R-estimators in GARCH models; asymptotics and applications

HANG LIU[†] AND KANCHAN MUKHERJEE[‡]

[†]*International Institute of Finance, School of Management,
University of Science and Technology of China, Hefei, China.*
E-mail: `hliu01@ustc.edu.cn`
[‡]*Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, UK.*
E-mail: `k.mukherjee@lancaster.ac.uk`

Summary   The quasi-maximum likelihood estimation is a commonly-used method for estimating the GARCH parameters. However, such estimators are sensitive to outliers and their asymptotic normality is proved under the finite fourth moment assumption on the underlying error distribution. In this paper, we propose a novel class of estimators of the GARCH parameters based on ranks of the residuals, called R-estimators, with the property that they are asymptotically normal under the existence of a finite $2+\delta$ moment of the errors and are highly efficient. We propose fast algorithm for computing the R-estimators. Both real data analysis and simulations show the superior performance of the proposed estimators under the heavy-tailed and asymmetric distributions.

Keywords: *R-estimation, Empirical process, GARCH models*

1. INTRODUCTION

1.1. Robust estimation based on ranks

Consider observations $\{X_t; 1 \leq t \leq n\}$ from a financial time series with the following representation

$$X_t = \sigma_t \epsilon_t,$$

where $\{\epsilon_t; t \in \mathbb{Z}\}$ are unobservable i.i.d. non-degenerate error r.v.'s with mean zero and unit variance and

$$\sigma_t = (\omega_0 + \sum_{i=1}^{p} \alpha_{0i} X_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j} \sigma_{t-j}^2)^{1/2}, \ t \in \mathbb{Z}, \tag{1.1}$$

with $\omega_0, \alpha_{0i}, \ \beta_{0j} > 0, \forall\, i, j$. In the literature, such models are known as the GARCH (p, q) model and we assume that $\{X_t; t \in \mathbb{Z}\}$ is stationary and ergodic.

Estimation of parameters based on ranks of the residuals was discussed by Koul and Ossiander (1994) for the homoscedastic autoregressive models and Mukherjee (2007) for the heteroscedastic models. Andrews (2012) proposed a class of R-estimators for the GARCH model using a log-transformation of the squared observations and then minimizing a rank-based residual dispersion function. However, such square-transformation, not being one-to-one transformation, may lead to loss of information under an asymmetric error distribution as demonstrated later in this paper. This motivates us to define

R-estimators for the GARCH model that uses the data directly without requiring such transformation.

Robust estimation of the GARCH parameters has been studied extensively in the literature although the attention has been focused exclusively on the class of M-estimators. See, for example, Berkes and Horváth (2004), Mukherjee (2008), Francq et al. (2011), Fan et al. (2014) and Zhu and Ling (2011) and the references in those papers. Some simulation study by Fan et al.(2014, Section 7.2) to compare M-estimators with the R-estimators proposed by Andrews (2012) revealed that these two classes of estimators have almost indistinguishable asymptotic performance while the rank-based estimators are slightly better and this provides another motivation for considering R-estimators. As shown in the motivating example of Section 1.2 and Table 2, our proposed R-estimators are highly efficient for symmetric error distributions while it can be more efficient than those of Andrews (2012) when the error distribution is asymmetric.

Since the M-estimators in GARCH models are obtained by maximizing a pseudo-likelihood of the data directly without requiring any (squared) transformation of the observations, we define R-estimators without transforming the original data. Consequently, similar to the linear regression and autoregressive models, the asymptotic normality of the R-estimators are derived under smoothness conditions on the error probability density function (pdf) while assumptions in Andrews (2012) are related to the pdf of the squared and logged error which are the errors of a transformed model. However, unlike for the R-estimators defined by Andrews (2012), the limiting covariance matrices of our proposed class of R-estimators do not have explicit forms and consequently, we can employ, for example, bootstrap methods to approximate their distributions.

Unlike the commonly-used quasi-maximum likelihood estimator (QMLE) which is asymptotically normal under the finite fourth moment assumption of the error distribution, the proposed class of R-estimators turn out to be asymptotically normal under the assumption of only a finite $2 + \delta$-th moment for some $\delta > 0$. The efficiency property of the R-estimators is studied based on the simulated data from the GARCH $(1, 1)$ model. Analysis of real data shows that the numerical values of R-estimates can be different from the QMLE and the subsequent analysis of the GARCH residuals shows that such difference may be attributed to the infinite fourth moment of the error distribution, which leads to the failure of the QMLE.

One conspicuous issue with previous studies is related to the estimation of an identifiable scale parameter that leads to often more than one stage of estimation. As pointed out by Fan et al. (2014, Section 7.2), such estimation is important for comparing the bias performance. Preminger and Storti (2017) also pointed out the importance of estimating the scale parameter although such estimates may not always be needed in some applications such as estimating the 'Value at Risk'. In Section 2.3 of this paper, we propose a simple consistent estimate of the scale based on R-estimators and the general principle can be applied to the M-estimation as well; see Andrews (2012, Remark 9) and Preminger and Storti (2017, Section 3) for other estimates of the scale.

The main contributions of the paper are as follows. First, a new class of robust and efficient estimators for the GARCH model parameters is proposed. Second, the asymptotic distributions of the proposed estimators are derived based on weak assumption on the error moments. Third, we propose an algorithm for computing the R-estimators which is computational friendly and easy to implement.

To illustrate the advantages of our proposed R-estimator over the commonly-used QMLE and R-estimator of Andrews (2012) (BA, henceforth), we consider below some simulation results corresponding to the GARCH $(1,1)$ model with underlying standardized error density being (i)the standard normal distribution and (ii) the skew normal distribution (see e.g. Azzalini and Dalla Valle (1996) for details of such distribution). We generate $R = 1000$ samples of size $n = 1000$ with parameter values $(\omega, \alpha, \beta)' = (6.50 \times 10^{-6}, 0.177, 0.716)'$ as in Section 3.2. Simulation results described below are similar to various other choices of the true parameters. To make a fair comparison, for both R-estimators we use the van der Waerden score (vdW, henceforth); our vdW score is given in Section 2.4 while BA's one has a different form; see Andrews (2012, Section 3). The resulting boxplots of all estimators are displayed in Figure 1 under the skew normal (upper panel) and standard normal (lower panel) error densities and the MSE ratios of the QMLE over other estimators are reported.

An inspection of these plots reveals superiority of our R-estimator over the QMLE and BA's. Under the normal error distribution, the distribution patterns of the R-estimators are quite similar to the QMLE around the true parameter value, and the MSE ratios of the QMLE over the R-estimators are all close to one. However, under the skew normal errors where asymmetry is present, our proposed R-estimator has the least dispersion and with a gain of around 44%-92% efficiency over the QMLE, while the BA's R-estimator has around 12%-23% gain in efficiency over the QMLE. Therefore, although the BA's R-estimator achieves efficiency similar to ours under the normal distribution, it is not as efficient as ours under an asymmetric distribution.

1.3. Outline of the paper

The rest of the paper is organized as follows. Section 2 defines a class of R-estimators of scaler-transformed parameters based on an asymptotic linearity result of a rank-based central sequence. A consistent estimator of the unknown scalar c_φ defined by (2.4) is provided and the asymptotic distribution and efficiency of the R-estimator are discussed. We describe an algorithm for computing the R-estimators. Section 3 contains empirical and simulation results of the R-estimators. Conclusion is given in Section 4. The technique used to establish the asymptotic distribution is included in the online Appendix A. The Appendix contains some additional numerical results.

## 2. THE CLASS OF R-ESTIMATORS FOR THE GARCH MODEL

In this section, we first define a central sequence based on ranks of the residuals of the GARCH model. We prove the asymptotic uniform linear expansion (2.7) of this central sequence, which enables us to define one-step R-estimator $\hat{\boldsymbol{\theta}}_{n\varphi}$ in (2.9) as a root-n consistent estimator of a scale-transformed GARCH parameter

$$\boldsymbol{\theta}_{0\varphi} = (c_\varphi \omega_0, c_\varphi \alpha_{01}, ..., c_\varphi \alpha_{0p}, \beta_{01}, ..., \beta_{0q})' \tag{2.1}$$

with a constant $c_\varphi > 0$ satisfying (2.5). Based on $\hat{\boldsymbol{\theta}}_{n\varphi}$ and $\{X_t\}$, we are able to derive a consistent estimator of c_φ. We discuss some computational aspects and propose a recursive algorithm for computation in Section 2.5.

NOTATIONS: Throughout the paper, for a function g, we use $\dot{g}$ and $\ddot{g}$ to denote

Figure 1. Boxplots of the QMLE, BA's and our proposed R-estimators (van der Waerden) under the skew normal (upper panel) and standard normal (lower panel) error densities. In each panel, the MSE ratios of the QMLE with respect to other estimators are reported. The horizontal line represents the actual parameter value.
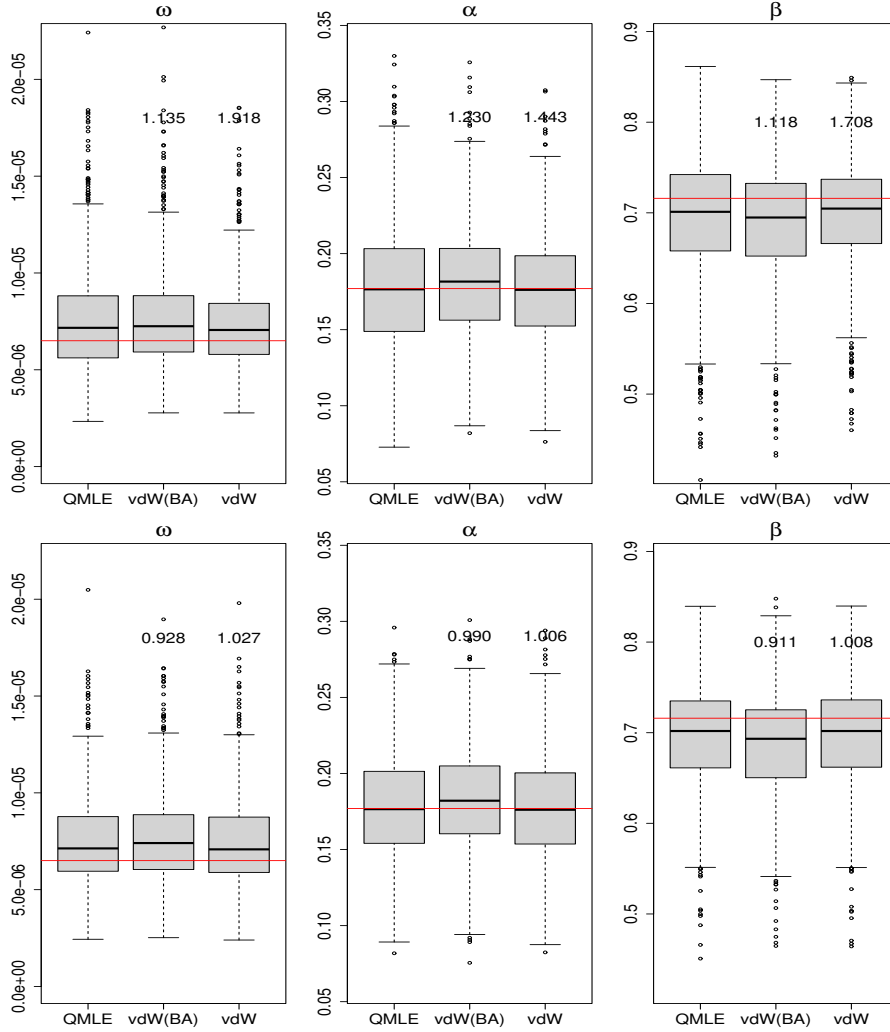
its first and second derivatives whenever they exist. We use c, b, c_1 to denote positive constants whose values can possibly change from line to line. Let ϵ be a generic random variable (r.v.) with the same distribution as $\{\epsilon_t\}$ and let F and f denote the cumulative distribution function (c.d.f.) and probability density function (p.d.f.) of ϵ, respectively. Let $\eta_t := \epsilon_t / \sqrt{c_\varphi}$ and η be a generic r.v. with the same distribution as $\{\eta_t\}$. Let G and g be the c.d.f. and p.d.f. of η, respectively. A sequence of stochastic processes $\{Y_n(\cdot)\}$ is said to be $u_{\mathrm{P}}(1)$ (denoted by $Y_n = u_{\mathrm{P}}(1)$) if for every $c > 0$, $\sup\{|Y_n(\boldsymbol{b})|; \|\boldsymbol{b}\| \le c\} = o_{\mathrm{P}}(1)$, where $\|\cdot\|$ stands for the Euclidean norm.

2.1. Rank-based central sequence

From Lemma 2.3 and Theorem 2.1 of Berkes et al. (2003), σ_t^2 of (1.1) has the unique almost sure representation $\sigma_t^2 = c_0 + \sum_{j=1}^{\infty} c_j X_{t-j}^2$, $t \in \mathbb{Z}$, where $\{c_j; j \geq 0\}$ are defined in (2.7)-(2.9) of Berkes et al. (2003).

Let $\boldsymbol{\theta}_0 = (\omega_0, \alpha_{01}, \ldots, \alpha_{0p}, \beta_{01}, \ldots, \beta_{0q})'$ denote the true parameter belonging to a compact subset $\boldsymbol{\Theta}$ of $(0, \infty)^{1+p} \times (0, 1)^q$. A typical element in $\boldsymbol{\Theta}$ is denoted by

$$\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)'.$$

Define the variance function by

$$v_t(\boldsymbol{\theta}) = c_0(\boldsymbol{\theta}) + \sum_{j=1}^{\infty} c_j(\boldsymbol{\theta}) X_{t-j}^2, \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}, t \in \mathbb{Z},$$

where the coefficients $\{c_j(\boldsymbol{\theta}); j \geq 0\}$ are given in (3.1) of Berkes et al. (2003) with the property $c_j(\boldsymbol{\theta}_0) = c_j$, $j \geq 0$, so that the variance functions satisfy $v_t(\boldsymbol{\theta}_0) = \sigma_t^2$, $t \in \mathbb{Z}$ and

$$X_t = \{v_t(\boldsymbol{\theta}_0)\}^{1/2} \epsilon_t, \quad 1 \leq t \leq n.$$

Let $\{\hat{v}_t(\boldsymbol{\theta})\}$ be observable approximation of $\{v_t(\boldsymbol{\theta})\}$, which is defined by

$$\hat{v}_t(\boldsymbol{\theta}) = c_0(\boldsymbol{\theta}) + I(2 \leq t) \sum_{j=1}^{t-1} c_j(\boldsymbol{\theta}) X_{t-j}^2, \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}, \ 1 \leq t \leq n.$$

Let $H^*(x) = x\{-\dot{f}(x)/f(x)\}$. From Berkes and Horvath (2004, Example 2.4) and Mukherjee (2008, p. 1534), the maximum likelihood estimator (MLE) is a solution of $\boldsymbol{\Delta}_{n,f}(\boldsymbol{\theta}) = \boldsymbol{0}$, where

$$\boldsymbol{\Delta}_{n,f}(\boldsymbol{\theta}) := n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta})}{v_t(\boldsymbol{\theta})} \left\{ 1 - H^* \left[\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta})} \right] \right\}.$$

However, f in H^* is usually unknown and we therefore consider an approximation to $\boldsymbol{\Delta}_{n,f}(\boldsymbol{\theta})$.

Let $\varphi : (0, 1) \to \mathbb{R}$ be a score function satisfying some regularity conditions which will be discussed later. Examples of φ are given in Section 2.4. Let $R_{nt}(\boldsymbol{\theta})$ denote the rank of $X_t/v_t^{1/2}(\boldsymbol{\theta})$ among $\{X_j/v_j^{1/2}(\boldsymbol{\theta}); 1 \leq j \leq n\}$. In linear regression models, the MLE has the same asymptotic efficiency as an R-estimator based on the score function $\varphi(u) = -\dot{f}(F^{-1}(u))/f(F^{-1}(u))$. For the estimation of the scale parameters, a correspondence to the MLE is through the *central sequence*

$$\boldsymbol{R}_n(\boldsymbol{\theta}) := \boldsymbol{R}_{n,\varphi}(\boldsymbol{\theta}) = n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta})}{v_t(\boldsymbol{\theta})} \left\{ 1 - \varphi \left[\frac{R_{nt}(\boldsymbol{\theta})}{n+1} \right] \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta})} \right\}. \tag{2.2}$$

However, since $v_t(\boldsymbol{\theta})$ is unobservable, we replace it by $\hat{v}_t(\boldsymbol{\theta})$. Let $\hat{R}_{nt}(\boldsymbol{\theta})$ denote the rank of $X_t/\hat{v}_t^{1/2}(\boldsymbol{\theta})$ among $\{X_j/\hat{v}_j^{1/2}(\boldsymbol{\theta}); 1 \leq j \leq n\}$. We define *rank-based central sequence* as

$$\hat{\boldsymbol{R}}_n(\boldsymbol{\theta}) := \hat{\boldsymbol{R}}_{n,\varphi}(\boldsymbol{\theta}) = n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta})}{\hat{v}_t(\boldsymbol{\theta})} \left\{ 1 - \varphi \left[\frac{\hat{R}_{nt}(\boldsymbol{\theta})}{n+1} \right] \frac{X_t}{\hat{v}_t^{1/2}(\boldsymbol{\theta})} \right\}. \tag{2.3}$$

2.2. One-step R-estimators and their asymptotic distributions

To define the R-estimator in terms of the classical Le Cam's one-step approach as in Hallin and La Vecchia (2017) and Hallin et al. (2020), we derive the asymptotic linearity of the rank-based central sequence under the following assumptions. Let $c_\varphi > 0$ be defined by

$$\sqrt{c_\varphi} = \mathrm{E}\left[\varphi\left(F\left(\epsilon_t\right)\right)\epsilon_t\right] \tag{2.4}$$

so that

$$\mathrm{E}\left\{\varphi\left[F\left(\epsilon_t\right)\right]\frac{\epsilon_t}{\sqrt{c_\varphi}}\right\} = 1. \tag{2.5}$$

Define $\mu(x) := \int_{-\infty}^x sg(s)ds$. Since $g(x) > 0$, $\mu(x)$ is strictly decreasing on $(-\infty, 0]$ with range $[\mu(0), 0]$ and strictly increasing on $[0, +\infty)$ with range $[\mu(0), 0]$. The functions $y \to \mu^{-1}(y)$ on $[\mu(0), 0]$ with ranges $(-\infty, 0]$ and $[0, +\infty)$ are well-defined when the ranges are considered separately.

The following conditions on the distribution of η_t are assumed for the proof of Theorem A.1 in Appendix on the approximation of a scale-perturbed weighted mixed-empirical process by its non-perturbed version.

ASSUMPTION 2.1. *(i). The function $x^2 g(x)$ is bounded on $x \in \mathbb{R}$ (and consequently so are the functions $g(x)$ and $xg(x)$); functions $y \to \mu^{-1}(y)g(\mu^{-1}(y))$ and $y \to (\mu^{-1}(y))^2 g(\mu^{-1}(y))$ are uniformly continuous on $[\mu(0), 0]$ when the ranges are considered separately as in the definition of μ above;(ii).*

$$\lim_{\delta \to 0} \sup\left\{|x| \int_0^1 |xg(x) - (x + hx\delta)g(x + hx\delta)|dh; x \in \mathbb{R}\right\} = 0;$$

(iii). There is a $\delta > 0$ such that $\mathrm{E}|\eta_t|^{2+\delta} < \infty$.

We remark that Assumption (i) entails that $\mu(x)$ is uniformly Lipschitz continuous in scale in the sense that for some constant $0 < c < \infty$ and for every $s \in \mathbb{R}$, we have $\sup_{x \in \mathbb{R}} |\mu(x + xs) - \mu(x)| \le c|s|$.

A more easily verifiable condition for Assumption (ii) can be obtained, for example, when g admits the derivative $\dot{g}$ which satisfies that for some $\delta > 0$,

$$\sup\{x^2 \sup |g(y) + y\dot{g}(y)|; x(1 - \delta) < y < x(1 + \delta)\} < \infty.$$

In particular, Assumptions (i), (ii) and (iii) hold for a wide range of error distributions, including normal, double-exponential, logistic and t-distributions with degrees of freedom more than 2 which are considered for simulation study.

We also need the following assumptions on the parameter space and the score function φ.

ASSUMPTION 2.2. *Denoting by $\boldsymbol{\Theta}_0$ the set of interior points of $\boldsymbol{\Theta}$, we assume that $\boldsymbol{\theta}_0, \boldsymbol{\theta}_{0\varphi} \in \boldsymbol{\Theta}_0$.*

ASSUMPTION 2.3. *The score function φ is non-decreasing, right-continuous with only a finite number of points of discontinuity and is bounded on $(0, 1)$.*

We now compare our assumptions with those made by Andrews (2012), to be called

as BA1, BA2 etc. Assumption BA1 states the stationarity and ergodicity of $\{X_t\}$ similar to what we assume but not necessarily finiteness of the variance of X_t. However, estimation of the intercept parameter α_{00} appearing in Andrews (2012, Equation (2.2)) is not considered there. In this paper, we estimate the intercept parameter under the finite variance assumption of X_t as in Assumption A4 below. Higher moment assumptions were also made in Fan et al. (2014) while estimating the equivalent scale parameter η_f.

Assumptions BA2 and BA3 are related to the uniqueness of parameters and the existence of the non-degenerate errors and observations. We assume non-degenerate models as is common in the literature. Assumptions BA4 and BA7 are on the score functions which are bounded, non-decreasing and left-continuous. In Assumption 2.3, we assume that the score functions are bounded, non-decreasing and right-continuous with finite number of discontinuity. Assumptions BA5 and BA6 are on the cdf and pdf of the log of the squared error distribution. We have made analogous assumptions on the error pdf itself and uniform continuity on the μ^{-1}-transformed axis in Assumption 2.1(i) and (ii); the later came up as a part of the technical assumptions in using some convergence results of empirical processes to derive the asymptotic distribution of the R-estimators. Assumptions BA8, BA9 and BA10 describe various scenarios related to some of the component parameters equal to zero as in Francq and Zakoian (2007). In Assumption 2.2, we assume that all parameters are positive as in Berkes and Horváth (2004), Mukherjee (2008) and Francq et al. (2011) in relation to the M-score and this corresponds to BA8.

To state the asymptotic linearity of $\hat{\boldsymbol{R}}_n(\boldsymbol{\theta})$, we introduce the following notation. Let

$$\gamma(\varphi) := \int_0^1 \int_0^1 G^{-1}(u)G^{-1}(v)\left[\min\{u,v\} - uv\right]d\varphi(u)d\varphi(v), \ \boldsymbol{J}(\boldsymbol{\theta}) := \mathrm{E}(\dot{\boldsymbol{v}}_t(\boldsymbol{\theta})\dot{\boldsymbol{v}}_t^{'}(\boldsymbol{\theta})/v_t^2(\boldsymbol{\theta}))$$

$$\rho(\varphi) := \int_0^1 \{G^{-1}(u)\}^2 g\{G^{-1}(u)\}d\varphi(u), \ \sigma^2(\varphi) := \mathrm{E}\left\{\varphi\left[G(\eta_t)\right]\eta_t\right\}^2 - 1,$$

$$\lambda(\varphi) := \int_0^1 \int_0^1 G^{-1}(u)I(v \leq u)(1 - G^{-1}(v)\varphi(v))dvd\varphi(u). \tag{2.6}$$

Let Z be the r.v. $Z := \int_0^1 G^{-1}(u)B(u)d\varphi(u)$, where $B(.)$ is the standard Brownian bridge. Then Z has mean zero and variance $\gamma(\varphi)$; see the proof in Lemma A.5 for details. Let $\tilde{G}_n(x)$, $x \in \mathbb{R}$ be the empirical distribution function of $\{\eta_t\}$ (which is unobservable),

$$\boldsymbol{Q}_n(\boldsymbol{\theta}) := \int_0^1 n^{-1/2}\sum_{t=1}^n \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta})}{v_t(\boldsymbol{\theta})}\left[\mu(G^{-1}(u)) - \mu(\tilde{G}_n^{-1}(u))\right]d\varphi(u),$$

$$\boldsymbol{N}_n(\boldsymbol{\theta}) := n^{-1/2}\sum_{t=1}^n \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta})}{v_t(\boldsymbol{\theta})}\left\{1 - \eta_t\varphi\left[G(\eta_t)\right]\right\}.$$

The following proposition states the asymptotic uniform linearity of $\hat{\boldsymbol{R}}_n(\boldsymbol{\theta})$.

PROPOSITION 2.1. *Let Assumptions 2.1-2.3 hold. Then for $\boldsymbol{b} \in \mathbb{R}^{1+p+q}$ with $||\boldsymbol{b}|| < c$,*

$$\hat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{0\varphi} + n^{-1/2}\boldsymbol{b}) - \hat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{0\varphi}) = (1/2 + \rho(\varphi)/2)\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b} + u_{\mathrm{P}}(1). \tag{2.7}$$

Moreover,

$$\hat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{0\varphi}) = \boldsymbol{Q}_n(\boldsymbol{\theta}_{0\varphi}) + \boldsymbol{N}_n(\boldsymbol{\theta}_{0\varphi}) + u_{\mathrm{P}}(1), \tag{2.8}$$

where $\boldsymbol{Q}_n(\boldsymbol{\theta}_{0\varphi})$ converges in distribution to $\mathrm{E}(\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})/v_1(\boldsymbol{\theta}_{0\varphi}))Z$ with mean zero and covariance matrix $\mathrm{E}(\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})/v_1(\boldsymbol{\theta}_{0\varphi}))\mathrm{E}(\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_{0\varphi})/v_1(\boldsymbol{\theta}_{0\varphi}))\gamma(\varphi)$ and $\boldsymbol{N}_n(\boldsymbol{\theta}_{0\varphi}) \to \mathcal{N}(\boldsymbol{0}, \boldsymbol{J}(\boldsymbol{\theta}_{0\varphi})\sigma^2(\varphi))$.

The above asymptotic linearity allows us to define a class of R-estimators through the one-step approach. Let $\{\hat{\boldsymbol{\Upsilon}}_n\}$ be a sequence of consistent estimators of $\boldsymbol{\Upsilon}_{\varphi,g}(\boldsymbol{\theta}_{0\varphi}) := (1/2 + \rho(\varphi)/2)\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi})$; see Section 2.5 for a construction of $\hat{\boldsymbol{\Upsilon}}_n$. Let $\bar{\boldsymbol{\theta}}_n$ be a root-n consistent estimator of $\boldsymbol{\theta}_{0\varphi}$ and, for technical reasons, we assume $\bar{\boldsymbol{\theta}}_n$ is asymptotically discrete. More precisely, a sequence $\{\bar{\boldsymbol{\theta}}_n\}$ is called discrete if there exists $K \in \mathbb{N}$ such that independent of $n \in \mathbb{N}$, $\bar{\boldsymbol{\theta}}_n$ takes on at most K different values in

$$\mathcal{Q}_n := \{\boldsymbol{\theta} \in \mathbb{R}^{1+p+q} : n^{-1/2}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq c\}, \quad c > 0 \text{ fixed};$$

see Kresis (1987, Section 4) for details. We remark that here asymptotically discreteness is only of theoretical interest since in practice $\bar{\boldsymbol{\theta}}_n$ always has a bounded number of digits; see Le Cam and Yang (2000, Chapter 6) and van der Vaart (1998, Section 5.7) for more details. Then the one-step R-estimator is defined as

$$\hat{\boldsymbol{\theta}}_{n\varphi} := \bar{\boldsymbol{\theta}}_n - n^{-1/2}\left(\hat{\boldsymbol{\Upsilon}}_n\right)^{-1}\hat{\boldsymbol{R}}_n(\bar{\boldsymbol{\theta}}_n). \tag{2.9}$$

Note that strictly speaking, the R-estimators based on this definition are not functions of the ranks of the residuals only. However, we borrow the terminology from the regression and the homoscedastic-autoregression settings and still call them (generalized) R-estimators. When, for example, $\varphi(u) = u - 1/2$, $\hat{\boldsymbol{\theta}}_{n\varphi}$ is an analogue of the Wilcoxon type R-estimator.

The following theorem shows that the R-estimator defined in (2.3) is $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}_{0\varphi}$. The proof is given in Appendix A.

THEOREM 2.1. *Let Assumptions 2.1-2.3 hold. Then, as $n \to \infty$,*

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{n\varphi} - \boldsymbol{\theta}_{0\varphi}\right) = -(1/2 + \rho(\varphi)/2)^{-1}(\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi}))^{-1}(\boldsymbol{Q}_n(\boldsymbol{\theta}_{0\varphi}) + \boldsymbol{N}_n(\boldsymbol{\theta}_{0\varphi})) + o_{\mathrm{P}}(1). \tag{2.10}$$

Hence as $n \to \infty$, $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{n\varphi} - \boldsymbol{\theta}_{0\varphi}\right)$ is normal with mean $\boldsymbol{0}$ and covariance matrix

$$(\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi}))^{-1}\frac{\left[4\gamma(\varphi) + 8\lambda(\varphi)\right]\mathrm{E}(\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})/v_1(\boldsymbol{\theta}_{0\varphi}))\mathrm{E}\left(\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_{0\varphi})/v_1(\boldsymbol{\theta}_{0\varphi})\right) + 4\sigma^2(\varphi)\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi})}{(1 + \rho(\varphi))^2}(\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi}))^{-1}.$$

2.3. Estimation of c_φ

To obtain a consistent estimator of $\boldsymbol{\theta}_0$, we estimate the unknown scalar c_φ under the following mild moment assumption on X_t.

ASSUMPTION 2.4. $\mathrm{E}(X_t^2) < \infty$

From Bollerslev (1986, Theorem 1), a sufficient condition for the GARCH model to be second-order stationary (hence to satisfy Assumption 2.4) is $\sum_{i=1}^p \alpha_{0i} + \sum_{j=1}^q \beta_{0j} < 1$. In this case,

$$\mathrm{E}(X_t^2) = \frac{\omega_0}{1 - \sum_{i=1}^p \alpha_{0i} - \sum_{j=1}^q \beta_{0j}} = \frac{c_\varphi \omega_0}{c_\varphi - \sum_{i=1}^p (c_\varphi \alpha_{0i}) - c_\varphi \sum_{j=1}^q \beta_{0j}}.$$

Using the ergodicity property, $\mathrm{E}(X_t^2)$ is estimated from the data by $\overline{X_n^2} := n^{-1}\sum_{t=1}^n X_t^2$.

Also, $c_\varphi \omega_0$ and $c_\varphi \alpha_{0i}$ are estimated by $\hat{\omega}_{n\varphi}$ and $\hat{\alpha}_{n\varphi i}$, respectively. Solving the following equation for c

$$\sum_{t=1}^{n} X_t^2 / n = \frac{\hat{\omega}_{n\varphi}}{c - \sum_{i=1}^{p} \hat{\alpha}_{n\varphi i} - c \sum_{j=1}^{q} \hat{\beta}_{nj}},$$

we obtain an estimate of c_φ as

$$\hat{c}_{n\varphi} := \left(1 - \sum_{j=1}^{q} \hat{\beta}_{nj}\right)^{-1} \left(\hat{\omega}_{n\varphi} / \overline{X_n^2} + \sum_{i=1}^{p} \hat{\alpha}_{n\varphi i}\right). \tag{2.11}$$

Consequently write $\hat{\boldsymbol{\theta}}_{n\varphi}$ in its component-wise form

$$\hat{\boldsymbol{\theta}}_{n\varphi} = \left(\hat{\omega}_{n\varphi}, \hat{\alpha}_{n\varphi 1}, ..., \hat{\alpha}_{n\varphi p}, \hat{\beta}_{n1}, ..., \hat{\beta}_{nq}\right)'$$

and let

$$\hat{\boldsymbol{\theta}}_n := \left(\hat{\omega}_{n\varphi} / \hat{c}_{n\varphi}, \hat{\alpha}_{n\varphi 1} / \hat{c}_{n\varphi}, ..., \hat{\alpha}_{n\varphi p} / \hat{c}_{n\varphi}, \hat{\beta}_{n1}, ..., \hat{\beta}_{nq}\right)'.$$

The following theorem states that $\hat{c}_{n\varphi}$ is a consistent estimator of c_φ and thus $\hat{\boldsymbol{\theta}}_n$ is a root-n consistent estimator of $\boldsymbol{\theta}_0$ with asymptotically normal distribution; see Appendix A for its proof.

THEOREM 2.2. *Let Assumptions 2.1-2.4 hold. Then, as $n \to \infty$,*

$$\hat{c}_{n\varphi} = c_\varphi + o_{\mathrm{P}}(1)$$

and

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) = -(1/2 + \rho(\varphi)/2)^{-1} (\boldsymbol{J}(\boldsymbol{\theta}_0))^{-1} (\boldsymbol{Q}_n(\boldsymbol{\theta}_0) + \boldsymbol{N}_n(\boldsymbol{\theta}_0)) + o_{\mathrm{P}}(1). \tag{2.12}$$

Hence as $n \to \infty$, $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right)$ is normal with mean $\mathbf{0}$ and covariance matrix

$$\boldsymbol{\Omega}_\varphi := (\boldsymbol{J}(\boldsymbol{\theta}_0))^{-1} \frac{[4\gamma(\varphi) + 8\lambda(\varphi)] \, \mathrm{E}(\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_0)/v_1(\boldsymbol{\theta}_0)) \mathrm{E}\left(\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_0)/v_1(\boldsymbol{\theta}_0)\right) + 4\sigma^2(\varphi)\boldsymbol{J}(\boldsymbol{\theta}_0)}{(1 + \rho(\varphi))^2} (\boldsymbol{J}(\boldsymbol{\theta}_0))^{-1}.$$

Note that by making Assumption 2.4 and following similar approach as in the proof of Theorem 2.2, one can also obtain consistent estimators of unknown scalars for robust estimators in GARCH models (e.g., M-estimators of Liu and Mukherjee (2020) and R-estimator of Andrews (2012)). For illustration, recall that the R-estimator of Andrews (2012) is root-n consistent estimator of

$$(\alpha_{01}/\omega_0, ..., \alpha_{0p}/\omega_0, \beta_{01}, ..., \beta_{0q})'$$

where ω_0 is the intercept parameter that plays the role of the unknown scalar. Denoting the R-estimator of Andrews (2012) by

$$\hat{\boldsymbol{\theta}}_{\mathrm{BA};n} := (\hat{a}_{\mathrm{BA};n1}, ..., \hat{a}_{\mathrm{BA};np}, \hat{\beta}_{\mathrm{BA};n1}, ..., \hat{\beta}_{\mathrm{BA};nq})',$$

a consistent estimator of ω_0 can be obtained as

$$\hat{\omega}_{\mathrm{BA};n} := \left(1 + \sum_{i=1}^{p} \hat{a}_{\mathrm{BA};ni} \overline{X_n^2}\right)^{-1} \left(1 - \sum_{j=1}^{q} \hat{\beta}_{\mathrm{BA};nj}\right) \overline{X_n^2}.$$

We remark that from Theorem 2.2, the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_n$ has a complicated form. Hence we can consider, for example, bootstrap methods to approximate the limit distribution of $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right)$.

2.4. Examples of the score functions

Below we cite examples of three commonly-used R-scores; for similar examples of scores in other models, see Mukherjee (2007) and Hallin and La Vecchia (2017).

Example 1 (sign score). Let $\varphi(u) = \text{sign}(u - 1/2)$. Then for symmetric error distribution, $c_\varphi = (\mathrm{E}|\epsilon|)^2$, which coincides with the scale factor of the LAD estimator in Mukherjee (2008). Therefore, the sign R-estimator is expected to be close to the LAD estimator. This is demonstrated later in the real data analysis.

Example 2 (Wilcoxon score). Let $\varphi(u) = u - 1/2$ so that the range of $\varphi(u)$ is symmetric.

Example 3 (van der Waerden (vdW) or normal score). One might also set $\varphi(u) = \Phi^{-1}(u)$, with $\Phi(\cdot)$ denoting the c.d.f. of the standard normal distribution. Notice that unlike the sign and Wilcoxon score, the vdW score is not bounded as $u \to 0$ and $u \to 1$. It thus does not satisfy Assumption 2.3. However, an approximating sequence of bounded score functions of φ on $(0, 1)$ can be constructed as in Andrews (2012). For example, letting

$$\varphi_m(u) = \Phi^{-1}(u)I(1/m \le u \le 1 - 1/m) + \Phi^{-1}(1/m)I(0 < u < 1/m)$$
$$+ \Phi^{-1}(1 - 1/m)I(u > 1 - 1/m), \ m > 2,$$

$\varphi_m(u)$ satisfies Assumption 2.3 and converges pointwise to the vdW score on $(0, 1)$. It is demonstrated later using both real data analysis and extensive simulation that the vdW has superior performance compared with the QMLE.

We now provide heuristics for the definition of the R-estimator in (2.2). When the underlying error distribution is known, one can obtain efficient R-estimator by choosing the score function as $\varphi(u) = -\dot{f}(F^{-1}(u))/f(F^{-1}(u))$. Since for large n, the empirical distribution function $R_{nt}(\boldsymbol{\theta}_{0\varphi})/(n+1)$ of $\{\epsilon_j; 1 \le j \le n\}$ evaluated at ϵ_t is close to $F(\epsilon_t)$, we have

$$\varphi\left[\frac{R_{nt}(\boldsymbol{\theta})}{n+1}\right]\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta})} \approx H^*\left[\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta})}\right].$$

Therefore, the criteria function of the R-estimator gets close to the MLE which is efficient. This leads to the choice of the vdW, sign and Wilcoxon under the normal, double exponential (DE) and logistic distributions, respectively. This is observed later in simulation study of the R-estimator.

2.5. Computational aspects

Here we discuss some key computational aspects and propose an algorithm to compute $\hat{\boldsymbol{\theta}}_{n\varphi}$ and $\hat{\boldsymbol{\theta}}_n$. Codes are available from the authors' `GitHub` page `https://github.com/HangLiu10/RestGARCH`.

First, since c_φ depends on the unknown density f, it is difficult to have a $\sqrt{n}$-consistent initial estimator $\bar{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}_{0\varphi}$. One possible choice of the initial consistent estimator $\bar{\boldsymbol{\theta}}_n$ is the two-stage type least squares estimator as proposed by Preminger and Storti (2017). In

practice, due to the finite sample size, the one-step procedure in (2.9) is usually iterated a number of times, taking $\hat{\boldsymbol{\theta}}_{n\varphi}$ as the new initial estimate, until it stabilizes numerically. This iteration process would mitigate the impact of different initial estimates; see van der Vaart (1998, Section 5.7) and Hallin and La Vecchia (2017) for similar comments. In fact, we observed during our extensive simulation study that irrespective of the choice of the QMLE, LAD or $\boldsymbol{\theta}_0$ as initial estimates, only few iterations result in the same estimates.

Second, to compute $\hat{\boldsymbol{\theta}}_{n\varphi}$ of (2.9), we need $\hat{\boldsymbol{\Upsilon}}_n$ which is a consistent estimator of $(1/2 + \rho(\varphi)/2)\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi})$. The matrix $\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi})$ can be consistently estimated by

$$\hat{\boldsymbol{J}}_n(\bar{\boldsymbol{\theta}}_n) := n^{-1}\sum_{t=1}^{n}\{\dot{\boldsymbol{v}}_t(\bar{\boldsymbol{\theta}}_n)\dot{\boldsymbol{v}}_t'(\bar{\boldsymbol{\theta}}_n)/\hat{v}_t^2(\bar{\boldsymbol{\theta}}_n)\}.$$

For estimating $\rho(\varphi)$ which is a function of the density g, we can use the asymptotic linearity in (2.7). Here with an arbitrarily chosen $\boldsymbol{b}$, we can substitute $\bar{\boldsymbol{\theta}}_n$ for $\boldsymbol{\theta}_{0\varphi}$ and then solve the equation for $\rho(\varphi)$ based on (2.7). A more delicate approach for estimating $\rho(\varphi)$ can be found in Cassart et al. (2010) and Hallin and La Vecchia (2017, Appendix C). Based on our extensive simulation study and real data analysis, it appears that different values of $\rho(\varphi)$ would finally lead to same estimate after some iterations. Consequently, we set $\rho(\varphi) = 1$ during the computation which is the value corresponding to the vdW score under the normal distribution. In summary, we propose the following iterative Algorithm 1 to compute $\hat{\boldsymbol{\theta}}_{n\varphi}$, with which we can obtain $\hat{c}_\varphi$ using (2.11) and hence $\hat{\boldsymbol{\theta}}_n$.

Algorithm 1: R-estimation for GARCH models

Input: a sample $\{X_t; 1 \leq t \leq n\}$, orders p and q of the GARCH process, number k of iterations in the one-step procedure.

Output: R-estimator $\hat{\boldsymbol{\theta}}_n$

STEP 1. Compute a preliminary root-n consistent estimator $\bar{\boldsymbol{\theta}}_n$ and set $\hat{\boldsymbol{\theta}}_{n\varphi} = \bar{\boldsymbol{\theta}}_n$.

STEP 2. **for** $i \leftarrow 1$ **to** k **do**

$$\hat{\boldsymbol{\theta}}_{n\varphi} \leftarrow \hat{\boldsymbol{\theta}}_{n\varphi} - \left[\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\hat{\boldsymbol{\theta}}_{n\varphi})\dot{\boldsymbol{v}}_t'(\hat{\boldsymbol{\theta}}_{n\varphi})}{\hat{v}_t^2(\tilde{\boldsymbol{\theta}}_{n\varphi})}\right]^{-1}$$
$$\times\left\{\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\hat{\boldsymbol{\theta}}_{n\varphi})}{\hat{v}_t(\tilde{\boldsymbol{\theta}}_{n\varphi})}\left[1 - \varphi\left(\frac{R_{nt}(\hat{\boldsymbol{\theta}}_{n\varphi})}{n+1}\right)\frac{X_t}{\hat{v}_t^{1/2}(\hat{\boldsymbol{\theta}}_{n\varphi})}\right]\right\} \qquad (2.13)$$

end

STEP 3. Compute $\hat{c}_\varphi$ using (2.11) and then $\hat{\boldsymbol{\theta}}_n$.

---

3. REAL DATA ANALYSIS AND SIMULATION RESULTS

This section examines the performance of the R-estimators and compare them with the QMLE by analyzing three financial time series and by carrying out extensive Monte Carlo simulation.

Table 1. The QMLE, LAD and R-estimates of the GARCH (1, 1) parameter for the EFCX, S&P 500 and GBP/USD data.

	fGarch	QMLE	LAD	sign	Wilcoxon	vdW
EFCX						
ω $(\times 10^3)$	0.19	0.63	1.22	1.23	1.22	1.24
α	0.05	0.07	0.17	0.17	0.16	0.13
β	0.92	0.84	0.66	0.65	0.67	0.69
S&P 500						
ω $(\times 10^5)$	0.65	0.70	0.53	0.53	0.53	0.62
α	0.18	0.18	0.19	0.19	0.19	0.18
β	0.72	0.70	0.73	0.73	0.73	0.72
GBP/USD						
ω $(\times 10^6)$	0.53	1.02	0.68	0.68	0.73	0.91
α	0.12	0.13	0.07	0.07	0.07	0.09
β	0.88	0.85	0.91	0.91	0.91	0.89

Note: The sign, Wilcoxon and vdW scores are used for the R-estimators. Estimates of ω are rescaled.

3.1. Real data analysis

In this section we fit GARCH $(1,1)$ model to three financial time series and compare the proposed three R-estimators with the M-estimators QMLE and LAD discussed in Mukherjee (2008), where the unknown scalar of the LAD can also be estimated by (2.11).

In an earlier work, Muller and Yohai (2008) fitted the the GARCH $(1,1)$ model to the Electric Fuel Corporation (EFCX) time series for the period of January 2000 to December 2001 with sample size $n = 498$. The parameters of the model are estimated by M-estimators based on various score functions. It turned out that the M-estimates of the parameter β differ widely depending on the score functions and so it is difficult to assess which estimate should be relied on in similar situations. Here we compare various M-estimates and R-estimates of the GARCH $(1,1)$ parameters for the EFCX series again shedding light on which could be some possible reasons for the difference in estimates and finally which estimation methods can be relied upon. We also compare M-estimates of the GARCH $(1,1)$ parameters when fitted to two other dataset, namely, the S&P 500 stock index from June 2013 to May 2017 with $n = 1005$ and the GBP/USD exchange rate from June 2013 to May 2017 with $n = 998$ to illustrate that the M- and R-estimates do not differ widely when the underlying theoretical assumptions hold in general.

In Table 1, we report the QMLE computed using the `fGarch` package in R program, the M-estimates QMLE and LAD and the R-estimates proposed in Examples 1-3 of Section 2.4. For the EFCX data, the R-estimates for all score functions are quite close to the LAD estimate, but they are very different than the QMLE. On the contrary, for the S&P 500 and GBP/USD data, M- and R-estimates are close to each other.

To investigate why the QMLE is different from the other R-estimates and LAD for the EFCX data, we check the assumption $E\epsilon^4 < \infty$ for this data by using the QQ-plots of the residuals (based on the vdW R-estimates) against t distributions. We consider the vdW score only because the R-estimates based on two other score functions and the LAD are close to the vdW estimates. For comparison, we have also provided QQ-plots for

the S&P 500 data. The main idea behind the QQ-plots of the residuals against the $t(d)$ distribution is simple: Suppose $\epsilon \sim t(d)$ distribution; then $E|\epsilon|^{\nu} < \infty$ if and only if $\nu < d$. Hence, residuals with heavier tail than the $t(d)$ distribution correspond to the errors with the infinite d-th moment while those with thinner tail than the $t(d)$ distribution have the finite d-th error moment.

The top-left panel of Figure 1 (see Appendix C) shows the QQ-plot of the residuals against the $t(4.01)$ distribution for the EFCX data. The residuals have heavier right tail than the $t(4.01)$ distribution which implies that the fourth moment of the error term may not exist. On the other hand, the QQ-plot against the $t(3.01)$ distribution reveals lighter tail as shown at the bottom-left panel of Figure 1 and this implies that $E|\epsilon|^3 < \infty$.

For the S&P 500 data, the QQ-plot against $t(4.01)$ distribution at the top-right panel of Figure 1 shows that the residuals have lighter tails than $t(4.01)$ distribution. For the QQ-plot against $t(6.01)$ distribution, as shown at the bottom-right panel of Figure 1, the residuals fit the distribution better. Therefore, we may conclude that $E|\epsilon|^4 < \infty$ holds for the S&P 500 data.

3.2. Simulation study of the R-estimators

We now evaluate the performance of the R-estimators based on simulated data from various error distributions in the GARCH $(1,1)$ model. Let R denote the number of replications and $\hat{\boldsymbol{\theta}}_{nr} = (\hat{\omega}_r, \hat{\alpha}_{r1}, ..., \hat{\alpha}_{rp}, \hat{\beta}_{r1}, ..., \hat{\beta}_{rq})'$ denote the R-estimator computed from the r-th data, $1 \leq r \leq R$. We throughout compare the R-estimators with the QMLE by using the bias and MSE based on averages over $1 \leq r \leq R$. We also compare the relative efficiency of the R-estimators wrt the QMLE under a finite sample size, as an estimate of the ARE, by using the formula

$$\widehat{\mathrm{ARE}}_{\mathrm{R/QMLE}} = \widehat{\mathrm{MSE}}_{\mathrm{QMLE}}/\widehat{\mathrm{MSE}}_{\mathrm{R}}.$$

Simulation for the GARCH $(1,1)$ model. Here we run simulation with $R = 500, n = 1000$ and $\boldsymbol{\theta}_0 = (6.50 \times 10^{-6}, 0.177, 0.716)'$, where our choice of $\boldsymbol{\theta}_0$ is motivated by the estimate given by the `fGarch` package for the S&P 500 data in Table 1. The estimates of the bias and MSE of the R-estimators and QMLE under various symmetric error distributions are reported in Table 2, where the estimates of the ARE with respect to the QMLE are shown in the parentheses. When a data is generated under the $t(3)$ error distribution, the fourth moment of the error is infinite and consequently the algorithm for computing the QMLE does not converge for many replications, while Algorithm 1 for computing the R-estimators always converges. Therefore, the bias and MSE are obtained using the data of those replication cases when the QMLE converges.

To explain our simulation results in Table 2 corresponding to the vdW score, recall the classical Chernoff and Savage (1958) phenomenon which appears in linear regression models as discussed in Jurečková and Sen (1996, Section 3.4) and linear autoregressive models as discussed in Mukherjee and Bai (2002). Accordingly, the ARE of the R-estimator (based on the unbounded vdW score $\varphi(u) = \Phi^{-1}(u)$) with respect to the least squares estimator is 1 at the standard normal error and more than 1 for all other symmetric error distributions. In this paper we consider a bounded approximating score of the vdW score in the nonlinear GARCH model where the QMLE is analogous to the least squares estimator. Simulation results in Table 2 reflects that phenomenon closely. In particular, the vdW achieves almost the same efficiency as the QMLE under the nor-

mal distribution and it is more efficient under symmetric heavier-tailed distributions. In general, the sign score is most efficient under the DE and $t(3)$ distributions, while the Wilcoxon score is optimal under the logistic distribution. Under the $t(3)$ distribution with infinite fourth moment, the R-estimators yield smaller bias and significantly smaller MSE than the QMLE.

For a demonstration of our asymptotic result with increasing sample size, we have reported simulation results for larger sample sizes $n = 3000$ and $n = 5000$ under $t(3)$ error distribution in Table C.1 of the Appendix. The QMLE failed to converge even for large sample size as the fourth moment was not finite; for example, with $n = 5000$, the QMLE did not converge for around 8% of replications. From Table C.1, when n increases, the performance of the R-estimators becomes even better in terms of both the ARE and MSE.

Table 2. The estimates of the bias, MSE and ARE (in parenthesis) of the R-estimators and the QMLE for the GARCH $(1, 1)$ model under symmetric error distributions.

	Bias			**MSE and ARE**		
	ω ($\times 10^7$)	α ($\times 10^3$)	β ($\times 10^2$)	ω ($\times 10^{12}$)	α ($\times 10^3$)	β ($\times 10^3$)
Normal						
QMLE	9.72	1.68	-1.92	6.05	1.21	3.8
Sign	9.08	1.07	-1.86	6.93 (0.87)	1.36 (0.89)	4.43 (0.86)
Wilcoxon	8.99	0.87	-1.83	6.42 (0.94)	1.28 (0.94)	4.12 (0.92)
vdW	9.19	0.96	-1.87	5.89 (1.03)	1.20 (1.01)	3.77 (1.01)
DE						
QMLE	9.97	3.02	-2.17	11.6	2.75	7.93
Sign	6.67	-1.76	-1.55	6.87 (1.70)	2.08 (1.33)	5.41 (1.47)
Wilcoxon	6.74	-1.72	-1.57	7.00 (1.66)	2.12 (1.30)	5.48 (1.45)
vdW	7.79	-0.76	-1.87	9.15 (1.27)	2.24 (1.23)	6.50 (1.22)
Logistic						
QMLE	11.34	5.69	-1.75	8.45	1.95	5.02
Sign	9.50	3.25	-1.46	6.81 (1.24)	1.68 (1.16)	4.30 (1.17)
Wilcoxon	9.40	3.15	-1.43	6.63 (1.28)	1.65 (1.18)	4.16 (1.21)
vdW	10.05	3.80	-1.61	7.23 (1.17)	1.71 (1.14)	4.42 (1.14)
$t(3)$						
QMLE	11.73	33.03	-5.14	42.24	81.55	39.51
Sign	0.88	-11.90	-2.04	8.20 (5.15)	3.53 (23.08)	8.86 (4.46)
Wilcoxon	1.16	-11.62	-2.15	8.55 (4.94)	3.61 (22.59)	9.33 (4.24)
vdW	3.10	-8.52	-2.91	11.99 (3.52)	4.77 (17.10)	13.26 (2.98)

Note: The sign, Wilcoxon and vdW scores are used for the R-estimators. Bias and MSE are rescaled. The sample size is $n = 1000$ and number of replications is $R = 500$.

4. CONCLUSION

We propose a new class of R-estimators for the GARCH model and derive the asymptotic normality of these estimators under mild moment and smoothness conditions on the error

distribution. We exhibit the robustness and efficiency of R-estimators with respect to the QMLE through simulation and real data analysis. To approximate the distribution of R-estimators, we can consider, for example, a general type of weighted bootstrap for such estimators which is computational-friendly and easy-to-implement. The theoretical analysis such as the asymptotic validity of the weighted bootstrap is an interesting but challenging problem that can be explored in the future. A 'Value at Risk' backtesting exercise could be performed to indirectly compare the predictive performances of various estimators under consideration.

ACKNOWLEDGEMENTS

References

Andrews, B. (2012). Rank-based estimation for GARCH processes. *Econometric Theory 28*, 1037–64.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society Series B*, 367–389.

Azzalini, A. and A. Dalla Valle (1996). The multivariate skew-normal distribution. *Biometrika 83*, 715–26.

Bahadur, R. R. (1966). A Note on Quantiles in Large Samples. *Ann. Math. Statist. 37*, 577–80.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 100*, 307–27.

Berkes, I., L. Horvath and P. Kokoszka (2003). GARCH processes: structure and estimation. *Bernoulli 9*, 201–28.

Berkes, I. and L. Horvath (2004). The efficiency of the estimators of the parameters in GARCH processes. *Annals of Statistics 32*, 633–55.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley Series in Probability and Mathematical Statistics. Wiley.

Cassart, D., M. Hallin and D. Paindaveine (2010). On the estimation of cross-information quantities in R-estimation. In J. Antoch, M. Hušková and P.K. Sen, Eds: *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in Honor of Professor Jana Jurečková*, I.M.S., 35–45.

Chernoff, H. and I. R. Savage (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *The Annals of Mathematical Statistics 29*, 972–94.

Fan, J., L. Qi and D. Xiu (2014). Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *Journal of Business & Economic Statistics 32*, 178–91.

Francq, C., G. Lepage, and J.-M. Zakoïan (2011). Two-stage non-Gaussian QML estimation of GARCH models and testing the efficiency of the Gaussian QMLE. *Journal of Econometrics 165*, 246–57.

Francq, C. and J.-M. Zakoïan (2007). Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero. *Stochastic Processes and their Applications 117*, 1265–84.

Ghosh, J. K. (1971). A new proof of the Bahadur representation of quantiles and an application. *Ann. Math. Statist. 42*, 1957–61.

Glosten, L.R., R. Jagannathan and D.E. Runkle (1993). On the relation between expected value and the volatility of the nominal excess return on stocks. *Journal of Finance 48*, 1779–1801.

Hall, P. and Q. Yao (2003) Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica 71*, 285–317.

Hallin, M. and D. La Vecchia (2017). R-estimation in semiparametric dynamic location-scale models. *Journal of Econometrics 196*, 233–47.

Hallin, M., D. La Vecchia and H. Liu (2020). Center-Outward R-Estimation for Semiparametric VARMA Models. *Journal of the American Statistical Association*, 1–14.

Huber, P.J. and E.M. Ronchetti (2011). *Robust Statistics*. Wiley Series in Probability and Statistics. Wiley.

Iqbal, F. and K. Mukherjee (2010). M-estimators of some GARCH-type models; computation and application. *Statistics and Computing 20*, 435–45.

Jurečková, J. and P. Sen (1996). *Robust Statistical Procedures: Asymptotics and Interrelations.* New York: Wiley.

Koul, H. and M. Ossiander (1994). Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Annals of Statistics 22*, 540–62.

Kreiss, J.-P. (1987). On adaptative estimation in stationary ARMA processes, *Annals of Statistics 15*, 112–33.

Le Cam, L. and G. L. Yang(2000). *Asymptotics in Statistics : Some basic concepts (2nd ed.).* New York: Springer.

Liu, H. and K. Mukherjee (2020). M-estimation in GARCH models without higher order moments. *arXiv preprint arXiv:2001.10782.*

Mukherjee, K. (2007). Generalized R-estimators under conditional heteroscedasticity. *Journal of Econometrics 141*, 383–415.

Mukherjee, K. (2008). M-estimation in GARCH models. *Econometric Theory 24*, 1530–53.

Mukherjee, K. and Z. Bai (2002). R-estimation in autoregression with square-integrable score function. *Journal of Multivariate Analysis 81*, 167-86.

Muller, N. and V. Yohai (2008). Robust estimates for GARCH models. *Journal of Statistical Planning and Inference 138*, 2918–40.

Preminger, A. and G. Storti (2017). Least-squares estimation of GARCH(1,1) models with heavy-tailed errors. *The Econometrics Journal 20*, 221–58.

Tsay, R. (2010). *Analysis of financial time series,* Hoboken, N.J.: Wiley. 3rd edition.

van der Vaart, A. (1998). *Asymptotic Statistics.* Cambridge University Press.

Zhu, K. and S. Ling (2011). Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. *Annals of Statistics 39*, 2131–2163.

R-estimators in GARCH models; asymptotics and applications

HANG LIU[†] AND KANCHAN MUKHERJEE[‡]

[†]*International Institute of Finance, School of Management,*
University of Science and Technology of China, Hefei, China.
E-mail: `hliu01@ustc.edu.cn`
[‡]*Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, UK.*
E-mail: `k.mukherjee@lancaster.ac.uk`

A. PROOFS OF PROPOSITION 2.1, THEOREM 2.1 AND 2.2

We will use the following facts from Berkes et al. (2003) for the proofs:

FACT 1. For any $\nu > 0$,

$$\mathrm{E}\left\{\sup\left[\left|\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta})}{v_1(\boldsymbol{\theta})}\right|^\nu; \boldsymbol{\theta} \in \boldsymbol{\Theta}_0\right]\right\} < \infty. \tag{A.1}$$

and

$$\mathrm{E}\left\{\sup\left[\left|\frac{\ddot{\boldsymbol{v}}_1(\boldsymbol{\theta})}{v_1(\boldsymbol{\theta})}\right|^\nu; \boldsymbol{\theta} \in \boldsymbol{\Theta}_0\right]\right\} < \infty.$$

FACT 2. There exist random variables Z_0, Z_1 and Z_2, all independent of $\{\epsilon_t; 1 \le t \le n\}$ and a number $0 < \rho < 1$, such that

$$0 < v_t(\boldsymbol{\theta}) - \widehat{v}_t(\boldsymbol{\theta}) \le \rho^t Z_0, \tag{A.2}$$

$$|\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}) - \dot{\widehat{\boldsymbol{v}}}_t(\boldsymbol{\theta})| \le \rho^t Z_1, \tag{A.3}$$

$$|\ddot{\boldsymbol{v}}_t(\boldsymbol{\theta}) - \ddot{\widehat{\boldsymbol{v}}}_t(\boldsymbol{\theta})| \le \rho^t Z_2.$$

FACT 3. Let $\{(A_t, B_t, C_t); t \ge 0\}$ be a sequence of identically distributed random variables. If $\mathrm{E}\log^+ A_0 + \mathrm{E}\log^+ B_0 + \mathrm{E}\log^+ C_0 < \infty$, then for any $|r| < 1$,

$$\sum_{t=0}^{\infty}(A_t + B_t C_t)r^t \quad \text{converges with probability 1.} \tag{A.4}$$

Idea of the proof of Theorem 2.1. We first derive the following Theorem A.1, Corollary A.1.1 and Theorem A.2 on empirical processes where a scale-perturbed weighted mixed-empirical process is approximated by its non-perturbed version. With $\boldsymbol{\theta}_{n\varphi} = \boldsymbol{\theta}_{0\varphi} + n^{-1/2}\boldsymbol{b}$, we derive asymptotic expansion of the difference between two quantities $\boldsymbol{T}_{1n}(\boldsymbol{\theta}_{n\varphi})$ and $\boldsymbol{T}_{2n}(\boldsymbol{\theta}_{n\varphi})$ which are defined later. We then show that $\boldsymbol{T}_{1n}(\boldsymbol{\theta}_{n\varphi})$ can be approximated by a r.v., which is asymptotic normal, plus a term linear in $\boldsymbol{b}$. Also, we use $\boldsymbol{T}_{2n}(\boldsymbol{\theta}_{n\varphi})$ to approximate $\boldsymbol{R}_n(\boldsymbol{\theta}_{n\varphi})$ and show that asymptotically their difference is a r.v. with mean zero. Finally, we prove that the difference of $\boldsymbol{R}_n(\boldsymbol{\theta}_{n\varphi})$ and $\widehat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{n\varphi})$ converges in probability to zero. Using these results, we are able to derive the asymptotic linearity of $\widehat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{n\varphi})$ as shown in Proposition 2.1. Finally, using the definition of

the one-step R-estimator in (2.9), we are able to derive the asymptotic distribution of $\widehat{\boldsymbol{\theta}}_{n\varphi}$.

Theorem A.1, Corollary A.1.1 and Theorem A.2.

Let $\{(\eta_t, \gamma_{nt}, \delta_{nt}), 1 \leq t \leq n\}$ be an array of 3-tuple r.v.'s defined on a probability space such that $\{\eta_t, 1 \leq t \leq n\}$ are i.i.d. with c.d.f. G and η_t is independent of $(\gamma_{nt}, \delta_{nt})$ for each $1 \leq t \leq n$. Let $\{\mathcal{A}_{nt}; 1 \leq t \leq n\}$ be an array of increasing sub-σ-fields in both n and t so that $\mathcal{A}_{nt} \subset \mathcal{A}_{n(t+1)}$, $\mathcal{A}_{nt} \subset \mathcal{A}_{(n+1)t}$, $1 \leq t \leq n-1$, $n \geq 2$. Assume also that $(\gamma_{n1}, \delta_{n1})$ is $\mathcal{A}_{n1}$ measurable, and $\{\{(\gamma_{nt}, \delta_{nt}); 1 \leq t \leq j\}, \eta_1, \eta_2, \ldots, \eta_{j-1}\}$ are $\mathcal{A}_{nj}$ measurable, $2 \leq j \leq n$. For $x \in \mathbb{R}$, recall that $\mu(x) = \mathrm{E}[\eta I(\eta < x)] = \int_{-\infty}^{x} sg(s)ds$ and consider the following weighted mixed-empirical processes

$$\tilde{V}_n(x) := n^{-1/2} \sum_{t=1}^{n} \gamma_{nt}\eta_t I(\eta_t < x + x\delta_{nt}), \tag{A.5}$$

$$\tilde{J}_n(x) := n^{-1/2} \sum_{t=1}^{n} \gamma_{nt}\mu(x + x\delta_{nt}),$$

$$V_n^*(x) := n^{-1/2} \sum_{t=1}^{n} \gamma_{nt}\eta_t I(\eta_t \leq x), \qquad J_n^*(x) := n^{-1/2} \sum_{t=1}^{n} \gamma_{nt}\mu(x),$$

$$\tilde{U}_n(x) := \tilde{V}_n(x) - \tilde{J}_n(x), \quad U_n^*(x) := V_n^*(x) - J_n^*(x).$$

Assume the following conditions on the weights $\{\gamma_{nt}\}$ and perturbations $\{\delta_{nt}\}$.

Let $C_n := \sum_{t=1}^{n} \mathrm{E}|\gamma_{nt}|^q$ for some $q > 2$. Let a with $0 < a < q/2$ be such that

$$C_n/n^{q/2-a} = o(1). \tag{A.6}$$

$$\left(n^{-1} \sum_{t=1}^{n} \gamma_{nt}^2\right)^{1/2} = \gamma + o_{\mathrm{P}}(1) \text{ for a positive r.v.}\gamma. \tag{A.7}$$

$$\mathrm{E}\left(n^{-1} \sum_{t=1}^{n} \gamma_{nt}^2\right)^{q/2} = O(1). \tag{A.8}$$

$$\max_{1 \leq t \leq n} n^{-1/2}|\gamma_{nt}| = o_{\mathrm{P}}(1). \tag{A.9}$$

$$\max_{1 \leq t \leq n} |\delta_{nt}| = o_{\mathrm{P}}(1). \tag{A.10}$$

$$\frac{n^{q/2-\epsilon}}{C_n}\mathrm{E}\left[n^{-1} \sum_{t=1}^{n}\{\gamma_{nt}^2|\delta_{nt}|\}\right]^{q/2} = o(1). \tag{A.11}$$

$$n^{-1/2} \sum_{t=1}^{n} |\gamma_{nt}\delta_{nt}| = O_{\mathrm{P}}(1). \tag{A.12}$$

The following theorem shows that *uniformly* over the entire real line, the perturbed process $\tilde{U}_n$ can be approximated by U_n^*.

THEOREM A.1. *Under the above set-up and Assumptions (A.6)-(A.12) and Assumption 2.1,*

$$\sup_{x \in \mathbb{R}} |\tilde{U}_n(x) - U_n^*(x)| = o_{\mathrm{P}}(1). \tag{A.13}$$

Proof. The proof is similar to the proof in Mukherjee (2007, Theorem 6.1). In particular, we show point-wise convergence for each x and then invoke the monotone structure of the mean processes to achieve the uniform convergence. For weighted empirical, the monotonically increasing mean process is given by the distribution function. Although μ in the present case is not a monotone function on $(-\infty, \infty)$, we use its monotone property separately on $(-\infty, 0]$ and $[0, \infty)$. $\qquad\square$

We remark that this theorem is different from Koul and Ossiander (1994, Theorem 1.1) and Mukherjee (2007, Theorem 6.1) where weighted empirical processes were considered for the estimation of the mean parameters. For the estimation of the scale parameters, in this paper we consider weighted mixed-empirical process which is a weighted sum of the mixture of error and its indicator process.

The following corollary describes a Taylor-type expansion of the weighted sum of indicator functions $\tilde{V}_n(x)$.

COROLLARY A.1.1. *Under the above setup and under the Assumptions (A.6)-(A.12) and Assumption 2.1,*

$$\sup_{x \in \mathbb{R}} |\tilde{J}_n(x) - J_n^*(x) - x^2 g(x) n^{-1/2} \sum_{t=1}^n \gamma_{nt} \delta_{nt}| = o_{\mathrm{P}}(1). \qquad (A.14)$$

Hence,

$$\sup_{x \in \mathbb{R}} |\tilde{V}_n(x) - V_n^*(x) - x^2 g(x) n^{-1/2} \sum_{t=1}^n \gamma_{nt} \delta_{nt}| = o_{\mathrm{P}}(1). \qquad (A.15)$$

Proof. Here (A.15) follows from (A.14) and (A.13). Therefore, it remains to prove (A.14). Notice that the LHS of (A.14) equals

$$\sup_{x \in \mathbb{R}} \left| n^{-1/2} \sum_{t=1}^n \gamma_{nt} \left[x \int_x^{x + x\delta_{nt}} sg(s)ds - x^2 g(x) \delta_{nt} \right] \right|$$

$$= \sup_{x \in \mathbb{R}} \left| n^{-1/2} \sum_{t=1}^n \gamma_{nt} \delta_{nt} \left[x \int_0^1 (x + hx\delta_{nt}) g(x + hx\delta_{nt}) dh - x^2 g(x) \right] \right|$$

$$= o_{\mathrm{P}}(1)$$

due to (A.12) and Assumption 2.1. $\qquad\square$

The next theorem provides an extended version of (A.13) when the weights are functions on appropriately scaled parameter space. We define the following processes of two arguments as follows.

Probabilistic framework: Let $\{\eta_t, 1 \le t \le n\}$ be i.i.d. with the c.d.f. G, $\{l_{nt}; 1 \le t \le n\}$ be an array of measurable functions from $\mathbb{R}^m$ to $\mathbb{R}$ such that for every $\boldsymbol{b} \in \mathbb{R}^m$

and $1 \leq t \leq n$, $(l_{nt}(\boldsymbol{b}), u_{nt}(\boldsymbol{b}))$ are independent of η_t. For $x \in \mathbb{R}$ and $\boldsymbol{b} \in \mathbb{R}^m$, let

$$\tilde{\mathcal{V}}(x, \boldsymbol{b}) := n^{-1/2} \sum_{t=1}^{n} l_{nt}(\boldsymbol{b})\eta_t I\Big(\eta_t < x + xu_{nt}(\boldsymbol{b})\Big),$$

$$\tilde{\mathcal{J}}(x, \boldsymbol{b}) := n^{-1/2} \sum_{t=1}^{n} l_{nt}(\boldsymbol{b})\, \mu\Big(x + xu_{nt}(\boldsymbol{b})\Big),$$

$$\tilde{\mathcal{U}}(x, \boldsymbol{b}) := \tilde{\mathcal{V}}(x, \boldsymbol{b}) - \tilde{\mathcal{J}}(x, \boldsymbol{b}),$$

$$\mathcal{V}^*(x, \boldsymbol{b}) := n^{-1/2} \sum_{t=1}^{n} l_{nt}(\boldsymbol{b})\eta_t I(\eta_t < x), \ \mathcal{J}^*(x, \boldsymbol{b}) := n^{-1/2} \sum_{t=1}^{n} l_{nt}(\boldsymbol{b})\, \mu(x),$$

$$\mathcal{U}^*(x, \boldsymbol{b}) := \mathcal{V}^*(x, \boldsymbol{b}) - \mathcal{J}^*(x, \boldsymbol{b}) = n^{-1/2} \sum_{t=1}^{n} l_{nt}(\boldsymbol{b})\Big[\eta_t I(\eta_t < x) - \mu(x)\Big].$$

Here $\mathcal{U}^*(\cdot, \cdot)$ is a sequence of ordinary non-perturbed weighted mixed-empirical processes with weights $\{l_{nt}(\cdot)\}$ and $\tilde{\mathcal{U}}(\cdot, \cdot)$ is a sequence of perturbed weighted mixed-empirical processes with scale perturbations $\{u_{nt}(\cdot)\}$. In Theorem A.2 below it is shown that $\tilde{\mathcal{U}}$ can be uniformly approximated by $\mathcal{U}^*$ under the following conditions (A.16)-(A.24) for $\{l_{nt}(\cdot)\}$ and $\{u_{nt}(\cdot)\}$. Note that the statements on assumptions and convergence hold point-wise for each fixed $\boldsymbol{b} \in \mathbb{R}^m$.

There exist numbers $q > 2$ and a (both free from $\boldsymbol{b}$) satisfying $0 < a < q/2$ such that with $C_n(\boldsymbol{b}) := \sum_{t=1}^{n} \mathrm{E}|l_{ni}(\boldsymbol{b})|^q$,

$$C_n(\boldsymbol{b})/n^{q/2-a} = o(1), \quad \text{for each } \boldsymbol{b} \in \mathbb{R}^m. \tag{A.16}$$

For some positive random process $\ell(\boldsymbol{b})$,

$$\left(n^{-1}\sum_{t=1}^{n}l_{nt}^{2}(\boldsymbol{b})\right)^{1/2}=\ell(\boldsymbol{b})+o_{\mathrm{P}}(1), \qquad\qquad \boldsymbol{b}\in\mathbb{R}^{m}. \qquad (A.17)$$

$$\mathrm{E}\left(n^{-1}\sum_{t=1}^{n}l_{ni}^{2}(\boldsymbol{b})\right)^{q/2}=O(1), \qquad\qquad \boldsymbol{b}\in\mathbb{R}^{m}. \qquad (A.18)$$

$$\max_{1\leq t\leq n}n^{-1/2}|l_{nt}(\boldsymbol{b})|=o_{\mathrm{P}}(1), \qquad\qquad \boldsymbol{b}\in\mathbb{R}^{m}. \qquad (A.19)$$

$$\max_{1\leq t\leq n}\{|u_{nt}(\boldsymbol{b})|\}=o_{\mathrm{P}}(1), \qquad\qquad \boldsymbol{b}\in\mathbb{R}^{m}. \qquad (A.20)$$

$$\frac{n^{q/2-a}}{C_{n}(\boldsymbol{b})}\mathrm{E}\left[n^{-1}\sum_{t=1}^{n}l_{nt}^{2}(\boldsymbol{b})|u_{nt}(\boldsymbol{b})|\right]^{q/2}=o(1), \qquad\qquad \boldsymbol{b}\in\mathbb{R}^{m}. \qquad (A.21)$$

$$n^{-1/2}\sum_{t=1}^{n}l_{nt}(\boldsymbol{b})u_{nt}(\boldsymbol{b})=O_{\mathrm{P}}(1), \quad \boldsymbol{b}\in\mathbb{R}^{m}. \qquad (A.22)$$

$$\forall\,b\,\text{and}\,\varepsilon>0,\ \exists\,\delta>0,\ \text{and}\,n_{1}\in\mathbb{N}\,\text{whenever}\,\|\boldsymbol{s}\|\leq b,\,\text{and}\,n>n_{1}, \qquad (A.23)$$

$$P\left(n^{-1/2}\sum_{t=1}^{n}|l_{nt}(\boldsymbol{s})|\left\{\sup_{\|\boldsymbol{t}-\boldsymbol{s}\|<\delta}|u_{nt}(\boldsymbol{t})-u_{nt}(\boldsymbol{s})|\right\}\leq\varepsilon\right)>1-\varepsilon.$$

$$\forall\,b\,\text{and}\,\varepsilon>0,\ \exists\,\delta>0,\ \text{and}\,n_{2}\in\mathbb{N}\,\text{whenever}\,\|\boldsymbol{s}\|\leq b,\,\text{and}\,n>n_{2}, \qquad (A.24)$$

$$P\left(\sup_{\|\boldsymbol{t}-\boldsymbol{s}\|\leq\delta}n^{-1/2}\sum_{t=1}^{n}|l_{nt}(\boldsymbol{t})-l_{nt}(\boldsymbol{s})|\leq\varepsilon\right)>1-\varepsilon.$$

Conditions (A.16)-(A.24) are regularity conditions on the weights and perturbations of the two-parameters empirical processes. Conditions (A.23)-(A.24) are smoothness conditions on the weights and perturbations. Under stationarity and ergodicity, many of these conditions reduce to much simpler conditions based on existence of the moments.

The following theorem generalizes (A.13) when the weights are functions of $\boldsymbol{b}$.

THEOREM A.2. *Under the above framework, suppose that conditions (A.16)-(A.24) and Assumption 2.1 hold. Then for every $0<b<\infty$,*

$$\sup_{x\in\mathbb{R},\|\boldsymbol{b}\|\leq b}|\tilde{\mathcal{U}}(x,\boldsymbol{b})-\mathcal{U}^{*}(x,\boldsymbol{b})|=o_{\mathrm{P}}(1). \qquad (A.25)$$

Proof. Clearly, under conditions (A.16)-(A.22), Theorem A.1 entails that for each fixed $\boldsymbol{b}$,

$$\sup_{x\in\mathbb{R}}|\tilde{\mathcal{U}}(x,\boldsymbol{b})-\mathcal{U}^{*}(x,\boldsymbol{b})|=o_{\mathrm{P}}(1).$$

The uniform convergence with respect to $\boldsymbol{b}$ over compact sets can be proved as in Mukherjee (2007, Lemma 3.2) using conditions (A.23) and (A.24). $\qquad\square$

The following facts are useful in the proofs of various results of this paper. Let $m=1+p+q$ be the total number of parameters and fix $\boldsymbol{b}\in\mathbb{R}^{m}$. Let $\boldsymbol{\theta}_{n\varphi}=\boldsymbol{\theta}_{0\varphi}+n^{-1/2}\boldsymbol{b}$,

$$u_{nt}(\boldsymbol{b})=\frac{v_{t}^{1/2}(\boldsymbol{\theta}_{n\varphi})}{v_{t}^{1/2}(\boldsymbol{\theta}_{0\varphi})}-1,\ \ v_{nt}(\boldsymbol{b})=\frac{v_{t}^{1/2}(\boldsymbol{\theta}_{0\varphi})}{v_{t}^{1/2}(\boldsymbol{\theta}_{n\varphi})}-1. \qquad (A.26)$$

Then $\{u_{nt}(\boldsymbol{b})\}$ satisfies (A.20) since

$$u_{nt}(\boldsymbol{b}) = \frac{v_t(\boldsymbol{\theta}_{n\varphi}) - v_t(\boldsymbol{\theta}_{0\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})\{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi}) + v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})\}} = \frac{n^{-1/2}\dot{\boldsymbol{v}}_t'(\boldsymbol{\theta}^*)\boldsymbol{b}}{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})\{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi}) + v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})\}},$$
(A.27)

for some $\boldsymbol{\theta}^* = \boldsymbol{\theta}^*(n,t,\boldsymbol{b})$ in the neighbourhood of $\boldsymbol{\theta}_{0\varphi}$ for large n. The $n^{-1/2}$-factor is used later for deriving convergence of some sequence of random vectors. Similarly, for some $\boldsymbol{\theta}^*$,

$$v_{nt}(\boldsymbol{b}) = \frac{v_t(\boldsymbol{\theta}_{0\varphi}) - v_t(\boldsymbol{\theta}_{n\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})\{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi}) + v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})\}} = \frac{-n^{-1/2}\dot{\boldsymbol{v}}_t'(\boldsymbol{\theta}^*)\boldsymbol{b}}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})\{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi}) + v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})\}} = n^{-1/2}\xi_{nt},$$
(A.28)

say. Let $a_{nt}(\boldsymbol{b}) = v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})/v_t^{1/2}(\boldsymbol{\theta}_{n\varphi}) = 1 + v_{nt}(\boldsymbol{b}) = 1 + n^{-1/2}\xi_{nt}$. Then

$$\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} = a_{nt}(\boldsymbol{b})\eta_t = \eta_t + n^{-1/2}\eta_t\xi_{nt} = \eta_t + n^{-1/2}z_{nt},$$

where

$$z_{nt} = \eta_t\xi_{nt} = \eta_t \times \frac{-\dot{\boldsymbol{v}}_t'(\boldsymbol{\theta}^*)\boldsymbol{b}}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})\{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi}) + v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})\}}.$$

For $\delta > 0$ in Assumption 2.1 and any $c > 0$,

$$P\left[n^{-1/2}\max_{1 \le t \le n}|z_{nt}| > c\right] \le \sum_{t=1}^{n} P\left[n^{-1/2}|z_{nt}| > c\right] \le n\frac{\mathrm{E}\left[n^{-1-\delta/2}|\eta_t|^{2+\delta}|\xi_{nt}|^{2+\delta}\right]}{c^{2+\delta}} = o(1)$$

since all moments of $\{|\xi_{nt}|\}$ are finite and η_t and ξ_{nt} are independent for all t. Therefore

$$\max_{1 \le t \le n}\left|\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} - \eta_t\right| = o_{\mathrm{P}}(1).$$
(A.29)

If $\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})/v_t(\boldsymbol{\theta}_{n\varphi})$ appears as the coefficients, we replace it by $\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})/v_t(\boldsymbol{\theta}_{0\varphi})$ and the difference is controlled as follows. Notice that all derivatives below exist with bounded moments and so

$$\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} - \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} = n^{-1/2}\boldsymbol{A}_t(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b} + n^{-1}\boldsymbol{A}_{tn}^*,$$
(A.30)

where $\boldsymbol{A}_t(\boldsymbol{\theta}_{0\varphi}) = \ddot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})/v_t(\boldsymbol{\theta}_{0\varphi}) - \dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})\dot{\boldsymbol{v}}_t'(\boldsymbol{\theta}_{0\varphi})/\{v_t(\boldsymbol{\theta}_{0\varphi})\}^2$. Only the term $n^{-1/2}\boldsymbol{A}_t(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}$ is of our interest since others are of higher order than n.

Take $l_{nt}(\boldsymbol{b})$ to be equal to the j-th coordinate $(1 \le j \le m = 1 + p + q)$ of

$$\boldsymbol{L}_{nt}(\boldsymbol{b}) = \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \times \frac{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})}$$
(A.31)

and $u_{nt}(\boldsymbol{b})$ as in (A.26). We now show that (A.17)-(A.24) hold with such choice.

For each t with $1 \le t \le n$, $\{\boldsymbol{L}_{nt}(\boldsymbol{b}), u_{nt}(\boldsymbol{b})\}$ are independent of η_t. Using a Taylor expansion of $l_{nt}(\boldsymbol{b})$ at $\boldsymbol{\theta}_{0\varphi}$ for each $1 \le t \le n$ and noting the existence of all moments of $v_t(\boldsymbol{\theta}_{0\varphi})$ and its derivatives of all higher orders, (A.17) and (A.18) hold. Existence of all higher moments of $\{l_{nt}(\boldsymbol{b}), u_{nt}(\boldsymbol{b})\}$ ensure conditions (A.19)-(A.21).

To verify (A.22), we use (A.27) and that for each t, $v_t(\cdot)$ is a smooth function with

derivatives of all order to conclude that

$$n^{-1/2} \sum_{t=1}^{n} \boldsymbol{L}_{nt}(\boldsymbol{b}) u_{nt}(\boldsymbol{b}) = \mathrm{E}[\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_{0\varphi})/v_1^2(\boldsymbol{\theta}_{0\varphi})](\boldsymbol{b}/2) + o_{\mathrm{P}}(1) = \boldsymbol{J}(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}/2 + o_{\mathrm{P}}(1).$$

Conditions (A.23) and (A.24) can be verified using the mean value theorem.

The following lemmas and their proofs represent the intermediate steps in the proofs of Proposition 2.1 and Theorem 2.1.

Lemma A.3, Lemma A.4, Lemma A.5 and Lemma A.6.

Let

$$\boldsymbol{T}_{n1}(\boldsymbol{\theta}_{n\varphi}) = n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \left\{ 1 - \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \varphi[G(\eta_t)] \right\},$$

$$\boldsymbol{T}_{n2}(\boldsymbol{\theta}_{n\varphi}) = n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \left\{ 1 - \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \varphi \left[G \left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right) \right] \right\}$$

and note that the difference in the definitions of these two quantities lies only in the argument of $\varphi(G(.))$. We show in Lemma A.3 below that $\int_0^1 [\tilde{\mathcal{V}}(u,\boldsymbol{b}) - \mathcal{V}^*(u,\boldsymbol{b})] d\varphi(u) = \boldsymbol{T}_{n1}(\boldsymbol{\theta}_{n\varphi}) - \boldsymbol{T}_{n2}(\boldsymbol{\theta}_{n\varphi})$. Using results on empirical processes in Theorem A.2, $\int_0^1 [\tilde{\mathcal{V}}(u,\boldsymbol{b}) - \mathcal{V}^*(u,\boldsymbol{b})] d\varphi(u)$ is linear in $\boldsymbol{b}$. Consequently, we obtain the following uniform approximations of $\boldsymbol{T}_{n1}(\boldsymbol{\theta}_{n\varphi}) - \boldsymbol{T}_{n2}(\boldsymbol{\theta}_{n\varphi})$ over $||\boldsymbol{b}|| \leq c$ where $c > 0$.

LEMMA A.3. *Let Assumptions 2.1-2.3 hold. Then, as $n \to \infty$,*

$$\boldsymbol{T}_{n2}(\boldsymbol{\theta}_{n\varphi}) - \boldsymbol{T}_{n1}(\boldsymbol{\theta}_{n\varphi}) = \boldsymbol{M}(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b} + u_{\mathrm{P}}(1), \tag{A.32}$$

where $\boldsymbol{M}(\boldsymbol{\theta}_{0\varphi}) = \boldsymbol{J}(\boldsymbol{\theta}_{0\varphi})\rho(\varphi)/2$.

Proof. To use Theorem A.2 in the proof, let $\boldsymbol{b} = n^{1/2}(\boldsymbol{\theta}_{n\varphi} - \boldsymbol{\theta}_{0\varphi})$ and $x = G^{-1}(u)$ for some $0 < u < 1$. For simplicity, we use the notation $\tilde{\mathcal{V}}(u,\boldsymbol{b})$ to denote $\tilde{\mathcal{V}}(G^{-1}(u),\boldsymbol{b})$ which is defined in the probabilistic framework above. Accordingly

$$\tilde{\mathcal{V}}(u,\boldsymbol{b}) := n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} I \left[\eta_t < G^{-1}(u) \frac{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})} \right]$$

and

$$\mathcal{V}^*(u,\boldsymbol{b}) = n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} I \left(\eta_t < G^{-1}(u) \right).$$

With the choice based on (A.31) and (A.26) and using

$$\frac{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \eta_t = \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})},$$

$$\tilde{\mathcal{V}}(u,\boldsymbol{b}) = n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} I \left[\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} < G^{-1}(u) \right]$$

$$= n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} I \left[G \left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right) < u \right].$$

Similarly,

$$\mathcal{V}^*(u, \boldsymbol{b}) = n^{-1/2} \sum_{t=1}^n \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} I\left(G(\eta_t) < u\right).$$

Since

$$\int_0^1 I\left\{ G\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})}\right) < u \right\} d\varphi(u) = \varphi(1) - \varphi\left[G\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})}\right)\right],$$

we get

$$\int_0^1 \tilde{\mathcal{V}}(u, \boldsymbol{b}) d\varphi(u) = n^{-1/2} \sum_{t=1}^n \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \left\{ \varphi(1) - \varphi\left[G\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})}\right)\right] \right\}$$

and

$$\int_0^1 \mathcal{V}^*(u, \boldsymbol{b}) d\varphi(u) = n^{-1/2} \sum_{t=1}^n \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \left(\varphi(1) - \varphi(G(\eta_t))\right).$$

Canceling $\varphi(1)$, $\int_0^1 [\tilde{\mathcal{V}}(u, \boldsymbol{b}) - \mathcal{V}^*(u, \boldsymbol{b})] d\varphi(u)$ equals

$$n^{-1/2} \sum_{t=1}^n \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \left\{ -\varphi\left[G\left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})}\right)\right] + \varphi(G(\eta_t)) \right\}$$
$$= \boldsymbol{T}_{n2}(\boldsymbol{\theta}_{n\varphi}) - \boldsymbol{T}_{n1}(\boldsymbol{\theta}_{n\varphi}).$$

Using (A.14) and (A.27) with

$$\tilde{\mathcal{J}}(u, \boldsymbol{b}) := n^{-1/2} \sum_{t=1}^n \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \frac{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \mu\left[G^{-1}(u) \frac{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})} \right]$$

$$\mathcal{J}^*(u, \boldsymbol{b}) := n^{-1/2} \sum_{t=1}^n \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \frac{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \mu\left(G^{-1}(u)\right),$$

we have

$$\sup_{u \in (0,1)} |\tilde{\mathcal{J}}(u, \boldsymbol{b}) - \mathcal{J}^*(u, \boldsymbol{b}) - [G^{-1}(u)]^2 g(G^{-1}(u)) n^{-1/2} \sum_{t=1}^n \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \frac{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} u_{nt}(\boldsymbol{b})| = u_{\mathrm{P}}(1).$$

Also,

$$|n^{-1/2} \sum_{t=1}^n \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \frac{v_t^{1/2}(\boldsymbol{\theta}_{0\varphi})}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} u_{nt}(\boldsymbol{b}) - \boldsymbol{J}(\boldsymbol{\theta}_{0\varphi})| = u_{\mathrm{P}}(1).$$

Hence,

$$\int_0^1 [\tilde{\mathcal{J}}(u, \boldsymbol{b}) - \mathcal{J}^*(u, \boldsymbol{b})] d\varphi(u) = \int_0^1 [G^{-1}(u)]^2 g(G^{-1}(u)) d\varphi(u) \boldsymbol{J} \boldsymbol{b}/2 + u_{\mathrm{P}}(1)$$
$$= \boldsymbol{M}(\boldsymbol{\theta}_{0\varphi}) \boldsymbol{b} + u_{\mathrm{P}}(1) \tag{A.33}$$

by recalling that $\boldsymbol{M}(\boldsymbol{\theta}_{0\varphi}) = \boldsymbol{J}(\boldsymbol{\theta}_{0\varphi}) \rho(\varphi)/2$. Finally, (A.32) follows from Theorem A.2. $\square$

The following lemma states that the difference between $\boldsymbol{T}_{n1}(\boldsymbol{\theta}_{n\varphi})$ and $\boldsymbol{N}_n(\boldsymbol{\theta}_{0\varphi})$ is asymptotically linear in $\boldsymbol{b}$.

LEMMA A.4. *Let Assumptions 2.1-2.3 hold. Then, as $n \to \infty$,*

$$T_{n1}(\boldsymbol{\theta}_{n\varphi}) - N_n(\boldsymbol{\theta}_{0\varphi}) = J(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}/2 + u_{\mathrm{P}}(1), \tag{A.34}$$

where

$$N_n(\boldsymbol{\theta}_{0\varphi}) \to \mathcal{N}(\boldsymbol{0}, J(\boldsymbol{\theta}_{0\varphi})\sigma^2(\varphi)), \tag{A.35}$$

with $\sigma^2(\varphi) = \mathrm{Var}\{\eta_1\varphi[G(\eta_1)]\}$.

Proof. The difference between $T_{n1}(\boldsymbol{\theta}_{n\varphi})$ and $N_n(\boldsymbol{\theta}_{0\varphi})$ lies in comparing $X_t/v_t^{1/2}(\boldsymbol{\theta}_{n\varphi}) = \eta_t + n^{-1/2}z_{nt}$ and η_t and involves smooth function of $\boldsymbol{b}$. So the proof follows easily with the details below. Notice that

$$T_{n1}(\boldsymbol{\theta}_{n\varphi}) - N_n(\boldsymbol{\theta}_{0\varphi}) = n^{-1/2}\sum_{t=1}^{n}\left[\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} - \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})}\right]\{1 - a_{nt}(\boldsymbol{b})\eta_t\varphi[G(\eta_t)]\}$$

$$- n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})}v_{nt}(\boldsymbol{b})\eta_t\varphi[G(\eta_t)] = \boldsymbol{F}_{n1} - \boldsymbol{F}_{n2}.$$

Using (A.30),

$$\boldsymbol{F}_{n1} = n^{-1}\sum_{t=1}^{n}\boldsymbol{A}_t(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}\{1 - a_{nt}(\boldsymbol{b})\eta_t\varphi[G(\eta_t)]\} + u_{\mathrm{P}}(1)$$

$$= n^{-1}\sum_{t=1}^{n}\boldsymbol{A}_t(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}\{1 - \eta_t\varphi[G(\eta_t)]\} - n^{-1}\sum_{t=1}^{n}\boldsymbol{A}_t(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}v_{nt}(\boldsymbol{b})\eta_t\varphi[G(\eta_t)] + u_{\mathrm{P}}(1).$$

Using the LLN, the first term in the above decomposition of $\boldsymbol{F}_{n1}$ is $u_{\mathrm{P}}(1)$ since

$$\mathrm{E}\{\boldsymbol{A}_t(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}\{1 - \eta_t\varphi[G(\eta_t)]\}\} = \mathrm{E}[\boldsymbol{A}_t(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}]\,\mathrm{E}\{1 - \eta_t\varphi[G(\eta_t)]\} = \boldsymbol{0}.$$

For the second term, using (A.28) we have $n^{-1/2}$ factor of $v_{nt}(\boldsymbol{b})$ and consequently it is $u_{\mathrm{P}}(1)$.

For $\boldsymbol{F}_{n2}$, we approximate $v_{nt}(\boldsymbol{b})$ by $-n^{-1/2}\dot{\boldsymbol{v}}_t'(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}/\{2v_t(\boldsymbol{\theta}_{0\varphi})\}$ and use $\mathrm{E}\{\eta_t\varphi[G(\eta_t)]\} = 1$ to obtain $\boldsymbol{F}_{n2} = -J(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}/2 + u_{\mathrm{P}}(1)$. Hence (A.34) is proved.

Using the independence of v_t and η_t for each t, $N_n(\boldsymbol{\theta}_{0\varphi})$ is a sum of the vectors of martingale differences and so (A.35) follows from the martingale CLT. $\qquad\square$

Now consider the rank-based counterpart of $T_{n2}(\boldsymbol{\theta}_{n\varphi})$

$$R_n(\boldsymbol{\theta}_{n\varphi}) = n^{-1/2}\sum_{t=1}^{n}\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})}\left\{1 - \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})}\varphi\left(\frac{R_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1}\right)\right\}.$$

The following lemma provides the difference between $T_{n2}(\boldsymbol{\theta}_{n\varphi})$ and $R_n(\boldsymbol{\theta}_{n\varphi})$. It shows that the effect of replacing observations in $T_{2n}(\boldsymbol{\theta}_{n\varphi})$ by ranks is asymptotically a r.v. with mean zero.

LEMMA A.5. *Let Assumptions 2.1-2.3 hold. Then, as $n \to \infty$,*

$$R_n(\boldsymbol{\theta}_{n\varphi}) - T_{n2}(\boldsymbol{\theta}_{n\varphi}) = Q_n(\boldsymbol{\theta}_{0\varphi}) + u_{\mathrm{P}}(1). \tag{A.36}$$

Also, $Q_n(\boldsymbol{\theta}_{0\varphi})$ converges in distribution to $\mathrm{E}(\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})/v_1(\boldsymbol{\theta}_{0\varphi}))Z$, where Z has mean zero and variance $\gamma(\varphi)$.

Proof. Consider the following decomposition

$$\boldsymbol{R}_n(\boldsymbol{\theta}_{n\varphi}) - \boldsymbol{T}_{n2}(\boldsymbol{\theta}_{n\varphi})$$

$$= n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \left\{ \varphi \left[G \left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right) \right] - \varphi \left[\frac{R_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \right] \right\}$$

$$= n^{-1/2} \sum_{t=1}^{n} \left[\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} - \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \right] \eta_t \left\{ \varphi \left[G \left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right) \right] - \varphi \left[\frac{R_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \right] \right\}$$

$$+ n^{-1/2} \sum_{t=1}^{n} \left[\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} - \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \right] v_{nt}(\boldsymbol{b}) \eta_t \left\{ \varphi \left[G \left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right) \right] - \varphi \left[\frac{R_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \right] \right\}$$

$$+ n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \eta_t \left\{ \varphi \left[G \left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right) \right] - \varphi \left[\frac{R_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \right] \right\}$$

$$+ n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} v_{nt}(\boldsymbol{b}) \eta_t \left\{ \varphi \left[G \left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right) \right] - \varphi \left[\frac{R_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \right] \right\}$$

$$= \boldsymbol{D}_{n1} + \boldsymbol{D}_{n2} + \boldsymbol{D}_{n3} + \boldsymbol{D}_{n4}.$$

Using the $n^{-1/2}$-factor in (A.30) and (A.28), $\boldsymbol{D}_{n1}$, $\boldsymbol{D}_{n2}$ and $\boldsymbol{D}_{n4}$ are $u_{\mathrm{P}}(1)$. We next prove that $\boldsymbol{D}_{n3} = \boldsymbol{Q}_n(\boldsymbol{\theta}_{0\varphi}) + u_{\mathrm{P}}(1)$ in detail. Recall that $\tilde{G}_n(x)$ is the empirical distribution function of $\{\eta_t\}$. Let $G_n(x), x \in \mathbb{R}$ be the empirical distribution function of $\{X_t/v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})\}$. Then

$$\boldsymbol{D}_{n3} = n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \eta_t \left\{ \varphi \left[G \left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right) \right] - \varphi \left[G_n \left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right) \right] \right\}$$

$$= n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \eta_t \left\{ \varphi \left[G \left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right) \right] - \varphi \left[G(\eta_t) \right] \right\}$$

$$- n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \eta_t \left\{ \varphi \left[G_n \left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right) \right] - \varphi \left[\tilde{G}_n(\eta_t) \right] \right\}$$

$$+ n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \eta_t \left\{ \varphi \left[G(\eta_t) \right] - \varphi \left[\tilde{G}_n(\eta_t) \right] \right\}$$

$$= \boldsymbol{D}_{n1}^* - \boldsymbol{D}_{n2}^* + \boldsymbol{D}_{n3}^*.$$

Since $\boldsymbol{D}_{n1}^*$ is the weighted sum of the difference of a c.d.f. evaluated at two different r.v.'s and integrated wrt φ, using the same technique for proving (A.33), $\boldsymbol{D}_{n1}^* = \boldsymbol{M}(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b} + u_{\mathrm{P}}(1)$.

Write $\boldsymbol{w}_t = \dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})/v_t(\boldsymbol{\theta}_{0\varphi})$. Since $\boldsymbol{D}_{n2}^*$ is the weighted sum of the difference of two

different c.d.f.'s evaluated at two different r.v.'s and integrated wrt φ,

$$\boldsymbol{D}_{n2}^* = \int_0^1 n^{-1/2} \sum_{t=1}^n \boldsymbol{w}_t \eta_t I \left[G_n \left(\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right) < u \right] - I \left[\tilde{G}_n(\eta_t) < u \right] d\varphi(u)$$

$$= \int_0^1 n^{-1/2} \sum_{t=1}^n \boldsymbol{w}_t \eta_t I \left[\frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} < G_n^{-1}(u) \right] - I \left[\eta_t < \tilde{G}_n^{-1}(u) \right] d\varphi(u)$$

$$= \int_0^1 n^{-1/2} \sum_{t=1}^n \boldsymbol{w}_t \eta_t I \left[\eta_t < G_n^{-1}(u) \frac{1}{1 + v_{nt}(\boldsymbol{b})} \right] - I \left[\eta_t < \tilde{G}_n^{-1}(u) \right] d\varphi(u).$$

Using (A.29), $\sup \left\{ \left| G_n^{-1}(u) - \tilde{G}_n^{-1}(u) \right|; u \in (0,1) \right\} = u_{\mathrm{P}}(1)$. Hence, by Theorem A.2,

$$\boldsymbol{D}_{n2}^* = \int_0^1 n^{-1/2} \sum_{t=1}^n \boldsymbol{w}_t \left[\mu \left(\tilde{G}_n^{-1}(u) \frac{1}{1 + v_{nt}(\boldsymbol{b})} \right) - \mu \left(\tilde{G}_n^{-1}(u) \right) \right] d\varphi(u) + u_{\mathrm{P}}(1)$$

$$= \int_0^1 n^{-1/2} \sum_{t=1}^n \boldsymbol{w}_t \dot{\mu} \left(\tilde{G}_n^{-1}(u) \right) \frac{-v_{nt}(\boldsymbol{b})}{1 + v_{nt}(\boldsymbol{b})} \tilde{G}_n^{-1}(u) \, d\varphi(u) + u_{\mathrm{P}}(1)$$

$$= \int_0^1 n^{-1/2} \sum_{t=1}^n \boldsymbol{w}_t \left(\tilde{G}_n^{-1}(u) \right)^2 g \left(\tilde{G}_n^{-1}(u) \right) \frac{-v_{nt}(\boldsymbol{b})}{1 + v_{nt}(\boldsymbol{b})} d\varphi(u) + u_{\mathrm{P}}(1)$$

$$= \int_0^1 n^{-1} \sum_{t=1}^n \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi}) \dot{\boldsymbol{v}}_t'(\boldsymbol{\theta}_{0\varphi})}{2 v_t^2(\boldsymbol{\theta}_{0\varphi})} \left(G^{-1}(u) \right)^2 g \left(G^{-1}(u) \right) d\varphi(u) \boldsymbol{b} + u_{\mathrm{P}}(1)$$

$$= \boldsymbol{M}(\boldsymbol{\theta}_{0\varphi}) \boldsymbol{b} + u_{\mathrm{P}}(1).$$

Finally consider $\boldsymbol{D}_{n3}^*$ written as

$$\boldsymbol{D}_{n3}^* = \int_0^1 n^{-1/2} \sum_{t=1}^n \boldsymbol{w}_t \eta_t \left\{ I \left[\eta_t \leq G^{-1}(u) \right] - I \left[\eta_t \leq \tilde{G}_n^{-1}(u) \right] \right\} d\varphi(u)$$

$$= \int_0^1 n^{-1/2} \sum_{t=1}^n \boldsymbol{w}_t \eta_t \left\{ I \left[\eta_t \leq G^{-1}(u) \right] - I \left[\eta_t \leq G^{-1}(G(\tilde{G}_n^{-1}(u))) \right] \right\} d\varphi(u)$$

$$= \int_0^1 \left[\boldsymbol{M}_n(u) - \boldsymbol{M}_n(G(\tilde{G}_n^{-1}(u))) \right] d\varphi(u)$$

$$+ \int_0^1 n^{-1/2} \sum_{t=1}^n \boldsymbol{w}_t \left[\mu(G^{-1}(u)) - \mu(\tilde{G}_n^{-1}(u)) \right] d\varphi(u),$$

where $\boldsymbol{M}_n(u) := n^{-1/2} \sum_{t=1}^n \boldsymbol{w}_t \left\{ \eta_t I \left[\eta_t \leq G^{-1}(u) \right] - \mu(G^{-1}(u)) \right\}$. We show that

$$\left| \int_0^1 \left[\boldsymbol{M}_n(u) - \boldsymbol{M}_n(G(\tilde{G}_n^{-1}(u))) \right] d\varphi(u) \right| = o_{\mathrm{P}}(1), \qquad (A.37)$$

$$\boldsymbol{Q}_n(\boldsymbol{\theta}_{0\varphi}) = \int_0^1 n^{-1/2} \sum_{t=1}^n \boldsymbol{w}_t \left[\mu(G^{-1}(u)) - \mu(\tilde{G}_n^{-1}(u)) \right] d\varphi(u) \to \mathrm{E}(\dot{\boldsymbol{v}}_1/v_1) Z. \qquad (A.38)$$

For (A.37), note that $\{\boldsymbol{M}_n(.)\}$ converges weakly to a Brownian Bridge since for each fixed u, $\boldsymbol{M}_n(u)$ converges to a normal distribution using the martingale CLT and it

is tight using the bound on the moment of the difference process in Billingsley (1968, Theorem 12.3).

Since $\sup\{|u - G(\tilde{G}_n^{-1}(u))|; u \in (0,1)\} = \sup\{|G(x) - \tilde{G}_n(x)|; x \in \mathbb{R}\} = o_P(1)$, by the Arzela-Ascoli theorem,

$$\sup\left\{\left|\boldsymbol{M}_n(u) - \boldsymbol{M}_n(G(\tilde{G}_n^{-1}(u)))\right|; u \in (0,1)\right\} = o_P(1),$$

and consequently, (A.37) is proved. For (A.38), we use the Bahadur representation; see Bahadur (1966) and Ghosh (1971) for details. Since $\dot{g}(x)$ is bounded and g is positive on $\mathbb{R}$,

$$n^{1/2}\left(G^{-1}(u)) - \tilde{G}_n^{-1}(u)\right) - n^{-1/2}\sum_{i=1}^n \frac{I\{\eta_i \leq G^{-1}(u)\} - u}{g\left(G^{-1}(u)\right)} = o(1) \quad \text{a.s.}.$$

Applying the mean value theorem,

$$n^{1/2}\left[\mu(G^{-1}(u)) - \mu(\tilde{G}_n^{-1}(u))\right] - \dot{\mu}(G^{-1}(u))n^{-1/2}\sum_{i=1}^n \frac{I\{\eta_i \leq G^{-1}(u)\} - u}{g\left(G^{-1}(u)\right)} = o(1) \quad \text{a.s.}. \tag{A.39}$$

Using $\dot{\mu}(x) = xg(x)$,

$$\boldsymbol{Q}_n(\boldsymbol{\theta}_{0\varphi}) = n^{-1}\sum_{t=1}^n \boldsymbol{w}_t \int_0^1 \left[\dot{\mu}(G^{-1}(u))n^{-1/2}\sum_{i=1}^n \frac{I\{\eta_i \leq G^{-1}(u)\} - u}{g\left(G^{-1}(u)\right)}\right]d\varphi(u) + o_P(1)$$

$$= n^{-1}\sum_{t=1}^n \boldsymbol{w}_t \int_0^1 \left[G^{-1}(u)n^{1/2}\left(\tilde{G}_n(G^{-1}(u)) - u\right)\right]d\varphi(u) + o_P(1).$$

Since $n^{-1}\sum_{t=1}^n \boldsymbol{w}_t \to \mathrm{E}(\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})/v_1(\boldsymbol{\theta}_{0\varphi}))$, and using van der Vaart (1998, Theorem 19.3),

$$n^{1/2}\left(\tilde{G}_n(G^{-1}(u)) - u\right) \to B(u),$$

we obtain (A.38) with the r.v. Z having mean zero. The variance of Z is given by

$$\mathrm{E}(Z^2) = \mathrm{E}\left[\int_0^1 \int_0^1 G^{-1}(u)G^{-1}(v)B(u)B(v)d\varphi(u)d\varphi(v)\right]$$

$$= \int_0^1 \int_0^1 G^{-1}(u)G^{-1}(v)\mathrm{E}[B(u)B(v)]d\varphi(u)d\varphi(v)$$

$$= \int_0^1 \int_0^1 G^{-1}(u)G^{-1}(v)\left[\min\{u,v\} - uv\right]d\varphi(u)d\varphi(v)$$

$$= \gamma(\varphi).$$

$\square$

Now recall the rank-based central sequence

$$\widehat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{n\varphi}) = n^{-1/2}\sum_{t=1}^n \frac{\dot{\hat{\boldsymbol{v}}}_t(\boldsymbol{\theta}_{n\varphi})}{\widehat{v}_t(\boldsymbol{\theta}_{n\varphi})}\left\{1 - \frac{X_t}{\widehat{v}_t^{1/2}(\boldsymbol{\theta}_{n\varphi})}\varphi\left(\frac{\widehat{R}_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1}\right)\right\},$$

which is an approximation to $\boldsymbol{R}_n(\boldsymbol{\theta}_{n\varphi})$. We have the following lemma dealing with the difference between $\boldsymbol{R}_n(\boldsymbol{\theta}_{n\varphi})$ and $\widehat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{n\varphi})$.

LEMMA A.6. *Let Assumptions 2.1-2.3 hold. Then, as $n \to \infty$,*

$$\boldsymbol{R}_n(\boldsymbol{\theta}_{n\varphi}) - \widehat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{n\varphi}) = u_{\mathrm{P}}(1). \tag{A.40}$$

Proof. Note that $\boldsymbol{R}_n(\boldsymbol{\theta}_{n\varphi}) - \widehat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{n\varphi})$ equals

$$n^{-1/2} \sum_{t=1}^{n} \left[\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} - \frac{\dot{\widehat{\boldsymbol{v}}}_t(\boldsymbol{\theta}_{n\varphi})}{\widehat{v}_t(\boldsymbol{\theta}_{n\varphi})} \right] \tag{A.41}$$

$$+ n^{-1/2} \sum_{t=1}^{n} \left[\frac{\dot{\widehat{\boldsymbol{v}}}_t(\boldsymbol{\theta}_{n\varphi})}{\widehat{v}_t(\boldsymbol{\theta}_{n\varphi})} \frac{X_t}{\widehat{v}_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} - \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right] \varphi \left(\frac{\widehat{R}_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \right) \tag{A.42}$$

$$- n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \left[\varphi \left(\frac{R_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \right) - \varphi \left(\frac{\widehat{R}_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \right) \right]. \tag{A.43}$$

Due to (A.2), (A.3) and $\widehat{v}_t(\boldsymbol{\theta}) \geq c_0(\boldsymbol{\theta}) > 0$, we have

$$\left| \frac{\dot{\widehat{\boldsymbol{v}}}_t(\boldsymbol{\theta}_{n\varphi})}{\widehat{v}_t(\boldsymbol{\theta}_{n\varphi})} - \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \right| = \left| \frac{\dot{\widehat{\boldsymbol{v}}}_t(\boldsymbol{\theta}_{n\varphi}) - \dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{\widehat{v}_t(\boldsymbol{\theta}_{n\varphi})} + \dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi}) \frac{v_t(\boldsymbol{\theta}_{n\varphi}) - \widehat{v}_t(\boldsymbol{\theta}_{n\varphi})}{\widehat{v}_t(\boldsymbol{\theta}_{n\varphi}) v_t(\boldsymbol{\theta}_{n\varphi})} \right|$$

$$\leq \frac{|\dot{\widehat{\boldsymbol{v}}}_t(\boldsymbol{\theta}_{n\varphi}) - \dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})|}{\widehat{v}_t(\boldsymbol{\theta}_{n\varphi})} + \frac{|v_t(\boldsymbol{\theta}_{n\varphi}) - \widehat{v}_t(\boldsymbol{\theta}_{n\varphi})|}{\widehat{v}_t(\boldsymbol{\theta}_{n\varphi})} \frac{|\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})|}{v_t(\boldsymbol{\theta}_{n\varphi})}$$

$$\leq C\rho^t \left[Z_1 + Z_0 \frac{|\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})|}{v_t(\boldsymbol{\theta}_{n\varphi})} \right]. \tag{A.44}$$

Hence, in view of (A.1) and (A.4), for every $0 < b < \infty$,

$$\sup_{||\boldsymbol{b}|| < b} \sum_{t=1}^{n} \left| \frac{\dot{\widehat{\boldsymbol{v}}}_t(\boldsymbol{\theta}_{n\varphi})}{\widehat{v}_t(\boldsymbol{\theta}_{n\varphi})} - \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \right| = O_{\mathrm{P}}(1),$$

which implies that (A.41) is $u_{\mathrm{P}}(1)$. Since φ is bounded, (A.42) is $u_{\mathrm{P}}(1)$. For (A.43), since there is a $n^{-1/2}$ factor from

$$\frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{n\varphi})}{v_t(\boldsymbol{\theta}_{n\varphi})} \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} - \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \eta_t \, ,$$

it suffices to prove that

$$\boldsymbol{K}_n := n^{-1/2} \sum_{t=1}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \eta_t \left[\varphi \left(\frac{R_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \right) - \varphi \left(\frac{\widehat{R}_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \right) \right] = u_{\mathrm{P}}(1).$$

Let $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x. We split the sum in $\boldsymbol{K}_n$ into two parts: in the first part, t runs till $\lfloor n^k \rfloor - 1$ where $0 < k < 1/2$. We show that this part is $u_p(1)$ by noting that its expectation is of the form $n^k n^{-1/2} = o(1)$ multiplied by expectation of $\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})/v_t(\boldsymbol{\theta}_{0\varphi})\eta_t$ and a bounded quantity because φ is bounded. The number of summands in the second term is $n - n^k$ which is large but there we bound expectation of the sum of by a quantity of the form $n\rho^{\lfloor n^k \rfloor}$ with $0 < k < 1/2$ and

$0 < \rho < 1$ and this is $o(1)$. Accordingly

$$
\begin{aligned}
\boldsymbol{K}_n = \quad & n^{-1/2} \sum_{t=1}^{\lfloor n^k \rfloor -1} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \eta_t \left[\varphi \left(\frac{R_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \right) - \varphi \left(\frac{\widehat{R}_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \right) \right] \\
& + n^{-1/2} \sum_{t=\lfloor n^k \rfloor}^{n} \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})} \eta_t \left[\varphi \left(\frac{R_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \right) - \varphi \left(\frac{\widehat{R}_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \right) \right].
\end{aligned}
\tag{A.45}
$$

To show (A.45) is $u_P(1)$, we prove that for every $0 < b < \infty$,

$$
\sup_{\substack{\lfloor n^k \rfloor \le t \le n \\ \|\boldsymbol{b}\| < b}} \left| \varphi \left(\frac{R_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \right) - \varphi \left(\frac{\widehat{R}_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \right) \right| = O_P(n^{k-1}).
\tag{A.46}
$$

Since sequences $\{R_{nt}(\boldsymbol{\theta}_{n\varphi})\}$ and $\{\hat{R}_{nt}(\boldsymbol{\theta}_{n\varphi})\}$ are permutations of $\{1, ..., n\}$, with the probability tending to one as $n \to \infty$, both $\{R_{nt}(\boldsymbol{\theta}_{n\varphi})\}$ and $\{\hat{R}_{nt}(\boldsymbol{\theta}_{n\varphi})\}$ are at points of continuity of φ that has a finite number of the points of discontinuity. Therefore, to prove (A.46), it suffices to prove

$$
\sup_{\substack{\lfloor n^k \rfloor \le t \le n \\ \|\boldsymbol{b}\| < b}} \left| \frac{R_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} - \frac{\widehat{R}_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \right| = O_P(n^{k-1}).
\tag{A.47}
$$

For $\lfloor n^k \rfloor \le t \le n$, we decompose ranks as

$$
\begin{aligned}
& \frac{R_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} - \frac{\widehat{R}_{nt}(\boldsymbol{\theta}_{n\varphi})}{n+1} \\
& = \frac{1}{n+1} \sum_{i=1}^{\lfloor n^k \rfloor -1} \left\{ I \left[\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta}_{n\varphi})} < \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right] - I \left[\frac{X_i}{\widehat{v}_i^{1/2}(\boldsymbol{\theta}_{n\varphi})} < \frac{X_t}{\widehat{v}_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right] \right\} \\
& \quad + \frac{1}{n+1} \sum_{i=\lfloor n^k \rfloor}^{n} \left\{ I \left[\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta}_{n\varphi})} < \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right] - I \left[\frac{X_i}{\widehat{v}_i^{1/2}(\boldsymbol{\theta}_{n\varphi})} < \frac{X_t}{\widehat{v}_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right] \right\},
\end{aligned}
\tag{A.48}
$$

where the first sum is $O_P(n^{k-1})$. For the second sum, writing

$$
I \left[\frac{X_i}{\widehat{v}_i^{1/2}(\boldsymbol{\theta}_{n\varphi})} < \frac{X_t}{\widehat{v}_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right] = I \left[\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta}_{n\varphi})} \frac{v_i^{1/2}(\boldsymbol{\theta}_{n\varphi}) \widehat{v}_t^{1/2}(\boldsymbol{\theta}_{n\varphi})}{\widehat{v}_i^{1/2}(\boldsymbol{\theta}_{n\varphi}) v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} < \frac{X_t}{v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} \right],
$$

the modulus of (A.48) is bounded above by

$$
\sup_{\substack{x \in \mathbb{R} \\ \|\boldsymbol{b}\| < b}} \frac{1}{n+1} \sum_{i=\lfloor n^k \rfloor}^{n} \left| I \left[\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta}_{n\varphi})} < x \right] - I \left[\frac{X_i}{v_i^{1/2}(\boldsymbol{\theta}_{n\varphi})} \frac{v_i^{1/2}(\boldsymbol{\theta}_{n\varphi}) \widehat{v}_t^{1/2}(\boldsymbol{\theta}_{n\varphi})}{\widehat{v}_i^{1/2}(\boldsymbol{\theta}_{n\varphi}) v_t^{1/2}(\boldsymbol{\theta}_{n\varphi})} < x \right] \right|.
$$

Using $|I(A) - I(B)| \le I(A \cap B^c) + I(A^c \cap B)$, this is bounded above by

$$
\sup_{\substack{x \in \mathbb{R} \\ \boldsymbol{\theta} \in \boldsymbol{\Theta}}} \frac{1}{n+1} \sum_{i=\lfloor n^k \rfloor}^{n} I(\mathcal{A}_{i,x,\boldsymbol{\theta}}),
$$

where the set $\mathcal{A}_{i,x,\boldsymbol{\theta}}$ is defined as

$$\mathcal{A}_{i,x,\boldsymbol{\theta}} := \left\{ \frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})} < x, \frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})} \frac{v_i^{1/2}(\boldsymbol{\theta})\widehat{v}_t^{1/2}(\boldsymbol{\theta})}{\widehat{v}_i^{1/2}(\boldsymbol{\theta})v_t^{1/2}(\boldsymbol{\theta})} \geq x \right\}$$
$$\cup \left\{ \frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})} \geq x, \frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})} \frac{v_i^{1/2}(\boldsymbol{\theta})\widehat{v}_t^{1/2}(\boldsymbol{\theta})}{\widehat{v}_i^{1/2}(\boldsymbol{\theta})v_t^{1/2}(\boldsymbol{\theta})} < x \right\}.$$

Therefore, it suffices to prove that $\sum_{i=\lfloor n^k \rfloor}^{n} I(\mathcal{A}_{i,x,\boldsymbol{\theta}}) = o_{\mathrm{P}}(1)$ uniformly with respect to both x and $\boldsymbol{\theta}$. We show this with sets containing $\mathcal{A}_{i,x,\boldsymbol{\theta}}$.

Recall that using (A.2), $\widehat{v}_t(\boldsymbol{\theta}) \geq c_0(\boldsymbol{\theta}) > c$ for a positive constant c and so

$$0 < v_t^{1/2}(\boldsymbol{\theta}) - \widehat{v}_t^{1/2}(\boldsymbol{\theta}) \leq \frac{\rho^t Z_0}{v_t^{1/2}(\boldsymbol{\theta}) + \widehat{v}_t^{1/2}(\boldsymbol{\theta})} \leq \frac{\rho^t Z_0}{2c_0^{1/2}(\boldsymbol{\theta})}.$$

Now using the triangular inequality,

$$\left| \frac{v_i^{1/2}(\boldsymbol{\theta})\widehat{v}_t^{1/2}(\boldsymbol{\theta})}{\widehat{v}_i^{1/2}(\boldsymbol{\theta})v_t^{1/2}(\boldsymbol{\theta})} - 1 \right|$$
$$\leq \left| \frac{\widehat{v}_t^{1/2}(\boldsymbol{\theta})\left(v_i^{1/2}(\boldsymbol{\theta}) - \widehat{v}_i^{1/2}(\boldsymbol{\theta})\right)}{\widehat{v}_i^{1/2}(\boldsymbol{\theta})v_t^{1/2}(\boldsymbol{\theta})} \right| + \left| \frac{\widehat{v}_i^{1/2}(\boldsymbol{\theta})\left(\widehat{v}_t^{1/2}(\boldsymbol{\theta}) - v_t^{1/2}(\boldsymbol{\theta})\right)}{\widehat{v}_i^{1/2}(\boldsymbol{\theta})v_t^{1/2}(\boldsymbol{\theta})} \right|. \tag{A.49}$$

Therefore (A.49) is bounded above by

$$\frac{\rho^i Z_0}{2c_0^{1/2}(\boldsymbol{\theta})} + \frac{\rho^t Z_0}{2c_0^{1/2}(\boldsymbol{\theta})}.$$

In view of (A.49), we get

$$\left| \frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})} \frac{v_i^{1/2}(\boldsymbol{\theta})\widehat{v}_t^{1/2}(\boldsymbol{\theta})}{\widehat{v}_i^{1/2}(\boldsymbol{\theta})v_t^{1/2}(\boldsymbol{\theta})} - \frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})} \right| \leq (\rho^i + \rho^t)Z_4 \left| \frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})} \right|,$$

where $Z_4 = Z_0/(2C^{1/2})$.

Therefore, $\mathcal{A}_{i,x,\boldsymbol{\theta}}$ is a subset of

$$
\begin{aligned}
\mathcal{B}_{i,x,\boldsymbol{\theta}} := {} & \left\{ \frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})} < x, \; \frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})} + (\rho^i + \rho^t) Z_4 \left| \frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})} \right| \geq x \right\} \\
& \cup \left\{ \frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})} \geq x, \; \frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})} - (\rho^i + \rho^t) Z_4 \left| \frac{X_i}{v_i^{1/2}(\boldsymbol{\theta})} \right| < x \right\} \\
= {} & \left\{ \eta_i < x \frac{v_i^{1/2}(\boldsymbol{\theta})}{v_i^{1/2}(\boldsymbol{\theta}_{0\varphi})}, \; \eta_i + (\rho^i + \rho^t) Z_4 |\eta_i| \geq x \frac{v_i^{1/2}(\boldsymbol{\theta})}{v_i^{1/2}(\boldsymbol{\theta}_{0\varphi})} \right\} \\
& \cup \left\{ \eta_i \geq x \frac{v_i^{1/2}(\boldsymbol{\theta})}{v_i^{1/2}(\boldsymbol{\theta}_{0\varphi})}, \; \eta_i - (\rho^i + \rho^t) Z_4 |\eta_i| < x \frac{v_i^{1/2}(\boldsymbol{\theta})}{v_i^{1/2}(\boldsymbol{\theta}_{0\varphi})} \right\} \\
= {} & \left\{ x \frac{v_i^{1/2}(\boldsymbol{\theta})}{v_i^{1/2}(\boldsymbol{\theta}_{0\varphi})} - (\rho^i + \rho^t) Z_4 |\eta_i| \leq \eta_i < x \frac{v_i^{1/2}(\boldsymbol{\theta})}{v_i^{1/2}(\boldsymbol{\theta}_{0\varphi})} \right\} \\
& \cup \left\{ x \frac{v_i^{1/2}(\boldsymbol{\theta})}{v_i^{1/2}(\boldsymbol{\theta}_{0\varphi})} \leq \eta_i < x \frac{v_i^{1/2}(\boldsymbol{\theta})}{v_i^{1/2}(\boldsymbol{\theta}_{0\varphi})} + (\rho^i + \rho^t) Z_4 |\eta_i| \right\}.
\end{aligned}
$$

Consider r.v.s X and $L \geq 0$ with X independent of η. Then $\mathrm{P}[X < \eta < X + L] \leq \sup_{y \in \mathbb{R}} \{g(y)\} \mathrm{E}(L)$ where the p.d.f. g is bounded. Consequently,

$$
\sup_{\substack{x \in \mathbb{R} \\ \boldsymbol{\theta} \in \boldsymbol{\Theta}}} \mathrm{E} \left[\sum_{i=\lfloor n^k \rfloor}^n I(\mathcal{A}_{i,x,\boldsymbol{\theta}}) \right] \leq \sup_{y \in \mathbb{R}} g(y) \sum_{i=\lfloor n^k \rfloor}^n \mathrm{E} \left\{ (\rho^i + \rho^t) Z_4 |\eta_i| \right\}.
$$

Notice that since $\lfloor n^k \rfloor \leq t \leq n$,

$$
\sum_{i=\lfloor n^k \rfloor}^n \mathrm{E} \left\{ \rho^t Z_4 |\eta_i| \right\} \leq n \rho^{\lfloor n^k \rfloor} \mathrm{E}(Z_4) \mathrm{E}|\eta| = o(1)
$$

due to $0 < \rho < 1$ and $\mathrm{E}|\eta| < \infty$. Hence, $\sum_{i=\lfloor n^k \rfloor}^n I(\mathcal{A}_{i,x,\boldsymbol{\theta}})$ converges in mean to zero uniformly with respect to both x and $\boldsymbol{\theta}$, which entails $\sum_{i=\lfloor n^k \rfloor}^n I(\mathcal{A}_{i,x,\boldsymbol{\theta}}) = o_\mathrm{P}(1)$ uniformly. $\quad\square$

With all the results above, we can easily prove Proposition 2.1 and the asymptotic result for the R-estimator as follows.

Proof of Proposition 2.1.

Proof. Combining (A.34), (A.36), (A.32) and (A.40), we get

$$
\widehat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{n\varphi}) - \boldsymbol{M}(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b} - \boldsymbol{Q}_n(\boldsymbol{\theta}_{0\varphi}) - \boldsymbol{N}_n(\boldsymbol{\theta}_{0\varphi}) - \boldsymbol{J}(\boldsymbol{\theta}_{0\varphi})\boldsymbol{b}/2 = u_\mathrm{P}(1), \qquad (\mathrm{A}.50)
$$

which, by letting $\boldsymbol{b} = \boldsymbol{0}$, entails

$$
\widehat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{0\varphi}) = \boldsymbol{Q}_n(\boldsymbol{\theta}_{0\varphi}) + \boldsymbol{N}_n(\boldsymbol{\theta}_{0\varphi}) + u_\mathrm{P}(1).
$$

Hence, (2.7) follows by recalling that $\boldsymbol{M}(\boldsymbol{\theta}_{0\varphi}) = \boldsymbol{J}(\boldsymbol{\theta}_{0\varphi})\rho(\varphi)/2$.

The proof of (2.8) follows directly from (A.35) and (A.38).

$\quad\square$

Proof of Theorem 2.1.

Proof. From the definition of $\widehat{\boldsymbol{\theta}}_{n\varphi}$ in (2.9), (2.7) and (2.8) in Proposition 2.1, consistency of $\widehat{\boldsymbol{\Upsilon}}_n$ and the asymptotic discreteness of $\bar{\boldsymbol{\theta}}_n$ (which allows us to treat $n^{1/2}(\bar{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta})$ as if it were a bounded constant: see Lemma 4.4 in Kreiss (1987)), we have

$$
\begin{aligned}
&n^{1/2}(\widehat{\boldsymbol{\theta}}_{n\varphi} - \boldsymbol{\theta}_{0\varphi}) \\
=&n^{1/2}\left\{\bar{\boldsymbol{\theta}}_n - n^{-1/2}\left(\widehat{\boldsymbol{\Upsilon}}_n\right)^{-1}\widehat{\boldsymbol{R}}_n(\bar{\boldsymbol{\theta}}_n) - \boldsymbol{\theta}_{0\varphi}\right\} \\
=&n^{1/2}\left\{\bar{\boldsymbol{\theta}}_n - n^{-1/2}\left(\widehat{\boldsymbol{\Upsilon}}_n\right)^{-1}\left[\widehat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{0\varphi}) + (1/2 + \rho(\varphi)/2)\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi})n^{1/2}(\bar{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0\varphi})\right] - \boldsymbol{\theta}_{0\varphi}\right\} + o_{\mathrm{P}}(1) \\
=&n^{1/2}\left\{\bar{\boldsymbol{\theta}}_n - n^{-1/2}(1/2 + \rho(\varphi)/2)^{-1}(\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi}))^{-1}\widehat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{0\varphi}) - (\bar{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0\varphi}) - \boldsymbol{\theta}_{0\varphi}\right\} + o_{\mathrm{P}}(1) \\
=& -(1/2 + \rho(\varphi)/2)^{-1}(\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi}))^{-1}\widehat{\boldsymbol{R}}_n(\boldsymbol{\theta}_{0\varphi}) + o_{\mathrm{P}}(1).
\end{aligned}
$$

In view of (2.8), we have

$$
n^{1/2}(\widehat{\boldsymbol{\theta}}_{n\varphi} - \boldsymbol{\theta}_{0\varphi}) = -(1/2 + \rho(\varphi)/2)^{-1}(\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi}))^{-1}(\boldsymbol{Q}_n(\boldsymbol{\theta}_{0\varphi}) + \boldsymbol{N}_n(\boldsymbol{\theta}_{0\varphi})) + o_{\mathrm{P}}(1).
$$

Now, it remains to obtain the asymptotic covariance matrix of $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{n\varphi} - \boldsymbol{\theta}_{0\varphi}\right)$. Recall (A.35) and (A.38). Since the asymptotic covariance matrices of $\boldsymbol{Q}_n(\boldsymbol{\theta}_{0\varphi})$ and $\boldsymbol{N}_n(\boldsymbol{\theta}_{0\varphi})$ have been derived, it remains to obtain the covariance matrix $\mathrm{Cov}(\boldsymbol{Q}_n(\boldsymbol{\theta}_{0\varphi}), \boldsymbol{N}_n(\boldsymbol{\theta}_{0\varphi}))$. Note that $\mathrm{E}[\boldsymbol{Q}_n(\boldsymbol{\theta}_{0\varphi})\boldsymbol{N}'_n(\boldsymbol{\theta}_{0\varphi})]$ equals

$$
\begin{aligned}
\mathrm{E}\Bigg\{&\left[\int_0^1 n^{-1/2}\sum_{t=1}^n \frac{\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})}\left[\mu(G^{-1}(u)) - \mu(\tilde{G}_n^{-1}(u))\right]d\varphi(u)\right] \\
&\times \left[n^{-1/2}\sum_{t=1}^n \frac{\dot{\boldsymbol{v}}'_t(\boldsymbol{\theta}_{0\varphi})}{v_t(\boldsymbol{\theta}_{0\varphi})}\left[1 - \eta_t\varphi\left[G(\eta_t)\right]\right]\right]\Bigg\}.
\end{aligned} \tag{A.51}
$$

Using (A.39) and $n^{-1}\sum_{t=1}^n \dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi})/v_t(\boldsymbol{\theta}_{0\varphi}) \to \mathrm{E}(\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})/v_1(\boldsymbol{\theta}_{0\varphi}))$, as $n \to \infty$, (A.51) has the same limit as

$$
\begin{aligned}
&\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right) \\
&\times \lim_{n\to\infty}\mathrm{E}\left\{\int_0^1 G^{-1}(u)\left\{\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\left[I\{\eta_i \le G^{-1}(u)\} - u\right]\frac{\dot{\boldsymbol{v}}'_j(\boldsymbol{\theta}_{0\varphi})}{v_j(\boldsymbol{\theta}_{0\varphi})}\left[1 - \eta_j\varphi\left[G(\eta_j)\right]\right]\right\}d\varphi(u)\right\} \\
=&\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right) \\
&\times \lim_{n\to\infty}\mathrm{E}\left\{\int_0^1 G^{-1}(u)\left\{\frac{1}{n}\sum_{i=1}^n\left[I\{\eta_i \le G^{-1}(u)\} - u\right]\frac{\dot{\boldsymbol{v}}'_i(\boldsymbol{\theta}_{0\varphi})}{v_i(\boldsymbol{\theta}_{0\varphi})}\left[1 - \eta_i\varphi\left[G(\eta_i)\right]\right]\right\}d\varphi(u)\right\} \\
=&\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}'_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\int_0^1 G^{-1}(u)\mathrm{E}\left\{\left[I\{\eta_1 \le G^{-1}(u)\} - u\right]\left[1 - \eta_1\varphi\left[G(\eta_1)\right]\right]\right\}d\varphi(u) \\
=&\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}'_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\int_0^1 G^{-1}(u)\mathrm{E}\left\{I\{\eta_1 \le G^{-1}(u)\}\left[1 - \eta_1\varphi\left[G(\eta_1)\right]\right]\right\}d\varphi(u),
\end{aligned}
$$

where the first equality is due to independence of η_i and η_j for $i \neq j$, independence of v_j and η_j, and Assumption 2.1. The second equality is due to independence of v_i and η_i. The last equality is due to Assumption 2.1.

Recall the definition of $\lambda(\varphi)$ in (2.6), which can also be written as

$$\lambda(\varphi) = \int_0^1 G^{-1}(u)\mathrm{E}\left\{I\{\eta_1 \leq G^{-1}(u)\}\left[1 - \eta_1\varphi\left[G(\eta_1)\right]\right]\right\} d\varphi(u).$$

We then have

$$\lim_{n\to\infty} \mathrm{Cov}(\boldsymbol{Q}_n(\boldsymbol{\theta}_{0\varphi}), \boldsymbol{N}_n(\boldsymbol{\theta}_{0\varphi})) = \mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\lambda(\varphi).$$

Hence, by recalling (A.38) and in view of (2.10), the asymptotic covariance matrix of $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{n\varphi} - \boldsymbol{\theta}_{0\varphi}\right)$ is

$$(\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi}))^{-1}\frac{\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\mathrm{Var}(Z) + \boldsymbol{J}(\boldsymbol{\theta}_{0\varphi})\sigma^2(\varphi) + 2\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\lambda(\varphi)}{(1/2 + \rho(\varphi)/2)^2}(\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi}))^{-1}$$

$$= (\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi}))^{-1}\frac{[4\gamma(\varphi) + 8\lambda(\varphi)]\,\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right)\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_{0\varphi})}{v_1(\boldsymbol{\theta}_{0\varphi})}\right) + 4\sigma^2(\varphi)\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi})}{(1 + \rho(\varphi))^2}(\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi}))^{-1}.$$

$\square$

Proof of Theorem 2.2.

Proof. From (2.11),

$$\hat{c}_{n\varphi} = \left(1 - \sum_{j=1}^q \hat{\beta}_{nj}\right)^{-1}\left(\hat{\omega}_{n\varphi}/\overline{X_n^2} + \sum_{i=1}^p \hat{\alpha}_{n\varphi i}\right) = \left(1 - \sum_{j=1}^q \beta_{0j}\right)^{-1}\left(c_\varphi\omega_0/\mathrm{E}(X_t^2) + c_\varphi\sum_{i=1}^p \alpha_{0i}\right) + o_{\mathrm{P}}(1).$$

Substituting $\mathrm{E}(X_t^2) = \omega_0/\{1 - \sum_{i=1}^p \alpha_{0i} - \sum_{j=1}^q \beta_{0j}\}$, this is equal to c_φ.

To prove (2.12), we define a $(1 + p + q) \times (1 + p + q)$ diagonal matrix

$$\boldsymbol{A}_\varphi := \mathrm{diag}(\underbrace{c_\varphi^{-1}, ..., c_\varphi^{-1}}_{1+p}, \underbrace{1, ..., 1}_{q}).$$

Then we have $\boldsymbol{\theta}_0 = \boldsymbol{A}_\varphi\boldsymbol{\theta}_{0\varphi}$ and $\boldsymbol{\theta}_n = \boldsymbol{A}_\varphi\boldsymbol{\theta}_{n\varphi}$. Using the forms of $\{c_j(\boldsymbol{\theta}); j \geq 0\}$ in (3.1) of Berkes et al. (2003), we have

$$v_t(\boldsymbol{\theta}_{0\varphi}) = c_\varphi v_t(\boldsymbol{\theta}_0) \quad \text{and} \quad \dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_{0\varphi}) = c_\varphi\boldsymbol{A}_\varphi\dot{\boldsymbol{v}}_t(\boldsymbol{\theta}_0)$$

which entail

$$\boldsymbol{J}(\boldsymbol{\theta}_{0\varphi}) = \boldsymbol{A}_\varphi\boldsymbol{J}(\boldsymbol{\theta}_0)\boldsymbol{A}_\varphi, \quad \boldsymbol{Q}_n(\boldsymbol{\theta}_{0\varphi}) = \boldsymbol{A}_\varphi\boldsymbol{Q}_n(\boldsymbol{\theta}_0), \quad \boldsymbol{N}_n(\boldsymbol{\theta}_{0\varphi}) = \boldsymbol{A}_\varphi\boldsymbol{N}_n(\boldsymbol{\theta}_0).$$

Hence (2.12) follows from (2.10) and consistency of $\hat{c}_{n\varphi}$.

$\square$

B. ASYMPTOTIC RELATIVE EFFICIENCY

Note that under assumption $\mathrm{E}\epsilon^4 < \infty$, the QMLE $\widehat{\boldsymbol{\theta}}_{\mathrm{QMLE}}$ is asymptotic normal with mean zero and covariance matrix $\boldsymbol{\Omega}_{\mathrm{QMLE}} = (\mathrm{E}\epsilon^4 - 1)(\boldsymbol{J}(\boldsymbol{\theta}_0))^{-1}$. Hence, in view of The-

orem 2.2, the ARE of the R-estimator wrt the QMLE is

$$\boldsymbol{\Omega}_\varphi^{-1}\boldsymbol{\Omega}_{\mathrm{QMLE}} = \boldsymbol{J}(\boldsymbol{\theta}_0)\left\{[4\gamma(\varphi) + 8\lambda(\varphi)]\,\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1(\boldsymbol{\theta}_0)}{v_1(\boldsymbol{\theta}_0)}\right)\mathrm{E}\left(\frac{\dot{\boldsymbol{v}}_1'(\boldsymbol{\theta}_0)}{v_1(\boldsymbol{\theta}_0)}\right) + 4\sigma^2(\varphi)\boldsymbol{J}(\boldsymbol{\theta}_0)\right\}^{-1}$$
$$\times (1 + \rho(\varphi))^2(\mathrm{E}\epsilon^4 - 1). \tag{B.1}$$

For the sign R-estimator, $\gamma(\varphi)$, $\lambda(\varphi)$ and $\rho(\varphi)$ are all zeros. Hence $\boldsymbol{\Omega}_\varphi^{-1}\boldsymbol{\Omega}_{\mathrm{QMLE}}$ reduces to

$$(\mathrm{E}\epsilon^4 - 1)/(4\sigma^2(\varphi))\boldsymbol{I}_{1+p+q},$$

where $\boldsymbol{I}_{1+p+q}$ is the $(1 + p + q) \times (1 + p + q)$ identity matrix. Consequently, the ARE of the sign R-estimator wrt the QMLE equals $(\mathrm{E}\epsilon^4 - 1)/(4\sigma^2(\varphi))$, which is 0.876 under the normal distribution. This corresponds to the classical result of the ARE of the mean absolute deviation wrt the mean square deviation; see, e.g., Huber and Ronchetti (2011, Chapter 1).

For the vdW and Wilcoxon R-estimators, the AREs are more difficult to calculate since $\gamma(\varphi)$ and $\lambda(\varphi)$ are non-zero. However, in the simulation study in Table 2, the estimated AREs reveal that the vdW R-estimator, compared with the QMLE, does not lose any efficiency which is a reflection of the well-known Chernoff-Savage phenomenon in the literature on the R-estimation in linear models.

C. SUPPLEMENTARY RESULTS FOR NUMERICAL ANALYSIS

Table C.1. The estimates of the bias, MSE and ARE (in parenthesis) of the R-estimators and the QMLE for the GARCH $(1,1)$ model under the $t(3)$ error distribution with larger sample sizes $n = 3000, 5000$.

	Bias			**MSE and ARE**		
	ω ($\times 10^7$)	α ($\times 10^3$)	β ($\times 10^3$)	ω ($\times 10^{12}$)	α ($\times 10^3$)	β ($\times 10^3$)
n = 3000						
QMLE	7.17	18.24	-30.56	19.17	20.49	16.31
Sign	-1.41	-9.80	-9.09	2.48 (7.74)	1.30 (15.70)	2.17 (7.50)
Wilcoxon	-1.26	-9.78	-9.41	2.62 (7.33)	1.33 (15.38)	2.29 (7.12)
vdW	-0.35	-8.18	-12.91	3.73 (5.15)	1.76 (11.66)	3.38 (4.82)
n = 5000						
QMLE	12.97	79.27	-77.61	12.72	36.91	19.81
Sign	4.87	6.01	-32.71	0.67 (18.90)	0.40 (91.15)	1.39 (14.29)
Wilcoxon	5.33	4.91	-32.34	0.84 (15.16)	0.37 (99.04)	1.40 (14.15)
vdW	5.32	3.43	-31.32	1.81 (7.02)	0.45 (81.28)	2.12 (9.35)

Note: The sign, Wilcoxon and vdW scores are used for the R-estimators. Bias and MSE are rescaled. Simulation results are based on $R = 500$ replications.

Figure C.1. QQ-plots of the residuals against t-distributions for the EFCX (left column) and S&P 500 data (right column).
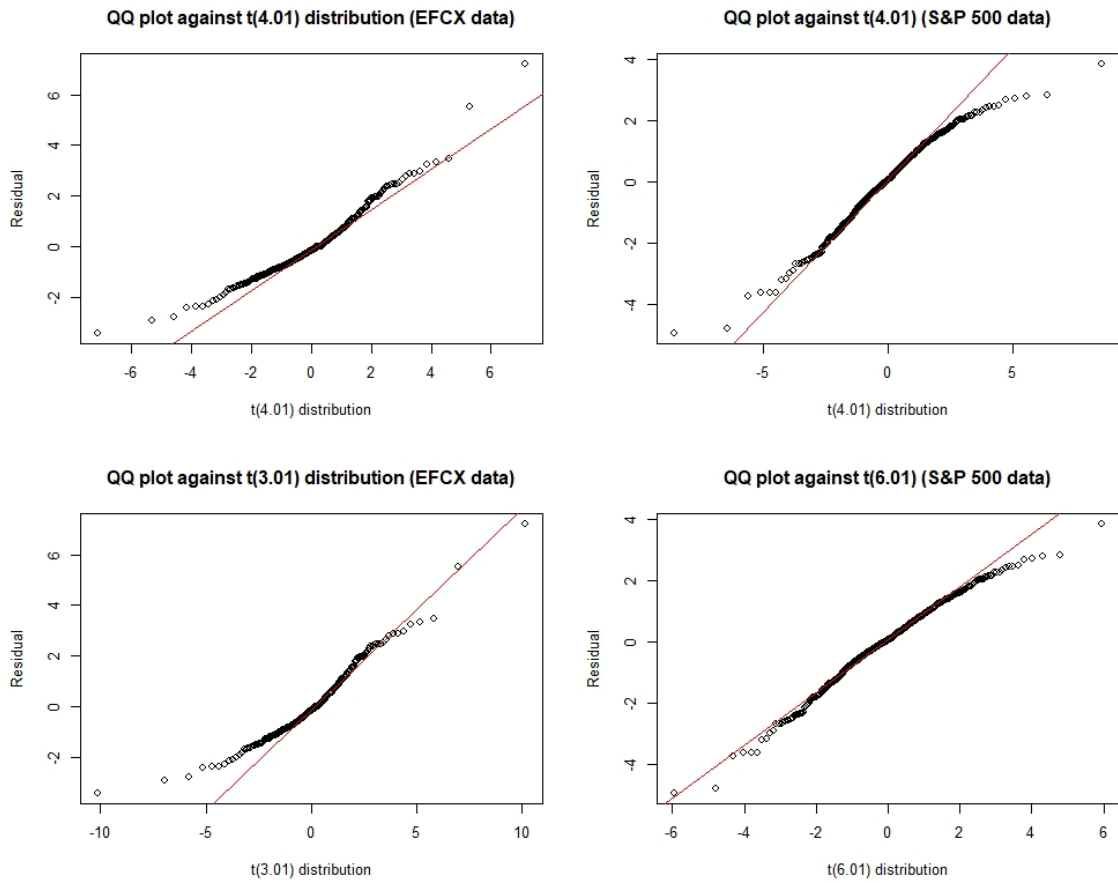
Note: The residuals are obtained by using the vdW R-estimator.