CHAPTER 7

AUTOREGRESSION

7.1. INTRODUCTION.

The purpose of this chapter is to offer a unified functional approach to some aspects of robust estimation and goodness-of-fit testing problems in pth order autoregression (AR(p)) models. This approach is similar to that of the previous chapters in connection with linear regression models, thereby extending a statistical methodology to one of the most applied models with dependent observations.

As before, let F be a d.f. on $\mathbb{R}$, $p \geq 1$ be an integer, $\epsilon_1, \epsilon_2, \ldots$ be i.i.d. F r.v.'s and $\mathbf{Y}_0 := (X_0, X_{-1}, \ldots, X_{1-p})'$ be an observable random vector independent of $\epsilon_1, \epsilon_2, \ldots$. In an AR(p) model one observes $\{X_i\}$ satisfying

$$(1) \qquad X_i = \rho_1 X_{i-1} + \cdots + \rho_p X_{i-p} + \epsilon_i, \qquad 1 \leq i \leq n, \ p \geq 1,$$

for some $\boldsymbol{\rho}' = (\rho_1, \rho_2, \ldots, \rho_p) \in \mathbb{R}^p$.

Processes that play a fundamental role in the robust estimation of $\boldsymbol{\rho}$ in this model are the *randomly weighted residual empirical* processes

$$(2) \qquad T_j(x, \mathbf{t}) := n^{-1} \sum_{i=1}^{n} g(X_{i-j})\, I(X_i - \mathbf{t}'\mathbf{Y}_{i-1} \leq x), \ x \in \mathbb{R}, \mathbf{t} \in \mathbb{R}^p, 1 \leq j \leq p,$$

where g is a measurable function from $\mathbb{R}$ to $\mathbb{R}$ and $\mathbf{Y}_{i-1} := (X_{i-1}, \ldots, X_{i-p})'$, $1 \leq i \leq n$. Let $\mathbf{T} := (T_1, \ldots, T_p)'$.

The generalized M- (GM) estimators of $\boldsymbol{\rho}$, as proposed by Denby and Martin (1979), are solution $\mathbf{t}$ of the p equations

$$(3) \qquad \mathcal{G}_j(\mathbf{t}) := \int \psi(x)\, T_j(dx, \mathbf{t}) = 0, \qquad 1 \leq j \leq p,$$

where ψ is a nondecreasing bounded measurable function from $\mathbb{R}$ to $\mathbb{R}$. These estimators are analogues of M-estimators of $\boldsymbol{\beta}$ in linear regression as discussed in Capter 4. Note that taking $g(x) \equiv xI[|x| \leq k] + kI[|x| > k] \equiv \psi(x)$ in (3) gives the Huber(k) estimators and taking $g(x) \equiv x \equiv \psi(x)$ gives the famous least square estimator.

The m.d. estimator $\boldsymbol{\rho}_g^+$ that is an analogue of $\boldsymbol{\beta}_D^+$ of (5.2.20) is defined as a minimizer, w.r.t. $\mathbf{t}$, of

$$(4) \qquad K_g(\mathbf{t}) = \sum_{j=1}^{p} \int [n^{-1/2} \sum_{i=1}^{n} g(X_{i-j})\{I(X_i \leq x + \mathbf{t}'\mathbf{Y}_{i-1}) - I(-X_i < x - \mathbf{t}'\mathbf{Y}_{i-1})\}]^2\, dG(x), \quad \mathbf{t} \in \mathbb{R}^p.$$

Observe that K involves $\mathbf{T}$. In fact, $\forall\ \mathbf{t} \in \mathbb{R}^p$,

$$K_g(\mathbf{t}) = \sum_{j=1}^{p} \int [n^{1/2}\{T_j(x, \mathbf{t}) - \sum_{i=1}^{n} g(X_{i-j}) + T_j(-x, \mathbf{t})\}]^2 \, dG(x).$$

Three members of this class of estimators are of special interest. They correspond to the cases $g(x) \equiv x \equiv G(x)$; $g(x) \equiv x$, $G \equiv \delta_0$, the measure degenerate at 0; $g(x) \equiv x$, $G \equiv F$ in the F known case. The first gives an analogue of the *Hodges–Lehmann* (h.l.) estimator of $\boldsymbol{\rho}$, the second gives the *least absolute deviation* (l.a.d.) estimator, while the third gives an estimator that is more efficient at logistic (double exponential) errors than l.a.d (h.l.) estimator.

Another important process in the model (1) is the ordinary residual empirical process

$$(5) \qquad F_n(x, \mathbf{t}) := n^{-1} \Sigma_i I(X_i - \mathbf{t}' \mathbf{Y}_{i-1} \le x), \qquad x \in \mathbb{R},\ \mathbf{t} \in \mathbb{R}^p.$$

An estimator of F or a test of goodness-of-fit pertaining to F are usually based on $F_n(x, \hat{\boldsymbol{\rho}})$, where $\hat{\boldsymbol{\rho}}$ is an estimator of $\boldsymbol{\rho}$.

Clearly F_n is a special case of (2). But, both F_n and T_j, $1 \le j \le p$, are special cases of

$$(6) \quad W_h(x, \mathbf{t}) := n^{-1} \Sigma_i h(\mathbf{Y}_{i-1}) I(X_i - \mathbf{t}' \mathbf{Y}_{i-1} \le x)$$
$$= n^{-1} \Sigma_i h(\mathbf{Y}_{i-1}) I(\epsilon_i \le x + (\mathbf{t} - \boldsymbol{\rho})' \mathbf{Y}_{i-1}), \qquad x \in \mathbb{R},\ \mathbf{t} \in \mathbb{R}^p,$$

where h is a measurable function from $\mathbb{R}^p$ to $\mathbb{R}$. Choosing $h(\mathbf{Y}_{i-1}) \equiv g(X_{i-j})$ in W_h gives T_j, $1 \le j \le p$ and the choice of $h \equiv 1$ yields F_n.

From the above discussion it is apparent that the investigation of the large sample behavior of various inferential procedures pertaining to $\boldsymbol{\rho}$ and F, based on $\{T_j\}$ and $F_n(\cdot, \hat{\boldsymbol{\rho}})$, is facilitated by the weak convergence properties of $\{W_h(x, \boldsymbol{\rho}+n^{-1/2}\mathbf{u}),\ x \in \mathbb{R},\ \mathbf{u} \in \mathbb{R}^p\}$. This will be investigated in Section 7.2, with the aid of Theorem 2.2b.1. In particular, this section contains an a.u.l. result about $\{W_h(x, \boldsymbol{\rho}+n^{-1/2}\mathbf{u}),\ x \in \mathbb{R},\ \|\mathbf{u}\| \le B\}$ which in turn yields a.u.l. results about $\{\mathbf{T}(x, \boldsymbol{\rho}+n^{-1/2}\mathbf{u}),\ x \in \mathbb{R},\ \|\mathbf{u}\| \le B\}$ and $\{F_n(x, \boldsymbol{\rho}+n^{-1/2}\mathbf{u}),\ x \in \mathbb{R},\ \|\mathbf{u}\| \le B\}$. These results are useful in studying GM- and R-estimators of $\boldsymbol{\rho}$, akin to Chapters 3 and 4 when dealing with linear regression models. They are also useful in studying the large sample behaviour of some tests of goodness-of-fit pertaining to F. Analogous results about the ordinary empirical of the residuals in autoregressive *moving average* models are briefly discussed in Remark 7.2.4.

Generalized M-estimators and analogues of Jaeckel's (1972) R-estimators are discussed in Section 7.3. In order to use R- or m.d.

estimators to construct confidence intervals one often needs consistent estimators of the functional $Q(f)$ of the error density f. Appropriate analogues of estimators of $Q(f)$ of Section 4.5 are shown to be consistent under **(F1)** and **(F2)**. This is also done in Section 7.3, with the help of the a.u.l. property of $\{F_n(x, \rho+n^{-1/2}u), x\in\mathbb{R}, \|u\|\le B\}$. This result is also used to prove the a.u.l. of serial rank correlations of the residuals in an AR(p) model. Such results should be useful in developing anlogues of the method of moment estimators or Yule-Walker equations based on ranks in AR(p) models.

Section 7.4 investigates the behaviour of two classes of m.d. estimators of ρ, including the class of estimators $\{\rho_g^+\}$. A crucial result needed to obtain the asymptotic distributions of these estimators is the asymptotic uniform quadraticity of their defining dispersions. This result is also proved in Section 7.4. Section 7.5 contains appropriate analogues of some of the goodness-of-fit tests of Capter 6 pertaining to F.

7.2. ASYMPTOTIC UNIFORM LINEARITY OF $\mathcal{W}_h$ and F_n.

Recall the definition of $\mathcal{V}_h$ process from (1.4.1) and the statement of Theorem 2.2b.1. In (1.4.1), let

$$(1)\qquad \zeta_{ni}\equiv \epsilon_i,\ h_{ni}\equiv h(\mathbf{Y}_{i-1}),\ \delta_{ni}\equiv n^{-1/2}\mathbf{u}'\mathbf{Y}_{i-1},\qquad \mathbf{u}\in\mathbb{R}^p,\ 1\le i\le n,$$

$$\mathcal{A}_{n1}=\sigma\text{–field }\{\mathbf{Y}_0\},\ \mathcal{A}_{ni}=\sigma\text{–field }\{\mathbf{Y}_0',\epsilon_1,\dots,\epsilon_{i-1}\},\qquad 2\le i\le n.$$

Then one readily sees that the corresponding $\mathcal{V}_h(x)$, $\mathcal{V}_h^*(x)$ are, respectively, equal to $\mathcal{W}_h(x, \rho+n^{-1/2}\mathbf{u})$, $\mathcal{W}_h(x, \rho)$ for each $\mathbf{u}\in\mathbb{R}^p$ and for all $x\in\mathbb{R}$. Consequently, if we let

$$(2)\qquad \nu_h(x,\mathbf{t}) := n^{-1}\Sigma_i h(\mathbf{Y}_{i-1})\, F(x+(\mathbf{t}-\rho)'\mathbf{Y}_{i-1}),$$

$$\mathcal{W}_h(x,\mathbf{t}) := n^{1/2}[\mathcal{W}_h(x,\mathbf{t})-\nu_h(x,\mathbf{t})],\qquad x\in\mathbb{R},\ \mathbf{t}\in\mathbb{R}^p,$$

then the corresponding $\mathcal{U}_h(x)$, $\mathcal{U}_h^*(x)$ are, respectively, equal to $\mathcal{W}_h(x, \rho+n^{-1/2}\mathbf{u})$, $\mathcal{W}_h(x,\rho)$ for each $\mathbf{u}\in\mathbb{R}^p$ and for all $x\in\mathbb{R}$. Recall the conditions **(F1)** and **(F2)** from Corollary 2.3.1. We are now ready to state and prove the following

Theorem 7.2.1. *In addition to* (7.1.1), *assume that the following conditions hold:*

(a1) $\quad h$ *is a bounded function.*

(a2) $\quad n^{-1/2} \max_{1 \le i \le n} \|Y_{i-1}\| = o_p(1).$

(a3) $\quad n^{-1} \sum_{i=1}^{n} \|h(Y_{i-1})Y_{i-1}\| = O_p(1).$

(a4) $\quad$ F *satisfies* (F1) *and* (F2).

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

$$\sup_{x \in \mathbb{R}, \|u\| \le B} |\mathscr{W}_h(x, \rho + n^{-1/2}u) - \mathscr{W}_h(x, \rho)| = o_p(1), \tag{3}$$

and

$$n^{1/2}[W_h(x, \rho + n^{-1/2}u) - W_h(x, \rho)] = -u' n^{-1} \Sigma_i h(Y_{i-1}) Y_{i-1} f(x) + \bar{o}_p(1). \tag{4}$$

where $\bar{o}_p(1)$ *is a sequence of stochastic processes that converges to zero, uniformly over the set* $x \in \mathbb{R}$, $\|u\| \le B$, *in probability.*

Proof. In view of the discussion preceeding the statement of the theorem it is clear that (2.2b.2) of Theorem 2.2b.1 applied to entities given in (1) above readily yields that

$$\sup\{|\mathscr{W}_h(x, \rho + n^{-1/2}u) - \mathscr{W}_h(x, \rho)|; x \in \mathbb{R}\} = o_p(1) \quad \text{for every fixed } u \in \mathbb{R}^p.$$

It is the uniformity with respect to u that requires an extra argument and that also turns out to be a consequence of another application of (2.2b.2) and a monotonic property inherent in these processes as we now show.

Since h is fixed, it will not be exhibited in the proof. Also, for convenience, write $\mathscr{W}(\cdot)$, $\mathscr{W}_u(\cdot)$, $W_u^{\pm}(\cdot)$, $\nu_u^{\pm}(\cdot)$ etc. for $\mathscr{W}_h(\cdot, \rho)$, $\mathscr{W}_h(\cdot, \rho + n^{-1/2}u)$, $W_h^{\pm}(\cdot, \rho + n^{1/2}u)$, $\nu_h^{\pm}(\cdot, \rho + n^{-1/2}u)$ etc. with $\pm$ signifying the fact that $h^{\pm}$ now appears in the place of h in these processes where $h^+ = 0 \vee h$, $h^- = h - h^+$. To avoid displays being broken into different lines often, write ξ_i, h_i, $h_i^{\pm}$ for Y_{i-1}, $h(Y_{i-1})$, $h^{\pm}(Y_{i-1})$, respectively, $i \ge 1$. Thus, e.g.,

$$\mathscr{W}_u^{\pm}(x) = n^{-1/2} \Sigma_i h_i^{\pm} \{I(\epsilon_i \le x + n^{-1/2} u' \xi_i) - F(x + n^{-1/2} u' \xi_i\}. \tag{5}$$

We also need the following processes:

$$\begin{aligned} T^{\pm}(x; u, a) &:= n^{-1/2} \Sigma_i h_i^{\pm} I(\epsilon_i \le x + n^{-1/2} u' \xi_i + n^{-1/2} a \|\xi_i\|), \\ m^{\pm}(x; u, a) &:= n^{-1/2} \Sigma_i h_i^{\pm} F(x + n^{-1/2} u' \xi_i + n^{-1/2} a \|\xi_i\|) \\ Z^{\pm} &:= T^{\pm} - m^{\pm}, \qquad x \in \mathbb{R},\ u \in \mathbb{R}^p,\ a \in \mathbb{R}. \end{aligned} \tag{6}$$

Observe that if in $\mathcal{U}_h$ of (2.2b.1) we take $\zeta_{ni} \equiv \epsilon_i$, $h_{ni} \equiv h^{\pm}(\xi_i)$, $\delta_{ni} \equiv n^{-1/2}\{\mathbf{u}'\xi_i + a\|\xi_i\|\}$ and $\mathcal{A}_{ni}$, $1 \le i \le n$, as in (1), we obtain

$$\mathcal{U}_h(\cdot) = Z^{\pm}(\cdot; \mathbf{u}, a), \qquad \text{for every } \mathbf{u} \in \mathbb{R}^p, \ a \in \mathbb{R}.$$

Similarly, if we take $\delta_{ni} \equiv n^{-1/2}\mathbf{u}'\xi_i$ and the rest of the quantities as above then

$$\mathcal{U}_h(\cdot) = Z^{\pm}(\cdot; \mathbf{u}, 0) = \mathcal{W}_{\mathbf{u}}^{\pm}(\cdot), \qquad \text{for every } \mathbf{u} \in \mathbb{R}^p.$$

It thus follows from two applications of (2.2b.2) and the triangle inequality that for every $\mathbf{u} \in \mathbb{R}^p$, $a \in \mathbb{R}$,

$$\text{(7a)} \qquad \sup_x |Z^{\pm}(x; \mathbf{u}, a) - Z^{\pm}(x; \mathbf{u}, 0)| = o_p(1),$$

$$\text{(7b)} \qquad \sup_x |\mathcal{W}_{\mathbf{u}}^{\pm}(x) - \mathcal{W}^{\pm}(x)| = o_p(1).$$

Thus, to prove (3), because of the compactness of $\mathcal{N}(B)$, it suffices to show that for every $\epsilon > 0$ there is a $\delta > 0$ such that for every $\|\mathbf{u}\| \le B$,

$$\text{(8)} \quad \limsup_n P\Big(\sup_{\|\mathbf{s}\| \le B, \|\mathbf{s}-\mathbf{u}\| \le \delta, x} |\mathcal{W}_{\mathbf{s}}(x) - \mathcal{W}_{\mathbf{u}}(x)| > 4\epsilon\Big) < \epsilon.$$

By the definition of $\mathcal{W}^{\pm}$ and the triangle inequality, for $x \in \mathbb{R}$, $\mathbf{s}, \mathbf{u} \in \mathbb{R}^p$,

$$\text{(9)} \qquad |\mathcal{W}_{\mathbf{s}}(x) - \mathcal{W}_{\mathbf{u}}(x)| \le |\mathcal{W}_{\mathbf{s}}^{+}(x) - \mathcal{W}_{\mathbf{u}}^{+}(x, \mathbf{u})| + |\mathcal{W}_{\mathbf{s}}^{-}(x) - \mathcal{W}_{\mathbf{u}}^{-}(x)|,$$

$$|\mathcal{W}_{\mathbf{s}}^{\pm}(x) - \mathcal{W}_{\mathbf{u}}^{\pm}(x)| \le n^{1/2}[|W_{\mathbf{s}}^{\pm}(x) - W_{\mathbf{u}}^{\pm}(x)| + |\nu_{\mathbf{s}}^{\pm}(x) - \nu_{\mathbf{u}}^{\pm}(x)|].$$

But $\mathbf{s} \in \mathcal{N}(B)$, $\|\mathbf{u}\| \le B$, $\|\mathbf{s} - \mathbf{u}\| \le \delta$ imply that for all $1 \le i \le n$,

$$\text{(10)} \quad n^{-1/2}\mathbf{u}'\xi_i - n^{-1/2}\delta\|\xi_i\| \le n^{-1/2}\mathbf{s}'\xi_i \le n^{-1/2}\mathbf{u}'\xi_i + n^{-1/2}\delta\|\xi_i\|.$$

From (10), the monotonicity of the indicator function and the nonnegativity of $h^{\pm}$, we obtain

$$T^{\pm}(x; \mathbf{u}, -\delta) - T^{\pm}(x; \mathbf{u}, 0) \le W_{\mathbf{s}}^{\pm}(x) - W_{\mathbf{u}}^{\pm}(x) \le T^{\pm}(x; \mathbf{u}, \delta) - T^{\pm}(x; \mathbf{u}, 0)$$

for all $x \in \mathbb{R}$, $\mathbf{s} \in \mathcal{N}(B)$, $\|\mathbf{s}-\mathbf{u}\| \le \delta$. Now center $T^{\pm}$ appropriately to obtain

$$(11) \quad n^{1/2}|W_s^{\pm}(x)-W_u^{\pm}(x)|$$
$$\leq |Z^{\pm}(x;\mathbf{u},\delta) - Z^{\pm}(x;\mathbf{u},0)| + |Z^{\pm}(x;\mathbf{u},-\delta) - Z^{\pm}(x;\mathbf{u},0)|$$
$$+ |m^{\pm}(x;\mathbf{u},\delta) - m^{\pm}(x;\mathbf{u},0)| + |m^{\pm}(x;\mathbf{u},-\delta) - m^{\pm}(x;\mathbf{u},0)|,$$

for all $x \in \mathbb{R}$, $\mathbf{s} \in \mathcal{N}(B)$, $\|\mathbf{s} - \mathbf{u}\| \leq \delta$.

But, by (a4), $\forall$ $\|\mathbf{u}\| \leq B$,

$$(12) \quad \sup_x |m^{\pm}(x;\mathbf{u},\pm\delta) - m^{\pm}(x;\mathbf{u},0)| \leq \delta\|f\|_\infty n^{-1}\Sigma_i\|h_i\xi_i\|,$$

$$(13) \quad \sup_{\|\mathbf{s}-\mathbf{u}\|\leq\delta,x} n^{1/2}|\nu_s^{\pm}(x) - \nu_u^{\pm}(x)| \leq \delta\|f\|_\infty n^{-1}\Sigma_i\|h_i\xi_i\|.$$

From (12), (11), (7a) applied with $a = \delta$ and $a = -\delta$ and the assumption (a3) one concludes that for every $\epsilon > 0$ there is a $\delta > 0$ such that for each $\|\mathbf{u}\| \leq B$,

$$\limsup_n P\Big(\sup_{\|\mathbf{s}-\mathbf{u}\|\leq\delta,x} n^{1/2}|W_s^{\pm}(x) - W_u^{\pm}(x)| > \epsilon\Big) \leq \epsilon/2.$$

From this, (13), (9), and (a3) one now concludes (8) in a routine fashion. Finally, (4) follows from (3) and (a4) by Taylor's expansion of F. □

An application of (4) with $h(Y_{i-1}) = g(X_{i-j})$ and the rest of the quantities as in (1) readily yields the a.u.l. property of T_j–processes, $1\leq j\leq p$ of (7.1.2). This together with integration by parts yields the following expansion of the M-scores $\mathcal{G}_j$, $1\leq j\leq p$ of (7.1.3).

Corollary 7.2.1. *In addition to* (7.1.1), (a2) and (a4), *assume that the following conditions hold.*

(b1) g *is bounded.*

(b2) ψ *is nondecreasing, bounded and* $\int \psi\, dF = 0$.

(b3) $n^{-1}\Sigma_i \|g(X_{i-j})Y_{i-1}\| = O_p(1)$, $1 \leq j \leq p$.

Then, $\forall$ $0 < k, B < \infty$,

$$\sup|n^{1/2}[\mathcal{G}_j(\rho + n^{-1/2}\mathbf{u}) - \mathcal{G}_j(\rho)] - \mathbf{u}' n^{-1}\Sigma_i g(X_{i-j})Y_{i-1}\int f d\psi| = o_p(1)$$

where the supremum is taken over all ψ *with* $\|\psi\|_{tv} \leq k < \infty$, $\|\mathbf{u}\| \leq B$, $1\leq j\leq p$. □

Upon choosing $h \equiv 1$ in (4) one obtains an analogous result for the *ordinary residual empirical process* $F_n(x, \mathbf{t})$. Because of its importance and for an easy reference later on we state it as a separate result. Observe that in the following corollary the assumption (a3*) is nothing but the assumption (a3) of Theorem 7.2.1 with $h \equiv 1$.

Corollary 7.2.2. *Suppose that* (7.1.1) *holds. In addition, assume that* (a2), (a3*) *and* (a4) *hold, where,*

(a3*) $\qquad n^{-1}\Sigma_i \|\mathbf{Y}_{i-1}\| = O_p(1).$

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

(14) $$\sup|n^{1/2}\{F_n(x, \rho+n^{-1/2}\mathbf{u}) - F_n(x, \rho)\} - \mathbf{u}' n^{-1}\Sigma_i \mathbf{Y}_{i-1} f(x)| = o_p(1),$$

where the supremum is taken over $x \in \mathbb{R}$, $\|\mathbf{u}\| \leq B$. □

Remark 7.2.1. Observe that non of the above results require that the process $\{X_i\}$ be stationary or any of the moments be finite. □

Remark 7.2.2. Consider the ***assumptions (a2) and (a3)***. If $\mathbf{Y}_0$ and $\{\epsilon_i\}$ are so chosen as to make $\{X_i\}$ stationary, ergodic and if $E(\|\mathbf{Y}_0\|^2+\epsilon_1^2)<\infty$ then (a2) is *a priori* satisfied and (a1) implies (a3). See, e.g., Anderson (1971; p 203). In paricular, (a3) holds for the h corresponding to the Huber function $h(x) \equiv |x|I(|x| \leq k) + \text{sign}(x)I(|x| > k)$, $k > 0$.

Of course if (a1) holds with the function h bounded in such a way that puts zero weight outside of compacts then (a3) is trivially satisfied.

Observe that (a2) is *weaker than requiring the finiteness of the second moment.* To see this, consider, for example, an AR(1) model where X_0 and $\epsilon_1, \epsilon_2, \ldots$ are independent r.v.'s and for some $|\rho| < 1$,

$$X_i = \rho X_{i-1} + \epsilon_i, \qquad i \geq 1.$$

Then,

$$X_i = \rho^i X_0 + \sum_{j=1}^{i} \rho^{j-i}\epsilon_j, \qquad Y_i = X_i, \qquad i \geq 1.$$

Thus, here (a2) is implied by

(*i*) $$\max_{1\leq i\leq n} n^{-1/2}|\epsilon_i| = o_p(1).$$

But, (*i*) is equivalent to showing that $x^2 \ln\{1-P(|\epsilon_1|>x)\} \longrightarrow 0$ as $x \longrightarrow \infty$, which, in turn is equivalent to requiring that $x^2P(|\epsilon_1|>x) \longrightarrow 0$ as $x \longrightarrow \infty$. This last condition is weaker than requiring that $E|\epsilon_1|^2 < \infty$. For example, let the right tail of the distribution of $|\epsilon_1|$ be given as follows:

$$P(|\epsilon_1|>x) = 1, \qquad x < 2,$$
$$= 1/(x^2 \ln x), \qquad x \geq 2.$$

Then, $E|\epsilon_1| < \infty$, $E\,\epsilon_1^2 = \infty$, yet $x^2 P(|\epsilon_1|>x) \longrightarrow 0$ as $x \longrightarrow \infty$. □

Remark 7.2.3. An analogue of (14) was first proved by Boldin (1982) requiring $\{X_i\}$ to be stationary, $E\epsilon_1 = 0$, $E(\epsilon_1^2) < \infty$ and a uniformly bounded second derivative of F. The Corollary 7.2.2 is an improvement of Boldin's result in the sense that F needs to be smooth only up to the first derivative and the r.v.'s need not have finite second moment.

Again, if $\mathbf{Y}_0$ and $\{\epsilon_i\}$ are so chosen that the Ergodic Theorem is applicable and $E(\mathbf{Y}_0) = \mathbf{0}$, then the coefficient $n^{-1}\Sigma_i \mathbf{Y}_{i-1}$ of the linear term in (14) will converge to $\mathbf{0}$, a.s.. Thus (14) becomes

$$(14^*) \qquad \sup_{\|\mathbf{u}\|\leq B} |n^{1/2}\{F_n(x, \rho + n^{-1/2}\mathbf{u}) - F_n(x, \rho)\}| = o_p(1).$$

In particular, this implies that if $\hat{\rho}$ is an estimator of ρ such that

$$\|n^{1/2}(\hat{\rho} - \rho)\| = O_p(1),$$

then

$$\|n^{1/2}\{F_n(\cdot, \hat{\rho}) - F_n(\cdot, \rho)\}\|_\infty = o_p(1).$$

Consequently, *the estimation of ρ has asymptotically negligible effect on the estimation of the error d.f.* F. This is similar to the fact, observed in the previous chapter, that the estimation of the slope parameters in linear regression has asymptotically negligible effect on the estimation of the error d.f. as long as the design matrix is centered at the origin. □

An important application of (14) occurs when proving the a.u.l property of the serial rank correlations of the residuals as functions of $\mathbf{t}$. More precisely, let $R_{i\mathbf{t}}$ denote the rank of $X_i - \mathbf{t}'\mathbf{Y}_{i-1}$ among $X_j - \mathbf{t}'\mathbf{Y}_{j-1}$, $1 \leq j \leq n$, $1 \leq i \leq n$. Define $R_{i\mathbf{t}} = 0$ for $i \leq 0$. Rank correlations of lag j, for $1 \leq j \leq p$, are defined as

$$(15) \qquad S_j(\mathbf{t}) := \frac{12}{n(n^2-1)} \sum_{i=j+1}^{n} \left(R_{i-j\mathbf{t}} - \frac{(n+1)}{2}\right)\left(R_{i\mathbf{t}} - \frac{(n+1)}{2}\right), \qquad \mathbf{t} \in \mathbb{R}^p,$$

$$\mathbf{S}' := (S_1, \ldots, S_p).$$

Simple algebra shows that

$$S_j(\mathbf{t}) = a_n[L_j(\mathbf{t}) - n(n+1)^2/4] + b_{nj}(\mathbf{t}), \qquad 1 \leq j \leq p,$$

where a_n is a nonrandom sequence not depending on $\mathbf{t}$, $|a_n| = O(1)$,

$$b_{nj}(\mathbf{t}) = \frac{6(n+1)}{\{n(n^2-1)\}}\Big(\sum_{i=n-j+1}^{n} + \sum_{i=1}^{j}\Big) R_{i\mathbf{t}},$$

and

$$L_j(\mathbf{t}) := n^{-3}\sum_{i=j+1}^{n} R_{i-j\mathbf{t}}\, R_{i\mathbf{t}}, \qquad 1 \le j \le p, \ \mathbf{t} \in \mathbb{R}^p.$$

Observe that $\sup\{|b_{nj}(\mathbf{t})|;\ \mathbf{t}\in\mathbb{R}^p\} \le 48p/n$, so that $n^{1/2}\sup\{|b_{nj}(\mathbf{t})|;\ \mathbf{t}\in\mathbb{R}^p\}$ tends to zero, a.s. It thus suffices to prove the a.u.l. of $\{L_j\}$ only, $1 \le j \le p$.

In order to state the a.u.l. result we need to introduce

$$\text{(16)} \qquad \begin{aligned} Z_{ij} &:= f(\epsilon_{i-j})F(\epsilon_i) + f(\epsilon_i)F(\epsilon_{i-j}), && i > j,\\ &:= 0, && i \le j.\\ \mathbf{U}_{ij} &:= \mathbf{Y}_{i-j-1}\, F(\epsilon_i)f(\epsilon_{i-j}) + \mathbf{Y}_{i-1}\, f(\epsilon_i)F(\epsilon_{i-j}), && i > j,\\ &:= 0, && i \le j.\\ \bar{Z}_j &:= n^{-1}\sum_{i=j+1}^{n} Z_{ij}, \qquad \bar{\mathbf{U}}_j := n^{-1}\sum_{i=j+1}^{n} \mathbf{U}_{ij}, && 1 \le j \le p.\\ \bar{\mathbf{Y}}_n &:= n^{-1}\sum_{i=1}^{n} \mathbf{Y}_{i-1}. \end{aligned}$$

Observe that $\{Z_{ij}\}$ are bounded r.v.'s with $EZ_{ij} = \int f^2(x)\,dx$ for all i and j. Moreover, $\{\epsilon_i\}$ i.i.d. F imply that $\{Z_{ij},\ j<i\le n\}$ are stationary and ergodic. By the Ergodic Theorem

$$\bar{Z}_j \longrightarrow b(f) := \int f^2(x)\,dx, \quad \text{a.s.}, \quad j = 1, \ldots, p.$$

We are now ready to state and prove

Theorem 7.2.2. *Assume that* (7.1.1), (a2), (a3*) *and* (a4) *hold. Then, for every* $0 < B < \infty$ *and for every* $1 \le j \le p$,

$$\text{(17)} \qquad \sup_{\|\mathbf{u}\|\le B} \big|n^{1/2}[L_j(\boldsymbol{\rho}+n^{-1/2}\mathbf{u}) - L_j(\boldsymbol{\rho})] - \mathbf{u}'[b(f)\bar{\mathbf{Y}}_n - \bar{\mathbf{U}}_j]\big| = o_p(1).$$

If (a2) *and* (a3*) *are strengthened to requiring* $E(\|\mathbf{Y}_0\|^2 + \epsilon_1^2) < \infty$ *and* $\{X_i\}$ *stationary and ergodic then* $\bar{\mathbf{Y}}_n$ *and* $\bar{\mathbf{U}}_j$ *may be replaced by their respective expectations in* (17).

Proof. Fix a j in $1\le j\le p$. For the sake of simplicity of the exposition, write $L(\mathbf{u})$, $L(0)$ for $L_j(\boldsymbol{\rho}+n^{-1/2}\mathbf{u})$, $L_j(\boldsymbol{\rho})$, respectively. Apply similar convention to other functions of $\mathbf{u}$. Also write $\epsilon_{i\mathbf{u}}$ for $\epsilon_i - n^{-1/2}\mathbf{u}'\mathbf{Y}_{i-1}$ and $F_n(\cdot)$ for $F_n(\cdot, \boldsymbol{\rho})$. With these conventions $R_{i\mathbf{u}}$ is now the rank of

$X_i - (\rho + n^{-1/2}\mathbf{u})' Y_{i-1} = \epsilon_{i\mathbf{u}}$. In other words, $R_{i\mathbf{u}} \equiv nF_n(\epsilon_{i\mathbf{u}}, \mathbf{u})$ and

$$L(\mathbf{u}) = n^{-1} \sum_{i=j+1}^{n} F_n(\epsilon_{i-j\mathbf{u}}, \mathbf{u})\, F_n(\epsilon_{i\mathbf{u}}, \mathbf{u}), \qquad \mathbf{u} \in \mathbb{R}^p.$$

The proof is based on the linearity properties of $F_n(\cdot, \mathbf{u})$ as given in (14) of Corollary 7.2.2 above. In fact if we let

$$B_n(x, \mathbf{u}) := F_n(x, \mathbf{u}) - F_n(x) - n^{-1/2}\mathbf{u}' \mathbf{Y}_n\, f(x), \qquad x \in \mathbb{R}.$$

then (14) is equivalent to

$$\sup n^{1/2} |B_n(x, \mathbf{u})| = o_p(1).$$

All supremums, unless specified otherwise, in the proof are over $x \in \mathbb{R}$, $1 \le i \le n$ and/or $\|\mathbf{u}\| \le B$. Rewrite

$$\begin{aligned} n^{1/2}&(L(\mathbf{u}) - L(0)) \\ &= n^{-1/2} \sum_{i=j+1}^{n} \{F_n(\epsilon_{i-j\mathbf{u}}, \mathbf{u})\, F_n(\epsilon_{i\mathbf{u}}, \mathbf{u}) - F_n(\epsilon_{i-j})\, F_n(\epsilon_i)\} \\ &= n^{-1/2} \sum_{i=j+1}^{n} [\{B_n(\epsilon_{i-j\mathbf{u}}, \mathbf{u}) + F_n(\epsilon_{i-j\mathbf{u}}) + n^{-1/2}\mathbf{u}' \mathbf{Y}_n\, f(\epsilon_{i-j\mathbf{u}})\} \cdot \\ &\qquad\qquad \cdot \{B_n(\epsilon_{i\mathbf{u}}, \mathbf{u}) + F_n(\epsilon_{i\mathbf{u}}) + n^{-1/2}\mathbf{u}' \mathbf{Y}_n\, f(\epsilon_{i\mathbf{u}})\} \\ &\qquad\qquad\qquad - F_n(\epsilon_{i-j})\, F_n(\epsilon_i)]. \end{aligned}$$

Hence, from (14), (a2) and (a3*),

$$\begin{aligned} (18) \quad n^{1/2}&(L(\mathbf{u}) - L(0)) \\ &= n^{-1/2} \sum_{i=j+1}^{n} [F_n(\epsilon_{i-j\mathbf{u}})\, F_n(\epsilon_{i\mathbf{u}}) - F_n(\epsilon_i)\, F_n(\epsilon_{i-j})] \\ &\quad + n^{-1} \sum_{i=j+1}^{n} [F_n(\epsilon_{i-j\mathbf{u}}) f(\epsilon_{i\mathbf{u}}) + F_n(\epsilon_{i\mathbf{u}}) f(\epsilon_{i-j\mathbf{u}})](\mathbf{u}' \mathbf{Y}_n) + \bar{o}_p(1), \end{aligned}$$

where, now, $\bar{o}_p(1)$ is a sequence of stochastic processes converging to zero uniformly, in probability, over the set $\mathcal{N}(B)$.

Now recall that (a4) and the asymptotic uniform continuity of the standard empirical process based on i.i.d. r.v.'s imply that

$$\sup_{|x-y| \le \delta} n^{1/2} |[F_n(x) - F(x)] - [F_n(y) - F(y)]| = o_p(1)$$

when first $n \to \infty$ and then $\delta \to 0$. Hence from (a2) and the fact that

$$\sup_{i,\mathbf{u}} |\epsilon_{i\mathbf{u}} - \epsilon_i| \le Bn^{-1/2}\max_i \|\mathbf{Y}_{i-1}\|,$$

one readily obtains

$$\sup_{i,\mathbf{u}} n^{1/2}|[F_n(\epsilon_{i\mathbf{u}}) - F(\epsilon_{i\mathbf{u}})] - [F_n(\epsilon_i) - F(\epsilon_i)]| = o_p(1).$$

From this and (a4) we obtain

$$\text{(19)} \qquad \sup_{i,\mathbf{u}} n^{1/2}|F_n(\epsilon_{i\mathbf{u}}) - F_n(\epsilon_i) + n^{-1/2}\mathbf{u}'\mathbf{Y}_{i-1} f(\epsilon_i)| = o_p(1).$$

From (18), (19), the uniform continuity of f and F, the Glivenko–Cantelli lemma, one obtains

$$\begin{aligned}\text{(20)} \quad & n^{1/2}(L(\mathbf{u}) - L(0)) \\ &= n^{-1}\sum_{i=j+1}^{n} [F(\epsilon_{i-j})\, f(\epsilon_i) + F(\epsilon_i)\, f(\epsilon_{i-j})](\mathbf{u}'\mathbf{Y}_n) \\ &\qquad - \mathbf{u}'n^{-1}\sum_{i=j+1}^{n}\{\mathbf{Y}_{i-j-1} f(\epsilon_{i-j})F(\epsilon_i) + \mathbf{Y}_{i-1} f(\epsilon_i)F(\epsilon_{i-j})\} + \bar{o}_p(1).\end{aligned}$$

In concluding (20) we also used the fact that by (a2) and (a3*),

$$\sup_{\mathbf{u}} |n^{-3/2}\sum_{i=j+1}^{n} |\mathbf{u}'\mathbf{Y}_{i-j}\cdot\mathbf{u}'\mathbf{Y}_{i-1}| \le B\, n^{-1/2}\max_i \|\mathbf{Y}_{i-1}\|\, n^{-1}\sum_{i=j+1}^{n} \|\mathbf{Y}_{i-j}\| = o_p(1).$$

Now (17) readily follows from (20) and the notation introduced just before the statement of the theorem. The rest is obvious. □

Remark 7.2.4. *Autoregressive moving average models.* Boldin (1989) and Kreiss (1991) give an analogue of (14*) for a moving average model of order q and an autoregressive moving average model of order (p,q) (ARMA(p,q)), respectively, when the error d.f. F has zero mean, finite second moment and bounded second derivative. Here we shall illustrate as to how Theorem 2.2b.1 can be used to yield the same result under weaker conditions on F. For the sake of clarity, the details are carried out for an ARMA(1,1) model only.

Let $\epsilon_0, \epsilon_1, \epsilon_2, \ldots,$ be i.i.d. F r.v.'s and X_0 be a r.v. independent of $\{\epsilon_i, i \ge 1\}$. Consider the process given by the relation

$$\text{(21)} \qquad X_i = \rho X_{i-1} + \epsilon_i + \beta\, \epsilon_{i-1}, \qquad i \ge 1,$$

where $|\rho| < 1$, $|\beta| < 1$. One can rewrite this model as

$$\begin{aligned}\text{(22)} \qquad \epsilon_i &= X_1 - (\rho X_0 + \beta\epsilon_0), & i = 1, \\ &= X_i - \sum_{j=1}^{i-1}(-\beta)^j(\rho+\beta)\, X_{i-j-1} + (-\beta)^{i-1}(\rho X_0 + \beta\epsilon_0), & i \ge 2.\end{aligned}$$

Let $\boldsymbol{\theta} := (s, t)'$ denote a point in the open square $(-1, 1)^2$ and $\boldsymbol{\theta}_0 := (\rho, \beta)'$ denote the true parameter value. Assume that $\boldsymbol{\theta}$'s are restricted to the following sequence of neighborhoods: For a $b \in (0, \infty)$,

(23) $$n^{1/2}\{|s-\rho| + |t-\beta|\} \le b.$$

Let $\{\tilde{\epsilon}_i, i \ge 1\}$ stand for the residuals $\{\epsilon_i, i \ge 1\}$ of (22) after ρ and β are replaced by s and t, respectively, in (22). Let $F_n(\cdot, \boldsymbol{\theta})$ denote the empirical process of $\{\tilde{\epsilon}_i, 1 \le i \le n\}$. This empirical can be rewritten as

(24) $$F_n(x, \boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n} I(\epsilon_i \le x + \delta_{ni}), \qquad x \in \mathbb{R},$$

where

(25) $$\begin{aligned} \delta_{ni} &:= (s-\rho)X_0 + (t-b)\,\epsilon_0, & i = 1,\\ &= \sum_{j=1}^{i-2}[(-t)^j(s+t) - (-\beta)^j(\rho+\beta)]\,X_{i-j-1} & \\ &\qquad + (-t)^{i-1}(sX_0 + t\epsilon_0) - (-\beta)^{i-1}(\rho X_0 + \beta\epsilon_0), & i \ge 2.\\ &= \delta_{ni1} + \delta_{ni2}, \quad \text{say}, & i \ge 2. \end{aligned}$$

From (25), it follows that for every $\boldsymbol{\theta} \in (-1, 1)^2$ satisfying (23),

$$|\delta_{n1}| \le bn^{-1/2}(|X_0| + |\epsilon_0|),$$

$$\max_{2\le i\le n}|\delta_{ni1}| \le 2b\,n^{-1/2}\max_{1\le i\le n}|X_i|\,(1 - bn^{-1/2} - \beta)^{-1}\{1 + (1-|\beta|)^{-1}\},$$

$$\max_{2\le i\le n}|\delta_{ni2}| \le 2bn^{-1/2}(1 - bn^{-1/2} - \beta)^{-1}(|X_0| + |\epsilon_0|).$$

Consequently, if $n^{-1/2}\max_{1\le i\le n}|X_i| = o_p(1)$, then the $\{\delta_{ni}\}$ of (25) would satisfy (2.2b.A2) for every $\boldsymbol{\theta} \in (-1, 1)^2$. But by (21),

(26) $$\begin{aligned} X_i &= \rho X_0 + \beta\epsilon_0 + \epsilon_1, & i = 1,\\ &= \rho^{i-1}(\rho X_0 + \beta\epsilon_0) + \sum_{j=1}^{i-2}\rho^j(\rho+\beta)\,\epsilon_{i-j-1} + \epsilon_i, & i \ge 2. \end{aligned}$$

Therefore, (2.2b.A2) will hold for the above $\{\delta_{ni}\}$ if

(27) $$n^{-1/2}\max_{1\le i\le n}|\epsilon_i| = o_p(1).$$

We now verify (2.2b.A3) for the above $\{\delta_{ni}\}$ and with $h_{ni} \equiv 1$. That is we must show that $n^{-1/2} \sum_i |\delta_{ni}| = O_p(1)$. We proceed as follows. Let $u = n^{1/2}(s-\rho)$, $v = n^{1/2}(t-\beta)$ and $Z_0 := |X_0| + |\epsilon_0|$. By (23), $|u| + |v| \le b$. From (25),

$$\begin{aligned}(28)\qquad n^{-1/2} \sum_i |\delta_{ni}| &\le n^{-1} b Z_0 \\ &+ n^{-1/2} \sum_{i=2}^{n} \Big| \sum_{j=0}^{i-2} [(-t)^j (s+t) - (-\beta)^j(\rho+\beta)]\, X_{i-j-1} \Big| \\ &+ n^{-1/2} \sum_{i=2}^{n} |(-t)^{i-1}(sX_0 + t\epsilon_0) - (-\beta)^{i-1} Z|, \\ &= A_{n1} + A_{n2} + A_{n3}, \quad \text{say.}\end{aligned}$$

Clearly, $|A_{n1}| = o(1)$, a.s. Rewrite

$$\begin{aligned} A_{n2} &= n^{-1/2} \sum_{i=2}^{n} \Big| \sum_{j=0}^{i-2} [(-t)^j \{(u+v)n^{-1/2} + \rho + \beta\} - (-\beta)^j (\rho+\beta)]\, X_{i-j-1} \Big| \\ &\le 2bn^{-1} \sum_{i=2}^{n} \sum_{j=0}^{i-2} |t|^j |X_{i-j-1}| + 2\, n^{-1/2} \sum_{i=2}^{n} \Big| \sum_{j=0}^{i-2} [(-t)^j - (-\beta)^j] X_{i-j-1} \Big| \\ (29)\qquad &= 2bA_{n21} + 2A_{n22}, \quad \text{say.}\end{aligned}$$

By a change of variables and an interchange of summations one obtains

$$(30)\qquad A_{n21} \le n^{-1} \sum_{i=1}^{n} |X_i|\, (1-|t|)^{-1}.$$

Next, use the expansion $a^j - c^j = (a-c) \sum_{k=0}^{j-1} a^{j-1-k} c^k$ for any real numbers a, c, to obtain

$$A_{n22} \le b\, n^{-1} \sum_{i=3}^{n} \sum_{j=1}^{i-2} \sum_{k=0}^{j-1} |t|^{j-1-k}\, |\beta|^k |X_{i-j-1}|.$$

Again, use change of variables and interchange of summations repeatedly and the fact that $|\beta| \vee |t| < 1$, to conclude that this upper bound is bounded above by

$$b(1-|\beta|)^{-1}[(1-|t|)^{-1} + 1]\, n^{-1} \sum_{i=1}^{n} |X_i|.$$

This, (28) and (29) together with (23) imply that

$$(31)\qquad A_{n2} \leq 2b\, n^{-1} \sum_{i=1}^{n} |X_i|\, [(1-bn^{-1/2}-|\beta|)^{-1}\{1+(1-|\beta|)^{-1}\} + (1-|\beta|)^{-1}].$$

Finally, similar calculations show that

$$(32)\qquad A_{n3} = O_p(n^{-1/2}).$$

From (28), (31) and (32) it thus follows that if $n^{-1}\sum_{i=1}^{n}|X_i| = O_p(1)$, then the $\{\delta_{ni}\}$ of (25) will satisfy (2.2b.A3) with $h_{ni} \equiv 1$. But in view of (26) and the assumption that $|\rho| \vee |\beta| < 1$, it readily follows that if

$$(33)\qquad n^{-1}\sum_{i=1}^{n}|\epsilon_i| = O_p(1),$$

then (2.2b.A3) with $h_{ni} \equiv 1$ holds for the $\{\delta_{ni}\}$ of (25). We have thus proved the following:

If (21) holds with the error d.f. F satisfying (**F1**), (**F2**), (27) and (33), then $\forall\ \boldsymbol{\theta} \in (-1, 1)^2$,

$$\sup_x |n^{-1/2} \sum_{i=1}^{n}\{I(\tilde{\epsilon}_i \leq x) - I(\epsilon_i \leq x) - F(x+\delta_{ni}) + F(x)\}| = o_p(1).$$

Now use an argument like the one used in the proof of Theorem 7.2.1 to conclude the following

Corollary 7.2.3. *In addition to* (21), *assume that the error d.f.* F *satisfies* (**F1**), (**F2**), (27) *and* (33). *Then,* $\forall\ 0 < b < \infty$,

$$\sup|n^{1/2}[F_n(x,\boldsymbol{\theta}) - F_n(x,\boldsymbol{\theta}_0)] - n^{-1/2}\Sigma_i\, \delta_{ni}\, f(x)| = o_p(1),$$

where the supremum is taken over $x \in \mathbb{R}$ *and* $\boldsymbol{\theta}$, $\boldsymbol{\theta}_0$ *satisfying* (23).

If (33) *is strengthened to assuming that* $E|\epsilon| < \infty$, *then*

$$\sup|n^{-1/2}\sum_i \delta_{ni} - n^{1/2}[(s-\rho)(1-\rho)^{-1} + (t-\beta)(1+\beta)^{-1}]\,\mu| = o_p(1),$$

where the supremum is taken over s, t *satisfying* (23) and $\mu = E\epsilon$. □

Consequently, if $E\epsilon = 0$ and $(\hat{\rho}, \hat{\beta})$ is an estimator of (ρ, β) such that $\|n^{1/2}(\hat{\rho}-\rho,\ \hat{\beta}-\beta)\| = O_p(1)$, then an analogue of (14*) holds in the present case also under weaker conditions than those given by Boldin or Kreiss.

The details for proving an anlogue of Corollary 7.2.3 for a general ARMA(p,q) model are similar but some what complicated to those given above. □

7.3. GM- and R- Estimators.

In this section we shall discuss the asymptotic distributions of GM- and R-estimators of ρ. In addition, some consistent estimators of the functional Q(f) will be also constructed. We begin with

7.3a. GM-Estimators.

Here we shall state the asymptotic normality of the GM-estimators. Let $\rho_{\mathbf{M}}$ stand for a solution of (7.1.3) such that $\|n^{1/2}(\rho_{\mathbf{M}} - \rho)\| = O_p(1)$. That such an estimator $\rho_{\mathbf{M}}$ exists can be seen by an argument similar to the one given in Huber (1981) in connection with the linear regression model. To state the asymptotic normality of $\rho_{\mathbf{M}}$ we need to introduce some more notation. Let

$$(1)\qquad \mathscr{X} := \begin{bmatrix} X_0 & X_{-1},\dots,X_{1-p} \\ X_1 & X_0,\dots\dots,X_{2-p} \\ \vdots & \vdots \qquad \vdots \\ X_{n-1} & X_{n-2},\dots,X_{n-p} \end{bmatrix}, \quad \mathscr{G} := \begin{bmatrix} g(X_0) & g(X_{-1}),\dots,g(X_{1-p}) \\ g(X_1) & g(X_0),\dots,g(X_{2-p}) \\ \vdots & \vdots \qquad \vdots \\ g(X_{n-1}) & g(X_{n-2}),\dots,g(X_{n-p}) \end{bmatrix},$$

$$\mathcal{G}' := n^{1/2}(\mathcal{G}_1(\rho),\dots, \mathcal{G}_p(\rho)), \quad \mathcal{B}_n := \mathscr{G}'\mathscr{X} = \Sigma_i\,(g(X_{i-1})Y'_{i-1},\dots, g(X_{i-p})Y'_{i-1})'.$$

Proposition 7.3.1. *In addition to* (7.1.1), (7.2.a2), (7.2.a4), (7.2.b1), (7.2.b2) and (7.2.b3) *assume that*

(b4) $n^{-1}\mathcal{B}_n = \mathcal{B} + o_p(1)$, *for some* $p\times p$ *non-random positive definite matrix* $\mathcal{B}$.

Then

$$n^{1/2}(\rho_{\mathbf{M}} - \rho) = -(\mathcal{B}a)^{-1}\mathcal{G} + o_p(1).$$

If, in addition, we assumes that

(b5) $n^{-1}\mathscr{G}'\mathscr{G} = \mathcal{G}^* + o_p(1)$, $\mathcal{G}^*$ *a* $p\times p$ *non-random positive definite matrix,*

then

$$n^{1/2}(\rho_{\mathbb{M}} - \rho) \underset{d}{\rightarrow} N(0, \mathbf{J}), \qquad \mathbf{J} := (\int f\, d\psi)^{-2}\cdot(\int \psi^2\, dF)\cdot \mathcal{B}^{-1} G^* \mathcal{B}^{-1}.$$

Proof. Follows from Corollary 7.2.1, the Cramer–Wold device and Lemma A.3 in the appendix applied to $\mathcal{G}$. □

Again, if Y_0 and $\{\epsilon_i\}$ are so chosen as to make $\{X_i\}$ stationary, ergodic and $E(\|Y_0\|^2+\epsilon_1^2)<\infty$ then (b4) and (b5) are *a priori* satisfied. See, e.g., Anderson (1971; p 203).

Note: For a more general class of GM–estimators see Bustos (1982) where a result analogous to the above corollary for smooth score functions ψ is obtained. □

7.3b. R-Estimators.

This section will discuss an analogue of Jaeckel's (1972) R-estimators of ρ and their large sample properties.

Recall that R_{it} is the rank of $X_i - \mathbf{t}'Y_{i-1}$ among $\{X_k - \mathbf{t}'Y_{k-1},\ 1 \leq k \leq n\}$, for $1 \leq i \leq n$. Also, $R_{it} \equiv 0$ for $i \leq 0$. Let φ be a nondecreasing score function from $[0, 1]$ to the real line such that

$$\text{(1)} \qquad \sum_{i=1}^{n} \varphi(i/(n+1)) = 0.$$

For example, if $\varphi(t) = -\varphi(1-t)$ for all $t \in [0. 1]$, i.e., if φ is skew symmetric, then it satisifies (1). Define

$$\mathcal{S}_j(\mathbf{u}) := n^{-1} \sum_{i=j+1}^{n} X_{i-j}\, \varphi(R_{i\mathbf{u}}/(n+1)), \qquad 1 \leq j \leq p,\ \mathbf{u} \in \mathbb{R}^p,$$
$$\mathcal{S}' := (\mathcal{S}_1, \ldots, \mathcal{S}_p).$$

The class of rank statistics $\mathcal{S}$, one for each φ, is an analogue of the class of rank statistics discussed in Section 4.3 above in connection with linear regression models where one replaces the weights $\{X_{i-j}\}$ by appropriate design points. A test of the hypothesis $\rho = \rho_0$ may be based on a suitably standardized $\mathcal{S}(\rho_0)$, the large values of the statistic being significant.

It is thus natural to define R-estimators of ρ by the relationship

$$\text{(2)} \qquad \tilde{\rho}_R = \arg\min\{\|\mathcal{S}(\mathbf{t})\|;\ \mathbf{t} \in \mathbb{R}^p\}.$$

An alternative way to define R-estimators of ρ is to adapt Jaeckel (1972) to the AR(p) situation. Accordingly, for a $\mathbf{t} \in \mathbb{R}^p$, let

$$Z_k(\mathbf{t}) = X_k - \mathbf{t}'Y_{k-1},\ 1 \leq k \leq n,$$

$$Z_{(i)}(\mathbf{t}) := \text{the ith largest residual among } \{Z_k(\mathbf{t}),\ 1 \leq k \leq n\},\ 1 \leq i \leq n,$$

$$\mathscr{J}(\mathbf{t}) := \sum_{i=1}^{n} \varphi(i/(n+1))\, Z_{(i)}(\mathbf{t}) \equiv \sum_{i=1}^{n} \varphi(R_{it}/(n+1))(X_i - \mathbf{t}'\mathbf{Y}_{i-1}).$$

Then Jaeckel's estimator $\tilde{\boldsymbol{\rho}}_J$ is defined by the relation

$$\tilde{\boldsymbol{\rho}}_J = \arg\min\{\mathscr{J}(\mathbf{t});\ \mathbf{t} \in \mathbb{R}^p\}.$$

Jaeckel's argument about the existence of an analogue of $\tilde{\boldsymbol{\rho}}_J$ in the context of linear regression model can be adapted to the present situation. This follows from the following *three* lemmas, the first of which is of a general interest.

Lemma 7.3b.1. *Let* $d_1, d_2, \ldots, d_n, v_1, v_2, \ldots, v_n$, *be real numbers such that not all* $\{d_i\}$ *are the same and no two* $\{v_i\}$ *are the same. Let* r_{iu} *denote the rank of* $v_i - ud_i$ *among* $\{v_j - ud_j;\ 1 \le j \le n\}$, $u \in \mathbb{R}$. *Let* $\{b_n(i);\ 1 \le i \le n\}$ *be a set of real numbers that are nondecreasing in* i. *Let*

$$T(u) := \sum_{i=1}^{n} d_i\, b_n(r_{iu}), \qquad u \in \mathbb{R}.$$

Then, $T(u)$ *is a nonincreasing step function in all those* $u \in \mathbb{R}$ *for which there are no ties among* $\{v_j - ud_j;\ 1 \le j \le n\}$.

Proof. See Theorem II.7E, p35 of Hájek (1969). □

Lemma 7.3b.2. *Assume that the model* (7.1.1) *holds with* $(\mathbf{Y}_0', X_1, X_2, \ldots, X_n)$ *having a continuous joint distribution. Then the following hold.*

(a) *For each realization* $(\mathbf{Y}_0', X_1, X_2, \ldots, X_n)$, *the assumption* (1) *implies that* $\mathscr{J}(\mathbf{t})$ *is nonnegative, continuous and convex function of* $\mathbf{t}$ *with its a.e. derivative equal to* $-n\mathcal{S}(\mathbf{t})$.

(b) *If the realization* $(\mathbf{Y}_0', X_1, X_2, \ldots, X_n)$ *is such that the rank of* $\mathscr{X}_c$ *is* p *then, for every* $0 < b < \infty$, *the set* $\{\mathbf{t} \in \mathbb{R}^p;\ \mathscr{J}(\mathbf{t}) \le b\}$ *is bounded, where* $\mathscr{X}_c$ *is the* $\mathscr{X}$ *of* (7.3a.1), *centered at the origin.*

Proof. (a). For any $\mathbf{x}' = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, let $x(1) \le x(2) \le \ldots \le x(n)$ denote the ordered $x_1, x_2, \ldots, x_n$. Let $\Pi := \{\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_n)';\ \boldsymbol{\pi}$ a permutation of the integers $1, 2, \ldots, n.\}$, $b_n(i) := \varphi(i/(n+1))$, $1 \le i \le n$, and define

$$D(\mathbf{x}) := \sum_{i=1}^{n} b_n(i)\, x(i), \qquad D_{\boldsymbol{\pi}}(\mathbf{x}) := \sum_{i=1}^{n} b_n(i)\, x_{\pi_i}, \qquad \mathbf{x} \in \mathbb{R}^n,$$

$$k := \min\{1 \le j \le n;\ b_n(j) > 0\}.$$

Observe that $\mathscr{J}(\mathbf{t}) = D(\mathbf{Z}(\mathbf{t}))$.

Now, (1) and φ nondecreasing implies that

$$
\begin{aligned}
D(\mathbf{x}) &= \sum_{i=1}^{n} b_n(i)\,(x(i) - x(k)) \\
&= \sum_{i=1}^{k-1} b_n(i)\,(x(i) - x(k)) + \sum_{i=k}^{n} b_n(i)\,(x(i) - x(k)) \\
&\geq 0, \qquad\qquad \forall\ \mathbf{x} \in \mathbb{R}^n,
\end{aligned}
$$

because each summand is nonnegative. This proves that $\mathcal{J}(\mathbf{t}) \geq 0$, $\mathbf{t} \in \mathbb{R}$.
By Theorem 368 of Hardy, Littlewood and Polya (1952),

$$D(\mathbf{x}) = \max_{\boldsymbol{\pi}\in\Pi} D_{\boldsymbol{\pi}}(\mathbf{x}), \qquad \forall\ \mathbf{x} \in \mathbb{R}^n.$$

Therefore, $\forall\ \mathbf{t} \in \mathbb{R}^p$,

$$
(*) \qquad
\begin{aligned}
\mathcal{J}(\mathbf{t}) = D(Z(\mathbf{t})) &= \max_{\boldsymbol{\pi}\in\Pi} D_{\boldsymbol{\pi}}(Z(\mathbf{t})) \\
&= \max_{\boldsymbol{\pi}\in\Pi} \sum_{i=1}^{n} b_n(i)(X_{\pi_i} - \mathbf{t}'\mathbf{Y}_{\pi_{i-1}}).
\end{aligned}
$$

This shows that $\mathcal{J}(\mathbf{t})$ is a maximal element of a finite number of continuous and convex functions, which itself is continuous and convex. The statement about a.e. differential being $-n\mathcal{S}(\mathbf{t})$ is obvious. This completes the proof of (a).

(b) Without the loss of generality assume $b > \mathcal{J}(\mathbf{0})$. Write a $\mathbf{t}\in\mathbb{R}^p$ as $\mathbf{t} = u\boldsymbol{\theta}$, $u\in\mathbb{R}$, $\boldsymbol{\theta}\in\mathbb{R}^p$, $\|\boldsymbol{\theta}\| = 1$. Let $d_i \equiv \boldsymbol{\theta}'\mathbf{Y}_{i-1}$. The assumptions about $\mathcal{X}$ imply that not all $\{d_i\}$ are equal. Rewrite

$$\mathcal{J}(\mathbf{t}) = \mathcal{J}(u\boldsymbol{\theta}) = \sum_{i=1}^{n} b_n(i)\,(X - ud)(i) = \sum_{i=1}^{n} b_n(r_{iu})(X_i - ud_i)$$

where now r_{iu} is the rank of $X_i - ud_i$ among $\{X_j - ud_j;\ 1 \leq j \leq n\}$. From (*) above, it follows that $\mathcal{J}(u\boldsymbol{\theta})$ is linear and convex in u, for every $\boldsymbol{\theta}\in\mathbb{R}^p$, $\|\boldsymbol{\theta}\| = 1$. Its a.e. derivative w.r.t. u is $-\sum_{i=1}^{n} d_i b_n(r_{iu})$, which by Lemma 7.3b.1 and because of the assumed continuity, is nondecreasing in u and eventually positive. Hence $\mathcal{J}(u\boldsymbol{\theta})$ will eventually exceed b, for every $\boldsymbol{\theta}\in\mathbb{R}^p$, $\|\boldsymbol{\theta}\| = 1$.

Thus, there exists a $u_{\boldsymbol{\theta}}$ such that $\mathcal{J}(u_{\boldsymbol{\theta}}\boldsymbol{\theta}) > b$. Since $\mathcal{J}$ is continuous, there is an open set $O_{\boldsymbol{\theta}}$ of unit vectors $\boldsymbol{\nu}$, containing $\boldsymbol{\theta}$ such that $\mathcal{J}(u_{\boldsymbol{\theta}}\boldsymbol{\nu}) > b$. Since $b > \mathcal{J}(\mathbf{0})$, and $\mathcal{J}$ is convex, $J(u\boldsymbol{\nu}) > b$, $\forall\ u \geq u_{\boldsymbol{\theta}}$ and $\forall\ \boldsymbol{\nu}\in O_{\boldsymbol{\theta}}$. Now, for each unit vector $\boldsymbol{\theta}$, there is an open set $O_{\boldsymbol{\theta}}$ covering it. Since the unit sphere is compact, a finite number of these sets covers it. Let m be the maximum of the corresponding finite set of $u_{\boldsymbol{\theta}}$. Then for all $u \geq m$, for all unit vectors $\boldsymbol{\nu}$, $\mathcal{J}(u\boldsymbol{\nu}) > b$. This proves the claim (b) and also the lemma. □

Note: Lemma 7.3b.2 *and its proof is an adaptation of Theorems 1 and 2 of Jaeckel* (1972) *to the present case.* □

From the above lemma it follows that if the r.v.'s $Y_0, X_1, X_2, \ldots, X_n$ are continuous and the matrix $n^{-1}\Sigma_i(Y_{i-1} - \bar{Y})(Y_{i-1} - \bar{Y})'$ is a.s. positive definite, then the rank of $\mathcal{K}$ is a.s. p and the set $\{t \in \mathbb{R}^p;\ \mathcal{J}(t) \leq b\}$ is a.s. bounded for every $0 \leq b < \infty$. Thus a minimizer $\tilde{\rho}_J$ of $\mathcal{J}$ exists a.s. and has the property that makes $\|\mathcal{S}\|$ small. As is shown in Jaeckel (1972) in connection with the linear regression model, it will follow from the linearity result given in Theorem 7.3b.1 below that $\tilde{\rho}_J$ and $\tilde{\rho}_R$ are asymptotically equivalent. Note that the score function φ need not satisfy (1) in this theorem.

Some steps of the proof of Theorem 7.3b.1 heavily depend on the representation of the AR(p) process $\{X_i\}$ in terms of the error variables $\{\epsilon_i\}$. For that reason we shall now extend the index i in the process $\{X_i\}$ to both sides of 0. Accordingly, assume that $\{\epsilon_i, i = 0, \pm 1, \pm 2, \ldots.\}$ are i.i.d. F r.v.'s and that

(3) $$X_i = \rho_1 X_{i-1} + \rho_2 X_{i-2} + \ldots + \rho_p X_{i-p} + \epsilon_i, \qquad i = 0, \pm 1, \pm 2, \ldots, \quad \rho \in \mathbb{R}^p.$$

In addition assume the following:

(4) *All roots of the equation*

$$x^p - \rho_1 x^{p-1} - \rho_2 x^{p-2} - \ldots - \rho_p = 0 \text{ are in the interval } (-1, 1).$$

It is well known that if $E|\epsilon|^2 < \infty$, there exist constants $\{\theta_j, j \geq 0\}$ such that $\theta_0 = 1$, $\Sigma_{j \geq 0} |\theta_j| < \infty$, and that

(5) $$X_i = \sum_{k \leq i} \theta_{i-k}\, \epsilon_k, \qquad i = 0, \pm 1, \pm 2, \ldots, \text{ in } L_2 \text{ and a.s.},$$

where the unspecified lower limit on the index of summation is $-\infty$. See, e.g., Anderson (1971) and Brockwell and Davis (1987, pp 76–86). Thus $\{X_i\}$ is stationary, ergodic and $E\|Y_0\|^2 < \infty$. Hence (7.2.a1) implies (7.2.a3). Moreover, the stationarity of $\{Y_{i-1}\}$ and $E\|Y_0\|^2 < \infty$ imply that $\forall\, \eta > 0$,

(6) $$P(\max_{1 \leq i \leq n} \|Y_{i-1}\| \geq \eta\, n^{1/2}) \leq \{\eta n^{1/2}\}^{-2} \sum_{i=1}^{n} E\|Y_{i-1}\|^2\, I(\|Y_{i-1}\| \geq \eta n^{1/2})$$

$$= \eta^{-2}\, E\|Y_0\|^2 I(\|Y_0\| \geq \eta n^{1/2}) = o(1).$$

Thus (7.2.a2) holds. These observations will be used in the sequel frequently, some times without mentioning.

With this preliminary background, we now state

Theorem 7.3b.1. (*A.U.L. of R–statistics*). *Assume that* (3) *and* (4) *above hold. In addition, assume that* F *satisfies* **(F1)**, **(F2)** *and that the following hold.*

(c1) (i) $E\epsilon = 0$. (ii) $0 < E\epsilon^4 < \infty$.

(c2) *φ is nondecreasing and differentiable with its derivative $\dot{\varphi}$ being uniformly continuous on* [0, 1].

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

$$\sup_{\|u\| \le B} \|n^{1/2}\{\mathcal{S}(\rho + n^{-1/2}u) - \hat{\mathcal{S}}\} + u' \Sigma Q\| = o_p(1), \tag{7}$$

where $\hat{\mathcal{S}}' := (\hat{\mathcal{S}}_1, \ldots, \hat{\mathcal{S}}_p)$ *with*

$$\hat{\mathcal{S}}_j := n^{-1} \sum_{i=j+1}^{n} (X_{i-j} - \overline{X}_j)[\varphi(F(\epsilon_i)) - \overline{\varphi}], \quad \overline{\varphi} = \int_0^1 \varphi(t)dt, \ \ 1 \le j \le p,$$

$$\overline{X}_j := n^{-1} \sum_{i=j+1}^{n} X_{i-j}, \qquad Q := \int f \, d\varphi(F),$$

$$\Sigma := ((\beta(k-j)), 1 \le k \le p; \ 1 \le i \le p; \ \beta(k) = \mathrm{Cov}\,(X_0, X_k), 1 \le k \le p.$$

Before proceeding to prove the above result, we shall state a lemma giving the asymptotic continuity of certain basic r.w.e.p.'s. Accordingly, let h be a nonnegative measurable function from [0, 1] to $\mathbb{R}$, U denote a uniform [0, 1] r.v., and define

$$\mathcal{Z}_j(t) := n^{-1/2} \sum_{i=1}^{n} X_{i-j} \, [h(F(\epsilon_i)) \, I(F(\epsilon_i) \le t) - H(t)], \ \ 0 \le t \le 1, \ \ 1 \le j \le p,$$

where $H(t) := E\, h(U) I(U \le t) = \int_0^t h(s)ds, \ \ 0 \le t \le 1.$

The proof of the following lemma will be given in the subsection 7.3d.

Lemma 7.3b.3. *In addition to* (3), (4) *and* (c1(i)) *assume that* $Eh^4(U) < \infty$. *Then,* $\forall \ \gamma > 0, \ \forall \ 1 \le j \le p$,

$$\lim_{\eta \to 0} \limsup_n P(\sup_{|t-v| \le \eta} |\mathcal{Z}_j(t) - \mathcal{Z}_j(v)| > \gamma) = 0. \qquad \square$$

Proof of Theorem 7.3b.1. Observe that with

$$\tilde{S}_j(\mathbf{u}) := n^{-1}\sum_{i=1}^{n} X_{i-j}\,\varphi(R_{i\mathbf{u}}/(n+1)), \quad \mathbf{u}\in\mathbb{R}^p,\ 1\le j\le p,$$

(8) $$\sup_{1\le j\le p,\,\mathbf{u}} n^{1/2}|S_j(\mathbf{u}) - \tilde{S}_j(\mathbf{u})| \le p \max_{1-p\le k\le 0} |X_k|\ \|\varphi\|_\infty\, n^{-1/2} \longrightarrow 0, \text{ a.s..}$$

Thus it suffices to prove the theorem with $\{S_j\}$ replaced by $\{\tilde{S}_j\}$. Let $\tilde{S}' := (\tilde{S}_1, \ldots, \tilde{S}_p)$. Observe that

$$\tilde{S}(\mathbf{u}) = n^{-1}\sum_{i=1}^{n} \mathbf{Y}_{i-1}\,\varphi(R_{i\mathbf{u}}/(n+1)), \qquad \mathbf{u}\in\mathbb{R}^p.$$

The proof is facilitated by centering $\tilde{S}$. Accordingly, define

$$\mathbf{M}(\mathbf{u}) := n^{-1}\sum_{i=1}^{n} \mathbf{Y}_{i-1}[\varphi(R_{i\mathbf{u}}/(n+1)) - \bar{\varphi}], \qquad \mathbf{u}\in\mathbb{R}^p,$$

$$\hat{\mathbf{M}} := n^{-1}\sum_{i=1}^{n} \mathbf{Y}_{i-1}[\varphi(F(\epsilon_i)) - \bar{\varphi}].$$

As in the proof of Theorems 7.2.1, 7.2.2, let $\mathbf{M}(\mathbf{u})$, $F_n(\cdot, \mathbf{u})$, etc. stand for $\mathbf{M}(\rho+n^{-1/2}\mathbf{u})$, $F_n(\cdot, \rho+n^{-1/2}\mathbf{u})$, etc. Thus, e.g., $F_n(\cdot, \mathbf{0})$ now stands for the empirical d.f. of ϵ_i, $1\le i\le n$. Write $F_n(\cdot)$ for $F_n(\cdot, \mathbf{0})$. Recall, from the proof of Theorem 7.2.2 that $\epsilon_{i\mathbf{u}} = \epsilon_i - n^{-1/2}\mathbf{u}'\mathbf{Y}_{i-1}$, $n^{-1}R_{i\mathbf{u}} \equiv F_n(\epsilon_{i\mathbf{u}}, \mathbf{u})$. Now, let

$$e_{ni\mathbf{u}} = n^{1/2}[(R_{i\mathbf{u}}/(n+1)) - F(\epsilon_i)], \qquad 1\le i\le n,\ \mathbf{u}\in\mathbb{R}^p.$$

We first prove the

(9) **Claim:** $$\sup_{i,\mathbf{u}} n^{-1/2}|e_{ni\mathbf{u}}| = o_p(1).$$

As in the proof of Theorem 7.2.2, the supremum w.r.t. i, $\mathbf{u}$ will be over $1\le i\le n$, $\mathcal{N}(B)$, respectively, unless mentioned otherwise.

To begin with, $|[n(n+1)^{-1} - 1]| = O(n^{-1})$ implies that

(10) $$\sup_{i,\mathbf{u}} |n^{-1/2}e_{ni\mathbf{u}} - [F_n(\epsilon_{i\mathbf{u}}, \mathbf{u}) - F(\epsilon_i)]| = O(n^{-1}), \quad \text{a.s.}$$

Now, in view of (3), (4) and the discussion preceeding the stament of this theorem, it follows that $\{X_i\}$ are stationary, ergodic and hence by (c1(i)) and the Ergodic Theorem, $(1/n)\Sigma_i \mathbf{Y}_{i-1} \longrightarrow E\mathbf{Y}_0 = \mathbf{0}$. This together with (6) above, Remark 7.2.3 and (7.2.14*) imply that

$$\sup_{|x|<\infty, \|\mathbf{u}\|\le B} n^{1/2}|F_n(x, \mathbf{u}) - F_n(x)| = o_p(1).$$

This together with (6), (10) and (7.2.19) readily imply that

$$n^{-1/2}e_{ni\mathbf{u}} = [F_n(\epsilon_{i\mathbf{u}}) - F(\epsilon_i)] + \bar{o}_p(n^{-1/2})$$

$$= [F_n(\epsilon_i) - F(\epsilon_i)] - n^{-1/2}\mathbf{u}'\mathbf{Y}_{i-1} f(\epsilon_i) + \bar{o}_p(n^{-1/2}), \tag{11}$$

where, $\bar{o}_p(n^{-1/2})$ is an array of processes in $(i, \mathbf{u})$ that converge to zero, uniformly in $(i, \mathbf{u})$, in probability, at a rate faster than $n^{-1/2}$.

Now the Claim (9) follows from (6), the Glivenko – Cantelli Lemma and the assumption (**F1**) that ensures $\|f\|_\infty < \infty$.

Next, define

$$\mathcal{T}(\mathbf{u}) := n^{-1} \Sigma_i \mathbf{Y}_{i-1}e_{ni\mathbf{u}}\, \dot{\varphi}(F(\epsilon_i)), \qquad \mathbf{u} \in \mathbb{R}^p.$$

Note that

$$\mathbf{M}(\mathbf{u}) = n^{-1} \Sigma_i \mathbf{Y}_{i-1} [\varphi(F(\epsilon_i) + n^{-1/2}e_{ni\mathbf{u}}) - \bar{\varphi}].$$

Therefore, from the uniform continuity of $\dot{\varphi}$, the facts that $n^{-1}\Sigma_i \|\mathbf{Y}_{i-1}\| = O_p(1) = n^{-1}\|\Sigma_i \mathbf{Y}_{i-1}\mathbf{Y}'_{i-1}\|$, which in turn follow from the assumption $E\epsilon^2 < \infty$ and the Erogodic Theorem, and from (9), one readily concludes that, with $U_i \equiv F(\epsilon_i)$,

$$\|n^{1/2}[M(\mathbf{u}) - \hat{M}] - \mathcal{T}(\mathbf{u})\|$$

$$= \|n^{-1/2}\Sigma_i \mathbf{Y}_{i-1}\{\varphi(U_i + n^{-1/2}e_{ni\mathbf{u}}) - \varphi(U_i) - n^{-1/2}e_{ni\mathbf{u}}\, \dot{\varphi}(U_i)\}\|$$

$$= \bar{o}_p(1). \tag{12}$$

Next, we approximate $\mathcal{T}(\mathbf{u})$. Again, by the Ergodic Theorem, the independence of $\mathbf{Y}_{i-1}$ from ϵ_i, $i \ge 1$, and $E\epsilon = 0$ imply that

$$n^{-1}\Sigma_i \mathbf{Y}_{i-1}\dot{\varphi}(U_i)\, f(\epsilon_i) \longrightarrow 0, \text{ a.s.}$$

Hence by (11),

$$\mathcal{T}(\mathbf{u}) = n^{-1/2}\Sigma_i \mathbf{Y}_{i-1}[\{F_n(\epsilon_i) - F(\epsilon_i)\} - n^{-1/2}\mathbf{u}'\mathbf{Y}_{i-1}f(\epsilon_i)]\dot{\varphi}(U_i) + \bar{o}_p(1) \tag{13}$$

$$= \mathcal{V}_n - \mathbf{u}'\mathbf{L}_n + \bar{o}_p(1),$$

where now $\bar{o}_p(1)$ is a sequence of stochastic processes converging to zero, uniformly in $\mathbf{u}$, in probability, and where

$$\mathcal{V}_n := n^{-1/2} \Sigma_i \mathbf{Y}_{i-1} [F_n(\epsilon_i) - F(\epsilon_i)] \dot{\varphi}(F(\epsilon_i)),$$

$$\mathbf{L}_n := n^{-1} \Sigma_i \mathbf{Y}_{i-1}\mathbf{Y}'_{i-1} f(\epsilon_i) \dot{\varphi}(F(\epsilon_i)).$$

Note that

$$E\mathbf{L}_n = E(n^{-1}\Sigma_i \mathbf{Y}_{i-1}\mathbf{Y}'_{i-1}) Q = \mathbf{\Sigma} Q, \qquad Q = \int f \, d\varphi(F).$$

By the Ergodic Theorem,

$$\mathbf{L}_n \longrightarrow \mathbf{\Sigma} Q, \text{ a.s.}. \tag{14}$$

Our next goal is to approximate $\mathcal{V}_n$. To that effect, let $\mathcal{V}_{nj}$ denote the jth component of $\mathcal{V}_n$. Define

$$\mathcal{U}_{nj}(x) := n^{-1/2} \Sigma_i X_{i-j} \dot{\varphi}(F(\epsilon_i)) I(\epsilon_i \le x)$$

$$\nu_{nj}(x) := n^{-1/2} \Sigma_i X_{i-j} \int_{-\infty}^{x} \dot{\varphi}(F(y)) \, dF(y) = n^{-1/2} \Sigma_i X_{i-j} \varphi(F(x)),$$

$$\mathcal{K}_{nj}(x) := \mathcal{U}_{nj}(x) - \nu_{nj}(x), \qquad x \in \mathbb{R},$$

$$\tilde{\mathcal{Z}}_j := \mathcal{K}_{nj}(F^{-1}(t)), \qquad 0 \le t \le 1, \ 1 \le j \le p.$$

Observe that

$$\begin{aligned}
\mathcal{V}_{nj} &= \int [F_n - F] \, d\,\mathcal{U}_{nj} = \int [F_n - F] \, d\mathcal{K}_{nj} + \int [F_n - F] \, d\nu_{nj} \\
&= -\int_0^1 [\mathcal{K}_{nj}(F_n^{-1}(t)) - \mathcal{K}_{nj}(F^{-1}(t))] \, dt - \int \nu_{nj} \, d[F_n - F] \\
&= -\int_0^1 [\tilde{\mathcal{Z}}_j(F(F_n^{-1}(t))) - \tilde{\mathcal{Z}}_j(t)] \, dt - \int \nu_{nj} \, d[F_n - F].
\end{aligned}$$

But, $\tilde{\mathcal{Z}}_j$ is a $\mathcal{Z}_j$–process of Lemma 7.3b.1 with $h = \dot{\varphi}$. Hence

$$\begin{aligned}
&\max_{1 \le j \le p} | \mathcal{V}_{nj} + \int \nu_{nj} \, d(F_n - F)| \\
&\quad \le \sup_{1 \le j \le p, \, 0 \le t \le 1} | \tilde{\mathcal{Z}}_j(F(F_n^{-1}(t))) - \tilde{\mathcal{Z}}_j(t)| = o_p(1),
\end{aligned} \tag{15}$$

by Lemma 7.3b.1 and the fact that $\sup[|F(F_n^{-1}(t)) - t|; \, 0 \le t \le 1] = o_p(1)$, which in turn follows from Lemma 3.4.1.

Next, observe that, with $\tilde{X}_j = n^{-1}\sum_{i=1}^{n} X_{i-j}$, $1 \le j \le p$,

$$\int \nu_{nj}\, d(F_n - F) = n^{-3/2}\, \Sigma_i X_{i-j}\, \Sigma_i [\varphi(F(\epsilon_i)) - \overline{\varphi}]$$
$$= \tilde{X}_j\, n^{-1/2}\, \Sigma_i [\varphi(F(\epsilon_i)) - \overline{\varphi}].$$

Let $\tilde{\mathbf{X}}' := (\tilde{X}_1, \ldots, \tilde{X}_p)$ and $\hat{T} = n^{-1/2}\, \Sigma_i [\varphi(F(\epsilon_i)) - \overline{\varphi}]$. Then from (13) – (15) we obtain

(16) $$\mathcal{V}_n = -\tilde{\mathbf{X}}\,\hat{T} + o_p(1), \qquad \mathcal{T}(\mathbf{u}) = -\tilde{\mathbf{X}}\,\hat{T} - \boldsymbol{\Sigma}\,\mathbf{u}\,Q + \overline{o}_p(1).$$

From (12), (16) and direct algebra one now readily concludes that

$$\mathbf{M}(\mathbf{u}) = \hat{\mathbf{M}} - \tilde{\mathbf{X}}\,\hat{T} - \boldsymbol{\Sigma}\,\mathbf{u}\,Q + \overline{o}_p(1)$$
$$= n^{-1/2}\sum_{i=1}^{n}(\mathbf{Y}_{i-1} - \tilde{\mathbf{X}})\,[\varphi(F(\epsilon_i)) - \overline{\varphi}] - \boldsymbol{\Sigma}\,\mathbf{u}\,Q + \overline{o}_p(1).$$

Now argue as for (5) to conclude that

$$\|n^{-1/2}\sum_{i=1}^{n}(\mathbf{Y}_{i-1} - \tilde{\mathbf{X}})\,[\varphi(F(\epsilon_i)) - \overline{\varphi}] - \hat{\mathcal{S}}\,\| = o_p(1),$$

thereby completing the proof of (3). □

Remark 7.3b.1. Note that the same proof shows that under the assumed conditions, for every $0 < B < \infty$,

$$\sup_{\|\mathbf{u}\| \le B} \|\mathcal{S}(\rho + n^{-1/2}\mathbf{u}) - \mathcal{S}(\rho) + \mathbf{u}'\boldsymbol{\Sigma}\int f\,d\varphi(F)\,\| = o_p(1). \qquad \square$$

Remark 7.3b.2. Argue either as in Section 3.4 or as in Jaeckal (1972) to conclude that $\|n^{1/2}(\hat{\rho}_R - \rho)\| = O_p(1)$ and that $\|n^{1/2}(\hat{\rho}_R - \hat{\rho}_J)\| = o_p(1)$. Consequently by Theorem 7.3b.1,

(17) $$n^{1/2}(\hat{\rho}_R - \rho) = n^{1/2}(\tilde{\rho}_J - \rho) + o_p(1) = Q^{-1}\,\boldsymbol{\Sigma}^{-1}\,\hat{\mathcal{S}} + o_p(1).$$

Observe that $\hat{\mathcal{S}}$ is a vector of square integrable mean zero martingales with $E\hat{\mathcal{S}}\hat{\mathcal{S}}' = \sigma_\varphi^2\,\Sigma$, $\sigma_\varphi^2 := \mathrm{Var.}(\varphi(U))$. Thus, by the routine Cramer–Wold device and by Lemma A.3 in the Appendix, one readily obtains

(18) $$\hat{\mathcal{S}} \xrightarrow[d]{} N(0, \sigma_\varphi^2\,\boldsymbol{\Sigma}),$$

(19) $n^{1/2}(\hat{\rho}_R - \rho) \xrightarrow[d]{} N(0, \mathcal{V}),\ n^{1/2}(\tilde{\rho}_J - \rho) \xrightarrow[d]{} N(0, \mathcal{V}),\ \mathcal{V} = Q^{-2}\sigma_\varphi^2\, \Sigma^{-1}$. □

Remark 7.3b.3. See the recent paper of Koul and Ossiander (1992) for an extension of the above results to any $\varphi \in \mathscr{C}$ of (3.2.1). □

7.3c. ESTIMATION OF $Q(f) := \int f\, d\varphi(F)$.

As is evident from (7.3b.19), the rank analysis of an AR(p) model via the above R-estimators will need a consistent estimator of the functional Q. In this subsection we give two classes of consistent estimators of this functional in the AR(p) model (7.3b.3), (7.3b.4). One class of estimators is obtained by replacing f and F in Q by a kernel density estimator and the empirical d.f. based on the estimated residuals, respectively. This is analogous to the class of estimators discussed in Theorem 4.5.3. The other class is an analogue of the class of estimators discussed in Theorem 4.5.1 in connection with the linear regression setup.

Accordingly, let $\tilde{\rho}$ be an estimator of ρ, K be a probability density in $\mathbb{R}$, h_n be a sequence of positive numbers, $h_n \longrightarrow 0$ and define, for $x \in \mathbb{R}$,

$$\tilde{\epsilon}_i := X_i - \tilde{\rho}' Y_{i-1},\ 1 \le i \le n; \qquad \tilde{F}_n(x) := F_n(x, \tilde{\rho}) = n^{-1} \Sigma_i\, I(\tilde{\epsilon}_i \le x),$$

$$\tilde{f}_n(x) := (nh_n)^{-1} \Sigma_i\, K\Big(\frac{x - \tilde{\epsilon}_i}{h_n}\Big), \qquad f_n(x) := (nh_n)^{-1} \Sigma_i\, K\Big(\frac{x - \epsilon_i}{h_n}\Big).$$

Finally, let

$$\tilde{Q}_n := \int \tilde{f}_n\, d\varphi(\tilde{F}_n).$$

Theorem 7.3c.1. *In addition to* (7.3.b3), (7.3.b4), *assume that* $E\epsilon_1 = 0$, $E\epsilon_1^2 < \infty$. *Moreover, assume that* **(F1)**, **(F2)** *and the following conditions hold.*

(i) $\varphi \in \mathscr{C} := \{\varphi$: φ *a nondecreasing function on* $[0, 1]$, $\varphi(0) = 0$, $\varphi(1) = 1\}$.

(ii) $h_n > 0$; $h_n \longrightarrow 0$, $n^{1/2}h_n \longrightarrow \infty$.

(iii) K *is absolutely continuous with its a.e derivative* $\dot{K}$ *satisfying* $\int |\dot{K}| < \infty$.

(iv) $\|n^{1/2}(\tilde{\rho} - \rho)\| = O_p(1)$.

Then,

(1) $$\sup_{\varphi \in \mathscr{C}} |\tilde{Q}_n - Q(f)| = o_p(1).$$

Proof. The proof is similar to that of Theorem 4.5.3, so we shall be brief, indicating only one major difference. Unlike in the linear regression setup, i.e., unlike (4.5.11), here we have from Remark 7.2.3,

$$\text{(2)} \qquad \sup_x n^{1/2}|\tilde{F}_n(x) - F_n(x)| = o_p(1), \quad \text{where } F_n(x) \equiv F_n(x, \rho).$$

In other words the linearity term involving $n^{1/2}(\tilde{\rho} - \rho)$ is not present in the approximation of $\tilde{F}_n$. Proceeding as in the proof of Theorem 4.5.3, (2) will yield

$$\begin{aligned} \|\tilde{f}_n - f_n\|_\infty &\le (n^{1/2}h_n)^{-1} \cdot \|n^{1/2}[\tilde{F}_n - F_n]\|_\infty \cdot \int |\dot{K}| \\ &= o_p((n^{1/2}h_n)^{-1}) = o_p(1). \end{aligned}$$

Compare this with (4.5.19) where $O_p((n^{1/2}h_n)^{-1})$ appears instead of $o_p((n^{1/2}h_n)^{-1})$. Rest of the proof is exactly the same as there with the proviso that one uses (2) instead of (4.5.11), whenever needed. □

The reader may wish to modify the above proof to see that $\tilde{Q}_n$ continues to be consistent for Q even when $E\epsilon \ne 0$, so that the term that is linear in $n^{1/2}(\tilde{\rho} - \rho)$ is now present in the expansion of $\tilde{F}_n$.

We shall now describe an analogue of $\hat{Q}_n^\alpha$ of (4.5.6). The motivation is the same as in Section 4.5, so we shall be brief on that also. Accordingly, let

$$\tilde{p}(y) := \int [\tilde{F}_n(y+x) - \tilde{F}_n(-y+x)]\, d\, \varphi(\tilde{F}_n(x)), \qquad y \ge 0.$$

Observe that $\tilde{p}$ is an estimator of the d.f. of the absolute difference $|\epsilon - \eta|$, where ϵ and η are independent r.v.'s with respective d.f.'s F and $\varphi(F)$. As in Section 4.5, one can use the following representation for the computational purposes.

$$\tilde{p}(y) = n^{-1} \sum_{j=1}^{n} [\varphi(j/n) - \varphi((j-1)/n)] \sum_{i=1}^{n} I(|\tilde{\epsilon}_{(i)} - \tilde{\epsilon}_{(j)}| \le y), \qquad y \ge 0,$$

where $\{\tilde{\epsilon}_{(i)}\}$ are the ordered residuals $\{\tilde{\epsilon}_i\}$ from the smallest to the largest.

Now let $\tilde{t}_n^\alpha$ denote an αth percentile of the d.f. $\tilde{p}(y)$ and define

$$\tilde{Q}_n^\alpha = n^{1/2}\, \tilde{p}(n^{-1/2}\, \tilde{t}_n^\alpha) \,/\, 2\tilde{t}_n^\alpha, \qquad 0 < \alpha < 1.$$

The consistency of these estimators may be proved using the method of the proof of Theorem 4.5.1 and the results given in Corollary 7.2.1. The discussion about the choice of α etc. that appears in Remark 4.5.1 is also pertinent here.

Another class of estimators is obtained by modifying $\tilde{Q}_n$ by replacing $\hat{F}_n$ by the estimator $\overline{F}_n(x) = \int \tilde{f}_n(y)\, I(-\infty<y\le x)\, dy$. The consistency of these estimators can be also proved by the help of Corollary 7.2.1. □

7.3d. PROOF OF LEMMA 7.3b.3.

The proof of Lemma 7.3b.3 is similar to that of Theorem 2.2a.1(i) and will be a consequence of the following *two* lemmas.

Lemma 7.3d.1. *In addition to* (7.3b.3), (7.3b.4) *and* (7.3b.c1) *assume that the following hold:*

(d1) *The d.f.* F *is continuous and strictly increasing.*

(d2) *The function h on* $[0, 1]$ *to* $\mathbb{R}$ *is nonnegative and* $\int |h(t)|^4 dt < \infty$.

Then the following hold:

(A) *For any* $0 \le u < v < w \le 1$ *and for all* $1 \le j \le p$,

(1) $\limsup_n E\{\mathcal{Z}_j(v) - \mathcal{Z}_j(u)\}^2\{(\mathcal{Z}_j(w) - \mathcal{Z}_j(v)\}^2 \le C\, m_1 m_2,$

where $m_1 := \int_u^v h^2(t)\, dt$, $m_2 := \int_v^w h^2(t)\, dt$, C *is a constant given in* (19) *below.*

(B) *For any* $0 \le u < v \le 1$, *and for* $1 \le j \le p$,

$$\limsup_n E\{\mathcal{Z}_j(v) - \mathcal{Z}_j(u)\}^4 \le C\, m_1^2.$$

Proof. (A). Since u, v, w, are fixed, we shall suppress these entities in the notation. Let $\mathcal{F}_k := \sigma$–field$\{\epsilon_i;\ i \le k\}$, $k = 0, \pm1, \pm2, \ldots$. Further, to simplify writing let $x = F^{-1}(u)$, $y = F^{-1}(v)$ and $z = F^{-1}(w)$ and define

(2) $p_1 = H(v) - H(u),\quad p_2 = H(w) - H(v);\quad q_j = 1 - p_j,\ j = 1, 2,$

$$\alpha_i := h(F(\epsilon_i))I(x < \epsilon_i \le y) - p_1,\quad \beta_i := h(F(\epsilon_i))I(y < \epsilon_i \le z) - p_2.$$

Then

(3) $\{\mathcal{Z}_j(v) - \mathcal{Z}_j(u)\}^2\{\mathcal{Z}_j(w) - \mathcal{Z}_j(v)\}^2 = n^{-2}(\Sigma_i X_{i-j}\alpha_i)^2\cdot(\Sigma_r X_{r-j}\beta_r)^2.$

In carrying out the computations that follow we have repeatedly used the following facts: α_i, β_i are centered; $\alpha_i\beta_i$ are $\mathcal{F}_{k-1}$ measurable for all $i < k$ and X_{i-j} is $\mathcal{F}_{i-1}$ measurable and independent of ϵ_i, $i \ge 1$. Thus,

(4) $\quad E\,\alpha_i = 0 = E\,\beta_k,$ for all i,k.

$$EX_{i-j}X_{k-j}\alpha_i\beta_k = E[X_{i-j}X_{k-j}\alpha_i E(\beta_k|\mathcal{F}_{k-1})] = E[X_{i-j}X_{k-j}\alpha_i]\,E(\beta_k) = 0, \quad i<k;$$

$$EX_{i-j}X_{k-j}X^2_{r-j}\alpha_i\,\alpha_k\,\beta_r^2 = E[X_{i-j}X_{k-j}X^2_{r-j}\beta_r^2\,\alpha_i E(\alpha_k|\mathcal{F}_{k-1})] = 0, \qquad i, r < k.$$

Using facts like these one can write

$$\begin{aligned}(5)\quad & E(\Sigma_i X_{i-j}\,\alpha_i)^2(\Sigma_r X_{r-j}\beta_r)^2 \\ &= \sum_i E\,X^4_{i-j}\,\alpha_i^2\,\beta_i^2 + \sum_{i\neq r}\sum EX^2_{i-j}\,X^2_{r-j}\alpha_i^2\,\beta_r^2 \\ &\quad + 4\sum_{i<k}\sum E\,X^2_{i-j}X^2_{k-j}\,\alpha_i\,\beta_i\,\alpha_k\,\beta_k \\ &\quad + 2\sum_{i<k<r}\sum\sum E\,X_{i-j}\,X_{k-j}\,X^2_{r-j}(\alpha_i\,\alpha_k\,\beta_r^2 + \beta_i\,\beta_k\,\alpha_r^2) \\ &\quad + 4\sum_{i<r<k}\sum\sum E\,X_{i-j}\,X_{r-j}\,X^2_{k-j}\,\alpha_i\,\beta_r\,\alpha_k\,\beta_k \\ &\quad + 4\sum_{r<i<k}\sum\sum E\,X_{i-j}\,X_{r-j}\,X^2_{k-j}\,\alpha_i\,\beta_r\,\alpha_k\,\beta_k \\ &= T_1 + T_2 + 4\,T_3 + 2(T_4 + T_5) + 4(T_6 + T_7), \text{ say.}\end{aligned}$$

We shall now show that $n^{-2}T_j \longrightarrow 0$, for $j = 1, 4, 5, 6, 7$, and that $\limsup n^{-2}(T_2 + 4\,T_3) \le C\,m_1m_2$. The basic idea of the proof is to exploit the hierarchal nature of the process. Observe that had the underlying observations been independent then T_j would have been equal to zero for $j = 4, 5, 6, 7$. However, under (7.3b.3), $\{X_j\}$ are not independent but asymptotically behave like independent r.v.'s. This is the reason to expect $n^{-2}T_j \longrightarrow 0$ for $j = 4, 5, 6, 7$.

The details of the proof of $n^{-2}T_j$ tending to zero for each $j = 4, 5, 6, 7$ are elementary and cumbersome but similar. So the details will be given only for $n^{-2}T_7 \longrightarrow 0$. To this effect, observe that

$$E(X_{i-j}X_{r-j}X^2_{k-j}\alpha_i\beta_r\alpha_k\beta_k) = E\,X_{i-j}X_{r-j}X^2_{k-j}\alpha_i\beta_r E(\alpha_k\,\beta_k|\mathcal{F}_{k-1}), \quad i, r < k.$$

Moreover, $\{\epsilon_i\}$ i. i. d. implies that for all $k \ge 1$,

$$E(\alpha_k\beta_k|\mathcal{F}_{k-1}) = (1-p_1)p_1(-p_2)+(-p_1)(1-p_2)p_2+p_1p_2(1-p_1-p_2) = -p_1p_2,$$

and, in addition, $E\epsilon = 0$ implies that $EX_{i-j}X_{r-j}X^2_{k-j}\alpha_i \equiv 0$, $r < i$, $k-j \le i-1$. Therefore,

$$(6)\quad n^{-2}T_7 = -n^{-2}p_1p_2\{\sum_{r<i}\sum EX_{i-j}\,X_{r-j}\,X_i^2\,\alpha_i\,\beta_r + \sum_{r<i}\sum\ \sum_{k-j>i} EX_{i-j}\,X_{r-j}\,X_{k-j}^2\,\alpha_i\,\beta_r\}$$

$$= -\,p_1p_2\,n^{-2}\{T_{71}+T_{72}\}, \qquad \text{say.}$$

Now, for a convenient reference we rewrite (7.3b.5) as

$$(7)\qquad X_i = \sum_{k\le i}\theta_{i-k}\,\epsilon_k, \qquad i\ge 1,\ \text{a.s.},$$

where, as in (7.3b.5), the unspecified lower limit on the index of summation is $-\infty$, and $\{\theta_k\}$ are real numbers satisfying $\theta_0 = 1$, $\Delta_1 < \infty$, with

$$\Delta_q := \Sigma_{k\ge 0}\,|\theta_k|^q, \quad q\ge 1.$$

Note that $\sup_k|\theta_k| \le \Delta_1$ and hence $\Delta_1 < \infty$ implies that

$$(8)\qquad \Delta_r \le \Delta_1^q < \infty, \qquad \text{for all } q\ge 1.$$

Next, define

$$(9)\qquad A_{m,n} := \sum_{r\le n}\theta_{m-r}\,\epsilon_r,$$

$$H_{m,k}^n := \sum_{r=k}^{n}\theta_{m-r}\,\epsilon_r, \qquad 0\le n\le m<\infty,\ k\le n.$$

$$a_r := E(\alpha\,\epsilon^r), \quad b_r := E(\beta\,\epsilon^r), \quad \mu_r := E\epsilon^r, \qquad 1\le r\le 4,$$

$$\sigma_1^2 := \text{Var.}(\alpha), \quad \sigma_2^2 := \text{Var.}(\beta)$$

where α, β are copies of α_1, β_1. Observe that

$$(10)\qquad H_{m,k}^n = A_{m,n} - A_{m,k-1}, \qquad k\le n\le m,$$

$$X_i = A_{i,i} = A_{i,i-j} + H_{i,i-j+1}^{i-1} + \epsilon_i = A_{i,i-1} + \epsilon_i, \qquad \forall\ i,$$

$$\sigma_k^2 \le m_k, \quad \text{where } m_k \text{ is as in (1)}, \qquad k = 1, 2.$$

Morover, $\{\epsilon_i\}$ i.i.d., $E(\epsilon) = 0$, and (8) imply that for all $n\le m<\infty$,

$$(11)\qquad EA_{m,n}^2 = \sum_{r\ge m-n}\theta_r^2\,\mu_2 \le \mu_2\,\Delta_1^2 < \infty,$$

$$EA_{m,n}^4 = \sum_{k\ge m-n}\theta_k^4\mu_4 + 3\sum_{k\ge m-n}\ \sum_{r\ge m-n,\,r\ne k}\theta_k^2\theta_r^2\mu_2 \le (\mu_4+\mu_2)\Delta_1^4 < \infty.$$

For the same reasons, from (10) it follows that

$$E\{X_i^2\ \alpha_i|\mathcal{F}_{i-1}\} = 2\ A_{i,i-1}\, a_1 + a_2, \qquad \text{for all } i.$$

Use this and argue as for (6) to obtain

$$\begin{aligned} T_{71} = &\ \Sigma_r\, E\, X_{r-j}\beta_r\, \{2a_1\, L_r + a_2\, X_r\} \\ &+ \sum_{r<i;\, i-j\geq r+1}\sum E\, X_{r-j}\beta_r\, \{2a_1 L_i + a_2 X_{i-j}\} \\ = &\ T_{711} + T_{712}, \ \text{ say, where } \quad L_i := X_{i-j}\, A_{i,i-1}. \end{aligned}$$

The C–S inequality, the staionarity of the process $\{X_i\}$ and (11) imply that for all r, j,

$$E|X_{r-j}\beta_r\, L_{r+j}| \leq \{E(X_{r-j}\beta_r)^2\, EL_{r+j}^2\}^{1/2} \leq D_4\, C_4 < \infty,$$

$$E|X_{r-j}\beta_r\, L_r| \leq D_4\, C_4 < \infty,$$

where D_k, C_k are constants depending on the kth moment of $h(U)$ and the kth moment of ϵ and Δ_k, respectively, $1 \leq k \leq 4$. These facts imply that

$$n^{-2}|T_{711}| = O(n^{-1}) = o(1). \tag{12}$$

Next, to handle T_{712}, use (11) to obtain that for $i-j \geq r+1$,

$$L_i = \{A_{i-j,r-1} + \theta_{i-j-r}\, \epsilon_r + H_{i-j,r+1}^{i-j}\}\{A_{i,r-1} + \theta_{i-r}\, \epsilon_r + H_{i,r+1}^{i-1}\}.$$

Use the above type of conditioning argument to obtain that

$$EX_{r-j}\ \beta_r\, L_i = EX_{r-j}\ \{[\theta_{i-r}\, A_{i-j,r-1} + \theta_{i-j-r}\, A_{i,r-1}]\, b_1 + \theta_{i-j-r}\, \theta_{i-r}\, b_2\},$$

$$EX_{r-j}\ \beta_r\, X_{i-j} = EX_{r-j}\{\theta_{i-j-r}^2\, b_2 + 2\theta_{i-j-r}\, A_{i-j,r-1}\, b_1\}, \qquad i-j \geq r+1.$$

Use these facts together with (11) and an argument like the one that led to (12) to conclude that $n^{-2}|T_{712}| = O(n^{-2})$. This and (12) yield

$$n^{-2}|T_{71}| = O(n^{-1}) = o(1). \tag{13}$$

Now we turn to T_{72}. Using (10) write

$$X_{k-j} = A_{k-j,i-1} + \theta_{k-j-i}\, \epsilon_i + H_{k-j,i+1}^{k-j}, \qquad k-j \geq i+1,$$

and use arguments like those above to obtain that

$$E\{X_{k-j}^2\ \alpha_i|\mathcal{F}_{i-1}\} = \theta_{k-j-i}^2\, a_2 + 2A_{k-j,i-1}\, \theta_{k-j-i}\, a_1, \qquad k-j \geq i+1,$$

so that

(14) $$T_{72} = \sum_{r<i<k;k-j\geq i+1}\sum\sum EX_{r-j}\,\beta_r\,X_{i-j}\,\{\theta^2_{k-j-i}\,a_2 + 2A_{k-j,i-1}\,\theta_{k-j-i}\,a_1\}$$
$$= a_2\,T_{721} + 2a_1\,T_{722}, \qquad \text{say.}$$

Arguing as above and using the stationarity and the fact that $EX_0 = 0$, one obtains

$$EX_{r-j}\,\beta_r\,X_{i-j} = EX_{r-j}\,\beta_r\,\{A_{i-j,r-1} + \theta_{i-j-r}\,\epsilon_r + H^{i-j}_{i-j,r+1}\}$$
$$= EX_{r-j}\,\theta_{i-j-r}\,b_1 = 0, \qquad i-j \geq r+1.$$

Thus,

$$|T_{721}| \leq a_2\,d_2 \sum_{r<i}\sum |EX_{r-j}\,\beta_r\,X_{i-j}| = 0.$$

Similar arguments show that $|n^{-2}T_{722}| = o(1)$ thereby completing the proof of $n^{-2}|T_{72}| = o(1)$. This together with (13) shows that

$$n^{-2}|T_7| = o(1).$$

Now consider T_2: Rewrite

$$T_2 = (\sum_{i<r}\sum + \sum_{r<i}\sum)\,(EX^2_{i-j}\,X^2_{r-j}\,\alpha^2_i\,\beta^2_r) = T_{21} + T_{22}, \quad \text{say.}$$

Again, by a conditionning argument,

(15) $$T_{21} = \sigma^2_2 \cdot \sum_{i<r}\sum EX^2_{i-j}\,X^2_{r-j}\,\alpha^2_i$$
$$= \sigma^2_2 \cdot (\sum_{i<r;r-i\leq j}\sum + \sum_{i<r;r-i\geq j+1}\sum)(EX^2_{i-j}\,X^2_{r-j}\,\alpha^2_i)$$
$$= \sigma^2_2 \cdot \{T_{211} + T_{212}\}, \qquad \text{say.}$$

Again, the C–S inequality, the stationarity of the process $\{X_i\}$, the assumptions (7.3b.c1) and (d2) imply that $0 < T_{211} \leq j\cdot n\cdot EX^4_0 = O(n)$, by (8) and (10), so that

(16) $$n^{-2}T_{211} = O(n^{-1}) = o(1).$$

Next, argue as for (14) to obtain

$$T_{212} = \sum_{i<r;r-i\geq j+1}\sum EX^2_{i-j}\,X^2_{r-j}\,\alpha^2_i$$
$$= \sum_{i<r;r-i\geq j+1}\sum EX^2_{i-j}\,\alpha^2_i\,\{A_{r-j,i-1} + \theta_{r-j-i}\,\epsilon_i + H^{r-j}_{r-j,i+1}\}^2$$

$$= \sigma_1^2 \cdot \sum_{i<r;\, r-i\geq j+1}\sum EX_{i-j}^2 \{A_{r-j,i-1}^2 + \sum_{m=i+1}^{r-j} \theta_{r-j-m}^2 \, E\epsilon^2\}$$

$$+ 2c \sum_{i<r;\, r-i\geq j+1}\sum EX_{i-j}^2 \, A_{r-j,i-1} \, \theta_{r-j-i}$$

(17) $$= \sigma_1^2 \cdot B_1 + 2c\cdot B_2 + O(n), \quad \text{say,} \qquad c = E(\epsilon\alpha)^2.$$

The C–S inequality and (11) yield that

(18) $$n^{-2}B_1 \leq C < \infty, \text{ for all } n \geq 1,$$

where C is a constant depending on μ_4 and Δ_1. A similar argument shows that $|B_2| = O(n^{-1})$. This together with (17) and (18) yield that

(19) $$\limsup_n n^{-2}|T_{212}| \leq C\, \sigma_1^2.$$

Hence, from (17) – (19) one readily obtains

$$\limsup_n n^{-2}|T_{21}| \leq C\, \sigma_1^2\, \sigma_2^2 \leq C\, m_1 m_2, \qquad \text{by (10).}$$

Similarly, one concludes a similar result for T_{22} thereby enabling one to conclude

(20) $$\limsup_n n^{-2}|T_2| \leq C\, m_1 m_2, \text{ where } C \text{ is as in (18).}$$

Finally, consider $n^{-2}T_3$: By arguments similar to those above we obtain

(21) $$n^{-2}T_3 = -n^{-2} \sum_{i<r}\sum EX_{i-j}^2\, X_{r-j}^2\, \alpha_i\, \beta_i\, p_1 p_2$$

$$= -p_1 p_2\, n^{-2} \sum_{i<r;\, r-i\geq j+1}\sum EX_{i-j}^2\, \alpha_i\, \beta_i\, X_{r-j}^2 + O(n^{-1}).$$

Let $c_r = E(\alpha\beta\, \epsilon^r)$. Use (10) and procced as before to obtain

$$EX_{i-j}^2\, \alpha_i\, \beta_i\, X_{r-j}^2 = EX_{i-j}^2\{-p_1 p_2\, (A_{r-j,i-1}^2 + \mu_2 \sum_{m=i+1}^{r-j} \theta_{r-j-m}^2\,)$$

$$+ \theta_{r-j-i}^2\, c_2 + 2c_1\, A_{r-j,i-1}\, \theta_{r-j-i}\}.$$

Combine this with (21), argue as above using (11) and the C–S inequality, to obtain

$$n^{-2}T_3 = (p_1 p_2)^2 n^{-2} \sum_{i<r;\, r-i\geq j+1}\sum E\, X_{i-j}^2[A_{r-j,i-1}^2 + \mu_2 \sum_{m=i+1}^{r-j} \theta_{r-j-m}^2] + O(n^{-1}).$$

Another application of (11) yields

$$\limsup_n |n^{-2}T_3| \le C\,(p_1p_2)^2 \le C\,m_1\,m_2\,,$$

where C is as in (18) above, because

$$p_1^2 = \left\{\int_u^v h(t)\,dt\right\}^2 \le (v-u)\int_u^v h^2(t)\,dt \le m_1$$

$$p_2^2 = \left\{\int_v^w h(t)\,dt\right\}^2 \le (w-v)\int_v^w h^2(t)\,dt \le m_2.$$

The proof of (A) is now terminated.

PROOF of (B). Fix j and define, for $r \ge 1$, $k \ge 1$,

$$u_{rk} = E\{(X_{k\text{-}j}\,\alpha_k)^r \,|\, \mathcal{F}_{k\text{-}1}\} = X_{k\text{-}j}^r\, E(\alpha_k^r \,|\, \mathcal{F}_{k\text{-}1});$$

$$U_{rk} = \sum_{i=1}^{k} u_{ri}; \qquad\qquad S_k = \sum_{i=1}^{k} X_{i\text{-}j}\,\alpha_i.$$

Now observe that

$$\mathscr{Z}_j(v) - \mathscr{Z}_j(u) = n^{-2}S_n. \tag{22}$$

Because $\{X_{i\text{-}j}\,\alpha_i\}$ are conditionally centered, gives $\mathcal{F}_{i\text{-}1}$, it readily follows that $\{S_n, \mathcal{F}_n\}$ is a mean zero martingale. Therefore, from Chow, Robbin and Teicher (1964),

$$E\,S_n^4 = E\{U_{4n} + 4\,S_n\,U_{3n} + 6\,S_n^2\,U_{2n} - 6\sum_{j=1}^{n} u_{2j}\,U_{2j}\}. \tag{23}$$

But,

$$E\sum_{k=1}^{n} u_{2k}\,U_{2k} = E\,\{\sum_{k=1}^{n} E(X_{k\text{-}j}^2\,\alpha_k^2 \,|\, \mathcal{F}_{k\text{-}1})\cdot \sum_{i=1}^{k} E(X_{i\text{-}j}^2\,\alpha_i^2 \,|\, \mathcal{F}_{i\text{-}1})\}$$
$$= E\sum_{k=1}^{n} X_{k\text{-}j}^2\,\sigma_1^2 \cdot \sum_{i=1}^{k} X_{i\text{-}j}^2\,\sigma_1^2 = \sum_{i\le k}\sum E\,X_{i\text{-}j}^2\,X_{k\text{-}j}^2 \cdot \sigma_1^4,$$

$$E\,S_nU_{3n} = E\{\sum_{i=1}^{n} X_{i\text{-}j}\,\alpha_i \cdot \sum_{k=1}^{n} E((X_{k\text{-}j}\,\alpha_k)^3 \,|\, \mathcal{F}_{k\text{-}1})\}$$
$$= \sum_{i\le k\text{-}j}\sum E X_{i\text{-}j}\,X_{k\text{-}j}^3\,\alpha_i \cdot E(\alpha^3),$$

$$E\,S_n^2\,U_{2n} = \sigma_1^2 \cdot E\{(\sum_i X_{i\text{-}j}^2\,\alpha_i^2 + 2\sum_{i<r}\sum X_{i\text{-}j}X_{r\text{-}j}\alpha_i\alpha_r)(\sum_k X_{k\text{-}j}^2)\}$$
$$= \sigma_1^2 \cdot [\sum_{i\le k\text{-}j}\sum EX_{i\text{-}j}^2\,\alpha_i^2\,X_{k\text{-}j}^2 + \sigma_1^2 \cdot \sum_{k\text{-}j\le i\text{-}1}\sum E\,X_{i\text{-}j}^2\,X_{k\text{-}j}^2$$
$$+ 2\sum_{i<r<k}\sum\sum EX_{i\text{-}j}\,X_{r\text{-}j}\,X_{k\text{-}j}^2\,\alpha_i\,\alpha_r],$$

$$\mathrm{E}\, U_{4n} = \sum_i \mathrm{E}\, X_{i-j}^4 \cdot \mathrm{E}(\alpha^4).$$

Combine the above with (23) to obtain

$$\begin{aligned} n^{-2}\mathrm{E}\, S_n^4 = n^{-2}\{&\sum_{i=1}^n \mathrm{E}X_{i-j}^4 \cdot \mathrm{E}(\alpha^4) + 4\sum_{i\le k}\sum_{-j} \mathrm{E}X_{i-j}X_{k-j}^3\, \alpha_i \cdot \mathrm{E}(\alpha^3) \\ &+ 6\sigma_1^2 \cdot [\sum_{i\le k}\sum_{-j} \mathrm{E}X_{i-j}^2\alpha_i^2 X_{k-j}^2 + \sigma_1^2 \cdot \sum_{i<k\le j}\sum_{+i-1} \mathrm{E}X_{i-j}^2 X_{k-j}^2 \\ &+ 2\sum_{i<r<k}\sum\sum \mathrm{E}X_{i-j}\, X_{r-j}\, X_{k-j}^2\, \alpha_i\, \alpha_r]\}. \\ = n^{-2}\{&K_1 + 4\, K_2 + 6\, \sigma_1^2\, [K_3 + \sigma_1^2\, K_4 + 2\, K_5]\}, \qquad \text{say.} \end{aligned}$$

Arguing as for the proof of (A), one can show that $n^{-2}K_j \longrightarrow 0$, $j = 1, 2, 5$, that $\limsup n^{-2}|K_3| \le C\, \sigma_1^2$, and that $\limsup n^{-2}K_4 \le C$. Hence (B). □

Lemma 7.3d.2. *In addition to* (7.3b.3) *and* (7.3b.4) *assume that* $\mathrm{E}\epsilon = 0$, $\mathrm{E}\epsilon^2 < \infty$, *and* $\mathrm{E}h^2(U) < \infty$. *Then the finite dimensional distribution of* $\mathcal{Z}_j$, *for every* $1 \le j \le p$, *converges weakly to that of* $\{\mathrm{E}(X_0)^2\}^{1/2}B(\cdot)$, *where B is the Brownian motion in* $\mathbb{C}[0, 1]$ *with the covariance function* $H(u) - H(u)H(v)$, $0 \le u \le v \le 1$.

Proof. The proof uses Corollory 3.1 of Hall and Heyde (1980; p 58)(see Lemma A.3 in the Appendix) and the Cramer–Wold device. Accordingly, fix j and let $0 \le u_1 \le u_2 \le \dots \le u_r \le 1$, $\theta \in \mathbb{R}^r$. Define

$$\alpha_n(\epsilon_i) := \sum_{k=1}^r \lambda_k\, \{h(F(\epsilon_i))I(F(\epsilon_i) \le u_k) - G(u_k)\},$$

$$\xi_{ni} := n^{-1/2}\, X_{i-j}\, \alpha_n(\epsilon_i), \qquad \mathcal{S}_{ni} := \sum_{k=1}^i \xi_{nk}, \qquad 1 \le i \le n.$$

Note that $\mathcal{S}_{nn} = \sum_{k=1}^r \theta_k\, \mathcal{Z}_j(u_k)$. Because of the given assumptions, and because ξ_{ni} is conditionally centered, given $\mathcal{F}_{i-1}$, $\{(\mathcal{S}_{ni}, \mathcal{F}_{i-1}), 1 \le i \le n\}$ is a mean zero square integrable martingale array. Next, for a $\theta > 0$, by the C–S inequality,

$$\begin{aligned} &\sum_{i=1}^n \mathrm{E}[\xi_{ni}^2\, I(|\xi_{ni}| > \theta)|\mathcal{F}_{i-1}] \\ &\quad = n^{-1}\sum_{i=1}^n \mathrm{E}X_{i-j}^2\, \{\mathrm{E}[\alpha_n^2(\epsilon_i)I(|X_{i-j}\, \alpha_n(\epsilon_i)| > \theta n^{1/2})|\mathcal{F}_{i-1}]\} \\ &\quad \le n^{-1}\sum_{i=1}^n \mathrm{E}X_{i-j}^2\, P^{1/2}(|X_{i-j}\, \alpha_n(\epsilon_i)| > \theta n^{1/2}|\mathcal{F}_{i-1}) \cdot D_{4,r} \end{aligned}$$

$$\leq (\theta n^{3/2})^{-1} \sum_{i=1}^{n} E|X_{i-j}|^3 E^{1/2}[|\alpha_n(\epsilon_i)|\mathcal{F}_{i-1}] \cdot D_{4,r}$$

$$\leq C_3 D_{4,r}(\theta n^{1/2})^{-1} = o(1)$$

where, in the above, $D_{4,r}$ is a constant dependin on r, D_4 and θ and C_3 is a constant dependin on μ_3 and d_1.

Next, from the definition of H in terms of h one readily sees that

$$\begin{aligned}\sum_{i=1}^{n} E(\xi_{ni}^2|\mathcal{F}_{i-1}) &= n^{-1} \sum_{i=1}^{n} X_{i-j}^2\, E[\alpha_n^2(\epsilon_i)|\mathcal{F}_{i-1}] \\ &= n^{-1} \sum_{i=1}^{n} X_{i-j}^2 \sum_{k=1}^{r} \sum_{m=1}^{r} \lambda_k \lambda_m\, [G(u_k \wedge u_m) - G(u_k)\, G(u_m)] \\ &= E(X_0^2) \sum_{k=1}^{r} \sum_{m=1}^{r} \lambda_k \lambda_m [G(u_k \wedge u_m) - G(u_k) G(u_m)] + o_p(1),\end{aligned}$$

by the Ergodic Theorem.

The above calculations show that $\{\mathcal{S}_{ni}, \mathcal{F}_{i-1}, 1 \leq i \leq n\}$ satisfy the conditions of Lemma A.3 and hence $\mathcal{S}_{nn}$ converges weakly to an appropriate normal r.v. This completes the proof of the Lemma. □

Proof of Lemma 7.3b.3. In view of the Lemmas 7.3d.1(A) and 7.3d.2 above, the proof uses Lemmas A.1, A.2 and Theorem A.1 in the Appendix and is exactly like that of Theorem 2.2a.1(i). □

7.4. M.D. ESTIMATION

In this section we shall discuss two classes of m.d. estimators. They are the analogues of the classes of estimators defined in the linear regression setup at (5.2.11) and (5.2.20). To be precise, consider the autoregression model (7.1.1) and define, for a $G \in \mathbb{D}\mathcal{I}(\mathbb{R})$,

$$\text{(1)} \quad K_g(\mathbf{t}) = \sum_{j=1}^{p} \int [n^{-1/2} \sum_{i=1}^{n} g(X_{i-j})\{I(X_i \leq x+\mathbf{t}'Y_{i-1}) - F(y)\}]^2\, dG(x),$$

$$K_g^+(\mathbf{t}) = \sum_{j=1}^{p} \int [n^{-1/2} \sum_{i=1}^{n} g(X_{i-j})\{I(X_i \leq x+\mathbf{t}'Y_{i-1}) - I(-X_i < x-\mathbf{t}'Y_{i-1})\}]^2\, dG(x), \quad \mathbf{t} \in \mathbb{R}^p,$$

In the case the error d.f. F is *known*, define a class of m.d. estimators of ρ to be

$$\text{(3)} \qquad \hat{\rho}_g := \operatorname{argmin}\{K_g(\mathbf{t});\ \mathbf{t} \in \mathbb{R}^p\}.$$

In the case the error distribution is *unknown but symmetric around* 0, define a class of m.d. estimators of ρ to be

(4) $$\rho_g^+ := \operatorname{argmin}\{K_g^+(\mathbf{t});\ \mathbf{t} \in \mathbb{R}^p\}.$$

Note that the role played by the vectors $\{n^{-1/2}[g(X_{i-1}), g(X_{i-2}), \ldots, g(X_{i-p})];\ 1 \le i \le n\}$ is similar to that of the vectors $\{\mathbf{d}_{ni};\ 1 \le i \le n\}$ of Chapter 5. To put it in matrices, the precise analogue of $\mathbf{D}$ is the matrix $n^{-1/2}\mathcal{G}$, where $\mathcal{G}$ is as in (7.3a.1).

The existence of these estimators has been discussed in Dhar (1991a) for $p = 1$ and in Dhar (1991c) for $p \ge 1$. For $p = 1$, these results are relatively easy to state and prove. We give an existence result for the estimator defined at (4) in the case $p = 1$.

Lemma 7.4.1. *In addition to* (7.1.1) *with* $p = 1$, *assume that either*

(5a) $xg(x) \ge 0,\ \forall\ x \in \mathbb{R}$, *or* (5b) $xg(x) \le 0,\ \forall\ x \in \mathbb{R}$,

Then, a minimizer of K_g^+ *exists if either* $G(\mathbb{R}) = \infty$ *or* $G(\mathbb{R}) < \infty$ *and* $g(0) = 0$.

The proof of this lemma is precisely similar to that of Lemma 5.3.1.

The discussion about the computation of their analogues that appears in Section 5.3 is also relevant here with appropriate modifications. Thus, for example, if G is continuous and symmetric around 0, i.e., satisfies (5.3.10), then, analogous to (5.3.7*),

$$K_g^+(\mathbf{t}) = \sum_{j=1}^{p}\sum_{i=1}^{n}\sum_{k=1}^{n} g(X_{i-j})g(X_{k-j})\{|G(X_i-\mathbf{t}'\mathbf{Y}_{i-1}) - G(-X_k+\mathbf{t}'\mathbf{Y}_{k-1})| - |G(X_i-\mathbf{t}'\mathbf{Y}_{i-1}) - G(X_k-\mathbf{t}'\mathbf{Y}_{k-1})|\}.$$

If G is degenerate at 0 then one obtains, assuming the continuity of the errors, that

(6) $$K_g^+(\mathbf{t}) = \sum_{j=1}^{p}\Big[\sum_{i=1}^{n} g(X_{i-j})\operatorname{sign}(X_i - \mathbf{t}'\mathbf{Y}_{i-1})\Big]^2, \text{ w.p.1.}$$

One has similar expressions for a general G. See (5.3.7) and (5.3.7′).

If $g(x) \equiv x \equiv G(x)$, $\hat{\rho}_g$ is m.l.e of ρ if F is logistic, while ρ_g^+ is an analogue of the Hodges–Lehmann estimator. Similarly, if $g(x) \equiv x$ and G is degenerate at 0 then ρ_g^+ is the l.a.d. estimator.

We shall now focus on proving their asymptotic normality. The approach is the same as that of Sections 5.4 and 5.5, i.e., we shall prove that these dispersions satisfy (5.4.A1) – (5.4.A5) by using the techniques that are similar to those used in Section 5.5. Only the tools are somewhat different because of the dependence structure.

To begin with we state the additional assumptions needed under which an asymptotic uniform quadraticity result for a general dispersion of the above type holds. Because here the weights are random, we have to be somewhat careful if we do not wish to impose more than necessary moment conditions on the underlying entities. For the same reason, unlike the linear regression setup where the asymptotic uniform quadraticity of the underlying dispersions was obtained in L_1, we shall obtain these results in probability only. This is also reflected in the formulation of the following assumptions.

(7) (a) $Eh^2(Y_0) < \infty$. (b) $0 < E\epsilon^2 < \infty$.

(8) $\forall\ \|u\| \le B$, $a \in \mathbb{R}$,

$$\int Eh^2(Y_0)\,|F(x+n^{-1/2}(u'Y_0+a\|Y_0\|)) - F(x)|\,dG(x) = o(1).$$

(9) There exists a constant $0 < k < \infty$, $\ni\ \forall\ \delta > 0$, $\forall\ \|u\| \le B$,

$$\liminf_n P\Big(\int n^{-1}\Big[\sum_{i=1}^n h^{\pm}(Y_{i-1})\{F(x+n^{-1/2}u'Y_{i-1}+n^{-1/2}\delta\|Y_{i-1}\|)- - F(x+n^{-1/2}u'Y_{i-1}-n^{-1/2}\delta\|Y_{i-1}\|)\}\Big]^2 dG(x) \le k\delta^2\Big) = 1,$$

where $h^{\pm}$ is as in the proof of Theorem 7.2.1.

(10) For every $\|u\| \le B$,

$$\int n^{-1}\Big[\sum_{i=1}^n h(Y_{i-1})\{F(x+n^{-1/2}u'Y_{i-1})-F(x)-n^{-1/2}u'Y_{i-1}f(x)\}\Big]^2 dG(x) = o_p(1),$$

and (5.5.68b) holds.

Now, recall the definitions of W_h, ν_h, $\mathscr{W}_h$, $\mathscr{W}^{\pm}$, $T^{\pm}$, $W^{\pm}$, $Z^{\pm}$, $m^{\pm}$ from (7.1.6), (7.2.2), (7.2.5) and (7.2.6). Let $|\cdot|_G$ denote the L_2–norm w.r.t. the measure G. In the *proofs below*, we have adopted the notation and conventions used in the proof of Theorem 7.2.1. Thus, e.g., $\xi_i \equiv Y_{i-1}$; $\mathscr{W}_u(\cdot)$, $\nu_u(\cdot)$ stand for $\mathscr{W}_h(\cdot, \rho+n^{-1/2}u)$, $\nu_h(\cdot, \rho+n^{-1/2}u)$, etc.

Lemma 7.4.2. *Suppose that the autoregression model* (7.3b.3) *and* (7.3b.4) *holds. Then the following hold.*

(11) *Assumption* (8) *implies that* $\forall\ 0 < B < \infty$,

$$E\int [Z^{\pm}(x; u, a)-Z^{\pm}(x; u, 0)]^2 dG(x) = o(1), \quad \forall\ \|u\| \le B,\ a \in \mathbb{R}.$$

(12) *Assumption* (9) *implies that* $\forall\ 0 < B < \infty$, $\forall\ \|u\| \le B$,

$$\liminf_n P\Big(\sup_{\|\mathbf{v}-\mathbf{u}\|\le\delta} n^{1/2}|\nu_h^{\pm}(x, \rho+n^{-1/2}\mathbf{v}) - \nu_h^{\pm}(x, \rho+n^{-1/2}\mathbf{u})|_G^2 \le k\delta^2\Big) = 1.$$

where k *and* δ *are as in* (9).

(13) *Assumptions* (7), (9) *and* (10) *imply that* $\forall\ 0 < B < \infty$,

$$\sup_{\|\mathbf{u}\|\le B} \int [n^{1/2}\{\nu_h(x, \rho+n^{-1/2}\mathbf{u}) - \nu_h(x, \rho)\} - \mathbf{u}'n^{-1}\sum_{i=1}^{n} h(Y_{i-1})Y_{i-1}\, f(x)]^2 dG(x) = o_p(1).$$

Proof. Let, for $x, a \in \mathbb{R}$; $\mathbf{u}, \mathbf{y} \in \mathbb{R}^p$,

$$p(x,\mathbf{u},a;\mathbf{y}) := |F(x+n^{-1/2}(\mathbf{u}'\mathbf{y}+a\|\mathbf{y}\|))-F(x+n^{-1/2}\mathbf{u}'\mathbf{y})|. \tag{14}$$

Now, observe that $n^{1/2}[Z^{\pm}(x; \mathbf{u}, a)-Z^{\pm}(x; \mathbf{u}, 0)]$ is a sum of n r.v.'s whose i^{th} summand is conditionally centered, given $\mathcal{F}_{i-1}$, and whose conditional variance, given $\mathcal{F}_{i-1}$, is $E[\{h^{\pm}(\xi_i)\}^2\, p(x,\mathbf{u},a;\xi_i)\{1- p(x,\mathbf{u},a;\xi_i)\}]$, $1 \le i \le n$. Hence, by Fubini, the stationarity of $\{\xi_i\}$ and the fact that $(h^{\pm})^2 \le h^2$, $\forall\ \mathbf{u}\in\mathcal{N}(B)$,

$$\text{l.h.s.}(11) \le \int Eh^2(Y_0)p(x,\mathbf{u},a;Y_0)\, dG(x) = o(1),$$

by (8) applied with the given a and with a = 0 and the triangle inequality.

To prove (12), use the nonnegativity of $h^{\pm}$, the monotonicity of F and (7.2.10), to obtain that $\|\mathbf{v}\| \le B$, $\|\mathbf{v} - \mathbf{u}\| \le \delta$ imply that $\forall\ \|\mathbf{u}\| \le B$,

$$n^{1/2}|\nu_{\mathbf{v}}^{\pm}(x) - \nu_{\mathbf{u}}^{\pm}(x)| \le |m^{\pm}(x; \mathbf{u}, \delta) - m^{\pm}(x; \mathbf{u}, -\delta)|, \qquad \forall\ x\in\mathbb{R}. \tag{15}$$

This and (9) readily imply (12) as the r.v. in the l.h.s. of (9) is precisely the $|\cdot|_G^2$ of the r.h.s. of (15) for each $n \ge 1$.

The proof of (13) is obtained from (7), (9) and (10) in the same way as that of (5.5.30) from (5.5.7), (5.5.8) and (5.5.9), hence no details are given. □

Lemma 7.4.3. *Suppose that the autoregression model* (7.3b.3) *and* (7.3b.4) *holds. In addition, assume that* (8) *and* (9) *hold.*
Then, $\forall\ 0 < B < \infty$,

$$(16)\qquad \sup_{\|\mathbf{u}\|\le B} \int [\, \mathcal{W}_h^{\pm}(x, \rho+n^{-1/2}\mathbf{u}) - \mathcal{W}_h^{\pm}(x, \rho)]^2 dG(x) = o_p(1).$$

$$(17)\qquad \sup_{\|\mathbf{u}\|\le B} \int [\, \mathcal{W}_h(x, \rho+n^{-1/2}\mathbf{u}) - \mathcal{W}_h(x, \rho)]^2 dG(x) = o_p(1).$$

Proof. Let $q(x,\mathbf{u};\mathbf{y}) := |F(x + n^{-1/2}\mathbf{u}'\mathbf{y}) - F(x)|$, $x \in \mathbb{R}$; $\mathbf{u}, \mathbf{y} \in \mathbb{R}^p$. The r.v. $n^{1/2}[\, \mathcal{W}_{\mathbf{u}}^{\pm}(\cdot) - \mathcal{W}^{\pm}(\cdot)]$ is a sum of n r.v.'s whose i^{th} summand is conditionally centered, given $\mathcal{F}_{i-1}$, and whose conditional variance, given $\mathcal{F}_{i-1}$, is $E[\{h^{\pm}(\xi_i)\}^2\, q(\cdot,\mathbf{u};\xi_i)\{1- q(\cdot,\mathbf{u};\xi_i)\}]$, $1 \le i \le n$. Hence, by Fubini, the stationarity of $\{\xi_i\}$ and the fact that $(h^{\pm})^2 \le h^2$, $\forall\ \|\mathbf{u}\| \le B$,

$$(18)\qquad E|\, \mathcal{W}_{\mathbf{u}}^{\pm} - \mathcal{W}^{\pm}|_G^2 \le \int n^{-1} \sum_{i=1}^{n} Eh^2(\mathbf{Y}_{i-1})|F(x+n^{-1/2}\mathbf{u}'\mathbf{Y}_{i-1}) - F(x)|\, dG(x)$$
$$\le \int Eh^2(\mathbf{Y}_0)|F(x+n^{-1/2}\mathbf{u}'\mathbf{Y}_0) - F(x)|\, dG(x).$$

Therefore, by (8) with $a = 0$ and the Markov inequality,

$$(19)\qquad |\, \mathcal{W}_{\mathbf{u}}^{\pm} - \mathcal{W}^{\pm}|_G^2 = o_p(1), \qquad \forall\ \|\mathbf{u}\| \le B.$$

Thus, to prove (16), because of the compactness of $\mathcal{N}(B)$, it suffices to show that for every $\eta > 0$ there is a $\delta > 0$ such that for every $\|\mathbf{u}\| \le B$,

$$(20)\qquad \liminf_n P\Big(\sup_{\|\mathbf{v}-\mathbf{u}\|\le\delta} |\mathcal{L}_{\mathbf{v}} - \mathcal{L}_{\mathbf{u}}| < \eta \Big) = 1,$$

where $\mathcal{L}_{\mathbf{u}} := |\, \mathcal{W}_{\mathbf{u}}^{\pm} - \mathcal{W}^{\pm}|_G^2$, $\|\mathbf{u}\| \le B$.

Expand the quadratic, apply the C–S inequality to the cross product terms, to obtain

$$(21)\qquad |\mathcal{L}_{\mathbf{u}} - \mathcal{L}_{\mathbf{v}}| \le |\, \mathcal{W}_{\mathbf{u}}^{\pm} - \mathcal{W}_{\mathbf{v}}^{\pm}|_G^2 + 2|\, \mathcal{W}_{\mathbf{u}}^{\pm} - \mathcal{W}_{\mathbf{v}}^{\pm}|_G\, |\, \mathcal{W}_{\mathbf{v}}^{\pm} - \mathcal{W}^{\pm}|_G.$$

Observe that $h^{\pm} \ge 0$, F nondecreasing and (7.2.10) imply that

$$0 \le |m^{\pm}(x; \mathbf{u}, \pm\delta) - m^{\pm}(x; \mathbf{u}, 0)| \le m^{\pm}(x; \mathbf{u}, \delta) - m^{\pm}(x; \mathbf{u}, -\delta),$$

for all $x \in \mathbb{R}$, $\mathbf{s} \in \mathcal{N}(B)$, $\|\mathbf{s}-\mathbf{u}\| \le \delta$. Use this, the second ineqality in (7.2.9), (7.2.10), (7.2.11), and the fact that $(a+b)^2 \le 2(a^2+b^2)$, $a, b \in \mathbb{R}$, to obtain

$$
\begin{aligned}
|\mathscr{W}_{\mathbf{v}}^{\pm} - \mathscr{W}_{\mathbf{u}}^{\pm}|_G^2 \le 16\{ & \int [Z^{\pm}(x; \mathbf{u}, \delta) - Z^{\pm}(x; \mathbf{u}, 0)]^2 dG(x) \\
& + \int [Z^{\pm}(x; \mathbf{u}, -\delta) - Z^{\pm}(x; \mathbf{u}, 0)]^2 dG(x) \\
& + \int [m^{\pm}(x; \mathbf{u}, \delta) - m^{\pm}(x; \mathbf{u}, -\delta)]^2 dG(x) \\
& + |n^{1/2}(\nu_{\mathbf{v}}^{\pm} - \nu_{\mathbf{u}}^{\pm})|_G^2 \},
\end{aligned}
$$

for all $\mathbf{v} \in \mathscr{N}(B)$, $\|\mathbf{v} - \mathbf{u}\| \le \delta$. This together with (9), (12), (13), (19), (21) and the C–S inequality proves (20) and hence, (16).

The proof of (17) follows from (16) and the first inequality in (7.2.9). □

Now define, for $\mathbf{t} \in \mathbb{R}^p$,

$$
\text{(22)} \quad K_h(\mathbf{t}) := \int [n^{-1/2} \sum_{i=1}^{n} h(\mathbf{Y}_{i-1})\{I(X_i \le x + \mathbf{t}'\mathbf{Y}_{i-1}) - F(y)\}]^2 \, dG(x),
$$

$$
\hat{K}_h(\mathbf{t}) := \int [\, \mathscr{W}_h(x, \rho) + n^{1/2}(\mathbf{t}-\rho)'\, n^{-1} \sum_{i=1}^{n} h(\mathbf{Y}_{i-1})\mathbf{Y}_{i-1} f(x)]^2 \, dG(x).
$$

Theorem 7.4.1. *Suppose that the autoregression model* (7.3b.3) *and* (7.3b.4) *holds and that* (5.5.69), (7) – (10) *hold. Then,* $\forall\ 0 < B < \infty$,

$$
\text{(23)} \qquad \sup_{\|\mathbf{u}\| \le B} |K_h(\rho + n^{-1/2}\mathbf{u}) - \hat{K}_h(\rho + n^{-1/2}\mathbf{u}))| = o_p(1).
$$

Proof. Observe that, by (5.5.69), (7),

$$
\text{(24)} \qquad E\int \mathscr{W}_h^2(x, \rho)\, dG(x) = Eh^2(Y_0) \int F(1-F)dG < \infty.
$$

The rest of the proof of (23) follows from Lemmas 7.4.2 and 7.4.3 in a similar way as that of (5.5.28) from Lemmas 5.5.1, 5.5.2 and the result (5.5.30). □

Now we shall apply this result to obtain the required quadraticity of the dispersion K_g and K_g^+. For that purpose recall the matrices $\mathscr{X}$, $\mathscr{G}$ and $\mathcal{B}_n$ from (7.3a.1). Note that X_{i-j}, $g(X_{i-j})$ are the $(i,j)^{th}$ entries of $\mathscr{X}$, $\mathscr{G}$, respectively, $1 \le i \le n$, $1 \le j \le p$. Also observe that the

$$
\text{(25)} \qquad j^{th} \text{ row of } \mathcal{B}_n \text{ is } \sum_{i=1}^{n} g(X_{i-j})\mathbf{Y}'_{i-1}, \quad 1 \le j \le p.
$$

To obtain the desired result about K_g, we need to apply the above theorem p times, j^{th} time with

$$
\text{(26)} \qquad h(\mathbf{Y}_{i-1}) \equiv g(X_{i-j}), \ j = 1, \ldots, p.
$$

Now write $\mathscr{W}_j$ for $\mathscr{W}_h$ when h is as in (26) and $\mathscr{W}_j(\cdot)$ for $\mathscr{W}_j(\cdot, \rho)$, $1 \le j \le p$. Note that

$$\mathscr{W}_j(x) := n^{-1/2} \sum_{i=1}^{n} g(X_{i-j})\{I(\epsilon_i \le x) - F(x)\}, \qquad 1 \le j \le p, \ x \in \mathbb{R}.$$

We also need to define the approximating quadratic forms: For $\mathbf{t} \in \mathbb{R}^p$, let

(27) $$\hat{K}_g(\mathbf{t}) := \sum_{j=1}^{p} \int [\mathscr{W}_j(x) + n^{1/2}(\mathbf{t}-\rho)'\, n^{-1} \sum_{i=1}^{n} g(X_{i-j})\mathbf{Y}_{i-1} f(x)]^2 dG(x),$$

(28) $$\hat{K}_g^{+}(\mathbf{t}) := \sum_{j=1}^{p} \int [\mathscr{W}_j^{+}(x) + 2n^{1/2}(\mathbf{t}-\rho)'\, n^{-1} \sum_{i=1}^{n} g(X_{i-j})\mathbf{Y}_{i-1} f(x)]^2 dG(x),$$

where

(29) $$\mathscr{W}_j^{+}(x) := n^{-1/2} \sum_{i=1}^{n} g(X_{i-j})\{I(\epsilon_i \le x) - I(-\epsilon_i < x)\}, \qquad 1 \le j \le p, \ x \in \mathbb{R}.$$

Before stating the desired results consider the conditions (7) – (10) when h is as in (26). Condition (7a) is now equal to requiring that $Eg^2(X_{1-j}) < \infty$ for all $j = 1, \ldots, p$. Because of the stationarity of $\{X_i\}$, this in turn is equal to

($7a_g$) $$Eg^2(X_0) < \infty.$$

Similarly, (8) is equal to

(8_g) $\forall \ \|\mathbf{u}\| \le B, a \in \mathbb{R}, 1 \le j \le p,$

$$\int Eg^2(X_{1-j})\,|F(x+n^{-1/2}(\mathbf{u}'\mathbf{Y}_0 + a\|\mathbf{Y}_0\|)) - F(x)|\,dG(x) = o(1).$$

Let (9_g) stand for the condition (9) after $h^{\pm}(\mathbf{Y}_{i-1})$ is replaced by $g^{\pm}(X_{i-j})$, $1 \le j \le p$, in (9), $1 \le i \le n$. Interpret (10_g) similarly. We are now ready to state

Theorem 7.4.2. *Suppose that the autoregression model* (7.3b.3) *and* (7.3b.4) *holds and that* (5.5.68a), (5.5.69), (7b), ($7a_g$) – (10_g) *hold.*
Then, $\forall \ 0 < B < \infty$,

(30) $$\sup_{\|\mathbf{u}\| \le B} |K_g(\rho + n^{-1/2}\mathbf{u}) - \hat{K}_g(\rho + n^{-1/2}\mathbf{u}))| = o_p(1). \qquad \square$$

Proof. Note that the j^{th} summand in K_g is a K_h with h as in (26). Hence (30) readily follows from (23). $\square$

Lemmas 7.4.2 and 7.4.3 can be directly used to obtain the following

Theorem 7.4.3. *In addition to the assumptions of Theorem* 7.4.2, *except* (5.5.69), *assume that* F *is symmetric around* 0, G *satisfies* (5.3.8) *and that* (5.6a.13) *holds.*

Then, $\forall\ 0 < B < \infty$,

(31) $$\sup_{\|\mathbf{u}\|\le B} |K_g^+(\rho+n^{-1/2}\mathbf{u}) - \hat{K}_g^+(\rho+n^{-1/2}\mathbf{u}))| = o_p(1). \qquad \square$$

Upon expanding the quadratic and using an appropriate analogue of (24) obtained when h is as in (26), one can rewrite

$$\hat{K}_g(\mathbf{t}) = \hat{K}_g(\rho) + 2(\mathbf{t}-\rho)' n^{-1/2}\mathcal{B}_n' \int \mathcal{W}(x)\, f(x)dG(x)$$
$$+ (\mathbf{t}-\rho)'\, n^{-1}\mathcal{B}_n'\mathcal{B}_n\, (\mathbf{t}-\rho)|f|_G^2, \quad \mathbf{t}\in\mathbb{R}^p,$$

where $\mathcal{W} := (\mathcal{W}_1, \ldots, \mathcal{W}_p)'$. Now consider the r.v.'s in the second term. Recalling the definition of ψ from (5.6a.2), one can rewrite

$$\mathcal{S}_n := \int \mathcal{W}(x)\, f(x)dG(x) = -n^{-1/2}\sum_{i=1}^{n} \mathbf{g}_i\, [\psi(\epsilon_i) - E\psi(\epsilon)],$$

where $\mathbf{g}_i'$ is the i^{th} row of $\mathcal{G}$, i.e.,

(32) $$\mathbf{g}_i' := (g(X_{i-1}), g(X_{i-2}), \ldots, g(X_{i-p})), \qquad 1 \le i \le n.$$

Since $\mathbf{g}_i$ is a function of Y_{i-1}, it is $\mathcal{F}_{i-1}$ – measurable. Therefore, in view of ($7a_g$) and (5.5.68a), $\{(\mathcal{S}_n, \mathcal{F}_{n-1}),\ n \ge 1\}$ is a mean zero square integrable martingale array. The same assumptions, and an argument like that in the proof of Lemma 7.3d.2, enable one to verify the applicability of Lemma A.3 in the Appendix to $\mathcal{S}_n$. Hence, it follows that

(33) $$\mathcal{S}_n \xrightarrow[d]{} N(0, G^*\tau^2 I_{p\times p}), \quad G^* = E\mathbf{g}_1\mathbf{g}_1', \quad \tau^2 = \mathrm{Var}\, \psi(\epsilon_1)/(\int f^2 dG)^2.$$

By the stationarity and the Ergodic Theorem, we also obtain

(34) $$n^{-1}\mathcal{B}_n \longrightarrow \mathcal{B}, \text{ a.s.}, \qquad \mathcal{B} := En^{-1}\mathcal{B}_n = E\mathbf{g}_1 Y_0'.$$

Consequently it follows that the dispersion K_g satisfied (5.4.A1) to (5.4.A3) with $\theta_0 = \rho$, $\delta_n \equiv n^{1/2}$, $S_n \equiv n^{-1/2}\mathcal{B}_n' \mathcal{S}_n$, $W_n \equiv n^{-1}\mathcal{B}_n'\mathcal{B}_n$, $W = \mathcal{B}$, $\Sigma = \mathcal{B}'\, G^*\mathcal{B}\, \tau^2$, and hence it is an u.l.a.n.q. dispersion.

In view of (24) applied to h as in (26), the condition (5.4.A4) is trivially implied by ($7a_g$) and (5.5.69).

Recall, from Section 5.5, that in the linear regression setup the condition (5.4.A5) was shown to be implied by (5.5.11) and (5.5.12). In the present situation, the role of Γ_n, $\mathbf{\Gamma}_n$ of (5.5.11) is being played by $n^{-1}\mathcal{B}_n f$, $n^{-1}\mathcal{B}_n \int f dG$, respectively. Thus, in view of (34) and (5.5.68a), an analogue of (5.5.11) would hold in the present case if we additionally assumed that $\mathcal{B}$ is positive definite. An exact analogue of (5.5.12) in the present case is

(35) Either

$$\theta' g_i Y'_{i-1}\theta \geq 0, \ \forall\ 1 \leq i \leq n, \forall\ \theta \in \mathbb{R}^p, \|\theta\| = 1, \text{ a.s.},$$

or

$$\theta' g_i Y'_{i-1}\theta \leq 0, \ \forall\ 1 \leq i \leq n, \forall\ \theta \in \mathbb{R}^p, \|\theta\| = 1, \text{ a.s..}$$

We are now ready to state the following

Theorem 7.4.4. *In addition to the assumptions of Theorem 7.4.2, assume that the* $\mathcal{B}$ *of* (34) *is positive definite and that* (35) *holds. Then,*

$$n^{1/2}(\hat{\rho}_g - \rho) = -\{n^{-1}\mathcal{B}_n \int f^2 dG\}^{-1} \mathcal{S}_n + o_p(1). \tag{36}$$

Consequently,

$$n^{1/2}(\hat{\rho}_g - \rho) \xrightarrow[d]{} N(0, (\mathcal{B})^{-1}\mathcal{G}^*(\mathcal{B}')^{-1}\tau^2). \tag{37}$$ □

Let $\hat{\rho}_x$ denote the estimator $\hat{\rho}_g$ when $g(x) \equiv x$. Observe that in this case $\mathcal{G}^* = \mathcal{B} = En^{-1}\mathcal{X}'\mathcal{X} = EY_0Y_0'$. Moreover, the assumption (35) is *a priori* satisfied and (7.3b.3), (7.3b.4) and (7b) imply that EY_0Y_0' is positive definite. Consequently, we have obtained

Corollary 7.4.1. *Suppose that the autoregression model* (7.3b.3) *and* (7.3b.4) *holds and that* (5.5.68), (5.5.69), (7b), (8_g) – (10_g) *with* $g(x) \equiv x$ *hold. Then,*

$$n^{1/2}(\hat{\rho}_x - \rho) \xrightarrow[d]{} N(0, (EY_0Y_0')^{-1}\tau^2). \tag{38}$$ □

Remark 7.4.1. *Asymptotic Optimality of* $\hat{\rho}_x$. Because $\mathcal{B}$ and EY_0Y_0' are positive definite, and because of $n^{-1}\mathcal{X}'\mathcal{X} \longrightarrow EY_0Y_0'$, a.s., and (34), there exists an N_0 such that $n^{-1}\mathcal{X}'\mathcal{X}$ and $n^{-1}\mathcal{B}_n$ are positive definite for all $n \geq N_0$.

Recall the inequality (5.6a.8). Take $\mathbf{J} = n^{-1/2}\mathcal{G}'$, $\mathbf{L} = n^{-1/2}\mathcal{X}'$ in that inequality to obtain

$$n^{-1}\mathcal{G}'\mathcal{G} \geq n^{-1}\mathcal{G}'\mathcal{X}(n^{-1}\mathcal{X}'\mathcal{X})^{-1}\cdot n^{-1}\mathcal{X}'\mathcal{G}, \quad \forall\ n \geq N_0, \text{ a.s.},$$

with equality holding if, and only if $\mathcal{X} \propto \mathcal{G}$. Letting n tend to infinity in this inequality yields

$$(\mathcal{B})^{-1}\mathcal{G}^*(\mathcal{B}')^{-1} \geq (EY_0Y_0')^{-1}.$$

We thus have proved the following:

(39) *Among all estimators* $\{\hat{\rho}_g;$ g *satifying* $(7a_g) - (10_g)$ *for the given* (F, G) *that satisfy* (7b), (5.5.68), (5.5.69)}, *the one that minimizes the asymptotic variance* is $\hat{\rho}_x$! □

We shall now state analogous results for ρ_g^+. Arguments for their proofs are similar to those appearing above and, hence, will not be given.

Theorem 7.4.5. *In addition to the assumptions of Theorem* 7.4.4, *except* (5.5.69), *assume that* F *is symmetric around* 0, G *satisfies* (5.3.8) *and that* (5.6a.13) *holds. Then,*

(40) $$n^{1/2}(\rho_g^+ - \rho) = -\{n^{-1}\mathcal{B}_n \int f^2 dG\}^{-1} \mathcal{S}_n^+ + o_p(1),$$

where

$$\mathcal{S}_n^+ := \int \mathcal{W}^+(x)\, f(x)dG(x) = n^{-1/2} \sum_{i=1}^{n} g_i\, [\psi(-\epsilon_i) - \psi(\epsilon_i)].$$

Consequently,

(41) $$n^{1/2}(\rho_g^+ - \rho) \xrightarrow[d]{} N(0, (\mathcal{B})^{-1} \mathcal{G}^*(\mathcal{B}')^{-1}\tau^2),$$

(42) $$n^{1/2}(\rho_x^+ - \rho) \xrightarrow[d]{} N(0, (EY_0Y_0')^{-1}\tau^2).$$ □

Obviously the optimality property like (39) holds here also.

Remark 7.4.2. *On assumptions for the asymptotic normality of* $\hat{\rho}_x$, ρ_x^+. If G is a finite measure and F has uniformly continuous density then it is not hard to see that $(8_g) - (10g)$, with $g(x) \equiv x$, are all implied by (7b).

Consider the following assumptions for general G:

(43) $E|\epsilon|^3 < \infty$, $E\epsilon^2 > 0$.

(44) As a function of $s \in \mathbb{R}$, $\int E|X_{1-j}|^2 \|Y_0\|\, f(x+s\|Y_0\|)\, dG(x)$ is continuous at 0, $1 \le j \le p$.

(45) For every $\delta > 0$, $\mathbf{u} \in \mathbb{R}^p$,

$$\int_{-1}^{1}\int E\{\|Y_0\|[f(x+n^{-1/2}(\mathbf{u}'Y_0+t\delta\|Y_0\|)) - f(x+n^{-1/2}\mathbf{u}'Y_0)]\}^2\, dG(x)dt = o(1).$$

(46) For every $\mathbf{u} \in \mathbb{R}^p$,

$$\int [n^{-1} \sum_{i=1}^{n} X_{i-j}^{\pm} \, \|Y_{i-1}\| \, f(x+n^{-1/2}\mathbf{u}'Y_{i-1})]^2 \, dG(x) = O_p(1), \quad 1 \le j \le p.$$

An argument similar to the one used in verifying the Claim 5.5.1 shows that (5.5.68a), (43) and (44) imply (8_g) while (5.5.68b), (45) and (46) imply (9g) and (10g).

In particular if $G(x) \equiv x$, then (5.5.68), (43) and f continuous imply all of the above conditions, (5.5.69) and (5.6a.13). This is seen with the help of a version of Scheffe's Theorem. □

Remark 7.4.3. *Asymptotic relative efficiency of* $\hat{\rho}_x$, ρ_x^+. Since their asymptotic variances are the same, we shall carry out the discussion in terms of $\hat{\rho}_x$ only, as the same applies to ρ_x^+ under the additional assumption of the symmetry of F and G.

Consider the case $p = 1$. Let $\sigma^2 = \mathrm{Var}(\epsilon)$ and $\hat{\rho}_{ls}$ denote the least square estimator of ρ_1. Then it is well known that under (7b), $n^{1/2}(\hat{\rho}_{ls} - \rho_1) \underset{d}{\longrightarrow} N(0, 1-\rho_1^2)$. See, e.g., Anderson (1971). Also note that in this case $(EY_0Y_0')^{-1} = (1-\rho_1^2)/\sigma^2$. Hence the asymptotic relative efficiency e of $\hat{\rho}_x$, relative to $\hat{\rho}_{ls}$, obtained by taking the ratio of the inverses of their asymptotic variances, is

$$e = e(\hat{\rho}_x, \hat{\rho}_{ls}) = \sigma^2/\tau^2. \tag{47}$$

Note that $e > 1$ means $\hat{\rho}_x$ is asymptotically more efficient than $\hat{\rho}_{ls}$. It follows that $\hat{\rho}_x$ is to be prefered to $\hat{\rho}_{ls}$ for the heavy tailed error d.f.'s F. Also note that if $G(x) \equiv x$ then $\tau^2 = 1/12\,[\int f^2(x)\,dx]^2$ and $e = 12\,\sigma^2\,[\int f^2(x)\,dx]^2$. If G is degenerate at 0, then $\tau^2 = 1/[4f^2(0)]$ and $e = 4\sigma^2 f^2(0)$. These expressions are well known in connection with the Wilcoxon and median rank estimators of the slope parameters in linear regression models. For example if F is $N(0, 1)$ then the first expression is $3/\pi$ while the second is $2/\pi$. See Lehmann (1975) for some bounds on these expressions. Similar conclusions remain valid for $p > 1$. □

Remark 7.4.4. *Least Absolute Deviation Estimator.* As mentioned earlier, if we choose $g(x) \equiv x$ and G to be degenerate at 0 then ρ_x^+ is the l.a.d. estimator, v.i.z.,

$$\rho_{\ell ad}^+ := \mathrm{argmin}\{\sum_{j=1}^{p} [\sum_{i=1}^{n} X_{i-j}\,\mathrm{sign}(X_i - \mathbf{t}'Y_{i-1})]^2;\ \mathbf{t} \in R^p\} \tag{48}$$

See also (6). Because of its importance we will now summarize sufficient conditions under which it is asymptotically normally distributed. Of course we could use the stronger conditions (43) – (46) but they do not use the given information about G.

Clearly, (7b) implies ($7a_g$) when $g(x) \equiv x$. Moreover, in this case the l.h.s. of (8_g) is

$$EX_{1-j}^2 \,|\, F(n^{-1/2}(\mathbf{u}'\mathbf{Y}_0 + a\|\mathbf{Y}_0\|)) - F(n^{-1/2}\mathbf{u}'\mathbf{Y}_0) \,|$$

which tends to 0 by the D.C.T., (7b) and the continuity of F, $1 \le j \le p$.

Now consider (9_g). Asssume the following:

(49) $\quad$ F *has a density* f *that is continuous at* 0 *and* $f(0) > 0$.

Recall from (7.3b.6) that under (7.3b.3), (7.3b.4) and (7b),

$$n^{-1/2}\max\{\|\mathbf{Y}_{i-1}\|;\ 1 \le i \le n\} = o_p(1). \tag{50}$$

The r.v.'s involved in the l.h.s. of (9_g) in the present case are

$$n^{-1}[\sum_{i=1}^{n} X_{i-j}^{\pm}\{F(n^{-1/2}\mathbf{u}'\mathbf{Y}_{i-1} + n^{-1/2}\delta\|\mathbf{Y}_{i-1}\|) \\ - F(n^{-1/2}\mathbf{u}'\mathbf{Y}_{i-1} - n^{-1/2}\delta\|\mathbf{Y}_{i-1}\|)\}]^2$$

which, in view of (49), can be bounded above by

$$4\delta^2\,[n^{-1}\sum_{i=1}^{n} X_{i-j}^{\pm}\|\mathbf{Y}_{i-1}\|\, f(\eta_{ni})]^2, \tag{51}$$

where $\{\eta_{ni}\}$ are r.v.'s, $\eta_{ni} \in n^{-1/2}[\mathbf{u}'\mathbf{Y}_{i-1} - \delta\|\mathbf{Y}_{i-1}\|,\ \mathbf{u}'\mathbf{Y}_{i-1} + \delta\|\mathbf{Y}_{i-1}\|]$, $1 \le i \le n$. Hence, by the stationarity and the ergodicity of the process $\{X_i\}$, (7b), (49) and (50) imply that the r.v.'s in (51) converge to $4\delta^2\,[EX_{1-j}^{\pm}\|\mathbf{Y}_0\|\, f(0)]^2$, a.s., $1 \le j \le p$. This verifies (9_g) in the present case. One similarly verifies (10_g).

Also note that here (5.5.68) is implied by (49) and (5.5.69) is trivially satisfied as $\int F(1-F)\, dG \le 1/4$ in the present case. We summarize the above discussion in

Corollary 7.4.3. *Assume that the autoregression model* (7.3b.3) *and* (7.3b.4) *holds. In addition, assume that the error d.f.* F *has finite second moment*, $F(0) = 1/2$ *and satisfies* (49). *Then,*

$$n^{1/2}(\rho_{\ell ad}^{+} - \rho) \underset{d}{\rightarrow} N(0, (E\mathbf{Y}_0\mathbf{Y}_0')^{-1}/4f^2(0)),$$

where $\rho_{\ell ad}^{+}$ *is defined at* (48). □

7.5. GOODNESS-OF-FIT TESTING.

Once again consider the AR(p) model given by (7.3b.3), (7.3b.4) and let F_0 be a known d.f.. Consider the problem of testing H_0: $F = F_0$. One of the common tests of H_0 is based on the Kolmogorov–Smirnov statistic

$$D_n := n^{1/2}\sup_x |F_n(x, \hat{\rho}) - F_0(x)|.$$

From Corollary 7.2.1 one readily has the following:

If F_0 *has a uniformly continuous density* f_0, $f_0 > 0$ a.e.; $\int x^2 dF_0(x) < \infty$, $\hat{\rho}$ *satisfies* (7.3c.(iv)) *under* F_0, *then, under* H_0,

$$D_n = \sup |B(F_0(x)) + n^{1/2}(\hat{\rho} - \rho)' n^{-1}\Sigma_i Y_{i-1} f_0(x)| + o_p(1).$$

In addition, if $EY_0 = 0 = E\epsilon_1$, *then* $D_n \xrightarrow[d]{} \sup\{|B(t)|,\ 0 \le t \le 1\}$, *thereby rendering* D_n *asymptotically distribution free.*

Next, consider, H_{01}: $F = N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$. In other words, H_{01} states that the AR(p) process is generated by some normal errors. Let $\hat{\mu}_n$, $\hat{\sigma}_n$ and $\hat{\rho}_n$ be estimators of μ, σ, ρ respectively. Define

$$\hat{F}_n(x) := n^{-1} \Sigma_i I(X_i \le x\hat{\sigma}_n + \hat{\mu}_n + \hat{\rho}_n' Y_i), \qquad x \in \mathbb{R},$$

$$\hat{D}_n := n^{1/2} \sup_x |\hat{F}_n(x) - \Phi(x)|, \quad \Phi = N(0, 1) \text{ d.f.}.$$

Corollary 7.2.1. can be readily modified in a routine fashion to yield that if

$$n^{1/2}|(\hat{\mu}_n - \mu) + (\hat{\sigma}_n - \sigma)|\ \sigma^{-1} + n^{1/2}\|\hat{\rho}_n - \rho\| = O_p(1)$$

then

$$\hat{D}_n := \sup_x|\ B(\Phi(x)) + n^{1/2}\{(\hat{\mu}_n - \mu) + (\hat{\sigma}_n - \sigma)\}\sigma^{-1}\ n(x)| + o_p(1),$$

where n is the density of Φ. Thus the asymptotic null distribution of $\hat{D}_n$ is similar to its analogue in the one sample location–scale model: *the estimation of* ρ *has no effect on the large sample null distribution of* $\hat{D}_n$.

Clearly, similar conclusions can be applied to other goodness-of-fit tests. In particular we leave it as *an exercise* for an interested reader to investigate the large sample behaviour of the goodness-of-fit tests based on L_2–distances, analogous to the results obtained in Section 6.3. Lemma 6.3.1 and the results of the previous section are found useful here. □□□□□