely used to study the asymptotic power and power efficiency of a class of two-sample tests. 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, $\{\xi^2_{\mathrm{x},t}\}$ and $\{\xi^2_{\mathrm{y},t}\}$ with possibly non-Gaussian distributions F and G. In order to highlight the possible differences between these distributions, a nonparametric technique is employed based on a class of Cramér-von Mises statistics. Such statistics serves as a basis for the comparison in terms of tests of goodness-of-fit.

For a two-sample Cramér-von Mises statistic in the i.i.d. settings, Anderson (1962) derived the exact distribution, compared it to the limiting distribution, and found it to be a good approximation 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 object of this paper 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). The same result is true for GARCH processes as well using the result by Berkes and Horváth (2003). 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 (and of course the i.i.d. case). More concretely, the paper is organized as follows. Section 2 gives the setting of $\{\widehat{T}_N\}$ pertaining to empirical processes of the squared residuals from two independent samples of ARCH(p) processes and establishes its asymptotic distribution. This result, in Section 3, facilitates the study of asymptotic performance of $\{\widehat{T}_N\}$, such as the asymptotic relative 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 and it is then illustrated by simulations based on such residual distributions. 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. These studies help to highlight some important features of ARCH residuals in comparison with the independent, identically distributed or ARMA settings.

2 Two-sample Cramér-von mises statistics and main result

In this section we study a class of Cramér-von Mises statistics (see e.g., Durbin (1974, p.44)) for two-sample problem pertaining to empirical processes based on the squared residuals from two classes of ARCH processes.

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

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

where $\{\varepsilon_t\}$ is a sequence of i.i.d.(0,1) random variables with fourth-order cumulant κ_4^{x}, $\boldsymbol{\theta}_{\mathrm{x}} = (\theta_{\mathrm{x}}^0, \theta_{\mathrm{x}}^1, ..., \theta_{\mathrm{x}}^{p_{\mathrm{x}}})^T \in \boldsymbol{\Theta}_{\mathrm{x}} \subset \mathbb{R}^{p_{\mathrm{x}}+1}$ is an unknown parameter vector satisfying $\theta_{\mathrm{x}}^0 > 0$, $\theta_{\mathrm{x}}^i \geq 0$, $i = 1, \ldots, p_{\mathrm{x}} - 1$, $\theta_{\mathrm{x}}^{p_{\mathrm{x}}} > 0$, and ε_t is independent of $X_s, s < t$. Denote by $F(x)$ the distribution function of ε_t^2 and we assume that $f(x) = F'(x)$ exists and is continuous on $(0, \infty)$.

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

$$Y_t = \begin{cases} \sigma_t(\boldsymbol{\theta}_{\mathrm{y}})\xi_t,\ \sigma_t^2(\boldsymbol{\theta}_{\mathrm{y}}) = \theta_{\mathrm{y}}^0 + \sum_{i=1}^{p_{\mathrm{y}}} \theta_{\mathrm{y}}^i Y_{t-i}^2, & t = 1, \ldots, n, \\ 0, & t = -p_{\mathrm{y}} + 1, \ldots, 0, \end{cases} \tag{2}$$

where $\{\xi_t\}$ is a sequence of i.i.d.(0,1) random variables with fourth-order cumulant κ_4^{y}, $\boldsymbol{\theta}_{\mathrm{y}} = (\theta_{\mathrm{y}}^0, \theta_{\mathrm{y}}^1, ..., \theta_{\mathrm{y}}^{p_{\mathrm{y}}})^T \in \boldsymbol{\Theta}_{\mathrm{y}} \subset \mathbb{R}^{p_{\mathrm{y}}+1}$, $\theta_{\mathrm{y}}^0 > 0$, $\theta_{\mathrm{y}}^i \geq 0$, $i = 1, \ldots, p_{\mathrm{y}} - 1$, $\theta_{\mathrm{y}}^{p_{\mathrm{y}}} > 0$, are unknown parameters, and ξ_t is independent of $Y_s, s < t$. The distribution function of ξ_t^2 will be denoted by $G(x)$ and we assume that $g(x) = G'(x)$ exists and is continuous on $(0, \infty)$. For (1) and (2), we assume that $\theta_{\mathrm{x}}^1 + \cdots + \theta_{\mathrm{x}}^{p_{\mathrm{x}}} < 1$ and $\theta_{\mathrm{y}}^1 + \cdots + \theta_{\mathrm{y}}^{p_{\mathrm{y}}} < 1$ for stationarity (see Milhøj (1985)).

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

$$H_0 : F(x) = G(x) \ \text{ for all } x \quad \text{against} \quad H_A : F(x) \neq G(x) \ \text{ for some } x. \tag{3}$$

First consider the estimation of $\boldsymbol{\theta}_{\mathrm{x}}$ and $\boldsymbol{\theta}_{\mathrm{y}}$. Write $Z_{\mathrm{x},t} = X_t^2$, $\mathbf{W}_{\mathrm{x},t} = (1, Z_{\mathrm{x},t}, \ldots, Z_{\mathrm{x},t-p_{\mathrm{x}}+1})^T$ and $\zeta_{\mathrm{x},t} = (\varepsilon_t^2 - 1)\boldsymbol{\theta}_{\mathrm{x}}^T \mathbf{W}_{\mathrm{x},t-1}$. Then the autoregressive representation is given by

$$Z_{\mathrm{x},t} = \boldsymbol{\theta}_{\mathrm{x}}^T \mathbf{W}_{\mathrm{x},t-1} + \zeta_{\mathrm{x},t}, \quad 1 \leq t \leq m,$$

and analogously for (2),

$$Z_{\mathrm{y},t} = \boldsymbol{\theta}_{\mathrm{y}}^T \mathbf{W}_{\mathrm{y},t-1} + \zeta_{\mathrm{y},t}, \quad 1 \leq t \leq n,$$

where $Z_{\mathrm{y},t} = Y_t^2$, $\mathbf{W}_{\mathrm{y},t} = (1, Z_{\mathrm{y},t}, \ldots, Z_{\mathrm{y},t-p_{\mathrm{y}}+1})^T$ and $\zeta_{\mathrm{y},t} = (\xi_t^2 - 1)\boldsymbol{\theta}_{\mathrm{y}}^T \mathbf{W}_{\mathrm{y},\mathbf{t-1}}$. 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. Then the corresponding conditional least squares estimators (see Tjøstheim (1986)) of $\boldsymbol{\theta}_{\mathrm{x}}$ and

$\boldsymbol{\theta}_{\mathrm{y}}$ are given by $\hat{\boldsymbol{\theta}}_{\mathrm{x},m} = (\hat{\theta}^0_{\mathrm{x},m}, \ldots, \hat{\theta}^{p_{\mathrm{x}}}_{\mathrm{x},m})^T = \arg\min_{\theta_{\mathrm{x}}} \mathcal{Q}_m(\boldsymbol{\theta}_{\mathrm{x}})$ and $\hat{\boldsymbol{\theta}}_{\mathrm{y},n} = (\hat{\theta}^0_{\mathrm{y},n}, \ldots, \hat{\theta}^{p_{\mathrm{y}}}_{\mathrm{y},n})^T = \arg\min_{\theta_{\mathrm{y}}} \mathcal{Q}_n(\boldsymbol{\theta}_{\mathrm{y}})$, where

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

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

$$m^{1/2} \|\hat{\boldsymbol{\theta}}_{\mathrm{x},m} - \boldsymbol{\theta}_{\mathrm{x}}\| = O_p(1) \quad \text{and} \quad n^{1/2} \|\hat{\boldsymbol{\theta}}_{\mathrm{y},n} - \boldsymbol{\theta}_{\mathrm{y}}\| = O_p(1), \tag{4}$$

where $\|\cdot\|$ denotes the Euclidean norm. For the validity of (4), Tjøstheim (1986), pp.254-256) gave a set of sufficient conditions. Conditions (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

$$\hat{\varepsilon}_t^2 = X_t^2 / \sigma_t^2(\hat{\boldsymbol{\theta}}_{\mathrm{x},m}), \quad 1 \le t \le m \quad \text{and} \quad \hat{\xi}_t^2 = Y_t^2 / \sigma_t^2(\hat{\boldsymbol{\theta}}_{\mathrm{y},n}), \quad 1 \le t \le n, \tag{5}$$

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

For the testing problem (3), we begin by describing our approach in line with Chernoff and Savage (1958). Put $N = m + n$ and $\lambda_N = m/N$. For (5), the sizes m and n are assumed to be such that $0 < \lambda_0 \le \lambda_N \le 1 - \lambda_0 < 1$ 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

$$\widehat{H}_N(x) = \lambda_N \hat{F}_m(x) + (1 - \lambda_N) \hat{G}_n(x). \tag{6}$$

Write $\hat{B}_m(x) = m^{1/2}(\hat{F}_m(x) - F(x))$ and $\hat{B}_n(x) = n^{1/2}(\hat{G}_n(x) - G(x))$. Then

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

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,

$$\hat{B}_m(x) = m^{1/2}(F_m(x) - F(x)) + A_{\mathrm{x}} x f(x) + \text{ higher order terms}, \tag{7}$$

where

$$F_m(x) = \frac{1}{m} \sum_{t=1}^{m} \mathbb{I}(\varepsilon_t^2 \le x) \quad \text{and} \quad A_{\mathrm{x}} = \sum_{0 \le i \le p_{\mathrm{x}}} m^{1/2} (\hat{\theta}^i_{\mathrm{x},m} - \theta^i_{\mathrm{x}}) \tau_{\mathrm{x},i} \tag{8}$$

with $\tau_{\mathrm{x},0} = E(1/\sigma_t^2(\boldsymbol{\theta}_{\mathrm{x}}))$ and $\tau_{\mathrm{x},i} = E(X_{t-i}^2 / \sigma_t^2(\boldsymbol{\theta}_{\mathrm{x}}))$, $1 \le i \le p_{\mathrm{x}}$. By analogy with (7), the corresponding representation of $\hat{B}_n(x)$ is given by

$$\hat{B}_n(x) = n^{1/2}(G_n(x) - G(x)) + A_{\mathrm{y}} x g(x) + \text{ higher order terms}, \tag{9}$$

where

$$G_n(x) = \frac{1}{n}\sum_{t=1}^{n} \mathbb{I}(\xi_t^2 \le x) \quad \text{and} \quad A_{\mathrm{y}} = \sum_{0\le i\le p_{\mathrm{y}}} n^{1/2}(\hat{\theta}^i_{\mathrm{y},n} - \theta^i_{\mathrm{y}})\tau_{\mathrm{y},i} \tag{10}$$

with $\tau_{\mathrm{y},0} = E(1/\sigma_t^2(\boldsymbol{\theta}_{\mathrm{y}}))$ and $\tau_{\mathrm{y},i} = E(Y_{t-i}^2/\sigma_t^2(\boldsymbol{\theta}_{\mathrm{y}}))$, $1 \le i \le p_{\mathrm{y}}$. Hence, from (7) and (9), the expression (6) becomes

$$\widehat{H}_N(x) = \mathcal{H}_N(x) + m^{-1/2}\lambda_N A_{\mathrm{x}} x f(x) + n^{-1/2}(1-\lambda_N)A_{\mathrm{y}} x g(x) + \text{ higher order terms}, \tag{11}$$

where $\mathcal{H}_N(x) = \lambda_N F_m(x) + (1-\lambda_N)G_n(x)$ with $0 < \mathcal{H}_N < 1$. The decomposition (11) is basic and will be used repeatedly in the sequel.

For the testing problem (3), let us consider a class of Cramér-von Mises statistics of the form

$$\widehat{T}_N = \int (\hat{F}_m(x) - \hat{G}_n(x))^2 d\widehat{H}_N(x). \tag{12}$$

Note that (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 (1974, p.47))

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

This statistic was essentially proposed by Lehmann (1951) and studied by many researchers (e.g. Anderson (1962), Ahmad (1996)) who contributed to its development, and numerous theorems have been formulated for different tests. Noting that under $H_0 : F = G$, the quantity $\mathcal{H}_N$ for S_N converges to F almost surely, and we may conclude under certain regularity conditions that $(mn/N)S_N$ converges in law to $\int_0^1 Z^2(t)dt$, where $\{Z(t);\ 0 \le t \le 1\}$ is a Gaussian process with $E(Z(t)) = 0$ and $E(Z(s)Z(t)) = \min(s,t) - st, \quad 0 \le s,t \le 1$ (see e.g. Sen and Singer (1993), pp.341-349). Our case based on $\{\widehat{T}_N\}$ is somewhat different.

The object of this section is to elucidate the asymptotics of (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 1.

(A.1) $xf(x)$, $xg(x)$, $xf'(x)$ and $xg'(x)$ are uniformly bounded continuous, and integrable functions on $(0,\infty)$.

(A.2) There exists $c > 0$ such that $F(x) \ge c\{xf(x)\}$ and $G(x) \ge c\{xg(x)\}$ for all $x > 0$.

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

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

and

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

Introduce the notation $\boldsymbol{\mathcal{A}}_{\mathrm{x},t}^{\otimes s} = \boldsymbol{\mathcal{A}}_{\mathrm{x},t} \overbrace{\otimes \cdots \otimes}^{s \text{ times}} \boldsymbol{\mathcal{A}}_{\mathrm{x},t}$ (e.g., Hannan (1970, p.518)), and define $\boldsymbol{\Sigma}_{\mathrm{x},s} = E(\boldsymbol{\mathcal{A}}_{\mathrm{x},t}^{\otimes s})$ and $\boldsymbol{\Sigma}_{\mathrm{y},s} = E(\boldsymbol{\mathcal{A}}_{\mathrm{y},t}^{\otimes s})$, where $\otimes$ denotes the tensor product.

Assumption 2.

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

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

where $\|\cdot\|$ is the spectral matrix norm. From (B.2) 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 $\|\boldsymbol{\Sigma}_{\mathrm{x},3}\| < 1$ implies $\theta_{\mathrm{x}}^1 < 15^{-\frac{1}{3}} \approx 0.4$.

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

$$\begin{aligned} \boldsymbol{\mathcal{U}}_{\mathrm{x}} &= E(\mathbf{W}_{\mathrm{x},t-1}\mathbf{W}_{\mathrm{x},t-1}^T), \quad \boldsymbol{\mathcal{U}}_{\mathrm{y}} = E(\mathbf{W}_{\mathrm{y},t-1}\mathbf{W}_{\mathrm{y},t-1}^T), \\ \boldsymbol{\mathcal{R}}_{\mathrm{x}} &= (\kappa_4^{\mathrm{x}} + 2)E(\sigma_t^4(\boldsymbol{\theta}_{\mathrm{x}})\mathbf{W}_{\mathrm{x},t-1}\mathbf{W}_{\mathrm{x},t-1}^T) \quad \text{and} \quad \boldsymbol{\mathcal{R}}_{\mathrm{y}} = (\kappa_4^{\mathrm{y}} + 2)E(\sigma_t^4(\boldsymbol{\theta}_{\mathrm{y}})\mathbf{W}_{\mathrm{y},t-1}\mathbf{W}_{\mathrm{y},t-1}^T) \end{aligned}$$

are positive definite. To justify $\boldsymbol{\mathcal{R}}_{\mathrm{x}}$ as an illustration, first note that it is evidently nonnegative definite, i.e., $\boldsymbol{\alpha}^T\boldsymbol{\mathcal{R}}_{\mathrm{x}}\boldsymbol{\alpha} = (\kappa_4^{\mathrm{x}} + 2)E(\boldsymbol{\alpha}^T\sigma_t^2(\boldsymbol{\theta}_{\mathrm{x}})\boldsymbol{W}_{\mathrm{x},t-1})^2 \geq 0$ for any $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_{p_{\mathrm{x}}})^T \in \mathbb{R}^{p_{\mathrm{x}}+1}$. Moreover, if we suppose that $\boldsymbol{\mathcal{R}}_{\mathrm{x}}$ 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 $\alpha_0 + \alpha_1 Z_{\mathrm{x},s-1} + \cdots + \alpha_{j_0} Z_{\mathrm{x},s-j_0} = 0$ a.e. Here, note that $\sigma_t^2(\boldsymbol{\theta}_{\mathrm{x}}) > 0$ 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 $\sigma_s^2(\boldsymbol{\theta}_{\mathrm{x}})$ in (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_{x}, leading to a contradiction.

Now recalling the definition of $\mathcal{Q}_m(\boldsymbol{\theta}_{\mathrm{x}})$ and $\mathcal{Q}_n(\boldsymbol{\theta}_{\mathrm{y}})$, we observe that

$$\begin{aligned} \frac{\partial \mathcal{Q}_m}{\partial \theta_{\mathrm{x}}^0} &= -2\sum_{t=1}^{m}(\varepsilon_t^2 - 1)\sigma_t^2(\boldsymbol{\theta}_{\mathrm{x}}), \qquad \frac{\partial \mathcal{Q}_m}{\partial \theta_{\mathrm{x}}^i} = -2\sum_{t=1}^{m}(\varepsilon_t^2 - 1)\sigma_t^2(\boldsymbol{\theta}_{\mathrm{x}})Z_{\mathrm{x},t-i}, \quad 1 \leq i \leq p_{\mathrm{x}}, \\ \frac{\partial \mathcal{Q}_n}{\partial \theta_{\mathrm{y}}^0} &= -2\sum_{t=1}^{n}(\xi_t^2 - 1)\sigma_t^2(\boldsymbol{\theta}_{\mathrm{y}}), \qquad \frac{\partial \mathcal{Q}_n}{\partial \theta_{\mathrm{y}}^i} = -2\sum_{t=1}^{n}(\xi_t^2 - 1)\sigma_t^2(\boldsymbol{\theta}_{\mathrm{y}})Z_{\mathrm{y},t-i}, \quad 1 \leq i \leq p_{\mathrm{y}}, \end{aligned}$$

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

$$\begin{aligned} \hat{\theta}_{\mathrm{x},m}^i - \theta_{\mathrm{x}}^i &= \frac{1}{m}\sum_{t=1}^{m} V_{\mathrm{x},t}^i(\varepsilon_t^2 - 1) + o_p(m^{-1/2}), \quad 0 \leq i \leq p_{\mathrm{x}} \quad \text{and} \qquad (13) \\ \hat{\theta}_{\mathrm{y},n}^i - \theta_{\mathrm{y}}^i &= \frac{1}{n}\sum_{t=1}^{n} V_{\mathrm{y},t}^i(\xi_t^2 - 1) + o_p(n^{-1/2}), \quad 0 \leq i \leq p_{\mathrm{y}}, \end{aligned}$$

where $V^i_{\mathrm{x},t}$ and $V^i_{\mathrm{y},t}$ are the ith elements of $\boldsymbol{\mathcal{U}}_{\mathrm{x}}^{-1}\sigma_t^2(\boldsymbol{\theta}_{\mathrm{x}})\mathbf{W}_{\mathrm{x},t-1}$ and $\boldsymbol{\mathcal{U}}_{\mathrm{y}}^{-1}\sigma_t^2(\boldsymbol{\theta}_{\mathrm{y}})\mathbf{W}_{\mathrm{y},t-1}$. Write $\delta_{\mathrm{x},i} = E(V^i_{\mathrm{x},t})$, $0 \le i \le p_{\mathrm{x}}$, $\delta_{\mathrm{y},i} = E(V^i_{\mathrm{y},t})$, $0 \le i \le p_{\mathrm{y}}$, and $\boldsymbol{\tau}_{\mathrm{x}} = (\tau_{\mathrm{x},0}, \ldots, \tau_{\mathrm{x},p_{\mathrm{x}}})^T$ and $\boldsymbol{\tau}_{\mathrm{y}} = (\tau_{\mathrm{y},0}, \ldots, \tau_{\mathrm{y},p_{\mathrm{y}}})^T$ (recall (8) and (10)). Then, under $H_A : F \neq G$, we have the following result, whose proof is given in Section 4.

Theorem 1. Suppose that Assumptions 1 and 2 hold and that, in addition, $\hat{\boldsymbol{\theta}}_{\mathrm{x},m}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{y},n}$ are the conditional least squares estimators of $\boldsymbol{\theta}_{\mathrm{x}}$ and $\boldsymbol{\theta}_{\mathrm{y}}$ satisfying (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 $\mu_N = \int (F(x) - G(x))^2 dH_N(x)$ and $\sigma_N^2 = \sigma_{1N}^2 + \sigma_{2N}^2 + \sigma_{3N}^2 + \zeta_N \neq 0$ with

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

where

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

Remark 2.1. It may be noted that the above results can easily be reformulated to the case of the one-sample as well as $c(\ge 2)$-sample problem.

Remark 2.2. Observe that σ_{2N}^2, σ_{3N}^2 and ζ_N depend on the volatility estimators $\hat{\boldsymbol{\theta}}_{\mathrm{x},m}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{y},n}$. Hence, the asymptotics of $\{\widehat{T}_N\}$ are greatly different in comparison with the independent, identically distributed or ARMA settings.

Remark 2.3. 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 $\boldsymbol{\mathcal{U}}_{\mathrm{x}}^{-1}\boldsymbol{\mathcal{R}}_{\mathrm{x}}\boldsymbol{\mathcal{U}}_{\mathrm{x}}^{-1}$ and $\boldsymbol{\mathcal{U}}_{\mathrm{y}}^{-1}\boldsymbol{\mathcal{R}}_{\mathrm{y}}\boldsymbol{\mathcal{U}}_{\mathrm{y}}^{-1}$ are the asymptotic covariance matrices of $\sqrt{m}(\hat{\boldsymbol{\theta}}_{\mathrm{x},m} - \boldsymbol{\theta}_{\mathrm{x}})$ and $\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathrm{y},n} - \boldsymbol{\theta}_{\mathrm{y}})$, respectively, and their estimation is discussed in Gouriéroux (1997).

3 Asymptotic performance of $\{\widehat{T}_N\}$

The limiting distribution of $\{\widehat{T}_N\}$ given in the preceding section 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 and it is then 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 subsection, we consider the assessment of asymptotic relative efficiency of the statistics S_N 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 S_N.

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

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

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

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

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

where $\{\xi_t\}$ is a sequence of i.i.d.(0,1) random variables with fourth-order cumulant κ_4^{y}, $\boldsymbol{\theta}_{\mathrm{y}} = (\theta_{\mathrm{y}}^0, \theta_{\mathrm{y}}^1)^T$ with $\theta_{\mathrm{y}}^0 > 0$ and $0 \le \theta_{\mathrm{y}}^1 < 1$, 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 consider the location model in the case of $G(x) = F(x+\delta)$ for some parameter δ. 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;

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

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

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

and the variance under $H_A : \delta = 1$ 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

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

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

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 $H_0 : \delta = \delta_0$ against $H_A : \delta > \delta_0$ with critical region $\widehat{T}_N \geq \lambda_{N,\alpha}$. Further, suppose

(i) $\lim_{N\to\infty} P_{\delta_0}(\widehat{T}_N \geq \lambda_{N,\alpha}) = \alpha$, where $0 < \alpha < 1$ is a given level;

(ii) there exist functions $\mu_N(\delta)$ and $\sigma_N(\delta)$ such that $N^{1/2}(\widehat{T}_N - \mu_N(\delta))/\sigma_N(\delta) \xrightarrow{d} \mathcal{N}(0,1)$ uniformly in $\delta \in [\delta_0, \delta_0 + \epsilon]$, $\epsilon > 0$;

(iii) $\mu_N'(\delta_0) > 0$;

(iv) for a sequence $\{\delta_N = \delta_0 + N^{-1/2}k,\ k > 0\}$,

$$\lim_{N\to\infty}[\mu_N'(\delta_N)/\mu_N'(\delta_0)] = 1, \quad \lim_{N\to\infty}[\sigma_N(\delta_N)/\sigma_N(\delta_0)] = 1;$$

(v) $\lim_{N\to\infty}[\mu_N'(\delta_0)/\sigma_N(\delta_0)] = c > 0$.

For $\alpha \in (0,1)$, write $\lambda_\alpha = \Phi^{-1}(1-\alpha)$, where $\Phi(x) = \int_0^x (2\pi)^{-1/2} e^{-t^2/2} dt$. 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 $S_N = S_N(F)$ 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

$$F(x) = P(\varepsilon_t^2 \le x) = \begin{cases} 2F^*(\sqrt{x}) - 1, & x > 0, \\ 0, & x \le 0. \end{cases} \tag{14}$$

We shall now compute (14) 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, $F_{\mathbb{N}}(x) = 2F_{\mathbb{N}}^*(\sqrt{x}) - 1$, $f_{\mathbb{N}}(x) = (\sqrt{2\pi x})^{-1} e^{-x/2}$, $x > 0$.

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

(iii) F^* (Logistic):

$$F_{\mathbb{L}}^*(x) = (1 + e^{-x})^{-1}, \quad f_{\mathbb{L}}^*(x) = e^{-x}((1 + e^{-x})^2)^{-1}, \quad x \in \mathbb{R}.$$

In this case, $F_{\mathbb{L}}(x) = (1 - e^{-\sqrt{x}}/(1 + e^{-\sqrt{x}})$, $f_{\mathbb{L}}(x) = e^{-\sqrt{x}}/\sqrt{x}(1 + e^{-\sqrt{x}})^2$, $x > 0$.

Recalling the definition of $\mu_F(\delta)$, σ_F^2 and $\sigma_1^2(F)$, and assuming that $\mu_F(\delta)$ is continuously differentiable with respect to δ at $\delta = 1$ 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 S_N and $\widehat{T}_N$ between distributions F_1 and F_2 is

$$e(F_2, F_1) = c_{F_2}^2 / c_{F_1}^2, \tag{15}$$

where $c_F = \mu_F'(1)/\sigma(F)$ with $\sigma(F) = \sigma_1(F)$ and σ_F.

In an attempt to evaluate (15), 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 generated realizations of X_t and Y_t. Note that the above choice of the parameter values satisfies necessary conditions. On the basis of the conditional least squares estimators $\hat{\theta}^0_m$ and $\hat{\theta}^1_m$ of θ^0 and θ^1, respectively, the quantities C_{x}, C_{y}, k_1 and k_2 are estimated by the corresponding averages. In the actual computation of $\mu'_F(1)$, $\sigma_1^2(F)$ and σ_F^2, 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 $S_N = S_N(F)$ and $\widehat{T}_N = \widehat{T}_N(F)$ based on $F = F_*$

ARE	$S_N(F)$	$\widehat{T}_N(F)$			
		$m = n = 100$, $\theta^0 = 1$		$m = n = 500$, $\theta^0 = 1$	
		$\theta^1 = 0.1$	$\theta^1 = 0.3$	$\theta^1 = 0.1$	$\theta^1 = 0.3$
$e(F_{\mathbb{N}}, F_{\mathbb{DE}})$	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 somes 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 $S_N(F)$ 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 $\hat{\theta}_m$. 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 $\widehat{T}_N$ 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 $S_N(F)$ 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 principle is best in the case of heavy-tailed ARCH residual distributions.

3.2 ARCH volatility effect

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

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

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

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

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

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 $F_{\mathbb{L}}$ in the preceding subsection, 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	0.0554
$\tilde{\alpha}_{F_{\mathbb{L}}}$	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 S_N 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 (12).

It was shown in the proof of Theorem 1 (see (16)) that

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

where

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

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

and

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

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

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

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

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

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

$$\boldsymbol{\gamma}_{S_{\mathrm{x}}} = E(S_{\mathrm{x},t}^{(1)} \mathbf{W}_{S_{\mathrm{x}},t-1}) \quad \text{and} \quad \boldsymbol{\mathcal{U}}_{S_{\mathrm{x}}} = E(\mathbf{W}_{S_{\mathrm{x}},t-1}\mathbf{W}_{S_{\mathrm{x}},t-1}^T),$$

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

$$\mathbf{S}_{\mathrm{x},t}^h = (1-h)\mathbf{S}_{\mathrm{x},t} + h\mathbf{K}_{\mathrm{x},t} \equiv \mathbf{S}_{\mathrm{x},t} + h\mathbf{L}_{\mathrm{x},t}.$$

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

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

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

$$\frac{d}{dh}\boldsymbol{\mathcal{U}}_{S_{\mathrm{x}}^h}^{-1}\Big|_{h=0} = -\boldsymbol{\mathcal{U}}_{S_{\mathrm{x}}}^{-1}(\boldsymbol{\Delta}_{\mathrm{x}} + \boldsymbol{\Delta}_{\mathrm{x}}^T)\boldsymbol{\mathcal{U}}_{S_{\mathrm{x}}}^{-1}, \quad \boldsymbol{\Delta}_{\mathrm{x}} = E\Big[\begin{pmatrix} 0 \\ \mathbf{L}_{\mathrm{x},t-1}\end{pmatrix}\mathbf{W}_{S_{\mathrm{x}},t-1}^T\Big].$$

Also,

$$\frac{d}{dh}\boldsymbol{\gamma}_{S_{\mathrm{x}}^h}\Big|_{h=0} = E(L_{\mathrm{x},t}^{(1)}\mathbf{W}_{S_{\mathrm{x}},t-1}) + E\Big[S_{\mathrm{x},t}^{(1)}\begin{pmatrix} 0 \\ \mathbf{L}_{\mathrm{x},t-1}\end{pmatrix}\Big] \equiv \boldsymbol{\gamma}'_{S_{\mathrm{x}}},$$

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

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

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

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

The quantities $IFH(F,G)$, $\mathbf{T}'_{\mathbf{S}_x}$ and $\boldsymbol{T}'_{S_y}$ will facilitate the fundamental description of sensitiveness or insensitiveness of $\widehat{T}_N$.

Returning to the setup of Subsection 3.1, we describe the quantitative information for $U_N(F_m, G_m)$, $V_{1N}(\hat{\boldsymbol{\theta}}_{x,m}; F, G)$ and $V_{2N}(\hat{\boldsymbol{\theta}}_{y,m}; F, G)$ by computing the quantities $IFH(F,G) = I(F)$, $V_{1N}(\mathbf{T}'_{S_x}, F)$ and $V_{2N}(\mathbf{T}'_{S_y}, F)$, 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 the results for Makro 5

Table 3. 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 $m = n = 100$, $\theta^0 = 1$

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	0.0647	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 $F = F_{\mathbb{N}}$, $F_{\mathbb{DE}}$ and $F_{\mathbb{L}}$. 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 the values of $V_{1N}(\cdot,\cdot)$ and $V_{2N}(\cdot,\cdot)$ tend to zero when θ^1 increases for each $m = n$. This behavior depends not only on the choice of parameters but also on other aspects of $F = F_*$. 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 Section 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 $V_{1N}(T'_{S_{\mathrm{x}}}, F)$ and $V_{2N}(T'_{S_{\mathrm{y}}}, F)$ based on various $F = F_*$.

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

4 Proof

In this section we give the proof of Theorem 1. 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 (12) after a little simplification becomes

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

where

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

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

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). For $\delta > 0$, we can find $K > 0$ such that $|(F - G)(x)| < K[H_N(x)(1 - H_N(x))]^{1/2-\delta}$ (see e.g., Puri and Sen (1991, p.43)). Thus,

$$
\left| \int (F-G)^2 dH_N(x) \right| \le K \int_0^1 [H_N(1-H_N)]^{1-(\delta/2)} dH_N(x) \le K < \infty.
$$

Next we show the statement (ii). From (11) and integrating B_{1N} by parts, we observe that

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

Then, from (7), (9) and (11), we obtain

$$
\begin{aligned}
N^{1/2}(B_{1N} + B_{2N}) &= 2N^{1/2}\Big\{ \int s(x) d(G_n - G)(x) - \int s^*(x) d(F_m - F)(x) \\
&\quad -m^{-1/2} A_{\mathrm{x}} \int [xf(x)](F-G)(x) dG(x) \\
&\quad -n^{-1/2} A_{\mathrm{y}} \int [zg(z)](F-G)(z) dF(z) \Big\} \\
&\quad + \text{higher order terms} \\
&\equiv a_N + b_N + c_N + d_N + \text{ higher order terms,} \qquad (16)
\end{aligned}
$$

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 (16), we shall first find a bound on the $(2+\delta')$th moments of $s(x)$ and $s^*(x)$, where $0 < \delta' \le 1$. Choose $\delta' > 0$ such that $(2+\delta')(\frac{1}{2}-\delta) > -1$. Then using $|(F-G)(x)| < K[H_N(x)(1-H_N(x))]^{1/2-\delta}$ and the fact that $dH_N \ge \lambda_0 dF$, we see that

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

and similarly, we can establish that $E\{|s^*(x)|\}^{2+\delta'} < \infty$.

We shall now find the variance of (16). 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

$$\sigma_{1N}^2 = Var(a_N + b_N). \tag{17}$$

Similarly, we can compute the same for c_N and d_N by first observing the result of Tjøstheim (1986) that

$$Var(m^{1/2}(\hat{\boldsymbol{\theta}}_{\mathrm{x},m} - \boldsymbol{\theta}_{\mathrm{x}})) = \boldsymbol{\mathcal{U}}_{\mathrm{x}}^{-1}\boldsymbol{\mathcal{R}}_{\mathrm{x}}\boldsymbol{\mathcal{U}}_{\mathrm{x}}^{-1} \quad \text{and} \quad Var(n^{1/2}(\hat{\boldsymbol{\theta}}_{\mathrm{y},n} - \boldsymbol{\theta}_{\mathrm{y}})) = \boldsymbol{\mathcal{U}}_{\mathrm{y}}^{-1}\boldsymbol{\mathcal{R}}_{\mathrm{y}}\boldsymbol{\mathcal{U}}_{\mathrm{y}}^{-1}.$$

Thus, recalling (8), (10) and (11), we get

$$\sigma_{2N}^2 = Var(c_N) \quad \text{and} \quad \sigma_{3N}^2 = Var(d_N). \tag{18}$$

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 (14), 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 (8) and (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 1. 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 1.

Hence, using the term ζ_N, (17), (18), and the central limit theorems given by Horváth *et al.* (2001) and Tjøstheim (1986), 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) $dH_N \geq \lambda_N dF \geq \lambda_0 dF.$

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

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

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

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

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

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

$$S_{N_\epsilon} = \{x : H_N(1-H_N) > \eta_\epsilon \lambda_0 N^{-1}\}. \tag{19}$$

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 C_{1N}. From (7) and (9), we obtain

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

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 (19) that

$$\begin{aligned}
E(|C_{11N}|) &= \frac{1}{N} \int_{S_{N_\epsilon}} \Big[\frac{1}{\lambda_N} F(1-F) + \frac{1}{1-\lambda_N} G(1-G)\Big] dH_N(x) \\
&\leq \frac{K}{N} \int_{S_{N_\epsilon}} H_N(1-H_N) dH_N(x) \\
&= \frac{1}{N} O[(H_N(\beta_N)(1-H_N(\beta_N)))^2] = o(N^{-1}).
\end{aligned} \tag{20}$$

Therefore, by the Dominated Convergence Theorem, we have $C_{11N} = o_p(N^{-1/2})$. Next we turn to C_{12N}. Let

$$I_N(\delta^*) = \sup_x m^{1/2}|F_m(x) - F(x)| \leq C^*[F(x)(1 - F(x))]^{(1/2)-\delta^*}, \; \delta^* > 0, \; C^* > 0, \quad (21)$$

so that $P(I_N(\delta^*)) \geq 1 - \epsilon$ (see Puri and Sen (1993, p. 401)). Notice that from (A.1) and (A.2), we can find $K > 0$ such that $|xf(x)| \leq KH_N(1 - H_N)$. Thus, using (E.5), (19) and the fact $m^{-1}|A_{\rm x}| = O_p(m^{-1})$, we obtain

$$\begin{aligned} |C_{12N}| &\leq m^{-1}|A_{\rm x}| \int_{S_{N_\epsilon}} |xf(x)||m^{1/2}(F_m(x) - F(x))|dH_N(x) \\ &\leq O_p(m^{-1}) \int_{S_{N_\epsilon}} [H_N(1 - H_N)]^{(3/2)-\delta'} dH_N(x) \\ &= O_p(m^{-1})O[(H_N(\beta_N)(1 - H_N(\beta_N)))^{(5/2)-\delta'}] = o_p(N^{-1}). \end{aligned} \quad (22)$$

Hence, $C_{12N} = o_p(N^{-1/2})$. The proof for $C_{13N} = C_{14N} = C_{15N} = o_p(N^{-1/2})$ is analogous to (22). Now we consider C_{16N}. Following the arguments of (22), it is seen that

$$\begin{aligned} |C_{16N}| &\leq m^{-1}|A_{\rm x}|^2 \int_{S_{N_\epsilon}} |xf(x)|^2 dH_N(x) \\ &\leq O_p(m^{-1}) \int_{S_{N_\epsilon}} (H_N(1 - H_N))^2 dH_N(x) = o_p(N^{-1}), \end{aligned} \quad (23)$$

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 (12), we obtain

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

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

$$\begin{aligned} C_{21N} &= \int ((F_m - F) - (G_n - G))^2 d(\mathcal{H}_N - H_N)(x) \\ &\quad + 2m^{-1/2}A_{\rm x} \int (xf(x))(F_m - F)(x)d(\mathcal{H}_N - H_N)(x) \\ &\quad - 2m^{-1/2}A_{\rm x} \int (xf(x))(G_n - G)(x)d(\mathcal{H}_N - H_N)(x) \end{aligned}$$

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

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

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

From the result by Chernoff and Savage (1958, p.990) and (20), it follows that

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

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

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

where

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

Let us first deal with C_{212N}^{*}. From (A.1), (A.2) and (20), it follows that

$$
\begin{aligned}
E(|C_{212N}^{*}|) &\le \frac{1}{cN\lambda_N}\int_{S_{N\epsilon}}F(1-F)dF(x)\\
&\le \frac{K}{N}\int_{S_{N\epsilon}}H_N(1-H_N)dH_N(x)=o(N^{-1}). \qquad (24)
\end{aligned}
$$

With probability greater than $1-\epsilon$, there are no observations $\bar{S}_{N_\epsilon}$ and

$$|C^{**}_{212N}| \le K \int_{\bar{S}_{N_\epsilon}} (H_N(1-H_N))^2 dH_N(x) = o(N^{-1}).$$

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

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

$$\begin{aligned}
C^{****}_{212N} &= \frac{2}{m} \iint_{\substack{x,y\varepsilon S_{N_\epsilon}\\ x<y}} xyf(x)f(y)F(x)(1-F(y))d((G_n-G)(x)(G_n-G)(y)),\\
E(|C^{****}_{212N}|) &\le \frac{2}{mn} \iint_{\substack{x,y\varepsilon S_{N_\epsilon}\\ x<y}} |xyf(x)f(y)|F(x)(1-F(y))dG(x)dG(y)\\
&\le \frac{K}{N^2} \iint_{\substack{x,y\varepsilon S_{N_\epsilon}\\ x<y}} |xyf(x)f(y)|H_N(x)(1-H_N(y))dH_N(x)dH_N(y)\\
&\le \frac{K}{N^2} \iint_{0<x<y<1} x^2(1-x)y(1-y)^2dxdy \le \frac{K}{N^2}. \qquad (25)
\end{aligned}$$

Thus, using the Dominated Convergence Theorem, $m^{-1/2}|A_{\mathrm{x}}| = O_p(m^{-1/2})$, (24) and (25), 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

$$C_{216N} = m^{-1}A_{\mathrm{x}}^2\{\lambda_N C^*_{216N} + (1-\lambda_N)C^{**}_{216N}\}, \qquad (26)$$

where

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

It suffices to show $C^*_{216N} = o_p(1)$. From (A.1), (A.2) and (19), we see that C^*_{216N} is dominated by

$$\begin{aligned}
|C^*_{216N}| &\le \int_{S_{N_\epsilon}} O\{[H_N(1-H_N)]^2\}|d(F_m-F)(x)|\\
&= m^{-1/2}\int_{S_{N_\epsilon}} O\{N^{-2}\}d[m^{1/2}(F_m-F)(x)]|\\
&= o_p(1) \quad \text{(eg., Theorem 2.11.6 of Puri and Sen (1993))},
\end{aligned}$$

which, together with $(m^{-1}|A_{\mathrm{x}}|^2) = O_p(m^{-1})$, implies $C^*_{216N} = o_p(N^{-1})$. Similarly, we can prove $C^{**}_{216N} = o_p(N^{-1})$, and $C_{217N} = C_{218N} = o_p(N^{-1})$. 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

$$\begin{aligned} C_{22N} &= m^{-1/2}\lambda_N A_{\rm x}\int((F_m-F)-(G_n-G))^2 d(xf(x)) \\ &+2m^{-1}\lambda_N A_{\rm x}^2\int(xf(x))(F_m-F)(x)d(xf(x)) \\ &-2m^{-1}\lambda_N A_{\rm x}^2\int(xf(x))(G_n-G)(x)d(xf(x)) \\ &-2m^{-1/2}n^{-1/2}\lambda_N A_{\rm x}A_{\rm y}\int(xg(x))(F_m-F)(x)d(xf(x)) \\ &+2m^{-1/2}n^{-1/2}\lambda_N A_{\rm x}A_{\rm y}\int(xg(x))(G_n-G)(x)d(xf(x)) \\ &+m^{-3/2}\lambda_N A_{\rm x}^3\int x^2 f^2(x)d(xf(x)) \\ &+m^{-1/2}n^{-1}\lambda_N A_{\rm x}A_{\rm y}^2\int x^2 g^2(x)d(xf(x)) \\ &-2m^{-1}n^{-1/2}\lambda_N A_{\rm x}^2 A_{\rm y}\int(xf(x))(xg(x))d(xf(x)) \\ &+\text{ higher order terms} \\ &\equiv \sum_{i=1}^{8} C_{22iN}+\text{ higher order terms.} \end{aligned}$$

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

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

Recalling (A.1), (A.2) and (20), we obtain

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

Therefore, by the Dominated Convergence Theorem and $(m^{-1/2}|A_{\rm x}|)=O_p(m^{-1/2})$, we have $C_{221N}=o_p(N^{-1})$. Next we evaluate C_{222N}. From (A.1), (A.2) and using the arguments of (22), it follows that

$$\begin{aligned} |C_{222N}| &\leq O_p(m^{-1})\int_{S_{N_\epsilon}}[H_N(1-H_N)]^{(3/2)-\delta'}dF(x) \\ &\leq O_p(m^{-1})O[(H_N(\beta_N)(1-H_N(\beta_N)))^{(5/2)-\delta'}]=o_p(N^{-1}), \end{aligned} \qquad (27)$$

hence, we have $C_{222N}=o_p(N^{-1/2})$. Similarly we can prove $C_{223N}=C_{224N}=C_{225N}=o_p(N^{-1/2})$. Next, we consider C_{226N}. In view of (A.1), (A.2) and (23), we observe that

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

Thus, $C_{226N} = o_p(N^{-1})$. Similarly, we can prove $C_{227N} = C_{228N} = o_p(N^{-1})$. 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

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

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 $|(F-G)(x)| < K[H_N(x)(1-H_N(x))]^{1/2-\delta}$ and the arguments of C_{212N}, we can easily show $C_{31N} = o_p(N^{-1})$, and analogously $C_{32N} = o_p(N^{-1})$. Next we turn to evaluate

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

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.2) and (26), we can show $C^*_{33N} = C^{**}_{33N} = o_p(1)$, which, combined with $(m^{-1/2}|A_{\text{x}}|) = O_p(m^{-1/2})$, 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}. In view of (A.1), (A.2), (E.1) and (22), we can show that $C_{35N} = o_p(N^{-1/2})$. 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 (23), we obtain

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

and 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.

Acknowledgement

The authors express their gratitude to the referees for their insightful comments, which improved the original version of the paper.

5 References

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. *Ann. Math. Stat.* 33, 1148-1159.

Berkes, I. and Horváth, L. (2003). Limit results for the empirical process of squared residuals in GARCH models. *Stoch. Proc. Appl.* 105, 271-298.

Brockwell, P. J. and Davis, R. A. (1994). *Time Series: Theory and Practice*, 2nd ed. New York: Springer-Verlag.

Chandra, S. A. and Taniguchi, M. (2005). Minimum α-divergence estimation for ARCH models. To appear in *J. Time Ser. Anal.*

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

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

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

Francq, C. and Zakoian, J-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, No.4, 605-637.

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

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

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

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

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 an ARCH sequence. *Ann. Stat.* 29, 445-469.

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. *Ann. Math. Stat.* 22, 165-179.

Milhøj, A. (1985). The moment structure of ARCH processes. *Scand. J. Stat.* 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. *Ann. Math. Stat.* 23, 617-623.

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

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

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