Generalized R-estimators under Conditional Heteroscedasticity

Kanchan Mukherjee

The University of Liverpool

Email: k.mukherjee@liverpool.ac.uk

Abstract

In this paper, we extend the classical idea of Rank-estimation of parameters from homoscedastic problems to heteroscedastic problems. In particular, we define a class of rank estimators of the parameters associated with the conditional mean function of an autoregressive model through a three-steps procedure and then derive their asymptotic distributions. The class of models considered includes Engel's ARCH model and the threshold heteroscedastic model. The class of estimators includes an extension of Wilcoxon-type rank estimator. The derivation of the asymptotic distributions depends on the uniform approximation of a randomly weighted empirical process by a perturbed empirical process through a very general weight-dependent partitioning argument.

Keywords: Rank estimation; heteroscedastic model; weighted empirical process; uniform approximation.

JEL Classifications: C14, C22.

Kanchan Mukherjee
Department of Mathematical Sciences
The University of Liverpool
Liverpool, L69 7ZL, UK
E-mail: k.mukherjee@liverpool.ac.uk

June 8, 2006

1 Introduction

Since the introduction of the autoregressive conditional heteroscedastic (ARCH) time series model of Engle (1982), there have been huge developments on the theory and application of this model and its various generalizations to economics and finance. ARCH models have been used to represent the volatility, i.e, the strong dependence of the instantaneous variability of a time series on its own past, in numerous economic and financial data sets. For a literature review, see Bollerslev, Chou, and Kroner (1992), Shephard (1996), and Gouriéroux (1997), among others. Most of the existing methodological literature have focused on developing estimation procedures for the parameters associated with the conditional variability using pseudo-likelihood methods. However, development of the estimation methods associated with the conditional mean component of a heteroscedastic problem is also important from the application point of view and this has been largely overlooked. In this paper, we aim to fill that gap by developing a rank-based robust procedure for estimating the mean parameter of an autoregressive model with conditional heteroscedastic errors.

In a parametric formulation, linearity of regression, independence and normality of errors, homoscedasticity or form of heteroscedasticity etc. are typically assumed for drawing conclusions about parameters of interest. However, there is no guarantee that such regularity assumptions will be valid in a given situation and therefore it is natural to investigate alternative procedures that can perform well under probable departures from model assumptions. Among different types of such robust procedures, estimators based on ranks or the so-called R-estimators are sometimes preferable to their other competitors for their global robustness property as they generally demand much less restrictive assumptions on the underlying distributions; see, for example, Jurečková and Sen (1996, Section 3.4) for a discussion on this. The need for using such robust estimators is even more for financial data due to the empirical finding that 'outliers' appear more often in asset returns than that implied by white noises having normal distribution. For more on this, see Tsay (2002, Section 3.3) and Engle and Gonzalez-Rivera (1991) who quantified the loss of efficiency resulting from the use of estimators arising from the first-order conditions for the normal MLE (called the quasi maximum likelihood estimator or the QMLE) on non-normal distributions and concluded that 'it is worthwhile searching for estimators that can improve on QMLE'.

There is a vast literature on the R-estimation of parameters in homoscedastic regression and autoregression models. For a glimpse, see Koul (1992, Section 4.4), Jurečková and Sen (1996, Section 3.4, Chapter 6) and Hájek, Šidák and Sen (1999, Section 10.3), among others. In linear regression model with i.i.d. or homoscedastic long memory errors, R-estimators are known to have highly desirable efficiency; see, e.g., Jurečková (1971), Koul (1971), Jaeckel (1972) and Koul and Mukherjee (1993). In the homoscedastic autoregressive time series model (1.1) with $\sigma \equiv 1$, analogs of the R-estimators are known to have similar efficiency and robustness properties as investigated by Koul and Ossiander (1994) and Mukherjee and Bai (2002). It is thus natural to investigate their behavior in the heteroscedastic set up.

Accordingly, consider the following autoregressive model with heteroscedastic error where for known integers s, p and r, $\{X_i,\ 1-s \le i \le n\}$ is an observable time series. Set $\boldsymbol{W}_{i-1} := (X_{i-1}, X_{i-2}, \ldots, X_{i-s})'$ and $\boldsymbol{Y}_{i-1} = c(\boldsymbol{W}_{i-1})$, $1 \le i \le n$, where $c : \mathbb{R}^s \to \mathbb{R}^p$ is a known function. Let Ω_j, $j = 1, 2$, be open subsets of $\mathbb{R}^p$, $\mathbb{R}^r$, respectively with $\Omega := \Omega_1 \times \Omega_2 \subset \mathbb{R}^m$, where $m = p + r$. Let σ be a known function from $\mathbb{R}^p \times \Omega_2$ to $\mathbb{R}^+ := (0, \infty)$, differentiable in its second argument. Consider the model

$$X_i = \boldsymbol{Y}_{i-1}'\boldsymbol{\alpha} + \sigma(\boldsymbol{Y}_{i-1}, \boldsymbol{\beta})\, \eta_i, \qquad 1 \le i \le n, \tag{1.1}$$

where $\boldsymbol{\alpha} \in \Omega_1$, $\boldsymbol{\beta} \in \Omega_2$ are the unknown parameters, and the unobservable errors $\{\eta_i,\ i \ge 1\}$ are i.i.d. with zero mean and finite variance having a distribution function (d.f.) G and probability density function (p.d.f.) g. Throughout, we also assume that $\{\eta_i,\ i \ge 1\}$ are independent of $\boldsymbol{W}_0 := (X_0, X_{-1}, \ldots, X_{1-s})'$ and hence independent of $\boldsymbol{Y}_0$; for each $y \in \mathbb{R}^p$,

$\dot{\sigma}(y, \boldsymbol{t})$ is the derivative of $\sigma(y, \boldsymbol{t})$ with respect to $\boldsymbol{t}$; and $\{X_i\}$ is strictly stationary and ergodic. All of these assumptions will be referred to as the model assumptions in the sequel. Although some sufficient conditions for the stationarity and ergodicity of $\{X_i\}$ in the full generality of the model (1.1) may not be possible at this stage, we discuss the relevant sufficient conditions for particular examples cited below. Our goal here is to develop the asymptotics of the R-estimators of the parameter $\boldsymbol{\alpha}$ in addition to the estimation of the entire parameter vector $\boldsymbol{\theta} := (\boldsymbol{\alpha}', \boldsymbol{\beta}')'$ based on the data $\boldsymbol{W}_0, X_1, X_2, \ldots, X_n$.

Note in this connection that model (1.1) is not the 'pure ARCH' model since the conditional variance depends on a lag of the observed dependent variable, rather than a lag of the error term. In the following, we cite some examples of (1.1).

Example 1. (Engle's ARCH model). In the ARCH model introduced by Engle (1982), one observes $\{Z_i, 1 - s \leq i \leq n\}$ such that

$$Z_i = (\alpha_0 + \alpha_1 Z_{i-1}^2 + \ldots + \alpha_s Z_{i-s}^2)^{1/2} \epsilon_i, \ 1 \leq i \leq n, \tag{1.2}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_s)' \in I\!R^{+(s+1)} := (0, \infty)^{(s+1)}$ is the unknown parameter and $\{\epsilon_i; 1 \leq i \leq n\}$ are unobservable i.i.d. with mean zero, variance 1 and finite fourth moment.

Squaring both sides of (1.2) and writing $\eta_i := \varepsilon_i^2 - 1$, $X_i = Z_i^2$, $\boldsymbol{W}_{i-1} = [X_{i-1}, \ldots, X_{i-s}]' = [Z_{i-1}^2, \ldots, Z_{i-s}^2]'$, and $\boldsymbol{Y}_{i-1}' = [1, \boldsymbol{W}_{i-1}']$, model (1.2) can be recast as

$$X_i = \boldsymbol{Y}_{i-1}'\boldsymbol{\alpha} + (\boldsymbol{Y}_{i-1}'\boldsymbol{\alpha})\, \eta_i, \qquad 1 \leq i \leq n. \tag{1.3}$$

This is an example of the model (1.1) with $\boldsymbol{\alpha} = \boldsymbol{\beta}$, $c(\boldsymbol{w}) = [1, \boldsymbol{w}]'$, $\boldsymbol{w} \in [0, \infty)^s$, $p = s + 1$, $r = s + 1$, and $\sigma(\boldsymbol{y}, \boldsymbol{t}) = \boldsymbol{t}'\boldsymbol{y}$. For various sufficient conditions related to the strict stationarity and ergodicity of the process $\{Z_i; 1 - s \leq i\}$, see Nelson (1990), Bougerol and Picard (1992) and Giraitis, Kokoszka and Lepius (2000).

Example 2. (Autoregressive Linear Square Conditional Heteroscedastic model) (ARLSCH). Consider the first order autoregressive model with heteroscedastic errors where one observes $\{X_i; 0 \le i \le n\}$ such that the conditional variance of the i-th observation X_i depends linearly on the squares of past as follows:

$$X_i = \alpha X_{i-1} + \{\beta_0 + \beta_1 X_{i-1}^2\}^{1/2}\, \eta_i, \qquad 1 \le i \le n, \tag{1.4}$$

where $\alpha \in I\!R$, $\boldsymbol{\beta} = (\beta_0, \beta_1)' \in (0, \infty)^2$ and $\{\eta_i\}$'s are i.i.d. with zero mean and unit variance. With the identification $s = 1 = p$, $c(w) = w$, $r = 2$, and

$$\sigma(y, \boldsymbol{t}) = (t_0 + t_1 y^2)^{1/2}, \;\; y \in I\!R,$$

model (1.4) can be seen as an example of (1.1).

The assumption needed on the parameters under which the process $\{X_i; i \ge 0\}$ of (1.4) is strictly stationary and ergodic is as follows:

$$|\alpha| + E|\eta_1| \max\{\beta_0^{1/2}, \beta_1^{1/2}\} < 1. \tag{1.5}$$

This follows by using Lemma 3.1 of Härdle and Tsybakov (1997, p 227) with $C_1 = |\alpha|$ and $C_2 = \max\{\beta_0^{1/2}, \beta_1^{1/2}\} = \sup\{(\beta_0 + \beta_1 x^2)^{1/2}/(1 + |x|); x \in I\!R\}$.

Example 3. (Autoregressive Threshold Conditional Heteroscedastic model) (ARTCH). Consider an s-th order autoregressive model with self exciting threshold heteroscedastic errors where the conditional standard deviation of the i-th observation X_i is piecewise linear on the past as follows:

$$\begin{aligned} X_i &= (\alpha_1 X_{i-1} + \ldots \alpha_s X_{i-p}) + \Big\{\beta_1 X_{i-1} I(X_{i-1} > 0) - \beta_2 X_{i-1} I(X_{i-1} \le 0) \\ &\qquad + \ldots + \beta_{2s-1} X_{i-s} I(X_{i-s} > 0) - \beta_{2s} X_{i-s} I(X_{i-s} \le 0)\Big\}\, \eta_i, \qquad 1 \le i \le n, \end{aligned}$$

where all β_j's are positive and $\{\eta_i\}$'s are i.i.d. with zero mean and unit variance. For applications and many probabilistic properties of this model including conditions on the stationarity and ergodicity, see Rabemananjara and Zakoian (1993). For a discussion on the difficulties associated with the asymptotics of the robust estimation in this model due to the lack of differentiability caused by threshold, see Rabemananjara and Zakoian (1993, p 38).

With the identification $p = s$, $c(\boldsymbol{w}) = \boldsymbol{w}$, $r = 2p$, and

$$\sigma(\boldsymbol{y}, \boldsymbol{t}) = \sum_{j=1}^{p} t_{2j-1} y_j I(y_j \geq 0) + \sum_{j=1}^{p} t_{2j}(-y_j) I(y_j < 0), \qquad \boldsymbol{y} \in \mathbb{R}^p,\ \boldsymbol{t} \in (0, \infty)^{2p},$$

this can be seen as an example of (1.1).

Some of the important findings on R-estimation under the model (1.1) are as follows. It turns out that efficiency properties similar to homoscedastic models continue to hold for the heteroscedastic setup also; see Remark 3.3 for details. In particular, for every fixed innovation density g satisfying some conditions, optimal R-estimator based on suitable score function exists. Also, the Wilcoxon R-estimator have asymptotic relative efficiency (ARE) of at least 0.864 with respect to the quasi maximum likelihood estimator for a large class of innovation density. Our simulation results reported in Tables 1 and 2 also confirm some of these theoretical efficiency results for a variety of innovation distributions. Moreover, using three well-known real data examples, the robustness of R-estimators against misspecified form of the heteroscedasticity is exhibited.

For estimation of the conditional mean parameters using the MLE and the least squares method in an autoregressive model with errors generated by an ARCH process itself, see Pantula (1988). See also Koenker and Zhao (1996) and Koul and Mukherjee (2002) for related work on the least absolute deviation and M-estimators.

The paper is organized as follows. The class of R-estimators is defined in Section 2. Section 3 states all distributional results and compares R-estimators with least squares estimator based on their asymptotic efficiencies. In Section 4, we verify that conditions of the theorems of Section 3 are satisfied for each of the above examples. Analysis of simulated and real data are reported in Section 5. Section 6 gives detail proofs of the theoretical results of Section 3.

2 Generalized R-estimators

To define the class of R-estimators, we proceed in three steps. First we estimate $\boldsymbol{\alpha}$ in (1.1) by a preliminary consistent estimator $\widehat{\boldsymbol{\alpha}}_p$ which only considers the linear additive autoregressive structure of (1.1) but does not take into account the conditional heteroscedasticity of the model. Next, we use $\widehat{\boldsymbol{\alpha}}_p$ to construct an estimator $\widehat{\boldsymbol{\beta}}$ of the parameter $\boldsymbol{\beta}$. Finally, an estimator $\widehat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$ based on the estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is defined which does take into account the heteroscedastic structure of the model (1.1). Throughout, $\dot{u}$ will denote the derivative of a function u.

Step 1: Define $\mathcal{H}(\boldsymbol{\tau}_1) := n^{-1/2}\sum_{i=1}^n \boldsymbol{Y}_{i-1}\Big(X_i - \boldsymbol{Y}'_{i-1}\boldsymbol{\tau}_1\Big)$. Since $E[\mathcal{H}(\boldsymbol{\alpha})] = 0$, a preliminary least squares estimator of $\boldsymbol{\alpha}$ is defined as a solution of $\mathcal{H}(\boldsymbol{\tau}_1) = 0$ and is given by

$$\widehat{\boldsymbol{\alpha}}_p := [\sum_{i=1}^n \boldsymbol{Y}_{i-1}\boldsymbol{Y}'_{i-1}]^{-1}[\sum_{i=1}^n X_i\boldsymbol{Y}_{i-1}].$$

Step 2: For $\boldsymbol{\tau} := (\boldsymbol{\tau}'_1, \boldsymbol{\tau}'_2)' \in \Omega := \Omega_1 \times \Omega_2$, let $\eta_i(\boldsymbol{\tau}) := [X_i - \boldsymbol{Y}'_{i-1}\boldsymbol{\tau}_1]/\sigma(\boldsymbol{Y}_{i-1}, \boldsymbol{\tau}_2)$ denote the i-th residual, $1 \le i \le n$. Let κ be a nondecreasing right continuous functions on $I\!R$ such that $E\{\eta_1\kappa(\eta_1)\} = 1$. This is automatically satisfied, for example, when the innovations have unit variance and κ is the identity function ($\kappa(x) \equiv x$) or when it is the score function for location of the maximum likelihood estimator at the error distribution G i.e., $\kappa(x) \equiv -\dot{g}(x)/g(x)$. Consider the statistic

$$M_s(\boldsymbol{\tau}) := n^{-1/2}\sum_{i=1}^n \frac{\dot{\sigma}(\boldsymbol{Y}_{i-1}, \boldsymbol{\tau}_2)}{\sigma(\boldsymbol{Y}_{i-1}, \boldsymbol{\tau}_2)}\Big[\eta_i(\boldsymbol{\tau})\kappa(\eta_i(\boldsymbol{\tau})) - 1\Big].$$

Since $E[M_s(\boldsymbol{\alpha}, \boldsymbol{\beta})] = 0$, an estimator of the scale parameter $\boldsymbol{\beta}$ is defined by the relation

$$\widehat{\boldsymbol{\beta}} := \operatorname{argmin}\{\sum_{j=1}^r |M_{sj}(\widehat{\boldsymbol{\alpha}}_p, \boldsymbol{\tau}_2)|; \boldsymbol{\tau}_2 \in \Omega_2\},$$

where $M_{sj}(\widehat{\boldsymbol{\alpha}}_p, \boldsymbol{\tau}_2)$ is the j-th coordinate of the vector $M_s(\boldsymbol{\tau})$, $1 \le j \le r$. This definition is motivated by the discussion in Huber (1981, Ch. 7, Eqns. 7.3-7.7) pertaining to the linear regression model. The idea is to obtain estimates of the location and concomitant scale parameters by solving a simultaneous system of equations. Estimates of the scale parameters are obtained by substituting those of the location parameters.

Step 3: Finally, based on $\widehat{\boldsymbol{\beta}}$, an improved estimator of $\boldsymbol{\alpha}$ can be motivated as follows. Note that (1.1) can be written as

$$X_i/\sigma(\boldsymbol{Y}_{i-1}, \boldsymbol{\beta}) = \boldsymbol{Y}'_{i-1}\boldsymbol{\alpha}/\sigma(\boldsymbol{Y}_{i-1}, \boldsymbol{\beta}) + \eta_i.$$

This in turn can be approximated by

$$X_i/\sigma(\boldsymbol{Y}_{i-1}, \widehat{\boldsymbol{\beta}}) \approx \{\boldsymbol{Y}_{i-1}/\sigma(\boldsymbol{Y}_{i-1}, \widehat{\boldsymbol{\beta}})\}'\boldsymbol{\alpha} + \eta_i. \tag{2.1}$$

This can be thought as a linear autoregressive model with homoscedastic errors. Hence, extending Koul and Ossiander (1994), a class of R-estimators generalized to the heteroscedastic model can be defined as follows. For $1 \le i \le n$, let

$$a_i(\boldsymbol{\tau}_2) := X_i/\sigma(\boldsymbol{Y}_{i-1}, \boldsymbol{\tau}_2),$$

and

$$\boldsymbol{Z}_{i-1}(\boldsymbol{\tau}_2) := \boldsymbol{Y}_{i-1}/\sigma(\boldsymbol{Y}_{i-1}, \boldsymbol{\tau}_2).$$

Let $\varphi : [0,1] \to I\!R$ be a (score) function belonging to the class

$$\begin{aligned}\mathcal{F} \;=\;& \{\varphi; \varphi : [0,1] \to I\!R \text{ is right continuous, non-decreasing, with} \\ & \varphi(1) - \varphi(0) = 1\}.\end{aligned}$$

The function $\varphi(u) = u - 1/2$ in this class corresponds to the Wilcoxon rank score.

Define a rank statistic as

$$\boldsymbol{S}_\varphi(\boldsymbol{\tau}) = n^{-1/2}\sum_{i=1}^{n}\{\boldsymbol{Z}_{i-1}(\boldsymbol{\tau}_2) - \bar{\boldsymbol{Z}}(\boldsymbol{\tau}_2)\}\varphi\left(\frac{R_i\boldsymbol{\tau}}{n+1}\right), \quad \boldsymbol{\tau} \in \Omega,$$

where $R_{i\boldsymbol{\tau}} = \sum_{j=1}^{n} I\{a_j(\boldsymbol{\tau}_2) - \boldsymbol{\tau}_1' \boldsymbol{Z}_{j-1}(\boldsymbol{\tau}_2) \le a_i(\boldsymbol{\tau}_2) - \boldsymbol{\tau}_1' \boldsymbol{Z}_{i-1}(\boldsymbol{\tau}_2)\}$ (the $\boldsymbol{\tau}$-residual rank of the i-th residual), $1 \le i \le n$ and $\bar{\boldsymbol{Z}}(\boldsymbol{\tau}_2) = \sum_{i=1}^{n} \boldsymbol{Z}_{i-1}(\boldsymbol{\tau}_2)/n$.

Note that $R_{i\boldsymbol{\tau}}$ is also the rank of $\eta_i(\boldsymbol{\tau})$ among $\{\eta_j(\boldsymbol{\tau}); 1 \le j \le n\}$. Hence, $E[\boldsymbol{S}_\varphi(\boldsymbol{\alpha}, \boldsymbol{\beta})] = 0$ and so a generalized R-estimator of $\boldsymbol{\alpha}$ corresponding to the score function φ is defined as

$$\widehat{\boldsymbol{\alpha}} = \operatorname{argmin}\{\sum_{j=1}^{p} |S_{\varphi j}(\boldsymbol{\tau}_1, \widehat{\boldsymbol{\beta}})|; \boldsymbol{\tau}_1 \in \Omega_1\},$$

where $S_{\varphi j}(\boldsymbol{\tau})$ is the j-th coordinate of the vector $\boldsymbol{S}_\varphi(\boldsymbol{\tau})$, $1 \le j \le p$.

See Section 5 of this paper and Mukherjee (2006 b) for some algebraic expressions for R-estimators based on Wilcoxon and the sign score function for simple linear model. Although an algebraic expression for an R-estimator for more complex models may not exist in general, fast computational algorithms for ranking are available. Using the initial estimator $\widehat{\boldsymbol{\alpha}}_p$ of $\boldsymbol{\alpha}$, a Newton-Raphson type method can be used to solve this minimization problem. For more on the existence of the solution to the above minimization problem and computation in the analogous setup, see Jaeckel (1972), Huber (1981, Section 7.3) and Koul (1992, Section 7.3b). Note that this minimization problem may not always have unique solution. However, as in Jurečková (1971, Section 4) for the analogous case of linear regression models, it can be shown using the asymptotic uniform linearity result (AUL) of Lemma 3.3 that all solutions are asymptotically equivalent.

Remark 2.1 Strictly speaking, these estimators are not functions of the ranks of the $\boldsymbol{\tau}$-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, $\varphi(u) = u - \frac{1}{2}$, $\widehat{\boldsymbol{\alpha}} = \widehat{\boldsymbol{\alpha}}_\varphi$ is an analogue of the Wilcoxon type R-estimator.

3 Main results

Our first result is on the asymptotic distribution of $\widehat{\boldsymbol{\alpha}}_p$. Here and in the sequel, the expectation of a random matrix is defined as the matrix of entry-wise expectations.

Theorem 3.1 *In the model (1.1), assume that* $E[\boldsymbol{Y}_0\boldsymbol{Y}_0'\sigma(\boldsymbol{Y}_0,\boldsymbol{\beta})^2] < \infty$. *Then*

$$n^{1/2}(\widehat{\boldsymbol{\alpha}}_p - \boldsymbol{\alpha}) \Longrightarrow N\Bigg[0, E(\eta_1^2)[E(\boldsymbol{Y}_0\boldsymbol{Y}_0')]^{-1}[E(\sigma^2(\boldsymbol{Y}_0,\boldsymbol{\beta})\boldsymbol{Y}_0\boldsymbol{Y}_0')][E(\boldsymbol{Y}_0\boldsymbol{Y}_0')]^{-1}\Bigg]. \qquad (3.1)$$

For the subsequent results, we need some additional notations and assumptions. Because of (2.1), we **standardize** the mean, variance and various other quantities by $\sigma(\boldsymbol{Y}_{i-1},\boldsymbol{\beta})$. Accordingly, for $\boldsymbol{t}_1 \in I\!R^p$, $\boldsymbol{t}_2 \in I\!R^r$, $1 \le i \le n$, let

$$\mu_{ni}(\boldsymbol{t}_1) := \frac{\boldsymbol{Y}_{i-1}'(\boldsymbol{\alpha} + n^{-1/2}\boldsymbol{t}_1)}{\sigma(\boldsymbol{Y}_{i-1},\boldsymbol{\beta})}, \qquad \dot{\mu}_{ni}(\boldsymbol{t}_1) := \frac{\boldsymbol{Y}_{i-1}}{\sigma(\boldsymbol{Y}_{i-1},\boldsymbol{\beta})},$$
$$\sigma_{ni}(\boldsymbol{t}_2) := \frac{\sigma(\boldsymbol{Y}_{i-1},\boldsymbol{\beta} + n^{-1/2}\boldsymbol{t}_2)}{\sigma(\boldsymbol{Y}_{i-1},\boldsymbol{\beta})}, \qquad \dot{\sigma}_{ni}(\boldsymbol{t}_2) := \frac{\dot{\sigma}(\boldsymbol{Y}_{i-1},\boldsymbol{\beta} + n^{-1/2}\boldsymbol{t}_2)}{\sigma(\boldsymbol{Y}_{i-1},\boldsymbol{\beta})},$$
$$s_{ni}(\boldsymbol{t}_1,\boldsymbol{t}_2) = \frac{\dot{\mu}_{ni}(\boldsymbol{t}_1)}{\sigma_{ni}(\boldsymbol{t}_2)} = \frac{\boldsymbol{Y}_{i-1}}{\sigma(\boldsymbol{Y}_{i-1},\boldsymbol{\beta} + n^{-1/2}\boldsymbol{t}_2)}, \qquad r_{ni}(\boldsymbol{t}_2) := \frac{\dot{\sigma}_{ni}(\boldsymbol{t}_2)}{\sigma_{ni}(\boldsymbol{t}_2)} = \frac{\dot{\sigma}(Y_{i-1},\boldsymbol{\beta} + n^{-1/2}\boldsymbol{t}_2)}{\sigma_{ni}(Y_{i-1},\boldsymbol{\beta} + n^{-1/2}\boldsymbol{t}_2)}.$$

Note that some of the above quantities, e.g., $\dot{\mu}_{ni}(\boldsymbol{t}_1)$, are free from both $\boldsymbol{t}_1$ and n; nevertheless, we retain these arguments for consistency. In the sequel, $\dot{\mu}_i$, μ_i, $\dot{\sigma}_i$, r_i will stand for $\dot{\mu}_{ni}(0)$, $\mu_{ni}(0)$, $\dot{\sigma}_{ni}(0)$ and $r_{ni}(0)$ respectively, as they also do not depend on n. Also, the probability and expectation are taken under the model (1.1) under $\boldsymbol{\theta} := (\boldsymbol{\alpha}',\boldsymbol{\beta}')'$. We assume the existence of the following limiting matrices as a consequence of the stationarity and ergodicity, where $\rightarrow$ denotes the convergence in probability. Also condition (3.4) below is a smoothness condition related to the heteroscedasticity.

There exist positive definite matrices $M(\boldsymbol{\theta})$, $\dot{\Sigma}(\boldsymbol{\theta})$ and matrices $\boldsymbol{G}(\boldsymbol{\theta})$ and $\boldsymbol{G}_c(\boldsymbol{\theta})$ such that

$$n^{-1}\sum_{i=1}^n [(\dot{\mu}_i - n^{-1}\sum_{i=1}^n \dot{\mu}_i)(\dot{\mu}_i - n^{-1}\sum_{i=1}^n \dot{\mu}_i)'] \rightarrow$$
$$E\Bigg\{\Bigg[\frac{\boldsymbol{Y}_0}{\sigma(\boldsymbol{Y}_0,\boldsymbol{\beta})} - E(\frac{\boldsymbol{Y}_0}{\sigma(\boldsymbol{Y}_0,\boldsymbol{\beta})})\Bigg]\Bigg[\frac{\boldsymbol{Y}_0}{\sigma(\boldsymbol{Y}_0,\boldsymbol{\beta})} - E(\frac{\boldsymbol{Y}_0}{\sigma(\boldsymbol{Y}_0,\boldsymbol{\beta})})\Bigg]'\Bigg\} = M(\boldsymbol{\theta}), \text{ say}, \quad (3.2)$$

$$n^{-1}\sum_{i=1}^n \dot{\sigma}_i\dot{\sigma}_i' \rightarrow E\Bigg\{\Bigg[\frac{\dot{\sigma}(\boldsymbol{Y}_0,\boldsymbol{\beta})}{\sigma(\boldsymbol{Y}_0,\boldsymbol{\beta})}\Bigg]\Bigg[\frac{\dot{\sigma}(\boldsymbol{Y}_0,\boldsymbol{\beta})}{\sigma(\boldsymbol{Y}_0,\boldsymbol{\beta})}\Bigg]'\Bigg\} = \dot{\Sigma}(\boldsymbol{\theta}), \text{ say},$$

$$n^{-1}\sum_{i=1}^n \dot{\sigma}_i\dot{\mu}_i' \rightarrow E\Bigg\{\Bigg[\frac{\dot{\sigma}(\boldsymbol{Y}_0,\boldsymbol{\beta})}{\sigma(\boldsymbol{Y}_0,\boldsymbol{\beta})}\Bigg]\Bigg[\frac{\boldsymbol{Y}_0}{\sigma(\boldsymbol{Y}_0,\boldsymbol{\beta})}\Bigg]'\Bigg\} = \boldsymbol{G}(\boldsymbol{\theta}), \text{ say, and}$$

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^{n}[(\dot{\mu}_i - n^{-1}\sum_{i=1}^{n}\dot{\mu}_i)\dot{\sigma}_i'] \to \\
&E\Big\{\Big[\frac{\boldsymbol{Y}_0}{\sigma(\boldsymbol{Y}_0,\boldsymbol{\beta})} - E(\frac{\boldsymbol{Y}_0}{\sigma(\boldsymbol{Y}_0,\boldsymbol{\beta})})\Big]\Big[\frac{\dot{\sigma}(\boldsymbol{Y}_0,\boldsymbol{\beta})}{\sigma(\boldsymbol{Y}_0,\boldsymbol{\beta})}\Big]'\Big\} = \boldsymbol{G}_c(\boldsymbol{\theta}),\ \text{say.}
\end{aligned}
$$

There exists a matrix-valued (of order $r \times r$) function $\dot{R}$ on $I\!R^p \times \Omega_2$ such that

$$
E\|\dot{R}(\boldsymbol{Y}_0,\boldsymbol{\beta})\| < \infty, \tag{3.3}
$$

and for every $\epsilon > 0$, $k > 0$, $\boldsymbol{s} \in \Omega_2$,

$$
\limsup_n P\left(\sup_{1\le i\le n, n^{1/2}\|\boldsymbol{t}-\boldsymbol{s}\|\le k} \frac{\|r_{ni}(\boldsymbol{t}) - r_{ni}(\boldsymbol{s}) - \dot{R}(\boldsymbol{Y}_{i-1},\boldsymbol{s})n^{-1/2}(\boldsymbol{t}-\boldsymbol{s})\|}{\|\boldsymbol{t}-\boldsymbol{s}\|} > \epsilon\right) = 0. \tag{3.4}
$$

The next theorem gives a one-step Taylor-type expansion of M_s around the true parameter $\boldsymbol{\theta}$, uniformly on its compact neighbourhood.

Lemma 3.1 *Suppose that in the model (1.1) assumptions (3.3) and (3.4) hold. Also, let κ be a nondecreasing twice differentiable function satisfying (i) $\int x\kappa(x)G(dx) = 1$, (ii) $\int x^2|\dot{\kappa}(x)|G(dx) < \infty$, and (iii) the second derivative of κ is bounded.*

Then, for every $0 < b < \infty$,

$$
\begin{aligned}
\sup_{\|\boldsymbol{t}\|\le b}\Big\| &M_s(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t}) - M_s(\boldsymbol{\theta}) + \Big[\int \kappa(x)G(dx) + \int x\dot{\kappa}(x)\,G(dx)\Big]\boldsymbol{G}(\boldsymbol{\theta})\ \boldsymbol{t}_1 \\
&+ \Big[\int x\kappa(x)G(dx) + \int x^2\dot{\kappa}(x)\,G(dx)\Big]\ \dot{\Sigma}(\boldsymbol{\theta})\ \boldsymbol{t}_2\Big\| = o_p(1).
\end{aligned}
$$

Therefore, substituting $\boldsymbol{t}_1 = n^{1/2}(\widehat{\boldsymbol{\alpha}}_p - \boldsymbol{\alpha})$ and $\boldsymbol{t}_2 = n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, and using the uniform convergence over compacta, we have the following theorem.

Theorem 3.2 *In addition to the assumptions of Theorem 3.1 and Lemma 3.1, assume that*

$$
\|n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\| = O_p(1). \tag{3.5}
$$

Then

$$
\begin{aligned}
&\Big[\int x\kappa(x)G(dx) + \int x^2\dot{\kappa}(x)\,G(dx)\Big]\dot{\Sigma}(\boldsymbol{\theta})\,n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
= \ &M_s(\boldsymbol{\theta}) - \Big[\int \kappa(x)G(dx) + \int x\dot{\kappa}(x)\,G(dx)\Big]\boldsymbol{G}(\boldsymbol{\theta})\ n^{1/2}(\widehat{\boldsymbol{\alpha}}_p - \boldsymbol{\alpha}) + o_p(1).
\end{aligned}
$$

If

$$\int \kappa(x)G(dx) = 0 = \int x\dot{\kappa}(x)G(dx) \tag{3.6}$$

also, then

$$\Big[\int x\kappa(x)G(dx) + \int x^2\dot{\kappa}(x)\,G(dx)\Big]\; n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = (\dot{\Sigma}(\boldsymbol{\theta}))^{-1}M_s(\boldsymbol{\theta}) + o_p(1).$$

Note that under (3.6), the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ does not depend on the preliminary estimator $\widehat{\boldsymbol{\alpha}}_p$ used in defining $\widehat{\boldsymbol{\beta}}$.

Remark 3.1. Conditions (i)-(iii) of Lemma 3.1 and (3.6) are satisfied by $\kappa(x) \equiv x$ when $E(\eta_1^2) = 1$. Another possible candidate is $\kappa(x) = -\dot{g}(x)/g(x)$. In this case $\int x\kappa(x)G(dx) = 1$ and when g is symmetric, $\int \kappa(x)G(dx) = 0$. Also for such choices conditions (i)-(iii) and (3.6) does not impose any extra moment condition for normal, logistic or double-exponential error densities since they are automatically satisfied.

The derivation of the asymptotic results on R-estimators depends on the uniform approximation of a randomly weighted empirical process by a perturbed empirical process. We define these processes under the following probabilistic framework.

Probabilistic framework: Let $\{\eta_i,\ 1 \le i \le n\}$ be i.i.d. with the d.f. G, $\{l_{ni},\ v_{ni},\ u_{ni}; 1 \le i \le n\}$ be an array of measurable functions from $\mathbb{R}^m$ to $\mathbb{R}$ such that for every $\boldsymbol{t} \in \mathbb{R}^m$, and $1 \le i \le n$, $(l_{ni}(\boldsymbol{t}),\ v_{ni}(\boldsymbol{t}),\ u_{ni}(\boldsymbol{t}))$ are independent of η_i. For $x \in \mathbb{R}$ and $\boldsymbol{t} \in \mathbb{R}^m$, let

$$\begin{aligned}
\tilde{\mathcal{V}}(x,t) &:= n^{-1/2}\sum_{i=1}^{n} l_{ni}(t)I\Big(\eta_i < x + xv_{ni}(t) + u_{ni}(t)\Big),\\
\tilde{\mathcal{J}}(x,t) &:= n^{-1/2}\sum_{i=1}^{n} l_{ni}(t)\,G\Big(x + xv_{ni}(t) + u_{ni}(t)\Big),\\
\tilde{\mathcal{U}}(x,t) &:= \tilde{\mathcal{V}}(x,t) - \tilde{\mathcal{J}}(x,t),\\
\mathcal{U}^*(x,t) &:= n^{-1/2}\sum_{i=1}^{n} l_{ni}(t)\Big[I(\eta_i < x) - G(x)\Big].
\end{aligned}$$

Here $\mathcal{U}^*(.,.)$ is a sequence of ordinary weighted empirical processes with weights $\{l_{ni}(.)\}$ and $\tilde{\mathcal{U}}(.,.)$ is a sequence of perturbed weighted empirical processes with location perturbations

$\{u_{ni}(.)\}$ and scale perturbations $\{v_{ni}(.)\}$. In Lemma 3.2 below it is shown that $\tilde{\mathcal{U}}$ can be uniformly approximated by $\mathcal{U}^*$ and this, in turn, will be applied to Lemma 3.3 to approximate empirical processes based on residuals that are different from actual errors by location and scale factors.

The following conditions (3.7)-(3.15) will be referred to as **Condition luv**. Here in (3.7)-(3.13), the assumptions/convergence hold pointwise for each fixed $\boldsymbol{t} \in I\!R^m$.

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

$$C_n(\boldsymbol{t})/n^{q/2-\epsilon} = o(1), \text{ for each } \boldsymbol{t} \in I\!R^m. \tag{3.7}$$

For some positive random process $\ell(t)$,

$$\left(n^{-1}\sum_{i=1}^n l_{ni}^2(\boldsymbol{t})\right)^{1/2} = \ell(\boldsymbol{t}) + o_p(1), \qquad \boldsymbol{t} \in I\!R^m. \tag{3.8}$$

$$E\left(n^{-1}\sum_{i=1}^n l_{ni}^2(\boldsymbol{t})\right)^{q/2} = O(1), \qquad \boldsymbol{t} \in I\!R^m. \tag{3.9}$$

$$\max_{1\le i\le n} n^{-1/2}|l_{ni}(\boldsymbol{t})| = o_p(1), \qquad \boldsymbol{t} \in I\!R^m. \tag{3.10}$$

$$\max_{1\le i\le n}\{|v_{ni}(\boldsymbol{t})| + |u_{ni}(\boldsymbol{t})|\} = o_p(1), \qquad \boldsymbol{t} \in I\!R^m. \tag{3.11}$$

$$\frac{n^{q/2-\epsilon}}{C_n(\boldsymbol{t})} E\left[n^{-1}\sum_{i=1}^n l_{ni}^2(\boldsymbol{t})\{|u_{ni}(\boldsymbol{t})| + |v_{ni}(\boldsymbol{t})|\}\right]^{q/2} = o(1), \qquad \boldsymbol{t} \in I\!R^m. \tag{3.12}$$

$$n^{-1/2}\sum_{i=1}^n |l_{ni}(\boldsymbol{t})|\,[|v_{ni}(\boldsymbol{t})| + |u_{ni}(\boldsymbol{t})|] = O_p(1), \quad \boldsymbol{t} \in I\!R^m. \tag{3.13}$$

$\forall\, b$ and $\epsilon > 0$, $\exists\, \delta > 0$, and an $n_1 \ni$ whenever $\|\boldsymbol{s}\| \le b$, and $n > n_1$, (3.14)

$$P\left(n^{-1/2}\sum_{i=1}^n |l_{ni}(\boldsymbol{s})|\Big\{\sup_{\|\boldsymbol{t}-\boldsymbol{s}\|<\delta}|v_{ni}(\boldsymbol{t}) - v_{ni}(\boldsymbol{s})| + \sup_{\|\boldsymbol{t}-\boldsymbol{s}\|<\delta}|u_{ni}(\boldsymbol{t}) - u_{ni}(\boldsymbol{s})|\Big\} \le \epsilon\right) > 1-\epsilon.$$

$\forall\, b$ and $\epsilon > 0$, $\exists$ a $\delta > 0$, and an n_2, $\ni$ whenever $\|\boldsymbol{s}\| \le b$, and $n > n_2$, (3.15)

$$P\left(\sup_{\|\boldsymbol{t}-\boldsymbol{s}\|\le\delta} n^{-1/2}\sum_{i=1}^n |l_{ni}(\boldsymbol{t}) - l_{ni}(\boldsymbol{s})| \le \epsilon\right) > 1-\epsilon.$$

Conditions (3.7)-(3.15) are regularity conditions on the weights and perturbations of the two-parameters empirical processes. Conditions (3.14)-(3.15) 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. These conditions will be verified for particular examples in Section 4.

We also make the following additional assumptions on the error d.f. G.

- (G.1) The d.f. G has Lebesgue density g satisfying the following: g is positive on the set $\{x : 0 < G(x) < 1\}$, $g(x)$ and $xg(x)$ are bounded in $x \in \mathbb{R}$, and the functions $u \mapsto g(G^{-1}(u))$ and $u \mapsto G^{-1}(u)g(G^{-1}(u))$ are uniformly continuous on $[0, 1]$.
- (G.2) The d.f. G is uniformly Lipschitz in scale: For some constant $0 < C < \infty$ and for every $s \in \mathbb{R}$, $\sup_{x\in\mathbb{R}} |G(x + xs) - G(x)| \le C\,|s|$.
- (G.3) $\lim_{\delta\to 0} \sup\{|x| \int_0^1 |g(x) - g(x + tx\delta)|dt; x \in \mathbb{R}\} = 0$.

We remark here that if the error density g has decreasing tails, then (G.2) is implied by $\sup_{x\in\mathbb{R}} |x|g(x) < \infty$, which in turn, is guaranteed by $E\eta^2 < \infty$. In this case, more easily verifiable conditions ensuring (G.3) can also be obtained. For example, if g is differentiable with the derivative $\dot{g}$ satisfying $\sup[x^2 \sup\{|\dot{g}(y)|; x(1 - \delta) < y < x(1 + \delta)\}, x \in \mathbb{R}] < \infty$, for some $\delta > 0$, then (G.3) holds. In particular, (G.1)-(G.3) hold for standardized normal, double-exponential logistic and t-distributions with degrees of freedom more than 2.

The following lemma is used for proving the needed result.

Lemma 3.2 *Under the above framework, suppose that* **Condition luv** *and assumptions (G.1)-(G.3) hold. Then for every* $0 < b < \infty$,

$$\sup_{x\in\mathbb{R}, \|t\|\le b} |\tilde{\mathcal{U}}(x,t) - \mathcal{U}^*(x,t)| \;\; = \;\; o_p(1). \tag{3.16}$$

Based on this lemma, the next result gives a Taylor-type expansion for the R-scores.

Lemma 3.3 *Suppose that the assumptions of Lemma 3.2 hold with* $l_{ni}(\boldsymbol{t})$ *equal to the* j*-th coordinate* $(1 \le j \le p)$ *of* $s_{ni}(\boldsymbol{t})$, $u_{ni}(\boldsymbol{t}) := \mu_{ni}(\boldsymbol{t}_1) - \mu_i$ *and* $v_{ni}(\boldsymbol{t}) := \sigma_{ni}(\boldsymbol{t}_2) - 1$, $1 \le i \le n$.

Then

$$\sup_{\|t\|\le b}\Bigg\|\boldsymbol{S}_\varphi(\boldsymbol{\theta}+n^{-1/2}\boldsymbol{t})-\boldsymbol{S}_\varphi(\boldsymbol{\theta})$$
$$-\quad\Big(\int g(x)\varphi(G(dx))\,M(\boldsymbol{\theta})\boldsymbol{t}_1+\int xg(x)\varphi(G(dx))\,\boldsymbol{G}_c(\boldsymbol{\theta})\boldsymbol{t}_2\Big)\Bigg\|=o_p(1).$$

Therefore, we have the following theorem on the asymptotic distribution of the R-estimator. Note that here the condition $n^{1/2}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})=O_p(1)$ is automatically satisfied as in Jurečková (1971, Theorem 1.1) and Koul (1996, Corollary 1.1, Remark 1.2) since the mean function in (1.1) is a linear function of the parameters.

Theorem 3.3 *In addition to the assumptions of Lemma 3.3, assume that (3.5) holds. Then*

$$(i)\qquad \int g(x)\varphi(G(dx))\,n^{1/2}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})$$
$$=-(M(\boldsymbol{\theta}))^{-1}\Big[\boldsymbol{S}_\varphi(\boldsymbol{\theta})+\boldsymbol{G}_c(\boldsymbol{\theta})n^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\int xg(x)\varphi(G(dx))\Big]+o_p(1).\quad(3.17)$$

(ii) If, in addition to (i), either $\int xg(x)\varphi(G(dx))=0$ *or* $\boldsymbol{G}_c(\boldsymbol{\theta})=0$*, then*

$$n^{1/2}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})=-\{\int g(x)\varphi(G(dx))M(\boldsymbol{\theta})\}^{-1}\boldsymbol{S}_\varphi(\boldsymbol{\theta})+o_p(1).$$

In order to get the asymptotic normality of the R-estimator $\widehat{\boldsymbol{\alpha}}$, we need to establish the same for

$$\boldsymbol{S}_\varphi(\boldsymbol{\theta})=n^{-1/2}\sum_{i=1}^n\{\boldsymbol{Z}_{i-1}(\boldsymbol{\beta})-\bar{\boldsymbol{Z}}(\boldsymbol{\beta})\}\varphi\left(\frac{R_i}{n+1}\right),$$

where R_i is the rank of η_i among $\{\eta_j; 1\le j\le n\}$. But, this is a randomly weighted sum of rank scores. Moreover, the random weights $\Big\{\boldsymbol{Z}_{i-1}(\boldsymbol{\beta})-\bar{\boldsymbol{Z}}(\boldsymbol{\beta}); 1\le i\le n\Big\}$ as well as $\{R_1,\ldots,R_n\}$ are dependent. However, extending an argument of Koul and Ossiander (1994, Theorem 1.2, Remark 1.1 and Lemma 1.2), $\boldsymbol{S}_\varphi(\boldsymbol{\theta})$ can be approximated by a randomly weighted sum of independent random variables defined by

$$\widehat{\boldsymbol{S}}_\varphi \;=\; n^{-1/2}\sum_{i=1}^n\{\boldsymbol{Z}_{i-1}(\boldsymbol{\beta})-\bar{\boldsymbol{Z}}(\boldsymbol{\beta})\}\varphi(G(\eta_i))$$

$$= n^{-1/2}\sum_{i=1}^{n}\{\boldsymbol{Z}_{i-1}(\boldsymbol{\beta}) - \bar{\boldsymbol{Z}}(\boldsymbol{\beta})\}\{\varphi(G(\eta_i)) - E[\varphi(G(\eta_1))]\}.$$

Then the asymptotic normality of $\widehat{\boldsymbol{S}}_\varphi$ can be established by using multivariate martingale central limit theorem on $\widehat{\boldsymbol{S}}_\varphi$. We state that formally in the following proposition whose proof is similar to Koul and Ossiander (1994, Lemma 1.2).

Proposition 3.1 *Under the model (1.1),*

$$\boldsymbol{S}_\varphi(\boldsymbol{\theta}) - \widehat{\boldsymbol{S}}_\varphi = o_p(1).$$

Moreover

$$\widehat{\boldsymbol{S}}_\varphi \Rightarrow \boldsymbol{N}_p[0, \sigma_\varphi^2 M(\boldsymbol{\theta})],$$

where $\sigma_\varphi^2 = Var[\varphi(G(\eta_1))]$. *Hence under the assumptions of Theorem 3.3(ii)*

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \Longrightarrow \mathcal{N}_p[0, \Sigma(\boldsymbol{\theta})], \tag{3.18}$$

where $\Sigma(\boldsymbol{\theta}) := (M(\boldsymbol{\theta}))^{-1}J(\varphi, G)$ *with* $J(\varphi, G) := \frac{\int \varphi^2(u)du - (\int \varphi(u)du)^2}{(\int g(x)\varphi(G(dx)))^2}$.

Remark 3.2. The conditions of Theorem 3.3(ii) ensures that the preliminary estimator and the scale estimator have no effect on the asymptotics of the final estimator. A sufficient condition for $\int xg(x)\varphi(G(dx)) = 0$ is that g is symmetric i.e., $g(-x) = g(x)$ and φ is skew symmetric, i.e., $\varphi(u) = -\varphi(1-u)$, $\forall u \in [0,1]$. Therefore, in practice, we recommend to use a skew symmetric φ to ensure that Theorem 3.3(ii) holds when the innovations are symmetrically distributed. For some model, e.g., in ARLSCH of Example 2, $\boldsymbol{G}_c(\boldsymbol{\theta}) = 0$ when X_0 is symmetrically distributed around zero. However, for Example 1 (Engle's ARCH) and Example 3 (ARTCH), $\boldsymbol{G}_c(\boldsymbol{\theta}) \neq 0$ and the use of a skew symmetric score function is essential. If the conditions of Theorem 3.3(ii) are not satisfied, then there will be extra terms in the variance-covariance matrix of the asymptotic distribution of $\widehat{\boldsymbol{\alpha}}$ that depend on $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and κ in a complex manner.

Under Theorem 3.3(ii), the asymptotic distribution of $\widehat{\boldsymbol{\alpha}}$ is the same as that of an R-estimator of $\boldsymbol{\alpha}$ for the model

$$\frac{X_i}{\sigma(\boldsymbol{Y}_{i-1},\boldsymbol{\beta})} = \frac{\boldsymbol{Y}'_{i-1}\boldsymbol{\alpha}}{\sigma(\boldsymbol{Y}_{i-1},\boldsymbol{\beta})} + \eta_i, \tag{3.19}$$

with $\boldsymbol{\beta}$ known. In general, an R-estimator is location invariant. However, since we compute R-estimator basically for the model (3.19), even though the original model (1.1) may have a location parameter like the ARCH model, (3.19) need not have that, unless σ is a constant. Thus we can estimate the intercept parameter of the original model through R-estimation.

Remark 3.3. Comparison with other estimators. (i) *Relative efficiency of an R-estimator with respect to (wrt) the optimal R-estimator*: From (3.18) it follows that for a fixed score function φ, the asymptotic dispersion of the standardized R-estimator is a scalar $J(\varphi, G)$ that depends only on the underlying error distribution, multiplied by a matrix which depends only on $\boldsymbol{\theta}$ and the error distribution. Hence, for a given innovation density g, the optimal R-estimator based on the score function $\varphi^*_g(u) = -\dot{g}(G^{-1}(u))/g(G^{-1}(u))$ exists, provided that $\varphi_g \in \mathcal{F}$. In particular, when g is the logistic density, $\varphi^*_g(u) = u - 1/2$ and when g is the double-exponential density, $\varphi^*_g(u) = (1/2)$ sign $(u - 1/2)$. Also

$$J(\varphi^*_g, G) = 1/I_g, \tag{3.20}$$

where I_g is the Fisher's information for g. See Jurečková and Sen (1996, Display 3.4.30) for a similar result under homoscedastic linear model. Note also that for the Wilcoxon R-estimator $\widehat{\boldsymbol{\alpha}}_W$ corresponding to the score function $\varphi(u) = u - 1/2$,

$$J(\varphi, G) = 1/\{12(\int g^2(x)dx)^2\} \tag{3.21}$$

and for the R-estimator $\widehat{\boldsymbol{\alpha}}_S$ based on the signed-score function $\varphi(u) = (1/2)$ sign $\{u - (1/2)\}$,

$$J(\varphi, G) = 1/\{4g^2(0)\}. \tag{3.22}$$

It is of natural interest to compare the performance of an R-estimator with the optimal R-estimator φ_g^*. Accordingly, one can define the absolute relative efficiency of an R-estimator based on φ as $1/[I_g J(\varphi, G)]$ which will be bounded above by one. From Mukherjee (2006 b), the absolute relative efficiency does not depend on the variance of G. Hence, from (3.18) and Lehmann (1983, Section 2.6, Table 6.2 and Section 5.6, Table 6.2), the absolute relative efficiencies of $\widehat{\boldsymbol{\alpha}}_W$ are $3/\pi = 0.955$, 1 and 0.75 at the normal, logistic and the double-exponential density, respectively. Also, from (3.18), Mukherjee (2006 b) and Lehmann (1983, Section 5.4, Table 4.4), the absolute relative efficiencies of $\widehat{\boldsymbol{\alpha}}_S$ are $2/\pi = 0.637$, 0.75 and 1 at the normal, logistic and the double-exponential density, respectively.

(ii) *Relative efficiency of an R-estimator wrt the quasi maximum likelihood estimator (QMLE)*: From (2.1) and (3.19), a maximum likelihood estimator of $\boldsymbol{\alpha}$ based on the normal distribution of the errors can be defined as a minimizer $\widehat{\boldsymbol{\alpha}}_{QMLE}$ of

$$\sum_{i=1}^{n}[X_i/\sigma(\boldsymbol{Y}_{i-1}, \widehat{\boldsymbol{\beta}}) - \{\boldsymbol{Y}_{i-1}/\sigma(\boldsymbol{Y}_{i-1}, \widehat{\boldsymbol{\beta}})\}'\boldsymbol{\tau}_1]^2$$

with respect to $\boldsymbol{\tau}_1$. This yields

$$\widehat{\boldsymbol{\alpha}}_{QMLE} = [\sum_{i=1}^{n}\boldsymbol{Y}_{i-1}\boldsymbol{Y}'_{i-1}/\sigma^2(\boldsymbol{Y}_{i-1}, \widehat{\boldsymbol{\beta}})]^{-1}[\sum_{i=1}^{n}\boldsymbol{Y}_{i-1}X_i/\sigma^2(\boldsymbol{Y}_{i-1}, \widehat{\boldsymbol{\beta}})]. \tag{3.23}$$

The estimator $\widehat{\boldsymbol{\alpha}}_{QMLE}$ can also be termed as the least squares estimator (LSE) and using standard techniques, its asymptotic distribution can be obtained as

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\alpha}}_{QMLE} - \boldsymbol{\alpha}) \Longrightarrow \mathcal{N}_p[0, (E[\boldsymbol{Y}_0\boldsymbol{Y}'_0/\sigma^2(\boldsymbol{Y}_0, \boldsymbol{\beta})])^{-1}]. \tag{3.24}$$

When, for example, $E[\boldsymbol{Y}_0/\sigma(\boldsymbol{Y}_0, \boldsymbol{\beta})] = 0$, we can use (3.18) and (3.24) to define the ARE of an R-estimator based on φ, with respect to the QMLE as $1/J(\varphi, G)$. Therefore from (3.21), the asymptotic relative efficiency (ARE) of the Wilcoxon R-estimator with respect to the QMLE is $12(\int g^2(x)dx)^2$ which is at least 0.864 for a large class of symmetric standardized error densities g; see, for example, Lehmann (1983, Section 5.6) for similar result under the location model. In particular, for the standardized normal, logistic and the double-exponential g, ARE equals $3/\pi = 0.955$, $\pi^2/9 = 1.10$ and 1.50, respectively. In a similar fashion, from (3.22), the ARE of the R-estimator based on signed score with respect to the QMLE is $4g^2(0)$ which is at least 1/3 for symmetric unimodal error densities g (with variance 1); see, for example, Lehmann (1983, Section 5.3) for similar result under the location model. In particular, for the standardized normal, logistic and double-exponential g, ARE equals $2/\pi = 0.637$, $\pi^2/12 = 0.82$ and 2, respectively.

A classical result due to Chernoff-Savage (1958), translated to our setup, asserts that there exists R-estimator that can ensure the ARE with respect to the QMLE to be at least one; in other words, such estimator is even better than the Wilcoxon-type R-estimator for which the minimum ARE is 0.864. Such R-estimator based on the *unbounded* normal score function (van der Waerden type R-estimator) is asymptotically efficient at the normal errors and has the ARE of at least 1 for all other error densities. In the homoscedastic autoregressive model with $\sigma \equiv 1$, Mukherjee and Bai (2002) derived (3.18) for unbounded but square-integrable score function and showed consequently that the Chernoff-Savage phenomenon holds for the autoregressive models. We conjecture that (3.18) holds for the unbounded score function under the heteroscedastic setup also, which, if proved, should give more motivation for considering the R-estimators.

(iii) *Relative efficiency of the optimal R-estimator wrt the QMLE*: Note from (3.20) and (3.24) that the ARE of the optimal R-estimator based on φ_g^* with respect the QMLE *at the error density* g is given by

$$1/(1/I_g) = I_g. \tag{3.25}$$

In particular, for the standardized normal, logistic and double-exponential g, this efficiency equals 1, $\pi^2/9 = 1.10$ and 2, respectively. However, in order to use the optimal estimator, the form of g should be known.

Remark 3.4. In order to use the result of Proposition 3.1 to construct, for example, confidence intervals, we need to estimate $\int g(x)\varphi(G(dx))$ appearing in $J(\varphi, G)$. For the R-estimation in the homoscedastic autoregressive model with $\sigma \equiv 1$ the same factor arise and an estimate can be obtained by replacing g and G by a kernel density estimator and the empirical distribution function based on the estimated residuals; see, for example, Koul (1992, Section 7.3c). In a similar fashion, we can obtain an estimate $\int g(x)\varphi(G(dx))$ by replacing g and G by a kernel density estimator and the empirical distribution function based on the estimated residuals $\{\eta_j(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}); 1 \le j \le n\}$; however, the performance of such estimator has been investigated here neither theoreticaly nor empirically. In the empirical study we use the Wilcoxon score function for which $\int g(x)\varphi(G(dx)) = \int g^2(x)dx$ and there we use simple histogram estimator of g which performs very well; see Section 5 for details.

4 Examples

This section contains some details for verifying the general conditions of the previous section in three examples. Here we check **Condition luv** with

$$u_{ni}(\boldsymbol{t}) = \frac{n^{-1/2}\boldsymbol{Y}'_{i-1}\boldsymbol{t}_1}{\sigma(\boldsymbol{Y}_{i-1}, \boldsymbol{\beta})}, \; v_{ni}(\boldsymbol{t}) = \frac{\sigma(\boldsymbol{Y}_{i-1}, \boldsymbol{\beta} + n^{-1/2}\boldsymbol{t}_2) - \sigma(\boldsymbol{Y}_{i-1}, \boldsymbol{\beta})}{\sigma(\boldsymbol{Y}_{i-1}, \boldsymbol{\beta})},$$

and

$$l_{ni}(\boldsymbol{t}) = \frac{j\text{-th coordinate of } \boldsymbol{Y}_{i-1}}{\sigma(\boldsymbol{Y}_{i-1}, \boldsymbol{\beta} + n^{-1/2}\boldsymbol{t}_2)}, \; 1 \le j \le p.$$

We will also use the following fact repeatedly which states that if $\boldsymbol{U} = [u_1, \ldots, u_k]'$, $\boldsymbol{V} = [v_1, \ldots, v_k]'$ and $\boldsymbol{W}$ are vectors with all entries nonnegative, then

$$\boldsymbol{W}'\boldsymbol{V}/\boldsymbol{W}'\boldsymbol{U} \le 1 + (v_1/u_1) + \ldots (v_k/u_k), \tag{4.1}$$

where we define $v_j/u_j = 0$ if $u_j = 0 = v_j$. See, for example, Mukherjee (2006 a, Lemma 2).

Example 1. (ARCH MODEL). In this example, $\boldsymbol{\alpha} = \boldsymbol{\beta}$, $\dot{\sigma}(\boldsymbol{Y}_{i-1}, \boldsymbol{t}) = \boldsymbol{Y}_{i-1}$ and (3.3)-(3.4) are satisfied with $\dot{R}(\boldsymbol{Y}_{i-1}, \boldsymbol{t}) = -\boldsymbol{Y}_{i-1}\boldsymbol{Y}'_{i-1}/(\boldsymbol{Y}'_{i-1}\boldsymbol{t})^2$. Now we check **Condition luv** with

$$l_{ni}(\boldsymbol{t}) = \text{ either } \frac{1}{\boldsymbol{Y}'_{i-1}(\boldsymbol{\alpha} + n^{-1/2}\boldsymbol{t}_2)}, \text{ or } \frac{X_{i-j}}{\boldsymbol{Y}'_{i-1}(\boldsymbol{\alpha} + n^{-1/2}\boldsymbol{t}_2)}, \ 1 \le j \le s,$$

$$u_{ni}(\boldsymbol{t}) = \frac{n^{-1/2}\boldsymbol{Y}'_{i-1}\boldsymbol{t}_1}{\boldsymbol{Y}'_{i-1}\boldsymbol{\alpha}} \text{ and } v_{ni}(\boldsymbol{t}) = \frac{n^{-1/2}\boldsymbol{Y}'_{i-1}\boldsymbol{t}_2}{\boldsymbol{Y}'_{i-1}\boldsymbol{\alpha}}.$$

Using (4.1), all coordinates of the vectors $\dot{\mu}_i$ and $\dot{\sigma}_i$ are uniformly bounded and consequently the existence of all the matrices in (3.2) is guaranteed. Also, there is a compact neighbourhood containing zero on which $\{l_{ni}(\boldsymbol{t})\}$'s are all uniformly bounded and by the stationarity, $\forall \boldsymbol{t}$, $C_n(\boldsymbol{t}) = O(n)$. Any choice of $q > 2$ and $0 < \epsilon < q/2$ with $1 < q/2 - \epsilon$ will satisfy (3.7).

By the stationarity and boundedness, $n^{-1}\sum_{i=1}^n E\{l_{ni}(\boldsymbol{t}) - l_{ni}(0)\}^2 = o(1)$. Therefore

$$\ell(\boldsymbol{t}) = \text{ either } \Big\{E\Big[\frac{1}{\boldsymbol{Y}'_0\boldsymbol{\alpha}}\Big]^2\Big\}^{1/2}, \text{ or } \Big\{E\Big[\frac{X_{-j}}{\boldsymbol{Y}'_0\boldsymbol{\alpha}}\Big]^2\Big\}^{1/2}, \ 1 \le j \le s,$$

and hence, condition (3.8) is satisfied. Conditions (3.9) and (3.10) are satisfied by boundedness which is a consequence (4.1). Condition (3.11) is also a consequence of (4.1) and so is (3.13) after taking expectation and using the stationarity.

For (3.12), the left hand side is bounded by a constant times $\frac{n^{q/2-\epsilon}}{n}[n^{-1/2}]^{q/2}$ which is $o(1)$ if $0 < q/2 - \epsilon - 1 < q/4$. In other words, any choice of q and ϵ satisfying $q/4 < 1 + \epsilon < q/2$ will satisfy (3.7) and (3.12). Verification of (3.14) and (3.15) are immediate by writing down the corresponding expressions.

Since here $\boldsymbol{\alpha} = \boldsymbol{\beta}$, for estimation in this model, we use just a two-step procedure, i.e., use $\widehat{\boldsymbol{\alpha}}_p$ instead of $\widehat{\boldsymbol{\beta}}$ to define final $\widehat{\boldsymbol{\alpha}}$. Therefore, from (3.18), if either $\int xg(x)\varphi(G(dx)) = 0$ or $\boldsymbol{G}_c(\boldsymbol{\theta}) = 0$, then

$$n^{1/2}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \Longrightarrow N_p(0, \Sigma(\boldsymbol{\alpha})), \quad \Sigma(\boldsymbol{\alpha}) := M^{-1}(\boldsymbol{\theta})J(\varphi, G).$$

Denote the estimator in (3.23) under a two-step procedure by $\widehat{\boldsymbol{\alpha}}_{QMLE}$ which is the most commonly-used estimator for this model. Introduced by Engle (1982), it is a maximizer of the normal likelihood

$$-(1/2)\sum_{i=1}^{n}[\{X_i^2/(\boldsymbol{Y}'_{i-1}\boldsymbol{\tau})\} + \log(\boldsymbol{Y}'_{i-1}\boldsymbol{\tau})].$$

Weiss (1986) proved that under the stationarity of $\{X_i\}$'s and the finite fourth moment assumption on the i.i.d. errors ϵ_i (which is the same as finiteness of the second moment of η_i), the asymptotic distribution of $\widehat{\boldsymbol{\alpha}}_{QMLE}$ is as follows:

$$n^{1/2}(\widehat{\boldsymbol{\alpha}}_{QMLE} - \boldsymbol{\alpha}) \Longrightarrow N_p(0, \Sigma_{QMLE}),\ \text{where } \Sigma_{QMLE} := \left(E\left[\boldsymbol{Y}_0\boldsymbol{Y}'_0/(\boldsymbol{\alpha}'\boldsymbol{Y}_0)^2\right]\right)^{-1} Var(\eta).$$

Since in this example $E[\boldsymbol{Y}_0/\sigma(\boldsymbol{Y}_0, \boldsymbol{\beta})]$ is non-null, computation of the ARE of a rank-estimator $\widehat{\boldsymbol{\alpha}}$, relative to the commonly-used quasi maximum likelihood estimator in Engle's ARCH model is not straight-forward. However, the ratio of the scalar-factors is exactly the same as that of the rank-estimator relative to the least squares estimator in the linear regression model; see Remark 3.3 for more on this.

Example 2. (ARLSCH MODEL). Letting $\tilde{\boldsymbol{Z}}_{i-1} = (1, X_{i-1}^2)'$, $\dot{\sigma}(\boldsymbol{Y}_{i-1}, \boldsymbol{t}) = \tilde{\boldsymbol{Z}}_{i-1}/\{2(\tilde{\boldsymbol{Z}}'_{i-1}\boldsymbol{t})^{1/2}\}$. Also, with $\dot{R}(\boldsymbol{Y}_{i-1}, \boldsymbol{t}) = -\tilde{\boldsymbol{Z}}_{i-1}\tilde{\boldsymbol{Z}}'_{i-1}/\{2(\tilde{\boldsymbol{Z}}'_{i-1}\boldsymbol{t})^2\}$, (3.3)-(3.4) are satisfied. Now we check **Condition luv** with

$$l_{ni}(\boldsymbol{t}) = \frac{X_{i-1}}{\left\{(\boldsymbol{\beta} + n^{-1/2}\boldsymbol{t}_2)'\tilde{\boldsymbol{Z}}_{i-1}\right\}^{1/2}},$$

$$u_{ni}(\boldsymbol{t}) = \frac{n^{-1/2}t_1' X_{i-1}}{\left\{\boldsymbol{\beta}'\tilde{\boldsymbol{Z}}_{i-1}\right\}^{1/2}},\ v_{ni}(\boldsymbol{t}) = \frac{\{(\boldsymbol{\beta} + n^{-1/2}\boldsymbol{t}_2)'\tilde{\boldsymbol{Z}}_{i-1}\}^{1/2} - (\boldsymbol{\beta}'\tilde{\boldsymbol{Z}}_{i-1})^{1/2}}{(\boldsymbol{\beta}'\tilde{\boldsymbol{Z}}_{i-1})^{1/2}}.$$

Using the boundedness of the function $x \to x/(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x^2)^{1/2}$ on $[0, \infty)$ and the stationarity and the ergodicity of $\{X_i\}$, (3.2) holds. To verify (3.8), note that by the stationarity

$$\begin{aligned} n^{-1}\sum_{i=1}^{n} E\{l_{ni}(\boldsymbol{t}) - l_{ni}(0)\}^2 &= E\Big[\frac{X_0}{\left\{(\boldsymbol{\beta} + n^{-1/2}\boldsymbol{t}_2)'\tilde{\boldsymbol{Z}}_0\right\}^{1/2}} - \frac{X_0}{\left\{\boldsymbol{\beta}'\tilde{\boldsymbol{Z}}_0\right\}^{1/2}}\Big]^2 \\ &= E\Big[\Big\{\frac{X_0}{\left\{\boldsymbol{\beta}'\tilde{\boldsymbol{Z}}_0\right\}^{1/2}}\Big\}\Big\{\frac{(\boldsymbol{\beta}'\tilde{\boldsymbol{Z}}_0)^{1/2}}{\{(\boldsymbol{\beta} + n^{-1/2}\boldsymbol{t}_2)'\tilde{\boldsymbol{Z}}_0\}^{1/2}} - 1\Big\}\Big]^2. \end{aligned}$$

By (4.1), the sequence of r.v.'s under the expectation is bounded and tends to 0, a.s. Therefore the above is $o(1)$ by the bounded Convergence Theorem. Also

$$\ell(\boldsymbol{t}) = \Big\{E\Big[\frac{X_0^2}{(\tilde{\boldsymbol{Z}}_0'\boldsymbol{\beta})}\Big]\Big\}^{1/2},$$

and hence, condition (3.8) is satisfied. Conditions (3.9) and (3.10) are satisfied by boundedness which is a consequence (4.1). Condition (3.11) is also a consequence of (4.1)

Next, we verify (3.13). Taking expectation, it is easy to see that $n^{-1/2}\sum_{i=1}^{n}|l_{ni}(\boldsymbol{t})u_{ni}(\boldsymbol{t})| = O_p(1)$; next we check that $n^{-1/2}\sum_{i=1}^{n}|l_{ni}(\boldsymbol{t})v_{ni}(\boldsymbol{t})| = O_p(1)$. First note that

$$n^{-1/2}\sum_{i=1}^{n}|l_{ni}v_{ni}| \le n^{-1/2}\sum_{i=1}^{n}\frac{|X_{i-1}|}{(\boldsymbol{\beta}'\tilde{\boldsymbol{Z}}_{i-1})^{1/2}}\left|\left\{\frac{\boldsymbol{\beta}'\tilde{\boldsymbol{Z}}_{i-1}}{(\boldsymbol{\beta}+n^{-1/2}\boldsymbol{t}_2)'\tilde{\boldsymbol{Z}}_{i-1}}\right\}^{1/2} - 1\right|$$

Next note that the derivative of the function $s \mapsto [x/(x+s)]^{1/2}$ at $s=0$ is $-1/(2x)$. Therefore above is bounded by

$$n^{-1/2}\sum_{i=1}^{n}\frac{|X_{i-1}|}{(\boldsymbol{\beta}'\tilde{\boldsymbol{Z}}_{i-1})^{1/2}}\left|\left\{\frac{\boldsymbol{\beta}'\tilde{\boldsymbol{Z}}_{i-1}}{(\boldsymbol{\beta}+n^{-1/2}\boldsymbol{t}_2)'\tilde{\boldsymbol{Z}}_{i-1}}\right\}^{1/2} - 1 + \frac{n^{-1/2}\tilde{\boldsymbol{Z}}_{i-1}'\boldsymbol{t}_2}{2\,\boldsymbol{\beta}'\tilde{\boldsymbol{Z}}_{i-1}}\right|$$
$$+\frac{1}{2}n^{-1}\sum_{i=1}^{n}\frac{|X_{i-1}|}{(\boldsymbol{\beta}'\tilde{\boldsymbol{Z}}_{i-1})^{1/2}}\frac{\tilde{\boldsymbol{Z}}_{i-1}'\boldsymbol{t}_2}{\tilde{\boldsymbol{Z}}_{i-1}'\boldsymbol{\beta}}.$$

Assuming $E\|X_0\|^4 < \infty$, we have $E\|\tilde{\boldsymbol{Z}}_0\|^2 < \infty$, and hence $\max_{1\le i\le n}|n^{-1/2}\tilde{\boldsymbol{Z}}_{i-1}'\boldsymbol{t}_2| = o_p(1)$. Using a two-step Taylor-type expansion of the function $s \mapsto [x/(x+s)]^{1/2}$ at $s=0$, we get a factor of $n^{-1/2}\times n^{-1}$ at the first term which together with the stationarity and ergodicity forces the first term to go to zero in probability. The n^{-1} factor implies that the r.v.'s in the second term converges in probability to $E[\{|X_{i-1}|\tilde{\boldsymbol{Z}}_{i-1}'\boldsymbol{t}_2\}/\{(\boldsymbol{\beta}'\tilde{\boldsymbol{Z}}_{i-1})^{1/2}\tilde{\boldsymbol{Z}}_{i-1}'\boldsymbol{\beta}\}]$, thereby verifying (3.13) here.

Finally, we can verify (3.7) and (3.12) an in Example 1 since all the underlying quantities are bounded. Verification of (3.14) and (3.15) can be done by writing down the corresponding expressions and invoking the smoothness of the derivatives.

Therefore, to summarize, we obtain that if either $\int xg(x)\varphi(G(dx)) = 0$ or $\boldsymbol{G}_c(\boldsymbol{\theta}) = 0$, then

$$n^{1/2}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \Longrightarrow N_1(0, \tau^2(\theta)J(\varphi, G)), \qquad \tau^2(\theta) := [\text{Var}\{X_0/(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 X_0^2)^{1/2}\}]^{-1}.$$

Example 3. (ARTCH MODEL). To verify the assumptions in this model, let

$$\tilde{\boldsymbol{Z}}_{i-1} = [X_{i-1}I(X_{i-1} > 0), -X_{i-1}I(X_{i-1} \le 0), \ldots, X_{i-p}I(X_{i-p} > 0), -X_{i-p}I(X_{i-p} \le 0)]'.$$

Then, in this example, $\dot{\sigma}(\boldsymbol{Y}_{i-1}, \boldsymbol{t}) = \tilde{\boldsymbol{Z}}_{i-1}$ and (3.3)-(3.4) are satisfied with $\dot{R}(\boldsymbol{Y}_{i-1}, \boldsymbol{t}) = -\tilde{\boldsymbol{Z}}_{i-1}\tilde{\boldsymbol{Z}}'_{i-1}(\tilde{\boldsymbol{Z}}'_{i-1}\boldsymbol{t})^{-2}$. Next we can check **Condition luv** with

$$l_{ni}(\boldsymbol{t}) = \frac{X_{i-j}}{\tilde{\boldsymbol{Z}}'_{i-1}(\boldsymbol{\beta} + n^{-1/2}\boldsymbol{t}_2)}, \ 1 \le j \le p,$$

$$u_{ni}(\boldsymbol{t}) = \frac{n^{-1/2}\boldsymbol{Y}'_{i-1}\boldsymbol{t}_1}{\tilde{\boldsymbol{Z}}'_{i-1}\boldsymbol{\beta}} \text{ and } v_{ni}(\boldsymbol{t}) = \frac{n^{-1/2}\tilde{\boldsymbol{Z}}'_{i-1}\boldsymbol{t}_2}{\tilde{\boldsymbol{Z}}'_{i-1}\boldsymbol{\beta}}.$$

The details are similar to those of Example 1 since the standard deviation is a linear function of the parameters; here one needs to use the fact that the functions $x \to x/(\boldsymbol{\beta}_{2j-1}xI(x \ge 0) - \boldsymbol{\beta}_{2j}xI(x < 0))$ are bounded. Hence, from Proposition 3.1, if $\int xg(x)\varphi(G(dx)) = 0$, then

$$n^{1/2}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \Longrightarrow N_p(0, \Sigma(\boldsymbol{\theta})). \tag{4.2}$$

5 Empirical study

In this section we first report Monte Carlo study comparing the Wilcoxon R-estimator ($\widehat{\boldsymbol{\alpha}}_W$), the R-estimator based on the signed score ($\widehat{\boldsymbol{\alpha}}_S$) and the QMLE ($\widehat{\boldsymbol{\alpha}}_{QMLE}$) at three error densities in terms of their average squared deviations from the true parameter. Consequently, the performance of some optimal R-estimators at certain error densities are compared with the Gaussian likelihood based MLE . Next we consider three important real data sets in the financial time series and study the robustness of R-estimators against misspecified form of the heteroscedasticity.

Model. Among many different models, we choose the ARTCH model of Example 3 with $p = s = 1$ and the ARLSCH model of Example 2 with $p = s = 1, r = 2$ with specific value of the underlying true parameters, when the errors are simulated from the standardized (i) normal (N), (ii) logistic (L) and (iii) double-exponential (D) distribution. For results with different combinations of the underlying true parameters for which the model could be even nonstationary, see Mukherjee (2006 b). To estimate the scale parameters, we use the score function $\kappa(u) = u$. The computations become relatively simpler under such choice of the score function with even closed form expressions for the scale estimators in the ARTCH model. For each model, we compute (i) the preliminary estimator $\widehat{\alpha}_p$, (ii) the MLE based on the normal distribution $\widehat{\alpha}_{QMLE}$, (iii) the Wilcoxon R-estimator $\widehat{\alpha}_W$ based on the score function $\varphi(u) = u - (1/2)$ and (iv) the R-estimator $\widehat{\alpha}_S$ based on the signed-score function $\varphi(u) = \text{sign}\ \{u - (1/2)\}$.

Formulae for the ARTCH model. From

$$X_i = \alpha X_{i-1} + \{\beta_1 X_{i-1} I(X_{i-1} > 0) - \beta_2 X_{i-1} I(X_{i-1} \leq 0)\}\, \eta_i, \qquad 1 \leq i \leq n,$$

note that $\widehat{\alpha}_p = \sum_{i=1}^n X_i X_{i-1} / \sum_{i=1}^n X_{i-1}^2$. Write $M_s(\boldsymbol{\tau}) = [p(\tau_1), n(\tau_2)]'$, where, for example,

$$p(\tau_1) = n^{-1/2} \sum_{i;X_{i-1}>0} X_{i-1}\{(\eta_i(\tau_1))^2 - 1\}/(\tau_1 X_{i-1}),$$

with $\eta_i(\tau_1) = (X_i - \widehat{\alpha}_p X_{i-1})/(\tau_1 X_{i-1})$. After some simplifications,

$$p(\tau_1) = c(n, \tau_1)\Big[\sum_{i;X_{i-1}>0} \{(X_i - \widehat{\alpha}_p X_{i-1})/X_{i-1}\}^2/\tau_1^2 - n_p\Big],$$

where $c(n, \tau_1)$ is a constant and n_p is the total number of positive X_{i-1}'s. Hence $p(\tau_1) = 0$ has the solution $\hat{\beta}_1 = \Big\{\sum_{i;X_{i-1}>0}\{(X_i - \widehat{\alpha}_p X_{i-1})/X_{i-1}\}^2/n_p\Big\}^{1/2}$ which estimates β_1. Similarly, $n(\tau_2) = c(n, \tau_2)\Big[\sum_{i;X_{i-1}<0}\{(X_i - \widehat{\alpha}_p X_{i-1})/X_{i-1}\}^2/\tau_2^2 - (n - n_p)\Big]$, which gives $\hat{\beta}_2 = \Big\{\sum_{i;X_{i-1}<0}\{(X_i - \widehat{\alpha}_p X_{i-1})/X_{i-1}\}^2/(n - n_p)\Big\}^{1/2}$.

To compute the Wilcoxon R-estimator, we apply the Hodges-Lehmann/Jaeckel (1972)'s formula in the approximating ARTCH model

$$\frac{X_i}{\widehat{\beta}_1 X_{i-1}} \approx \frac{1}{\widehat{\beta}_1} \times \alpha + \eta_i, \text{ when } X_{i-1} > 0,$$

and

$$\frac{X_i}{-\widehat{\beta}_2 X_{i-1}} \approx \frac{-1}{\widehat{\beta}_2} \times \alpha + \eta_i, \text{ when } X_{i-1} < 0,$$

to get

$$\widehat{\alpha}_W = \text{median } \Big\{ \frac{X_i I(X_{i-1} > 0)}{\widehat{\beta}_1 X_{i-1}} + \frac{X_j I(X_{j-1} < 0)}{\widehat{\beta}_2 X_{j-1}} \Big\} / \{(\widehat{\beta}_1)^{-1} + (\widehat{\beta}_2)^{-1}\}. \tag{5.1}$$

For $\widehat{\alpha}_S$, first order m number of $\{X_i/\widehat{\beta}_2 X_{i-1}\}$'s corresponding to negative X_{i-1}'s and call them $\{y_1, y_2, \cdots y_m\}$; here we assume that all of $\{y_1, y_2, \cdots y_m\}$ are distinct and m equals $n - n_p$ with probability one. Next order n_p number of $\{X_i/\widehat{\beta}_1 X_{i-1}\}$'s corresponding to positive X_{i-1}'s and call them $\{y_{m+1}, y_{m+2}, \cdots y_n\}$. Then from Mukherjee (2006 b), we get that if n is odd,

$$\widehat{\alpha}_S = \text{median } \{(y_j + y_i)/[(\widehat{\beta}_1)^{-1} + (\widehat{\beta}_2)^{-1}]; i + j = (n+1)/2 + m + 1, 1 \le i \le m, m+1 \le j \le n\}, \tag{5.2}$$

whereas if n is even,

$$\widehat{\alpha}_S = \text{median } \{(y_j + y_i)/[(\widehat{\beta}_1)^{-1} + (\widehat{\beta}_2)^{-1}]; i + j = (n)/2 + m + 1, 1 \le i \le m, m+1 \le j \le n\}.$$

Finally, from (3.23), the QMLE for the ARTCH model is obtained as

$$\widehat{\alpha}_{QMLE} = \{n_p(\widehat{\beta}_1)^{-2} + (n - n_p)(\widehat{\beta}_2)^{-2}\}^{-1} \Big[\sum_{i; X_{i-1} > 0} \{ \frac{X_i}{X_{i-1}\widehat{\beta}_1^2} \} + \sum_{j; X_{j-1} < 0} \{ \frac{X_j}{X_{j-1}\widehat{\beta}_2^2} \} \Big]. \tag{5.3}$$

Formulae for the ARLSCH model. From

$$X_i = \alpha X_{i-1} + \{\beta_0 + \beta_1 X_{i-1}^2\}^{1/2} \eta_i, \qquad 1 \le i \le n,$$

note that $\widehat{\alpha}_p = \sum_{i=1}^n X_i X_{i-1} / \sum_{i=1}^n X_{i-1}^2$. To estimate the scale parameters β_0 and β_1, write $M_s(\boldsymbol{\tau}) = [m_1(\boldsymbol{\tau}), m_2(\boldsymbol{\tau})]'$, where $\boldsymbol{\tau} = [\tau_0, \tau_1]'$ and with $c_i = (X_i - \widehat{\alpha}_p X_{i-1})^2$,

$$m_1(\boldsymbol{\tau}) = \sum_{i=1}^n \{c_i/(\tau_0 + \tau_1 X_{i-1}^2) - 1\}/(\tau_0 + \tau_1 X_{i-1}^2),$$

and

$$m_2(\boldsymbol{\tau}) = \sum_{i=1}^n X_{i-1}^2 \{c_i/(\tau_0 + \tau_1 X_{i-1}^2) - 1\}/(\tau_0 + \tau_1 X_{i-1}^2).$$

Write $\tilde{r} = \tau_1/\tau_0$. Then the equations $m_1(\tau) = 0 = m_2(\tau)$ can be rewritten as

$$\sum_{i=1}^n \frac{c_i}{(1 + \tilde{r} X_{i-1}^2)^2} = \tau_0 \sum_{i=1}^n \frac{1}{(1 + \tilde{r} X_{i-1}^2)}$$

and

$$\sum_{i=1}^n \frac{c_i X_{i-1}^2}{(1 + \tilde{r} X_{i-1}^2)^2} = \tau_0 \sum_{i=1}^n \frac{X_{i-1}^2}{(1 + \tilde{r} X_{i-1}^2)}.$$

Now eliminating τ_0 one can get an equation in $\tilde{r}$ which can be solved using numerical method.

To compute the Wilcoxon R-estimator, we apply Jaeckel (1972)'s formula in the approximating ARLSCH model

$$\frac{X_i}{(\widehat{\beta}_0 + \widehat{\beta}_1 X_{i-1}^2)^{1/2}} \approx \frac{X_{i-1}}{(\widehat{\beta}_0 + \widehat{\beta}_1 X_{i-1}^2)^{1/2}} \times \alpha + \eta_i$$

to get $\widehat{\alpha}_W$ as the median of the set of numbers $\{\alpha_{ij}\}$ with corresponding probability proportional to $\{p_{ij}\}$ where

$$\alpha_{ij} = \frac{Y_i - Y_j}{d_i - d_j} \text{ and } p_{ij} = d_i - d_j, \tag{5.4}$$

with $Y_i = X_i/(\widehat{\beta}_0 + \widehat{\beta}_1 X_{i-1}^2)^{1/2}$, $d_i = X_{i-1}/(\widehat{\beta}_0 + \widehat{\beta}_1 X_{i-1}^2)^{1/2}$; here p_{ij}'s are defined only for those $\{(i,j)\}$ for which $d_i - d_j > 0$.

For computing $\widehat{\alpha}_S$ we obtain from Mukherjee (2006 b) that it is the median of the set of numbers $\{\alpha_{ij}\}$ with corresponding probability proportional to $\{p_{ij}\}$ where p_{ij}'s are defined positive only for those $1 \le i, j \le n$ for which $d_i - d_j > 0$ and for which $Y_i - d_i\alpha_{ij}$ (also equal to $Y_j - d_j\alpha_{ij}$ by the definition of α_{ij}) is the "median" of the n numbers $\{Y_u - d_u\alpha_{ij}; u \neq i, j, Y_i -$

$d_i\alpha_{ij}, Y_j - d_j\alpha_{ij}\}$; for this later "median", the definition is the $n/2$-th ordered observation when n is even and as usual the $(n+1)/2$-th ordered observation when n is even.

Simulation results and analysis. For simulation, we use $r = 100$ replications. For each of the k-th replication $(1 \leq k \leq r)$, we generate a sample of size $n = 100$ from the underlying model with parameters $\alpha = 0.1, \beta_1 = 0.2, \beta_2 = 0.3$ for the ARTCH model and $\alpha = 0.1, \beta_0 = 0.2, \beta_1 = 0.3$ for the ARLSCH model and compute $\widehat{\alpha}_p(k) = \widehat{\alpha}_p$, $\widehat{\alpha}_W$, $\widehat{\alpha}_S$ and $\widehat{\alpha}_{QMLE}$. For each estimator (denoted generically by $\widehat{\alpha}(k)$), we also compute $r^{-1}\sum_{k=1}^{r}(\widehat{\alpha}(k) - \alpha)^2$ which is the average (over all replications) squared deviation of the estimate from the true parameter value α and this is an estimate of mean squared error (MSE) of $\widehat{\alpha}$. These are reported in columns (2)-(5) in Tables 1 and 2 below. Columns (6) and (8) are obtained from dividing Column (5) by Columns (3) and (4) respectively and represent the estimated ARE of $\widehat{\alpha}_W$ and $\widehat{\alpha}_S$ with respect to $\widehat{\alpha}_{QMLE}$ (denoted by E($\widehat{\alpha}_W$) etc.); Columns (7) and (9) represent the corresponding theoretical ARE of $\widehat{\alpha}_W$ and $\widehat{\alpha}_S$ as explained in Remark 3.3(ii) (denoted by T($\widehat{\alpha}_W$) etc.). For each scenario (corresponding to a particular row in the tables), we have run simulations five times under identical setup and have reported the result of that simulation which has best estimated ARE (in the sense that it is either more than or the closest to the theoretical ARE); for simulation results of all five runs and also the results when the observations were generated under different true parameters, see Mukherjee (2006 b).

Table 1 : Estimated MSE's and ARE's of the different estimators of α (ARTCH model)

g	MSE($\widehat{\alpha}_p$)	MSE($\widehat{\alpha}_W$)	MSE($\widehat{\alpha}_S$)	MSE($\widehat{\alpha}_{QMLE}$)	E($\widehat{\alpha}_W$)	T($\widehat{\alpha}_W$)	E($\widehat{\alpha}_S$)	T($\widehat{\alpha}_S$)
N	0.0544951888	0.0005477203	0.0005744816	0.0005382787	0.983	.96	0.940	.64
L	0.0458744400	0.0006679956	0.0006526314	0.0007891252	**1.181**	1.1	1.209	.82
D	0.0415501346	0.0004704636	0.0004387167	0.0007328313	1.558	1.5	**1.670**	2

Table 2 : Estimated MSE's and ARE's of the different estimators of α (ARLSCH model)

g	MSE($\widehat{\alpha}_p$)	MSE($\widehat{\alpha}_W$)	MSE($\widehat{\alpha}_S$)	MSE($\widehat{\alpha}_{QMLE}$)	E($\widehat{\alpha}_W$)	T($\widehat{\alpha}_W$)	E($\widehat{\alpha}_S$)	T($\widehat{\alpha}_S$)
N	0.01827685	0.02077171	0.02907107	0.01875711	0.903	.96	0.645	.64
L	0.02324641	0.01353362	0.02136932	0.01540861	**1.139**	1.1	0.721	.82
D	0.02168655	0.01279244	0.01333624	0.01732438	1.354	1.5	**1.300**	2

Simulation results as well as several histograms conform with our theoretical finding on the asymptotic distributions of the different estimators. In several cases, the estimated ARE is more than the theoretical ARE even at much smaller value of n. In particular, the estimated AREs of $\widehat{\alpha}_W$ at the logistic density are 1.181 and 1.139 for the ARTCH and ARLSCH models respectively, exceeding the theoretical ARE of 1.10 which, from (3.25), represents the relative efficiency of the optimal R-estimator with respect to the QMLE. However, the estimated AREs of $\widehat{\alpha}_S$ at the double-exponential density are 1.670 and 1.300 for the ARTCH and ARLSCH models respectively which are far below the theoretical relative efficiency of 2 of the optimal R-estimator with respect to the QMLE. A plausible reason for this could be that $n = 100$ may not be 'large enough' for asymptotics to hold at the double-exponential density.

In many other simulations not reported here with different combinations of the underlying parameters, it was observed that the ARE-results for $\widehat{\alpha}_W$ and $\widehat{\alpha}_S$ approximately hold even when the models are nonstationary. In general, to a practitioner, we recommend the use of $\widehat{\alpha}_W$ as a good alternative to the QMLE which has high ARE for a wide number of distributions with a 'small sacrifice' at the normal distribution. Hence, in the real data examples below, we use only $\widehat{\alpha}_W$ and $\widehat{\alpha}_{QMLE}$ for our analysis.

Financial Data. Tsay (2002, Chapter 3 on Conditional Heteroscedastic Models) have analyzed three important data sets, namely, (A) The monthly log stock returns of the Intel Corporation from 1973 to 1997 (300 observations with first value 0.010050 and last value -0.095008), (B) The monthly excess returns of S & P 500 from 1926 to 1991 (792 observations with first value 0.0225 and last value 0.1116) and (C) The monthly log returns of IBM stock from 1926 to 1999 (888 observations with first value 1.0434 and last value 4.5633) and fitted various types of conditional heteroscedastic models to them. These data can be found in

http://www.gsb.uchicago.edu/fac/ruey.tsay/teaching/fts/m-intc.dat

For Data A, denoted by $\{X_i; 0 \leq i \leq n = 299\}$, Tsay's analysis of the autocorrelation function (ACF) of log returns, absolute log returns and squared log returns suggests that monthly returns are serially uncorrelated but dependent. The mean, median, standard deviation and kurtosis of $\{X_i\}$ are 0.0286162, 0.019202, 0.1297513 and 3.370, respectively. Other exploratory analysis show presence of heavy tails.

Next we fitted the centered $\{X_i\}$ with the ARLSCH model. We use (5.4) and other related formulae from the previous subsection to compute $\widehat{\alpha}_W$. For estimating its standard error (SE), we estimate (i) $M(\boldsymbol{\theta})$ using the lhs of (3.2) with $\boldsymbol{\beta}$ replaced by $\widehat{\boldsymbol{\beta}}$ and (ii) $\int g^2(x)dx$ using the standardized residuals $\{\eta_i(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}); 1 \leq i \leq n\}$. For the integral, we use simple histogram estimator of g by dividing $[\min\{\eta_i(\widehat{\alpha}, \widehat{\boldsymbol{\beta}})\}, \max\{\eta_i(\widehat{\alpha}, \widehat{\boldsymbol{\beta}})\}]$ into an ad hoc choice of $m = 15$ equal intervals over each of which the estimate of g is constant and then estimate the integral based on the integral of the step function. For estimating the SE of $\widehat{\alpha}_{QMLE}$ using (3.24), we use a formula similar to (3.2). The efficiency of the R-estimator is defined as the square of the ratio of two estimated SE's. The estimates are reported in Table 3 below.

Table 3 : Estimates of α for the Intel Corporation data based on the ARLSCH model.

Auxiliary Estimates	$\widehat{\alpha}_{QMLE}$	$\widehat{\alpha}$	Efficiency
$\widehat{\alpha}_p = 0.05654418$	0.05174328	0.05043456	1.18223914
$\hat{\beta}_0 = 0.01052003$	SE=0.05779476	SE=0.05315395	
$\hat{\beta}_1 = 0.4322009$			

Tsay (2002, Example 3.1) used a standard ARCH model (where $p = 1$) with intercept to analyze this data. Using $X_i = \mu + a_i$ with $a_i = \sigma_{i-1}(\boldsymbol{\beta})\varepsilon_t$, where $\sigma^2_{i-1}(\boldsymbol{\beta}) = \beta_0 + \beta_1 a^2_{i-1}$,

$1 \leq i \leq n = 299$, Tsay (2002) obtained $\hat{\mu} = 0.0213, \hat{\beta}_0 = 0.00998$ and $\hat{\beta}_1 = 0.4437$ using the QMLE. Note that our estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ of the variance parameters are quite close to those of Tsay. The differences are due to the fact that we used centered (mean-subtracted) observations and used preliminary estimate of α before estimating the variance parameters with $\kappa(x) = x$. Introduction of the autoregressive term 'α' seems to have misspecified the model for this data. This is reflected in the studentized ratio of $\hat{\alpha}_W$ which equals 0.95 and hence the null hypothesis $\alpha = 0$ is not significant. The conclusion remains same using the studentized ratio of $\hat{\alpha}_{QMLE}$ also.

Asymmetry is an inherent feature in the financial market as the market seems to be more sensitive to a negative news. Usual ARCH model of volatility may not capture this feature because of its symmetric dependence on the past values in the form of squares. Sometimes an ARTCH model with $\{\beta_{2j-1} \neq \beta_{2j}; 1 \leq j \leq p\}$ may be a reasonable model to capture such asymmetry. Hence we now fit an ARTCH model with $p = s = 1$ to Data A where we estimate the parameter α using $\hat{\alpha}_W$ and $\hat{\alpha}_{QMLE}$. Formulae (5.1) and (5.3) yield the following estimates.

Table 4 : Estimates of α for the Intel Corporation data based on the ARTCH model.

Auxiliary Estimates	$\widehat{\alpha}_{QMLE}$	$\widehat{\alpha}_W$	Efficiency
$\widehat{\alpha}_p = 0.05654418$	0.40840697	0.03742153	8.73970912
$\hat{\beta}_1 = 8.84610615$	SE=0.57839845	SE=0.19564945	
$\hat{\beta}_2 = 11.46261069$			

The asymmetric feature of the data set is reflected by the fact that $\hat{\beta}_1 < \hat{\beta}_2$. Since $\hat{\beta}_2/\hat{\beta}_1 = 1.296$, impact of a negative shock is about 29.6% higher than that of a positive shock of the same magnitude. Also, similar to the ARLSCH fitting, α is not significant using both $\widehat{\alpha}_W$ and $\widehat{\alpha}_{QMLE}$ as the model is misspecified. For both models, the R-estimator turned out to be much more efficient (in the sense of smaller estimated MSE) than the commonly-used $\widehat{\alpha}_{QMLE}$.

Next consider Data B denoted by $\{X_i; 0 \le i \le n = 791\}$. Similar analysis with the ARLSCH and ARTCH models yields the following estimates of the parameters reported in Tables 5 and 6. Tsay (2002, Example 3.3) fitted an AR(3)-GARCH(1, 1) model to this data and the joint estimation of the parameters in the model yields 0.021 as the estimate of the intercept at lag 1. Clearly, in the ARLSCH model, $\widehat{\alpha}_W$ is closer to this estimate than $\widehat{\alpha}_{QMLE}$. However, as in Tsay, the coefficient is insignificant using both $\widehat{\alpha}_W$ and $\widehat{\alpha}_{QMLE}$.

Table 5 : Estimates of α for the S & P 500 data based on the ARLSCH model.

Auxiliary Estimates	$\widehat{\alpha}_{QMLE}$	$\widehat{\alpha}_W$	Efficiency
$\widehat{\alpha}_p = 0.09023211$	0.03311225	0.01982906	1.27527764
$\hat{\beta}_0 = 0.002768820$	SE=0.03558038	SE=0.03150709	
$\hat{\beta}_1 = 0.1657376$			

Table 6 : Estimates of α for the S & P 500 data based on the ARTCH model.

Auxiliary Estimates	$\widehat{\alpha}_{QMLE}$	$\widehat{\alpha}_W$	Efficiency
$\widehat{\alpha}_p = 0.09023211$	-0.52288761	0.04611074	2.31178836
$\hat{\beta}_1 = 12.51359399$	SE= 0.51621701	SE=0.33951446	
$\hat{\beta}_2 = 18.27277865$			

For both Data sets A and B, we observe that the R-estimate and the QMLE of the autoregressive parameter α are small and turned out to be 'not significant' while fitting the ARLSCH model; hence there was little for the R-estimator to target other than concluding that the model is misspecified. However, under the ARTCH model, the absolute values of the QMLE are higher than $\widehat{\alpha}_W$ for both data sets. As the inclusion of the autoregressive parameter seems to have misspecified the model, the R-estimate resulted in a small value rightfully while the QMLE resulted in high value. Moreover, R-estimators are highly efficient compared to the QMLE in terms of smaller estimated MSE for both models and data sets with estimated relative efficiency well above one.

Finally, we consider Data (C). Tsay (2002, Example 3.5) fitted an AR(1) model with GARCH error to this data to obtain the estimate of the autoregressive parameter as 0.099

with SE 0.037 and the model seemed to be adequate. We use the ARLSCH model to get the preliminary estimate $\widehat{\alpha}_p = 0.10601551$ and the R-estimate $\widehat{\alpha}_W = 0.10864080$ with SE 0.01903097. Therefore the intercept parameter is close to Tsay's estimate and is significant in accord with Tsay's result. However, the QMLE turns out to be $\widehat{\alpha}_{QMLE} = 0.31733076$ with SE 0.09571206 and is very different than the estimate obtained by Tsay using the QMLE of AR(1)-GARCH model. This shows that $\widehat{\alpha}_W$ is more robust to the specification between the ARCH or GARCH model than $\widehat{\alpha}_{QMLE}$. Moreover, the estimated ARE of the R-estimator wrt the QMLE is as high as 25.29363788.

Let $L(k)$ denote the Ljung-Box statistic with lag k for the portmanteau test of the randomness of the residuals. Using the R-estimate for residuals, the Ljung-Box statistics turn out to be $L(10) = 6.8387$ and $L(20) = 15.0339$ while using the QMLE for residuals, $L(10) = 6.9607$ and $L(20) = 14.7694$. Since the Ljung-Box statistics have high p-values, the ARLSCH model seems to be adequate using both R-estimate and the QMLE.

Next we appeal to the asymmetric feature of Data C. Tsay (2002, Section 3.7.2) fitted an AR(1)-EGARCH model to this data to obtain the estimate of the autoregressive parameter as 0.092. Fitting an ARTCH model to this data, we obtain the preliminary estimate 0.10601551 and $\widehat{\alpha}_W = 0.09289947$ with SE 0.14118706. However, the QMLE is very different from the R-estimate and Tsay's comparable estimate with value $\widehat{\alpha}_{QMLE} = 0.41444369$ and SE 0.26747658. Note that the intercept parameter appears to be not significant using both estimates. Using the Ljung-Box statistics, with rank-estimate for residuals $L(10) = 7.0857$ and $L(20) = 31.7230$ while with the QMLE, $L(10) = 7.4309$ and $L(20) = 31.3810$ and the ARTCH model seems to be adequate. This shows, as before, that the R-estimator performs better with model misspecification between the ARTCH and the EGARCH models. Moreover, the estimated ARE of the R-estimator is 3.58906858.

6 Proofs

Proof of Theorem 3.1. Clearly

$$n^{1/2}(\widehat{\boldsymbol{\alpha}}_p - \boldsymbol{\alpha}) = \left[n^{-1}\sum_{i=1}^n \boldsymbol{Y}_{i-1}\boldsymbol{Y}'_{i-1}\right]^{-1}\left[n^{-1/2}\sum_{i=1}^n \boldsymbol{Y}_{i-1}\sigma(\boldsymbol{Y}_{i-1},\boldsymbol{\beta})\eta_i\right],$$

and so the result follows by applying the martingale CLT on the second term. ⊥⊥

In the following, for two sequences of vector-valued stochastic processes $\{X_n(.)\}$ and $\{Y_n(.)\}$, we write $X_n(\boldsymbol{t}) = u_p(1)$, if $\forall b > 0 \;\; \epsilon > 0,\;\; P[\sup\{||X_n(\boldsymbol{t})||; ||\boldsymbol{t}|| \le b\} > \epsilon] = o(1)$ and $X_n(\boldsymbol{t}) = Y_n(\boldsymbol{t}) + u_p(1)$ if $X_n(\boldsymbol{t}) - Y_n(\boldsymbol{t}) = u_p(1)$.

Proof of Lemma 3.1. The proof of this uses a simple Taylor expansion of the function $x\kappa(x)$ as follows. Fix a $0 < b < \infty$. Let $h(x) = x\kappa(x)$, and for a $\boldsymbol{t} = (\boldsymbol{t}'_1, \boldsymbol{t}'_2)' \in I\!R^m$, $\|\boldsymbol{t}\| \le b$, let

$$\tilde{\eta}_i(\boldsymbol{t}) := \eta_i(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t}) = [X_i - \boldsymbol{Y}'_{i-1}(\boldsymbol{\alpha} + n^{-1/2}\boldsymbol{t}_1)]/\sigma(\boldsymbol{Y}_{i-1}, \boldsymbol{\beta} + n^{-1/2}\boldsymbol{t}_2).$$

Recall that $r_{ni}(\boldsymbol{t}_2) = \dot{\sigma}_{ni}(\boldsymbol{t}_2)/\sigma_{ni}(\boldsymbol{t}_2)$. Then

$$\begin{aligned} & M_s(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t}) - M_s(\boldsymbol{\theta}) \\ = \;& n^{-1/2}\sum_{i=1}^n r_{ni}(\boldsymbol{t}_2)[h(\tilde{\eta}_i(\boldsymbol{t})) - h(\eta_i)] + n^{-1/2}\sum_{i=1}^n [r_{ni}(\boldsymbol{t}_2) - r_i][h(\eta_i) - 1] \\ = \;& M_1(\boldsymbol{t}) + M_2(\boldsymbol{t}), \qquad \text{say.} \end{aligned}$$

Using the second differentiability of κ, $M_1(\boldsymbol{t}) = n^{-1/2}\sum_{i=1}^n r_{ni}(\boldsymbol{t}_2)[\tilde{\eta}_i(\boldsymbol{t}) - \eta_i]\dot{h}(\eta_i) + u_p(1)$, where $\dot{h}(\eta_i) = \eta_i\dot{\kappa}(\eta_i) + \kappa(\eta_i)$. Next, using $\sigma_{ni}(0) = 1$, rewrite

$$\begin{aligned} \eta_i(\boldsymbol{t}) - \eta_i &= \frac{[\eta_i - (\mu_{ni}(\boldsymbol{t}_1) - \mu_i(0))]}{\sigma_{ni}(\boldsymbol{t}_2)} - \eta_i \\ &= -\frac{\sigma_{ni}(\boldsymbol{t}_2) - 1}{\sigma_{ni}(\boldsymbol{t}_2)}\,\eta_i - \frac{\mu_{ni}(\boldsymbol{t}_1) - \mu_i(0)}{\sigma_{ni}(\boldsymbol{t}_2)}. \end{aligned}$$

Therefore, the leading term in the above approximation of M_1 can be further rewritten a $M_{11}(\boldsymbol{t}) + M_{12}(\boldsymbol{t})$, where

$$M_{11}(\boldsymbol{t}) \;=\; -n^{-1/2}\sum_{i=1}^n r_{ni}(\boldsymbol{t}_2)\frac{\sigma_{ni}(\boldsymbol{t}_2) - 1}{\sigma_{ni}(\boldsymbol{t}_2)}\,\eta_i\dot{h}(\eta_i) = -\dot{\Sigma}(\boldsymbol{\theta})\,t_2\,E[\eta\dot{h}(\eta)] + u_p(1),$$

$$M_{12}(\boldsymbol{t}) \;=\; -n^{-1/2}\sum_{i=1}^{n} r_{ni}(\boldsymbol{t}_2)\frac{\mu_{ni}(\boldsymbol{t}_1)-\mu_i}{\sigma_{ni}(\boldsymbol{t}_2)}\,\dot{h}(\eta_i) = -\boldsymbol{G}(\boldsymbol{\theta})\,\boldsymbol{t}_1\,E[\dot{h}(\eta)] + u_p(1).$$

In the above approximations, the conditions (3.3)-(3.4) are used. Similarly, one obtains $M_2(\boldsymbol{t}) = u_p(1)$, thereby completing the proof of the Lemma. $\perp\perp$

The proof of Lemma 3.2 depends on the following technical result.

Let $\{(\eta_i,\, \gamma_{ni}, \delta_{ni}, \xi_{ni}),\ 1 \le i \le n\}$ be an array of 4-tuple r.v.'s defined on a probability space such that $\{\eta_i,\ 1 \le i \le n\}$ are i.i.d. according to a d.f. G, and for each $1 \le i \le n$, η_i is independent of $(\gamma_{ni}, \delta_{ni}, \xi_{ni})$. Let $\{\mathcal{A}_{ni}; 1 \le i \le n\}$ be an array of sub-σ-fields such that $\mathcal{A}_{ni} \subset \mathcal{A}_{ni+1},\ \mathcal{A}_{ni} \subset \mathcal{A}_{n+1i},\ 1 \le i \le n,\ n \ge 1$; $(\gamma_{n1}, \delta_{n1}, \xi_{n1})$ is $\mathcal{A}_{n1}$ measurable, and $\{\{(\gamma_{ni}, \delta_{ni}, \xi_{ni}); 1 \le i \le j\}, \eta_1, \eta_2, \ldots, \eta_{j-1}\}$ are $\mathcal{A}_{nj}$ measurable, $2 \le j \le n$. Define the following processes for $x \in I\!R$.

$$\tilde{V}_n(x) := n^{-1/2}\sum_{i=1}^{n}\gamma_{ni} I(\eta_i < x + x\delta_{ni} + \xi_{ni}), \tag{6.1}$$

$$\tilde{J}_n(x) := n^{-1/2}\sum_{i=1}^{n}\gamma_{ni} G(x + x\delta_{ni} + \xi_{ni}),$$

$$V_n(x) := n^{-1/2}\sum_{i=1}^{n}\gamma_{ni} I(\eta_i < x + \xi_{ni}), \qquad J_n(x) := n^{-1/2}\sum_{i=1}^{n}\gamma_{ni} G(x + \xi_{ni}),$$

$$V_n^*(x) := n^{-1/2}\sum_{i=1}^{n}\gamma_{ni} I(\eta_i \le x), \qquad J_n^*(x) := n^{-1/2}\sum_{i=1}^{n}\gamma_{ni} G(x),$$

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

We assume that the following conditions are satisfied by the weights $\{\gamma_{ni}\}$ and the perturbations $\{\delta_{ni}, \xi_{ni}\}$.

Let $C_n := \sum E|\gamma_{ni}|^q$. Then for some $q > 2$ and ϵ, with $0 < \epsilon < q/2$,

$$C_n / n^{q/2-\epsilon} = o(1). \tag{6.2}$$

$$\left(n^{-1}\sum_{i=1}^{n}\gamma_{ni}^2\right)^{1/2} = \gamma + o_p(1), \quad \gamma \text{ a positive r.v.} \tag{6.3}$$

$$E\left(n^{-1}\sum_{i=1}^{n}\gamma_{ni}^2\right)^{q/2} = O(1). \tag{6.4}$$

$$\max_{1\le i\le n} n^{-1/2}|\gamma_{ni}| = o_p(1). \tag{6.5}$$

$$\text{(a)}\quad \max_{1\le i\le n}|\xi_{ni}| = o_p(1), \qquad \text{(b)}\quad \max_{1\le i\le n}|\delta_{ni}| = o_p(1). \tag{6.6}$$

$$\frac{n^{q/2-\epsilon}}{C_n} E\left[n^{-1}\sum_{i=1}^{n}\{\gamma_{ni}^2(|\xi_{ni}|+|\delta_{ni}|)\}\right]^{q/2} = o(1). \tag{6.7}$$

$$\text{(a)}\quad n^{-1/2}\sum_{i=1}^{n}|\gamma_{ni}\xi_{ni}| = O_p(1), \qquad \text{(b)}\quad n^{-1/2}\sum_{i=1}^{n}|\gamma_{ni}\delta_{ni}| = O_p(1). \tag{6.8}$$

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

Theorem 6.1. *Under the above setup and under the assumptions (6.2)-(6.8) and (G.1)-(G.3),*

$$\sup_{x\in I\!R}|\tilde{U}_n(x) - U_n(x)| = o_p(1), \tag{6.9}$$

$$\sup_{x\in I\!R}|\tilde{U}_n(x) - U_n^*(x)| = o_p(1). \tag{6.10}$$

Proof. The proof of such uniform approximation theorem depends on efficient partitioning of the real line; here pointwise convergence can be shown easily and then we invoke the monotone structure of the empirical processes to achieve the uniform convergence. The uniform closeness of the processes U_n and U_n^* was proved in Koul and Ossiander (1994, Theorem 1.1), under the assumption that G has uniformly continuous positive density g, and under (6.3), (6.5), (6.6)(a) and (6.8)(a). Thus, the claim (6.10) is a consequence of that theorem and (6.9).

To prove (6.9), assume without loss of generality that all γ_{ni} are non-negative. Next, write $\tilde{U}_n(x) = \tilde{U}_n^+(x) + \tilde{U}_n^-(x)$, where $\tilde{U}_n^+(x)$, $\tilde{U}_n^-(x)$ correspond to that part of the sum in $\tilde{U}_n(x)$ which has $\delta_{ni} \ge 0$, $\delta_{ni} < 0$, respectively. Decompose $U_n(x)$ similarly. It thus suffices to show that

$$\sup_{x\in I\!R}|\tilde{U}_n^+(x) - U_n^+(x)| = o_p(1), \tag{6.11}$$

$$\sup_{x\in I\!R}|\tilde{U}_n^-(x) - U_n^-(x)| = o_p(1). \tag{6.12}$$

Details will be given only for (6.11), they being similar for (6.12).

Fix a $\delta > 0$ and let $C_n = C_n(q) := \sum_{i=1}^{n} E\gamma_{ni}^q$. Let $-\infty = x_0 < x_1 \leq \ldots \leq x_{r_n-1} \leq x_{r_n} = \infty$ be a *weight-dependent partition* of $I\!R$ where $x_j = G^{-1}(j\delta C_n/n^{(q/2)-\epsilon})$, $0 \leq j \leq r_n - 1$ and $r_n := [n^{q/2-\epsilon}/(C_n\delta)] + 1$, with $[x]$ denoting the integer part of x. Note that

$$[G(x_j) - G(x_{j-1})] \leq \delta C_n/n^{q/2-\epsilon}, \ \forall\, 1 \leq j \leq r_n. \tag{6.13}$$

The dependence of x_j's on n, δ and q is suppressed for the sake of convenience.

Using the monotonicity of the indicator function and the d.f. G, we obtain that for $x_{j-1} < x \leq x_j$,

$$\begin{aligned} &|\tilde{U}_n^+(x) - U_n^+(x)| \\ \leq\ & |\tilde{U}_n^+(x_j) - U_n^+(x_{j-1})| + |\tilde{U}_n^+(x_{j-1}) - U_n^+(x_j)| \\ &\quad + 2\,|\tilde{J}_n^+(x_j) - \tilde{J}_n^+(x_{j-1})| + 2\,|J_n^+(x_j) - J_n^+(x_{j-1})| \\ =\ & |\mathcal{A}_{nj,1}| + |\mathcal{A}_{nj,2}| + 2|\mathcal{A}_{nj,3}| + 2|\mathcal{A}_{nj,4}|, \qquad \text{say.} \end{aligned} \tag{6.14}$$

Note that the number of partitions varies with n; nevertheless, intuitively, we show the convergence of the j-th partition and consequently, the uniform convergence over it. First, consider $\mathcal{A}_{nj,1}$. For the sake of brevity, let $t_{ni} = \delta_{ni} + 1$. Then, one can rewrite $\mathcal{A}_{nj,1}$ as

$$n^{-1/2}\sum_{i=1}^{n}\gamma_{ni}\Big\{I(\eta_i < x_j t_{ni} + \xi_{ni}) - I(\eta_i < x_{j-1} + \xi_{ni}) - G(x_j t_{ni} + \xi_{ni}) + G(x_{j-1} + \xi_{ni})\Big\},$$

which is a sum of martingale differences. We need the following inequality on the tail probability of a sum of martingale differences; see Hall and Heyde (1980, Corollary 2.1 and Theorem 2.12).

Rosenthal Inequality. *Suppose $M_j = \sum_{i=1}^{j} D_i$ is a sum of martingale differences with respect to the underlying increasing filtration $\{\mathcal{D}_i\}$ and $q \geq 2$. Then, there exists a constant $C = C(q)$ such that for any $\epsilon > 0$,*

$$\begin{aligned} P[|M_n| > \epsilon] &\leq P\left[\max_{1\leq j\leq n}|M_j| > \epsilon\right] \\ &\leq\ C\epsilon^{-q}\left[\sum_{i=1}^{n}E|D_i|^q + E\left\{\sum_{i=1}^{n}E(D_i^2|\mathcal{D}_{i-1})\right\}^{q/2}\right]. \end{aligned}$$

Apply the above inequality with $\mathcal{D}_0 = \sigma < \gamma_{n1}, \delta_{n1}, \xi_{n1} >$ and for $2 \le i \le n$, $\mathcal{D}_{i-1} = \sigma < \eta_1, \ldots, \eta_{i-1}; (\gamma_{nj}, \delta_{nj}, \xi_{nj}), 1 \le j \le i >$; $D_i = n^{-1/2}\gamma_{ni}\Big\{I(x_{j-1} + \xi_{ni} \le \eta_i < x_j t_{ni} + \xi_{ni}) - G(x_j t_{ni} + \xi_{ni}) + G(x_{j-1} + \xi_{ni})\Big\}$. Use $|D_i| \le n^{-1/2}|\gamma_{ni}|$, and the fact $E(D_i^2|\mathcal{D}_{i-1}) \le n^{-1}\gamma_{ni}^2\{|G(x_j t_{ni} + \xi_{ni}) - G(x_{j-1} + \xi_{ni})|\}$, to obtain

$$
\begin{aligned}
& P[|\mathcal{A}_{nj,1}| > \epsilon] \\
& \le \; C\epsilon^{-q}n^{-q/2}C_n + C\epsilon^{-q}E\left[n^{-1}\sum_{i=1}^n \gamma_{ni}^2\{|G(x_j t_{ni} + \xi_{ni}) - G(x_{j-1} + \xi_{ni})|\}\right]^{q/2}.
\end{aligned}
$$

The first term in the above inequality is free from j. Next, we shall obtain an upper-bound (free of j) for the second term using (i) the Taylor expansion of G and the boundedness of g, and (ii) assumptions (G.2) as follows.

$$
\begin{aligned}
& \sum_{i=1}^n \gamma_{ni}^2\{|G(x_j t_{ni} + \xi_{ni}) - G(x_{j-1} + \xi_{ni})|\} \\
& \le \; \sum_{i=1}^n \gamma_{ni}^2\{|G(x_j) - G(x_{j-1})|\} + \sum_{i=1}^n \gamma_{ni}^2\{|G(x_j t_{ni} + \xi_{ni}) - G(x_j t_{ni})|\} \\
& + \quad \sum_{i=1}^n \gamma_{ni}^2\{|G(x_j t_{ni}) - G(x_j)|\} + \sum_{i=1}^n \gamma_{ni}^2\{|G(x_{j-1} + \xi_{ni}) - G(x_{j-1})|\} \\
& \le \; K_1 \sum_{i=1}^n \gamma_{ni}^2 \Big[\delta\, C_n n^{-(q/2-1)} \; + \; 2\,|\xi_{ni}| \; + \; |\delta_{ni}|\,\Big].
\end{aligned}
$$

The above bound is obtained by using (6.13) for the first term, the boundedness of g for the 2nd and 4th terms, and (G.2) for the 3rd term. Now using the so called 'C_r'-inequality

$$
\begin{aligned}
& \Big[n^{-1}\sum_{i=1}^n \gamma_{ni}^2\{|G(x_j t_{ni} + \xi_{ni}) - G(x_{j-1} + \xi_{ni})|\}\Big]^{q/2} \\
& \le \; K_2[C_n n^{-(q/2-\epsilon)}\delta n^{-1}\sum_{i=1}^n \gamma_{ni}^2]^{q/2} + K_2\Big[n^{-1}\sum_{i=1}^n \gamma_{ni}^2\{|\xi_{ni}| \; + \; |\delta_{ni}|\}\Big]^{q/2}.
\end{aligned}
$$

Hence, using $r_n = O(n^{q/2-\epsilon}/C_n)$, for some constant $C(\delta) > 0$,

$$
\begin{aligned}
& P\left(\max_{1 \le j \le r_n} |\mathcal{A}_{nj,1}| > \epsilon\right) \\
& \le \; C(\delta)\Bigg[C\epsilon^{-q}n^{-q/2}C_n \times \frac{n^{q/2-\epsilon}}{C_n} + \left\{\frac{C_n}{n^{(q/2-\epsilon)}}\right\}^{q/2-1} (\delta)^{q/2}\, E[n^{-1}\sum_{i=1}^n \gamma_{ni}^2]^{q/2} \\
& + \; \frac{n^{q/2-\epsilon}}{C_n}E\Big\{n^{-1}\sum_{i=1}^n \gamma_{ni}^2(|\xi_{ni}| + |\delta_{ni}|)\Big\}^{q/2}\Bigg] = o(1), \qquad (6.15)
\end{aligned}
$$

using (6.2), (6.4) and (6.7). This implies that $\max_{1\le j\le r}|\mathcal{A}_{nj,1}| = o_p(1)$. Note that for (6.15) to hold, the order of the total number of partitions r_n is carefully chosen. A similar statement holds for $\mathcal{A}_{nj,2}$. Next,

$$\begin{aligned}
\mathcal{A}_{nj,3} &= n^{-1/2}\sum_{i=1}^{n}\gamma_{ni}\Big[G(x_j t_{ni}+\xi_{ni}) - G(x_{j-1}t_{ni}+\xi_{ni})\Big] \\
&= \Big\{n^{-1/2}\sum_{i=1}^{n}\gamma_{ni}[G(x_j) - G(x_{j-1})] + n^{-1/2}\sum_{i=1}^{n}\gamma_{ni}[G(x_j t_{ni}+\xi_{ni}) - G(x_j t_{ni})] \\
&\quad +n^{-1/2}\sum_{i=1}^{n}\gamma_{ni}[G(x_j t_{ni}) - G(x_j)] - n^{-1/2}\sum_{i=1}^{n}\gamma_{ni}[G(x_{j-1}t_{ni}) - G(x_{j-1})] \\
&\quad -n^{-1/2}\sum_{i=1}^{n}\gamma_{ni}[G(x_{j-1}t_{ni}+\xi_{ni}) - G(x_{j-1}t_{ni})]\Big\}
\end{aligned}$$

Hence

$$\begin{aligned}
|\mathcal{A}_{nj,3}| &\le \Big\{(n^{-1/2}\sum_{i=1}^{n}\gamma_{ni})\frac{C_n}{n^{q/2-\epsilon}}\delta + n^{-1/2}\sum_{i=1}^{n}\gamma_{ni}|G(x_j t_{ni}+\xi_{ni}) - G(x_j t_{ni}) - \xi_{ni}g(x_j t_{ni})| \\
&\quad +n^{-1/2}\sum_{i=1}^{n}\gamma_{ni}|G(x_j t_{ni}) - G(x_j) - \delta_{ni}x_j g(x_j)| \\
&\quad + n^{-1/2}\sum_{i=1}^{n}\gamma_{ni}|G(x_{j-1}t_{ni}) - G(x_{j-1}) - \delta_{ni}x_{j-1}g(x_{j-1})| \\
&\quad +n^{-1/2}\sum_{i=1}^{n}\gamma_{ni}|G(x_{j-1}t_{ni}+\xi_{ni}) - G(x_{j-1}t_{ni}) - \xi_{ni}g(x_{j-1}t_{ni})| \\
&\quad +n^{-1/2}\sum_{i=1}^{n}\gamma_{ni}\xi_{ni}|g(x_j t_{ni}) - g(x_{j-1}t_{ni})| \\
&\quad +n^{-1/2}\sum_{i=1}^{n}\gamma_{ni}\delta_{ni}|x_j g(x_j) - x_{j-1}g(x_{j-1})|\Big\}.
\end{aligned}$$

Now, let $m_n := \max_{1\le i\le n}|\xi_{ni}|$, $\mu_n := \max_{1\le i\le n}|\delta_{ni}|$. Note that the sum of the absolute values of the second and fifth term in the right hand side of the above equation is bounded above by

$$n^{-1/2}\sum_{i=1}^{n}|\gamma_{ni}\xi_{ni}| \sup_{|x-y|\le m_n}|g(x)-g(y)| = o_p(1),$$

uniformly in $j = 1,\ldots,m$, by the uniform continuity of g and (6.8)(a).

Next we handle the third term; the fourth term can be handled similarly. By the one-step Taylor expansion of G with remainder in the integral form, for all large n such that

$\max\{|\delta_{ni}|; 1 \le i \le n\}$ is sufficiently small, the absolute value of the third term is bounded by

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^{n} |\gamma_{ni}\delta_{ni}x_j| \int_0^1 |g(x_j + tx_j\delta_{ni}) - g(x_j)|dt \\ & \le n^{-1/2} \sum_{i=1}^{n} |\gamma_{ni}\delta_{ni}| \sup\{|x| \int_0^1 |g(x + tx\delta) - g(x)|dt; x \in I\!R\} = o_p(1), \end{aligned}$$

by (G.3) and (6.8)(b).

Finally, consider the sixth term; the seventh one can be dealt with similarly. To begin with observe that by (G.2), $\max_{1\le i\le n, 1\le j\le r_n} |G(x_j t_{ni}) - G(x_j)| \le C \max_{1\le i\le n} |\delta_{ni}|$, and hence by (G.1), (6.6)(b) and (6.8)(b), $\max_{1\le i\le n, 1\le j\le r_n} |g(x_j t_{ni}) - g(x_j)| = o_p(1)$. Upon combining all these bounds and using $E(n^{-1}\sum_{i=1}^n \gamma_{ni}) = O(1)$, we obtain

$$\max_{1\le j\le m} |\mathcal{A}_{nj,3}| \le O_p(1)\, o(1)\, \delta + o_p(1).$$

A similar result holds for $\mathcal{A}_{nj,4}$. All the above facts together with the arbitrariness of δ thus imply (6.9), thereby completing the proof of the lemma. $\perp\perp$

Remark 6.1. Boldin (1998) proved an analog of (6.10) for the ordinary residual empirical processes in Engle's ARCH model with $p = 1$, using a different method of proof. Koul and Mukherjee (2002) also proved an anologous result using more stringent moment assumptions.

Proof of Lemma 3.2.. Fix a $0 < b < \infty$. Observe that if in (6.1), we take

$$\gamma_{ni} = l_{ni}(\boldsymbol{t}), \quad \delta_{ni} = v_{ni}(\boldsymbol{t}), \quad \xi_{ni} = u_{ni}(\boldsymbol{t}), \quad 1 \le i \le n, \tag{6.16}$$

then, $\tilde{U}_n(x)$ and $U_n^*(x)$ are, respectively equal to $\tilde{\mathcal{U}}(x, \boldsymbol{t})$ and $\mathcal{U}^*(x, \boldsymbol{t})$, for all $x \in I\!R$, $\boldsymbol{t} \in I\!R^m$. Clearly the assumptions (3.7)-(3.13) for each fixed $\boldsymbol{t}$ imply (6.2)-(6.8). Hence, (6.10) implies that for each $\boldsymbol{t} \in I\!R^m$,

$$\sup_{x\in I\!R} |\tilde{\mathcal{U}}(x, \boldsymbol{t}) - \mathcal{U}^*(x, \boldsymbol{t})| = o_p(1). \tag{6.17}$$

The uniform convergence with respect to $\boldsymbol{t}$ over compact sets can be proved as in Koul (1996) and Koul and Mukherjee (2002) using the last two assumptions (3.14) and (3.15) which are related to the smoothness assumptions on the weights. $\perp\perp$

Proof of Lemma 3.3. Using $\varphi(y)-\varphi(0)=\int_0^1 I(y\geq u)\varphi(du)$, and $R_{i\boldsymbol{\tau}}=nG_{n\boldsymbol{\tau}}\{\eta_i(\boldsymbol{\tau})\}$, where $G_{n\boldsymbol{\tau}}$ is the empirical distribution function based on $\{\eta_j(\boldsymbol{\tau}),\,1\leq j\leq n\}$ we get

$$\begin{aligned}\varphi\left(\frac{R_{i\boldsymbol{\tau}}}{n+1}\right)-\varphi(0)=\int_0^1 I\left(\frac{R_{i\boldsymbol{\tau}}}{n+1}\geq u\right)\varphi(du) &= \int_0^1 I\Bigg(G_{n\boldsymbol{\tau}}(\eta_i(\boldsymbol{\tau}))\geq (n+1)u/n\Bigg)\varphi(du)\\ &= \int_0^1 I\Bigg(\eta_i(\boldsymbol{\tau})\geq G^{-1}_{n\boldsymbol{\tau}}\{(n+1)u/n\}\Bigg)\varphi(du),\end{aligned}$$

where for any distribution function H, $H^{-1}(u)=\inf\{x;u\leq H(x)\}$, $0<u<1$. In the following, suppressing the dependence of $\boldsymbol{Z}_{i-1}$ on $\boldsymbol{\tau}_2$, we get

$$\begin{aligned}&n^{1/2}\boldsymbol{S}_\varphi(\boldsymbol{\tau})\\ &=\sum_{i=1}^n(\boldsymbol{Z}_{i-1}(\boldsymbol{\tau})-\bar{\boldsymbol{Z}}(\boldsymbol{\tau}))\varphi\left(\frac{R_{i\boldsymbol{\tau}}}{n+1}\right)\\ &=\sum_{i=1}^n(\boldsymbol{Z}_{i-1}-\bar{\boldsymbol{Z}})\{\varphi\left(\frac{R_{i\boldsymbol{\tau}}}{n+1}\right)-\varphi(0)\}\\ &=\int\sum_{i=1}^n(\boldsymbol{Z}_{i-1}-\bar{\boldsymbol{Z}})I\Bigg(\eta_i(\boldsymbol{\tau})\geq G^{-1}_{n\boldsymbol{\tau}}\{(n+1)u/n\}\Bigg)\varphi(du)\\ &=\int\sum_{i=1}^n(\boldsymbol{Z}_{i-1}-\bar{\boldsymbol{Z}})I\Bigg(X_i\geq\sigma(\boldsymbol{Y}_{i-1},\boldsymbol{\tau}_2)G^{-1}_{n\boldsymbol{\tau}}\{(n+1)u/n\}+\boldsymbol{Y}'_{i-1}\boldsymbol{\tau}_1\Bigg)\varphi(du)\\ &=\int\sum_{i=1}^n(\boldsymbol{Z}_{i-1}-\bar{\boldsymbol{Z}})I\Bigg(\boldsymbol{Y}'_{i-1}\boldsymbol{\alpha}+\sigma(Y_{i-1},\boldsymbol{\beta})\eta_i\geq\sigma(Y_{i-1},\boldsymbol{\tau}_2)G^{-1}_{n\boldsymbol{\tau}}\{(n+1)u/n\}+\boldsymbol{Y}'_{i-1}\boldsymbol{\tau}_1\Bigg)\varphi(du)\\ &=\int\sum_{i=1}^n(\boldsymbol{Z}_{i-1}-\bar{\boldsymbol{Z}})I\Bigg(\eta_i\geq G^{-1}_{n\boldsymbol{\tau}}\{(n+1)u/n\}\frac{\sigma(\boldsymbol{Y}_{i-1},\boldsymbol{\tau}_2)}{\sigma(Y_{i-1},\boldsymbol{\beta})}+\frac{(\boldsymbol{\tau}_1-\boldsymbol{\alpha})'\boldsymbol{Y}_{i-1}}{\sigma(Y_{i-1},\boldsymbol{\beta})}\Bigg)\varphi(du)\\ &=-\int\sum_{i=1}^n(\boldsymbol{Z}_{i-1}-\bar{\boldsymbol{Z}})I\Bigg(\eta_i< G^{-1}_{n\boldsymbol{\tau}}\{(n+1)u/n\}\frac{\sigma(\boldsymbol{Y}_{i-1},\boldsymbol{\tau}_2)}{\sigma(Y_{i-1},\boldsymbol{\beta})}+\frac{(\boldsymbol{\tau}_1-\boldsymbol{\alpha})'\boldsymbol{Y}_{i-1}}{\sigma(Y_{i-1},\boldsymbol{\beta})}\Bigg)\varphi(du).\end{aligned}$$

Substituting $\boldsymbol{\tau}=\boldsymbol{\theta}_n$ in the above where $\boldsymbol{\theta}_n=\boldsymbol{\theta}+n^{-1/2}\boldsymbol{t}$, and using Lemma 3.2, and the boundedness of φ,

$$\begin{aligned}&\boldsymbol{S}_\varphi(\boldsymbol{\theta}+n^{-1/2}\boldsymbol{t})\\ =\;& n^{-1/2}\sum_{i=1}^n\{\boldsymbol{Z}_{i-1}(\boldsymbol{\beta}+n^{-1/2}\boldsymbol{t}_2)-\bar{\boldsymbol{Z}}(\boldsymbol{\beta}+n^{-1/2}\boldsymbol{t}_2)\}\varphi\left(\frac{R_{i(\boldsymbol{\theta}+n^{-1/2}\boldsymbol{t})}}{n+1}\right)\\ =\;& -n^{-1/2}\int\sum_{i=1}^n\Big(\boldsymbol{Z}_{i-1}(\boldsymbol{\theta}_n)-\bar{\boldsymbol{Z}}(\boldsymbol{\theta}_n)\Big)\\ & I\Bigg(\eta_i< G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\}\frac{\sigma(\boldsymbol{Y}_{i-1},\boldsymbol{\beta}+n^{-1/2}\boldsymbol{t}_2)}{\sigma(Y_{i-1},\boldsymbol{\beta})}+\frac{n^{-1/2}\boldsymbol{t}_1'\boldsymbol{Y}_{i-1}}{\sigma(Y_{i-1},\boldsymbol{\beta})}\Bigg)\varphi(du)\end{aligned}$$

$$= \quad -n^{-1/2}\int\sum_{i=1}^{n}\Big(\boldsymbol{Z}_{i-1}(\boldsymbol{\theta}_n)-\bar{\boldsymbol{Z}}(\boldsymbol{\theta}_n)\Big)$$

$$I\Bigg(\eta_i < G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\}+G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\}(\sigma_{ni}(\boldsymbol{t}_2)-1)+(\mu_{ni}(\boldsymbol{t}_1)-\mu_i)\Bigg)\varphi(du)$$

$$= \quad -n^{-1/2}\int\sum_{i=1}^{n}\Bigg(\boldsymbol{Z}_{i-1}(\boldsymbol{\theta}_n)-\bar{\boldsymbol{Z}}(\boldsymbol{\theta}_n)\Bigg)$$

$$G\Bigg(G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\}+G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\}(\sigma_{ni}(\boldsymbol{t}_2)-1)+(\mu_{ni}(\boldsymbol{t}_1)-\mu_i)\Bigg)\varphi(du)$$

$$+ \quad n^{-1/2}\int\sum_{i=1}^{n}\Bigg(\boldsymbol{Z}_{i-1}(\boldsymbol{\theta}_n)-\bar{\boldsymbol{Z}}(\boldsymbol{\theta}_n)\Bigg)$$

$$\Bigg\{I(\eta_i < G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\})-G(G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\})\Bigg\}\varphi(du)+u_p(1)$$

$$= -T_1+T_2+u_p(1), \text{ say.}$$

Similarly, substituting $\boldsymbol{\tau}=\boldsymbol{\theta}$, we have

$$\boldsymbol{S}_{\varphi}(\boldsymbol{\theta}) = -n^{-1/2}\int\sum_{i=1}^{n}\Bigg(\boldsymbol{Z}_{i-1}(\boldsymbol{\theta})-\bar{\boldsymbol{Z}}(\boldsymbol{\theta})\Bigg)G(G^{-1}_{n\boldsymbol{\theta}}\{(n+1)u/n\})\varphi(du)$$

$$+n^{-1/2}\int\sum_{i=1}^{n}\Bigg(\boldsymbol{Z}_{i-1}(\boldsymbol{\theta})-\bar{\boldsymbol{Z}}(\boldsymbol{\theta})\Bigg)$$

$$\{I(\eta_i < G^{-1}_{n\boldsymbol{\theta}}\{(n+1)u/n\})-G(G^{-1}_{n\boldsymbol{\theta}}\{(n+1)u/n\})\}\varphi(du)+o_p(1) = -T_3+T_4+o_p(1), \text{ say.}$$

For the terms T_2 and T_4 involving centered empirical processes, note that T_2-T_4 equals

$$n^{-1/2}\int\sum_{i=1}^{n}(\boldsymbol{Z}_{i-1}(\boldsymbol{\theta}_n)-\bar{\boldsymbol{Z}}(\boldsymbol{\theta}_n))\Bigg\{I(\eta_i < G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\})-G(G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\})\Bigg\}\varphi(du)$$

$$- \quad n^{-1/2}\int\sum_{i=1}^{n}(\boldsymbol{Z}_{i-1}(\boldsymbol{\theta})-\bar{\boldsymbol{Z}}(\boldsymbol{\theta}))\{I(\eta_i < G^{-1}_{n\boldsymbol{\theta}}\{(n+1)u/n\})-G(G^{-1}_{n\boldsymbol{\theta}}\{(n+1)u/n\})\}\varphi(du)$$

$$= \quad n^{-1/2}\int\sum_{i=1}^{n}\Bigg\{(\boldsymbol{Z}_{i-1}(\boldsymbol{\theta}_n)-\bar{\boldsymbol{Z}}(\boldsymbol{\theta}_n))-(\boldsymbol{Z}_{i-1}(\boldsymbol{\theta})-\bar{\boldsymbol{Z}}(\boldsymbol{\theta}))\Bigg\}$$

$$\Bigg\{I(\eta_i < G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\})-G(G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\})\Bigg\}\varphi(du)$$

$$- \quad n^{-1/2}\int\sum_{i=1}^{n}(\boldsymbol{Z}_{i-1}(\boldsymbol{\theta})-\bar{\boldsymbol{Z}}(\boldsymbol{\theta}))$$

$$\Bigg\{I(\eta_i < G^{-1}_{n\boldsymbol{\theta}}\{(n+1)u/n\})-I(\eta_i < G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\})$$

$$+G(G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\})-G(G^{-1}_{n\boldsymbol{\theta}}\{(n+1)u/n\})\Bigg\}\varphi(du) = T_5-T_6, \text{ say.}$$

Since $|I(\eta_i < G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\}) - G(G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\})| \le 1$, T_5 is $u_p(1)$ by assumption (3.15) with $l_{ni}(\boldsymbol{t}) = s_{ni}(\boldsymbol{t})$. Next we heavily use the result of Koul and Ossiander (1994, Theorem 1.1) on the tightness of U_n^* and the following fact from Koul and Ossiander (1994, Eqns (3.11), (3.12))

$$\sup\{|G(G^{-1}_{n\boldsymbol{\theta}+n^{-1/2}\boldsymbol{t}}\{(n+1)u/n\}) - u|; u \in [0,1], \|\boldsymbol{t}\| \le b\} = o_p(1), \tag{6.18}$$

which entails $\sup\{|G(G_{n\boldsymbol{\theta}_n}\{(n+1)u/n\}) - G(G_{n\boldsymbol{\theta}}\{(n+1)u/n\})|; u \in [0,1], \|\boldsymbol{t}\| \le b\} = o_p(1)$. Therefore $T_6 = o_p(1)$.

Next, it remains to examine $T_1 - T_3$ which equals

$$\begin{aligned} & n^{-1/2}\int \sum_{i=1}^n (\boldsymbol{Z}_{i-1}(\boldsymbol{\theta}_n) - \bar{\boldsymbol{Z}}(\boldsymbol{\theta}_n)) \\ & G\Big(G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\} + G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\}(\sigma_{ni}(\boldsymbol{t}_2)-1) + (\mu_{ni}(\boldsymbol{t}_1) - \mu_i)\Big)\varphi(du) \\ - \quad & n^{-1/2}\int \sum_{i=1}^n (\boldsymbol{Z}_{i-1}(\boldsymbol{\theta}) - \bar{\boldsymbol{Z}}(\boldsymbol{\theta}))G(G^{-1}_{n\boldsymbol{\theta}}\{(n+1)u/n\})\varphi(du). \end{aligned}$$

Now subtracting and adding

$$n^{-1/2}(\boldsymbol{Z}_{i-1}(\boldsymbol{\theta}) - \bar{\boldsymbol{Z}}(\boldsymbol{\theta}))G(G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\} + G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\}(\sigma_{ni}(\boldsymbol{t}_2)-1) + (\mu_{ni}(\boldsymbol{t}_1) - \mu_i))$$

to the i-th summand, and using

$$\begin{aligned} & n^{-1/2}\int \sum_{i=1}^n \Big((\boldsymbol{Z}_{i-1}(\boldsymbol{\theta}_n) - \bar{\boldsymbol{Z}}(\boldsymbol{\theta}_n) - \boldsymbol{Z}_{i-1}(\boldsymbol{\theta}) + \bar{\boldsymbol{Z}}(\boldsymbol{\theta})\Big) \\ & \times G\Big(G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\} + G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\}(\sigma_{ni}(\boldsymbol{t}_2)-1) + (\mu_{ni}(\boldsymbol{t}_1) - \mu_i)\Big)\varphi(du) = u_p(1), \end{aligned}$$

and

$$n^{-1/2}\int \sum_{i=1}^n (\boldsymbol{Z}_{i-1}(\boldsymbol{\theta}) - \bar{\boldsymbol{Z}}(\boldsymbol{\theta}))\{G(G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\}) - G(G^{-1}_{n\boldsymbol{\theta}}\{(n+1)u/n\})\}\varphi(du) = u_p(1),$$

$T_1 - T_3$ equals

$$n^{-1/2}\int \sum_{i=1}^n \Big(\boldsymbol{Z}_{i-1}(\boldsymbol{\theta}) - \bar{\boldsymbol{Z}}(\boldsymbol{\theta})\Big)$$

$$\times G\Big(G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\} + G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\}(\sigma_{ni}(\boldsymbol{t}_2)-1) + (\mu_{ni}(\boldsymbol{t}_1)-\mu_i)\Big)\varphi(du)$$

$$-n^{-1/2}\int\sum_{i=1}^{n}(\boldsymbol{Z}_{i-1}(\boldsymbol{\theta})-\bar{\boldsymbol{Z}}(\boldsymbol{\theta}))G(G^{-1}_{n\boldsymbol{\theta}}\{(n+1)u/n\})\varphi(du)+u_p(1).$$

Next we use the mean value theorem on G around the point $G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\}$, and write $g(G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\}+\xi_{inu}\boldsymbol{t}) = gG^{-1}\Big(G(G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\}+\xi_{inu}\boldsymbol{t})\Big)$. We also use the uniform continuity of the function gG^{-1} and $G(G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\}+\xi_{inu}\boldsymbol{t}) = g(G^{-1}_{n\boldsymbol{\theta}_n}\{(n+1)u/n\})+\xi_{inu}\boldsymbol{t})$, (6.18) and

$$\sup_{\|\boldsymbol{t}\|\le b}\Big\|n^{-1/2}\sum_{i=1}^{n}\left\{\frac{\dot\mu_{ni}(\boldsymbol{t}_1)}{\sigma_{ni}(\boldsymbol{t}_2)} - n^{-1}\sum_{i=1}^{n}(\frac{\dot\mu_{ni}(\boldsymbol{t}_1)}{\sigma_{ni}(\boldsymbol{t}_2)})\right\}\{\sigma_{ni}(\boldsymbol{t}_2)-1\} - \boldsymbol{G}_c(\theta)\boldsymbol{t}_2\Big\| = o_p(1)$$

$$\sup_{\|\boldsymbol{t}\|\le b}\Big\|n^{-1/2}\sum_{i=1}^{n}\left\{\frac{\dot\mu_{ni}(\boldsymbol{t}_1)}{\sigma_{ni}(\boldsymbol{t}_2)} - n^{-1}\sum_{i=1}^{n}(\frac{\dot\mu_{ni}(\boldsymbol{t}_1)}{\sigma_{ni}(\boldsymbol{t}_2)})\right\}\{\mu_{ni}(\boldsymbol{t}_1)-\mu_i\} - M(\theta)\boldsymbol{t}_1\Big\| = o_p(1),$$

we get that $T_1 - T_3$ equals

$$\int G^{-1}(u)g(G^{-1}(u))\varphi(du)\boldsymbol{G}_c(\boldsymbol{\theta})\boldsymbol{t}_2 + \int g(G^{-1}(u))\varphi(du)M(\boldsymbol{\theta})\boldsymbol{t}_1 + u_p(1).$$

Hence Lemma 3.3 follows. ⊥⊥

Acknowledgment. I am grateful to the Editor, the Associate Editor, and the two anonymous referees for making many insightful and constructive comments. Their extraordinarily careful reading and suggestions have changed and improved the paper to a great extent compared to the original manuscript.

The research was supported partly by the Association of Commonwealth Universities during my visit to the London School of Economics and Political Sciences under the Commonwealth Fellowship.

REFERENCES

Boldin, M. V., 1998, On residual empirical distribution functions in ARCH models with applications to testing and estimation. Mathem. Seminar Giessen, 49-66.

Bollerslev, T., Chou, R. Y. and K. F. Kroner, 1992, ARCH modeling in finance; a review of the theory and empirical evidence. Journal of Econometrics 52, 115-127.

Bougerol, P. and N. Picard, 1992, Stationarity of GARCH processes and of some nonnegative time series. Journal of Econometrics 52, 115-127.

Chernoff, H. and I. R. Savage, 1958, Asymptotic normality and efficiency of certain nonparametric test statistics. Annals of Mathematical Statistics 29, 972-994.

Engle, R.F., 1982, Autoregressive conditional heteroscedasticity and estimates of the variance of UK inflation. Econometrica 50, 987-1008.

Engle, R.F. and G. Gonzalez-Rivera, 1991, Semiparametric ARCH models. Journal of Business and Economic Statistics 9, 345-349.

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

Gouriéroux, C., 1997, ARCH models and Financial Applications. Springer-Verlag, New York.

Hájek, J., Šidák, Z. and P. Sen, 1999, Theory of Rank Tests. Academic Press, San Diego.

Hall P. and C. Heyde, 1980, Martingale Limit Theory and its Application. Academic Press, New York.

Härdle, W. and A. Tsybakov, 1997, Local polynomial estimators of the volatility function in nonparametric autoregressive. Journal of Econometrics 81, 223-242.

Huber, P.J., 1981, Robust Statistics. John Wiley and Sons Inc., New York.

Jaeckel, L. A., 1972, Estimating regression coefficients by minimizing the dispersion of the residuals. Annals of Mathematical Statistics 43, 1449-1458.

Jurečková, J., 1971, Nonparametric estimates of regression coefficients. Annals of Mathematical Statistics 42, 1328-1338.

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

Koenker, R. and Q. Zhao, 1996, Conditional quantile estimation and inference for ARCH models. Econometric Theory 12, 793-813.

Koul, H., 1971, Asymptotic behavior of a class of confidence region based on rank in regression. Annals of Mathematical Statistics 42, 466-476.

Koul H., 1992, Weighted empiricals and linear models. IMS Lecture Notes-Monograph Ser. 21, Hayward, CA.

Koul, H., 1996, Asymptotics of some estimators and sequential residual empiricals in nonlinear time series. Annals of Statistics 24, 380-404.

Koul, H. and K. Mukherjee, 1993, Asymptotics of R-, MD- and LAD-estimators in linear regression models with long range dependent errors. Probability Theory and Related Fields 95, 535-553.

Koul, H. and K. Mukherjee, 2002, Some estimation procedures in ARCH models. Technical Report 9-2000. National University of Singapore.

Koul H.L. and M. Ossiander, 1994, Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. Annals of Statistics 22, 540-562.

Lehmann, E. L., 1983, Theory of Point Estimation. Wiley, New York.

Mukherjee, K. (2006 a). Pseudo-likelihood estimation in ARCH models. To appear in Canadian Journal of Statistics, 34,?-?.

Mukherjee, K. (2006 b). Computation of R-estimators and some related topics. Unpublished manuscript with link from http://www.liv.ac.uk/maths/SP/HOME/K.Mukherjee.html.

Mukherjee, K. and Z. D. Bai, 2002, R-estimation in autoregression with square-integrable score function. Journal of Multivariate Analysis 81, 167-186.

Nelson, D. B., 1990, Stationarity and persistence in the GARCH (1, 1) model. Econometric Theory 6, 318-334.

Pantula, S. G., 1988, Estimation of autoregressive models with ARCH errors. Sankhyā, Ser B 50, 119-138.

Rabemananjara, R. and J. M. Zakoian, 1993, Threshold ARCH models and asymmetry in volatility. Journal of Applied Econometrics 8, 31-49.

Shephard, N., 1996, Statistical aspects of ARCH and stochastic volatility, in: D. Cox, D.

Hinkley and O. Barndorff-Nielsen, (Eds.), Time Series Models in Econometric, Finance and other fields, Chapman and Hall Ltd., London, pp. 1-67.

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

Weiss, A. A., 1986, Asymptotic Theory for ARCH models: estimation and testing. Econometric Theory 2, 107-131.