The Annals of Statistics
1993, Vol. 21, No. 1, 534–551

R-ESTIMATION OF THE PARAMETERS OF AUTOREGRESSIVE [AR(p)] MODELS

BY HIRA L. KOUL[1] AND A. K. MD. E. SALEH[2]

Michigan State University and Carleton University

In an AR(p) model, R-estimation of a subset of parameters is considered when the complementary subset is possibly redundant. Along with the rank test of the full hypothesis and the subhypothesis of the parameters, both preliminary test and shrinkage R-estimators are considered. In the light of asymptotic distributional risks, the relative asymptotic risk-efficiency results are given. Though, the *shrinkage R-estimator* may dominate their classical versions, they do not in general dominate the preliminary test R-estimators.

1. Introduction. Let F be a d.f. on $\mathscr{R}$, $p \geq 1$ be an integer, and $\varepsilon_0, \varepsilon_{\pm 1}, \varepsilon_{\pm 2}, \dots$ be i.i.d. F r.v.'s. Let $\mathbf{Y}_0' = (X_0, X_{-1}, \dots, X_{1-p})$ be an observable random vector independent of $\varepsilon_1, \varepsilon_2, \dots$. Consider the pth order autoregressive, [AR(p)], model where the observations $X_1, \dots, X_n$ satisfy the relation

$$X_i = \rho_1 X_{i-1} + \cdots + \rho_p X_{i-p} + \varepsilon_i, \qquad 1 \leq i \leq n, p \geq 1 \tag{1.1}$$

and we assume that all roots of

$$x^p - \rho_1 x^{p-1} - \rho_2 x^{p-2} - \cdots - \rho_p = 0, \quad \text{are in } (-1, 1). \tag{1.2}$$

Here, $\boldsymbol{\rho}' = (\rho_1, \dots, \rho_p) \in \mathscr{R}^p$ is a vector of *unknown* autoregressive parameters. Let $1 \leq p_1 < p$ and $p_2 = p - p_1$ and let $\boldsymbol{\rho}' = (\boldsymbol{\rho}_1', \boldsymbol{\rho}_2')$ where $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ are p_1 and p_2 vectors, respectively. Our primary interest is to develop the theory of R-estimation of $\boldsymbol{\rho}_1$ when it is suspected that $\boldsymbol{\rho}_2$ may be equal to $\mathbf{0}$. The situation may arise when the experimentor has over-modelled up to order p but suspects that the p_2-vector $\boldsymbol{\rho}_2$ may be negligible. Under this situation he wants to estimate $\boldsymbol{\rho}_1$. To this end we first consider the theory of R-estimation of $\boldsymbol{\rho}$ based on a class of rank statistics. To obtain the asymptotic properties of these R-estimators, we prove the asymptotic uniform linearity (AUL) result for the class of rank statistics. This result is then used to investigate the R-estimates of $\boldsymbol{\rho}_1$ when $\boldsymbol{\rho}_2$ is suspected to be $\mathbf{0}$.

For the AR(p) model (1.1) let the R-estimator of $\boldsymbol{\rho}' = (\boldsymbol{\rho}_1', \boldsymbol{\rho}_2')$ be denoted by $\tilde{\boldsymbol{\rho}}_n' = (\tilde{\boldsymbol{\rho}}_{1n}', \tilde{\boldsymbol{\rho}}_{2n}')$ so that $\tilde{\boldsymbol{\rho}}_{1n}$ is the *unrestricted* R-estimator (URE) of $\boldsymbol{\rho}_1$. When

---

Received May 1990; revised May 1992.

[1]Research partially supported by NSF Grant DMS-91-02841.

[2]Research supported by NSERC Grant A3088 and GR-5 Grant from the Faculty of Graduate Studies and Research, Carleton University.

AMS 1991 *subject classifications*. Primary 62M10, 62M05; secondary 62G05, 62G10.

Key words and phrases. Robust estimation, rank-statistics, asymptotic distributional risk, Stein phenomenon, R-estimation.

$\boldsymbol{\rho}_2 = \mathbf{0}$, we consider the restricted AR(p_1) model, namely,

$$(1.3) \qquad X_i = \rho_1 X_{i-1} + \cdots + \rho_{p_1} X_{i-p_1} + \varepsilon_i, \qquad 1 \le i \le n, p_1 \ge 1.$$

Let $\hat{\boldsymbol{\rho}}_{1n}$ be the restricted R-estimator (RRE) of $\boldsymbol{\rho}_1$. This RRE performs better than URE when $\boldsymbol{\rho}_2 = \mathbf{0}$ holds. But, as soon as $\boldsymbol{\rho}_2$ deviates from $\mathbf{0}$, the RRE may be considerably biased, inefficient and even inconsistent, while URE retains all the performance characteristics steadily for the variation of $\boldsymbol{\rho}_2$.

This paper considers the preliminary test R-estimators (PTRE) and the shrinkage R-estimators (SRE) of $\boldsymbol{\rho}_1$ when $\boldsymbol{\rho}_2$ is suspected to be near $\mathbf{0}$ as a compromise between URE and RRE. The relative merits of the estimators of $\boldsymbol{\rho}_1$ are studied in terms of the asymptotic distributional risk (ADR) as in Saleh and Sen (1986) and Sen and Saleh (1987). It is shown that when $\boldsymbol{\rho}_2$ is near $\mathbf{0}$, both PTRE and SRE outperform URE, though RRE may still dominate either of these. However, when $\boldsymbol{\rho}_2$ is away from $\mathbf{0}$, RRE performs rather poorly, while both PTRE and SRE are robust.

The proposed R-estimators of $\boldsymbol{\rho}$ along with their asymptotic properties are discussed in Section 2. This section also contains the proof of an AUL result for the class of rank statistics considered. Section 3 discusses the RRE, PTRE and SRE with details. Section 4 contains the ADR and ADRE of various R-estimators of $\boldsymbol{\rho}_1$ with concluding remarks in Section 5. Some proofs of the results in Section 2 appear in the Appendix.

2. *R*-estimation of $\boldsymbol{\rho}$ for the AR(*p*)-model. In this section, we introduce a class of R-estimators of $\boldsymbol{\rho}$ for the AR(p)-model (1.1) and discuss their large sample properties. Let $\mathbf{Y}_i' = (X_i, \dots, X_{i-p+1})$, $1 \le i \le n$ and define $R_{i\mathbf{u}}$ to be the rank of $X_i - \mathbf{u}'\mathbf{Y}_{i-1}$ among $\{X_j - \mathbf{u}'\mathbf{Y}_{j-1}, 1 \le j \le n\}$, for $1 \le i \le n$. Set $R_{i\mathbf{u}} \equiv 0$ for $i \le 0$. Let φ be a nondecreasing function from $[0, 1]$ to the real line and define $\mathbf{S}' = (S_1, \dots, S_p)$, where

$$(2.1) \quad S_j(\mathbf{u}) = n^{-1/2} \sum_{i=j+1}^{n} X_{i-j}\varphi(R_{i\mathbf{u}}/(n+1)), \qquad 1 \le j \le p, \mathbf{u} \in R^p.$$

The class of rank statistics $\mathbf{S}$, one for each φ, is analogous to a class of similar rank statistics in the linear regression model where one replaces the weights $\{X_{i-j}\}$ by the appropriate design points as discussed in Hájek and Šidák (1967). A test of the null hypothesis H_0: $\boldsymbol{\rho} = \boldsymbol{\rho}_0$ may be based on a suitably standardized $\mathbf{S}(\boldsymbol{\rho}_0)$, the large values of the statistic being significant. For an alternative class of rank tests useful in testing one ARMA model against another, see Hallin and Puri (1988).

It is thus natural to define R-estimator $\tilde{\boldsymbol{\rho}}_n$ of $\boldsymbol{\rho}$ by the relationship

$$(2.2) \qquad \inf_{\mathbf{u} \in R^p} \|\mathbf{S}(\mathbf{u})\| = \|\mathbf{S}(\tilde{\boldsymbol{\rho}}_n)\|.$$

An alternative way to define R-estimator of $\boldsymbol{\rho}$ is to adapt Jaeckel (1972) to the AR(p) model. Accordingly, set $a_n(i) = \varphi(i/(n+1))$, $Z_{(i)}(\mathbf{u}) =$ the ith largest residual among $\{X_k - \mathbf{u}'\mathbf{Y}_{k-1},\ 1 \le k \le n\}$, $1 \le i \le n$, and

$$\mathbf{T}(\mathbf{u}) = \sum_{i=1}^{n} a_n(i) Z_{(i)}(\mathbf{u}), \qquad \mathbf{u} \in \mathscr{R}^p. \tag{2.3}$$

Arguing as in Jaeckel (1972), if $\Sigma_{i=1}^{n} a_n(i) = 0$, then $\mathbf{T}(\mathbf{u})$ can be shown to be convex on $\mathscr{R}^p$ with its a.e. differential equal to $-n^{1/2}\mathbf{S}$. Thus, a minimizer $\boldsymbol{\rho}_J$ of $\mathbf{T}(\mathbf{u})$ exists and has the property that makes $\|\mathbf{S}\|$ small. As is shown in Jaeckel in connection with linear regression model, it will follow from the linearity result given in Theorem 2.1 that $\boldsymbol{\rho}_J$ and $\tilde{\boldsymbol{\rho}}_n$ are asymptotically equivalent.

THEOREM 2.1 (AUL of R-statistics). *Assume that* (1.1) *and* (1.2) *hold. In addition, assume that the following conditions hold.*

(a) (i) $E\varepsilon = 0$, $0 < E\varepsilon^4 < \infty$. (ii) *$F$ has uniformly continuous density f, $f > 0$ a.e.*

(b) *φ is nondecreasing and differentiable with its derivative φ' being uniformly continuous on* $[0, 1]$.

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

$$\sup_{\|\boldsymbol{\Delta}\| \le b} \|\mathbf{S}(\boldsymbol{\rho} + n^{-1/2}\boldsymbol{\Delta}) - \hat{\mathbf{S}} + \boldsymbol{\Delta}'\boldsymbol{\Sigma}\gamma\| = o_p(1), \tag{2.4}$$

where $\hat{\mathbf{S}}' = (\hat{S}_1, \dots, \hat{S}_p)$ *with*

$$\hat{\mathbf{S}}_j := n^{-1/2} \sum_{i=j+1}^{n} \left(X_{i-j} - \bar{X}_j\right)\left[\varphi(F(\varepsilon_i)) - \bar{\varphi}\right], \qquad 1 \le j \le p,$$

$$\bar{\varphi} := \int_0^1 \varphi(t)\, dt,$$

$$\bar{X}_j := n^{-1} \sum_{i=j+1}^{n} X_{i-j}, \qquad \gamma = \int f \, d\varphi(F), \tag{2.5}$$

$$\boldsymbol{\Sigma} := ((\beta(i-j))), \qquad i = 1, \dots p;\ j = 1, \dots p;$$

$$\beta(k) := \operatorname{Cov}(X_0, X_k), \qquad 1 \le k \le p.$$

Note that the above theorem covers the Wilcoxon-type score but not the normal score.

Before proceeding to prove the above result, we state two lemmas. Accordingly, let g be a nonnegative measurable function on $[0, 1]$, U denote a uniform $[0, 1]$ r.v. and define

$$Z_j(t) = n^{-1/2} \sum_{i=1}^{n} X_{i-j}\left[g(F(\varepsilon_i)) I(F(\varepsilon_i) \le t) - G(t)\right], 0 \le t \le 1, 1 \le j \le p,$$

where $G(t) = Eg(U)I(U \le t) = \int_0^t g(s)\, ds$, $0 \le t \le 1$, and $I(A)$ denotes the indicator function of the event A.

LEMMA 2.1. *In addition to* (1.1), (1.2) *and* (a)(i), *assume that* $Eg^4(U) < \infty$. *Then* $\forall\ \eta > 0$,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P\Big(\sup_{|t-v| \le \delta,\ 1 \le j \le p} |Z_j(t) - Z_j(v)| > \eta \Big) = 0.$$

A proof of this lemma is given in the Appendix.

In order to state the next lemma we need to define the empirical process of the residuals

$$F_n(x, \mathbf{u}) = n^{-1} \sum_{i=1}^{n} \mathbf{I}(X_i - \mathbf{u}'\mathbf{Y}_{i-1} \le x), \qquad x \in \mathscr{R}, \mathbf{u} \in \mathscr{R}^p.$$

Note that $F_n(x, \boldsymbol{\rho})$ is the empirical process of $\{\varepsilon_i\}$ and that

$$F_n(x, \boldsymbol{\rho} + n^{-1/2}\boldsymbol{\Delta}) = n^{-1} \sum_{i=1}^{n} \mathbf{I}(\varepsilon_i \le x + n^{-1/2}\boldsymbol{\Delta}'\mathbf{Y}_{i-1}), \qquad x \in \mathscr{R}, \mathbf{u} \in \mathscr{R}^p.$$

The next lemma gives the AUL result for F_n.

LEMMA 2.2. *Let* (1.1) *and* (1.2) *hold*, $E\varepsilon^2 < \infty$ *and* (a)(ii) *hold*. *Then for every* $0 < b < \infty$,

$$\begin{aligned} &\sup_{|x|<\infty,\ \|\boldsymbol{\Delta}\| \le b} \Big| n^{1/2}\big[F_n(x, \boldsymbol{\rho} + n^{-1/2}\boldsymbol{\Delta}) - F_n(x, \boldsymbol{\rho})\big] - \boldsymbol{\Delta}' n^{-1} \sum_{i=1}^{n} \mathbf{Y}_{i-1} f(x) \Big| \\ &\quad = o_p(1). \end{aligned} \tag{2.6}$$

If, in addition, $E\varepsilon_1 = 0$, *then*

$$\sup_{|x|<\infty,\ \|\boldsymbol{\Delta}\| \le b} n^{1/2} |F_n(x, \boldsymbol{\rho} + n^{-1/2}\boldsymbol{\Delta}) - F_n(x, \boldsymbol{\rho})| = o_p(1). \tag{2.7}$$

PROOF. The first claim of the lemma follows from Corollary 1.1 of Koul (1991). The last claim follows because by the ergodic theorem $n^{-1}\Sigma_i \mathbf{Y}_{i-1} \to E\mathbf{Y}_0$. But $E\varepsilon = 0$ implies $E\mathbf{Y}_0 = 0$. □

PROOF OF THEOREM 2.1. Observe that with

$$\begin{aligned} &\tilde{S}_j(\mathbf{u}) = n^{-1/2} \sum_{i=1}^{n} X_{i-j} \varphi(R_{i\mathbf{u}}/(n+1)), \qquad \mathbf{u} \in \mathscr{R}^p, 1 \le j \le p, \\ &\sup_{1 \le j \le p,\ \mathbf{u} \in \mathscr{R}^p} |S_j(\mathbf{u}) - \tilde{S}_j(\mathbf{u})| \le p \max_{1-p \le k \le 0} |X_k| \, \|\varphi\|_\infty n^{-1/2} \to 0, \quad \text{a.s.} \end{aligned} \tag{2.8}$$

Thus, it suffices to prove the theorem with $\{S_j\}$ replaced by $\{\tilde{S}_j\}$. Let

$$\tilde{\mathbf{S}}(\mathbf{u}) = n^{-1/2} \sum_{i=1}^{n} \mathbf{Y}_{i-1} \varphi(R_{i\mathbf{u}}/(n+1)).$$

Define for $\mathbf{u} \in \mathscr{R}^p$,

$$
\begin{aligned}
\mathbf{M}(\mathbf{u}) &= n^{-1/2} \sum_{i-1}^{n} \mathbf{Y}_{i-1}[\varphi(R_{i\mathbf{u}}/(n+1)) - \bar{\varphi}], \\
\hat{\mathbf{M}} &= n^{-1/2} \sum_{i=1}^{n} \mathbf{Y}_{i-1}[\varphi(F(\varepsilon_i)) - \bar{\varphi}].
\end{aligned} \tag{2.9}
$$

For the sake of convenience we write $\mathbf{M}(\boldsymbol{\Delta})$, $F_n(\cdot, \boldsymbol{\Delta})$ and so on for $\mathbf{M}(\boldsymbol{\rho} + n^{-1/2}\boldsymbol{\Delta})$, $F_n(\cdot, \boldsymbol{\rho} + n^{-1/2}\boldsymbol{\Delta})$ and so on. Thus, for example, $F_n(\cdot, \mathbf{0})$ now stands for the empirical of ε_i, $1 \le i \le n$. Write $F_n(\cdot)$ for $F_n(\cdot, \mathbf{0})$. The index i in the summation or maximum will vary from 1 to n, unless specified otherwise. The supremum with respect to $\boldsymbol{\Delta}$ will be always over the ball $\{\boldsymbol{\Delta};\ \|\boldsymbol{\Delta}\| \le b\}$. We shall sometimes write U_i for $F(\varepsilon_i)$, $1 \le i \le n$.

Now, let

$$
e_{ni\boldsymbol{\Delta}} = n^{1/2}[(R_{i\boldsymbol{\Delta}}/(n+1)) - F(\varepsilon_i))], \qquad 1 \le i \le n, \boldsymbol{\Delta} \in \mathscr{R}^p.
$$

We shall first show that

$$
\sup_{i, \boldsymbol{\Delta}} n^{-1/2} |e_{ni\boldsymbol{\Delta}}| = o_p(1). \tag{2.10}
$$

Observe that with $\varepsilon_{i\boldsymbol{\Delta}} = \varepsilon_i - n^{-1/2}\boldsymbol{\Delta}'\mathbf{Y}_{i-1}$, $n^{-1}R_{i\boldsymbol{\Delta}} \equiv F_n(\varepsilon_{i\boldsymbol{\Delta}}, \boldsymbol{\Delta})$, $1 \le i \le n$. Hence by (2.7), and the fact $|n^{1/2}[n(n+1)^{-1} - 1]| \to 0$, it follows that

$$
\begin{aligned}
e_{ni\boldsymbol{\Delta}} &= n^{1/2}[F_n(\varepsilon_{i\boldsymbol{\Delta}}) - F(\varepsilon_i)] + \bar{o}_p(1) \\
&= D_n(F(\varepsilon_{i\boldsymbol{\Delta}})) + n^{1/2}[F(\varepsilon_{i\boldsymbol{\Delta}}) - F(\varepsilon_i)] + \bar{o}_p(1),
\end{aligned} \tag{2.11}
$$

where

$$
D_n(t) := n^{1/2}[F_n(F^{-1}(t)) - t], \qquad 0 \le t \le 1, \tag{2.12}
$$

and $\bar{o}_p(1)$ is an array of processes in $(i, \boldsymbol{\Delta})$ that converge to zero, uniformly in $(i, \boldsymbol{\Delta})$, in probability.

Next, $E\varepsilon^2 < \infty$ and the stationarity of the sequence $\{\|\mathbf{Y}_{i-1}\|\}$ implies $E\|\mathbf{Y}_{i-1}\|^2 = E\|\mathbf{Y}_0\|^2 < \infty$ and hence

$$
n^{-1/2} \max_i \|\mathbf{Y}_{i-1}\| = o_p(1). \tag{2.13}
$$

This, together with the continuity of F, implies that

$$
\sup_{i, \boldsymbol{\Delta}} |F(\varepsilon_{i\boldsymbol{\Delta}}) - F(\varepsilon_i)| = o_p(1). \tag{2.14}
$$

From (2.11), (2.14), the asymptotic uniform continuity of the D_n-process [see, e.g., Billingsley (1968), page 105] and the assumption (a)(ii) of Theorem 2.1, one readily concludes that

$$
e_{ni\boldsymbol{\Delta}} = D_n(F(\varepsilon_i)) - \boldsymbol{\Delta}'\mathbf{Y}_{i-1} f(\varepsilon_i) + \bar{o}_p(1). \tag{2.15}
$$

Now (2.10) follows from (2.15), (2.13) and the fact that $\sup\{|D_n(t)|, 0 \le t \le 1\} = O_p(1)$ and the assumption (a)(ii) that ensures $\|f\|_\infty < \infty$.

Next define

$$T(\mathbf{\Delta}) := n^{-1}\Sigma_i \mathbf{Y}_{i-1} e_{ni\mathbf{\Delta}} \varphi'(F(\varepsilon_i)), \qquad \mathbf{\Delta} \in \mathscr{R}^p.$$

Note that

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

Therefore, from the uniform continuity of φ', 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 follow from the ergodic theorem and the assumption $E\varepsilon^2 < \infty$, and from (2.10) and (2.15) one readily concludes that

$$\begin{aligned} &\sup_{\mathbf{\Delta}} \|\mathbf{M}(\mathbf{\Delta}) - \hat{\mathbf{M}} - \mathbf{T}(\mathbf{\Delta})\| \\ (2.16) \qquad &= \sup_{\mathbf{\Delta}} \Big\| n^{-1/2}\Sigma_i \mathbf{Y}_{i-1}\big\{\varphi\big(U_i + n^{-1/2}e_{ni\mathbf{\Delta}}\big) \\ &\qquad\qquad - \varphi(U_i) - n^{-1/2}e_{ni\mathbf{\Delta}}\varphi'(U_i)\big\}\Big\| = o_p(1). \end{aligned}$$

Next, we approximate $\mathbf{T}(\mathbf{\Delta})$. Again, by the ergodic theorem, $E\varepsilon = 0$ implies that $n^{-1}\Sigma_i \mathbf{Y}_{i-1}\varphi'(U_i) f(\varepsilon_i) \to 0$, a.s. Hence

$$\begin{aligned} \mathbf{T}(\mathbf{\Delta}) &= n^{-1/2}\Sigma_i \mathbf{Y}_{i-1}\left[\{F_n(\varepsilon_i) - F(\varepsilon_i)\} - n^{-1/2}\mathbf{\Delta}'\mathbf{Y}_{i-1} f(\varepsilon_i)\right]\varphi'(U_i) + \bar{o}_p(1) \\ &= \mathbf{V}_n - \mathbf{L}_n\mathbf{\Delta}' + \bar{o}_p(1), \end{aligned}$$

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

$$\begin{aligned} \mathbf{V}_n &= n^{-1/2}\Sigma_i \mathbf{Y}_{i-1}\left[F_n(\varepsilon_i) - F(\varepsilon_i)\right]\varphi'(U_i), \\ \mathbf{L}_n &= n^{-1}\Sigma_i \mathbf{Y}_{i-1}\mathbf{Y}'_{i-1} f(\varepsilon_i)\varphi'(U_i). \end{aligned}$$

Note that $E\mathbf{L}_n = E(n^{-1}\Sigma_i \mathbf{Y}_{i-1}\mathbf{Y}'_{i-1})\gamma = \mathbf{\Sigma} \cdot \gamma$, $\gamma = \int f \, d\varphi(F)$. By the ergodic theorem, $\mathbf{L}_n \to \mathbf{\Sigma}\gamma$, a.s.

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

$$\begin{aligned} U_{nj}(x) &= n^{-1/2}\Sigma_i X_{i-j}\varphi'(F(\varepsilon_i))\mathbf{I}(\varepsilon_i \le x), \\ \nu_{nj}(x) &= n^{-1/2}\Sigma_i X_{i-j}\int_{-\infty}^{x}\varphi'(F(y))\,dF(y) = n^{-1/2}\Sigma_i X_{i-j}\varphi(F(x)), \\ K_{nj}(x) &= U_{nj}(x) - \nu_{nj}(x), \qquad x \in \mathscr{R}, 1 \le j \le p. \end{aligned}$$

Observe that

$$V_{nj} = \int_{-\infty}^{\infty} [F_n - F]\, dU_{nj} = \int_{-\infty}^{\infty} [F_n - F]\, dK_{nj} = \int_{-\infty}^{\infty} [F_n - F]\, d\nu_{nj}$$

$$= -\int_0^1 \big[\tilde{Z}_j\big(F(F_n^{-1}(t))\big) - \tilde{Z}_j(t)\big]\, dt - \int_{-\infty}^{\infty} \nu_{nj}\, d[F_n - F],$$

where $\tilde{Z}_j = K_{nj}(F^{-1})$. Thus

$$\max_{1 \le j \le p} \left| V_{nj} + \int_{-\infty}^{\infty} \nu_{nj}\, d(F_n - F) \right|$$

$$\le \sup_{0 \le t \le 1,\, 1 \le j \le p} |\tilde{Z}_j\big(F(F_n^{-1}(t))\big) - \tilde{Z}_j(t)| = o_p(1),$$

by Lemma 2.1 applied with $g \equiv \varphi'$ and by using the well known fact that $\sup_{0 \le t \le 1} |(F(F_n^{-1}(t)) - t| = o_p(1)$.

Now, observe that with $\tilde{X}_j = n^{-1}\Sigma_{i=1}^n X_{i-j}$, $\hat{T} = n^{-1/2}\Sigma_i[\varphi(U_i) - \bar{\varphi}]$,

$$\int_{-\infty}^{\infty} \nu_{nj}\, d(F_n - F) = n^{-3/2}\Sigma_i X_{i-j} \Sigma_i [\varphi(U_i) - \bar{\varphi}] = \tilde{X}_j \cdot \hat{T}, \qquad 1 \le j \le p,$$

so that from the above we have $\mathbf{V}_n = -\tilde{\mathbf{X}}\hat{T} + o_p(1)$ and also

$$\text{(2.17)} \qquad \mathbf{T}(\boldsymbol{\Delta}) = -\tilde{\mathbf{X}}\hat{T} - \boldsymbol{\Sigma\Delta}\gamma + \bar{o}_p(1), \quad \text{with } \tilde{\mathbf{X}}' = \big(\tilde{X}_1, \dots, \tilde{X}_p\big).$$

From (2.16), (2.17) and direct algebra one now readily concludes that

$$\mathbf{M}(\boldsymbol{\Delta}) = n^{-1/2} \sum_{i=1}^{n} \big(\mathbf{Y}_{i-1} - \tilde{\mathbf{X}}\big)[\varphi(U_i) - \bar{\varphi}] - \boldsymbol{\Sigma\Delta}\gamma + \bar{o}_p(1).$$

Now argue as for (2.8) to conclude that $\|n^{-1/2}\Sigma_{i=1}^n(\mathbf{Y}_{i-1} - \tilde{\mathbf{X}})[\varphi(U_i) - \bar{\varphi}] - \hat{\mathbf{S}}\| = o_p(1)$, thereby completing the proof of (2.4). □

REMARK 2.1. Note that the same proof shows that under (a) and (b), for every $0 < b < \infty$,

$$\sup_{\|\boldsymbol{\Delta}\| \le b} \left\| \mathbf{S}(\boldsymbol{\rho} + n^{-1/2}\boldsymbol{\Delta}) - \mathbf{S}(\boldsymbol{\rho}) + \boldsymbol{\Delta\Sigma} \int f\, d\varphi(F) \right\| = o_p(1).$$

REMARK 2.2. Argue either as in Koul [(1985), Lemma 3.1] or as in Jaeckel (1972) to conclude that $\|n^{1/2}(\tilde{\boldsymbol{\rho}}_n - \boldsymbol{\rho})\| = O_p(1)$. Consequently by Theorem 2.1,

$$\text{(2.18)} \qquad n^{1/2}(\tilde{\boldsymbol{\rho}}_n - \boldsymbol{\rho}) = \gamma^{-1}\boldsymbol{\Sigma}^{-1}\hat{\mathbf{S}} + o_p(1), \qquad \gamma = \int f\, d\varphi(F).$$

Observe that $\hat{\mathbf{S}}$ is a vector of square integrable mean zero martingales with $E\hat{\mathbf{S}}\hat{\mathbf{S}}' = \sigma_\varphi^2\boldsymbol{\Sigma}$, $\sigma_\varphi^2 = \operatorname{Var}(\varphi(U))$. Thus, by the routine Cramér–Wold device and Corollary 3.1 of Hall and Heyde (1980), one readily obtains

$$\text{(2.19)} \quad \hat{\mathbf{S}} \Rightarrow N_p\big(\mathbf{0}, \sigma_\varphi^2\boldsymbol{\Sigma}\big); \quad \text{and hence } n^{1/2}(\tilde{\boldsymbol{\rho}}_n - \boldsymbol{\rho}) \Rightarrow N_p\big(\mathbf{0}, \gamma^{-2}\sigma_\varphi^2\boldsymbol{\Sigma}^{-1}\big).$$

3. The proposed PTRE and SRE. Let $1 \le p_1 < p$ and $p_2 = p - p_1$. Partition $\boldsymbol{\rho}$ into two subsets, namely, $\boldsymbol{\rho}' = (\boldsymbol{\rho}_1', \boldsymbol{\rho}_2')$ where $\boldsymbol{\rho}_i$ is a p_i, $i = 1, 2$, vector. We are primarily interested in the R-estimation of $\boldsymbol{\rho}_1$ when it is suspected but not sure that, H_0: $\boldsymbol{\rho}_2 = \mathbf{0}$ holds. For this we partition $\mathbf{S}' = (\mathbf{S}_1', \mathbf{S}_2')$ and $\mathbf{Y}_{i-1}' = (\mathbf{Y}_{i,1}', \mathbf{Y}_{i,2}')$, where

$$\begin{aligned} \mathbf{S}_1' &= (S_1, \dots, S_{p_1}), \quad \mathbf{S}_2' = (S_{p_1+1}, \dots, S_p), \\ \mathbf{Y}_{i,1}' &= (X_{i-1}, \dots, X_{i-p_1}), \quad \mathbf{Y}_{i,2}' = (X_{i-p_1-1}, \dots, X_{i-p}), \quad 1 \le i \le n. \end{aligned} \tag{3.1}$$

Similarly, partition the matrix $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}, \tag{3.2}$$

where $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12}$ and $\boldsymbol{\Sigma}_{22}$ are $p_1 \times p_1$, $p_1 \times p_2$ and $p_2 \times p_2$ matrices, respectively, with $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}$.

Further, we consider the estimator $\boldsymbol{\Sigma}_n$ of $\boldsymbol{\Sigma}$, where

$$\boldsymbol{\Sigma}_n = ((\sigma_{nij})) \text{ with } \sigma_{nij} = \sum_{k=1}^{n-\max(i,j)} (X_{k-i} - \bar{X}_i)(X_{k-j} - \bar{X}_j), \quad 1 \le i, j \le n.$$

Partitioning $\boldsymbol{\Sigma}_n$ similar to $\boldsymbol{\Sigma}$ we let $\boldsymbol{\Sigma}_{nrr\cdot s} = \boldsymbol{\Sigma}_{nrr} - \boldsymbol{\Sigma}_{nrs}\boldsymbol{\Sigma}_{nss}^{-1}\boldsymbol{\Sigma}_{nsr}$, $r \ne s = 1, 2$. Define $\boldsymbol{\Sigma}_{rr\cdot s}$ analogously, $r \ne s = 1, 2$ using (3.2). Note that $E\varepsilon^2 < \infty$ and the ergodic theorem implies that $n^{-1}\boldsymbol{\Sigma}_n \to \boldsymbol{\Sigma}$ and $n^{-1}\boldsymbol{\Sigma}_{nrr\cdot s} \to \boldsymbol{\Sigma}_{rr\cdot s}$, a.s. as $n \to \infty$. Since $\boldsymbol{\Sigma}^{-1}$ exists, $\boldsymbol{\Sigma}_n^{-1}$ and $\boldsymbol{\Sigma}_{nrr\cdot s}^{-1}$ also exist, for sufficiently large n.

Now, write $\tilde{\boldsymbol{\rho}}_n$ of (2.2) as $\tilde{\boldsymbol{\rho}}_n' = (\tilde{\boldsymbol{\rho}}_{1n}', \tilde{\boldsymbol{\rho}}_{2n}')$ and call $\tilde{\boldsymbol{\rho}}_{1n}$ the URE of $\boldsymbol{\rho}_1$. The restricted R-estimator (RRE) $\hat{\boldsymbol{\rho}}_{1n}$ of $\boldsymbol{\rho}_1$, is defined to be a solution of

$$\inf_{\mathbf{u} \in \mathscr{R}^{p_1}} \|\mathbf{S}_1(\mathbf{u}, \mathbf{0})\| = \|\mathbf{S}_1(\hat{\boldsymbol{\rho}}_{1n}, \mathbf{0})\|. \tag{3.3}$$

For the PTRE and SRE, we need to introduce a suitable rank test for testing the null hypothesis H_0: $\boldsymbol{\rho}_2 = \mathbf{0}$. For this define $\bar{\mathbf{S}}_2 = S_2(\hat{\boldsymbol{\rho}}_{1n}, 0)$, and let the test statistic be

$$\mathscr{L}_n = n\sigma_\varphi^{-2}\bar{\mathbf{S}}_2'\boldsymbol{\Sigma}_{n22\cdot 1}^{-1}\bar{\mathbf{S}}_2. \tag{3.4}$$

It can be shown that under H_0, $\mathscr{L}_n$ has asymptotically the central chi-square distribution with p_2 degrees of freedom (DF). Let $\chi^2_{p_2,\alpha}$ be the upper $100\alpha\%$ point of the central chi-squared distribution with p_2 DF. The preliminary test R-estimator (PTRE) of $\boldsymbol{\rho}_1$ is defined by

$$\hat{\boldsymbol{\rho}}_{1n}^{PT} = \hat{\boldsymbol{\rho}}_{1n} + I(\mathscr{L}_n \ge \chi^2_{p_2,\alpha})(\tilde{\boldsymbol{\rho}}_{1n} - \hat{\boldsymbol{\rho}}_{1n}). \tag{3.5}$$

The shrinkage R-estimator (SRE) is defined following Berger, Bock, Brown, Casella and Gleser (1977) as

$$(3.6)\qquad \hat{\boldsymbol{\rho}}_{1n}^{S} = \hat{\boldsymbol{\rho}}_{1n} + \left(\mathbf{I}_{p_1} - c d_n n^{-1}\mathscr{L}_n^{-1}\mathbf{W}^{-1}\boldsymbol{\Sigma}_{n11\cdot 2}\right)\left(\tilde{\boldsymbol{\rho}}_{1n} - \hat{\rho}_{1n}\right),$$

where $d_n = ch_{p_1}(n\mathbf{W}\boldsymbol{\Sigma}_{11\cdot 2}^{-1})$ = smallest characteristic root of $n\mathbf{W}\boldsymbol{\Sigma}_{n11\cdot 2}^{-1}$, and c is a positive shrinkage factor. If one chooses $\mathbf{W} = n^{-1}\boldsymbol{\Sigma}_{n11\cdot 2}$, then we get the simpler form of (3.6), viz.,

$$(3.7)\qquad \hat{\boldsymbol{\rho}}_{1n}^{S} = \hat{\boldsymbol{\rho}}_{1n} + \left(1 - c\mathscr{L}_n^{-1}\right)\left(\tilde{\boldsymbol{\rho}}_{1n} - \hat{\boldsymbol{\rho}}_{1n}\right), \qquad 0 < c < 2(p_2 - 2).$$

The formula (3.7) is justified when we evaluate the risk using Mahalanobis distance

$$(3.8)\qquad L(\boldsymbol{\rho}_{1n}^{*}; \boldsymbol{\rho}_1) = n(\boldsymbol{\rho}_{1n}^{*} - \boldsymbol{\rho}_1)'\boldsymbol{\Sigma}_{n11\cdot 2}(\boldsymbol{\rho}_{1n}^{*} - \boldsymbol{\rho}_1)\gamma^2\sigma_{\psi}^{-2}.$$

A further improvement over (3.7) is obtained by defining the positive-rule shrinkage R-estimator (PRSRE) due to Sclove, Morris and Radhakrishnan (1972)

$$(3.9)\qquad \hat{\boldsymbol{\rho}}_{1n}^{S+} = \hat{\boldsymbol{\rho}}_{1n} + \left(1 - c\mathscr{L}_n^{-1}\right)I(\mathscr{L}_n > c)\left(\tilde{\boldsymbol{\rho}}_{1n} - \hat{\boldsymbol{\rho}}_{1n}\right).$$

Note that $\hat{\boldsymbol{\rho}}_{1n}^{PT}$ and $\hat{\boldsymbol{\rho}}_{1n}^{S+}$ are convex combinations of $\tilde{\boldsymbol{\rho}}_{1n}$ and $\hat{\boldsymbol{\rho}}_{1n}$ while $\hat{\boldsymbol{\rho}}_{1n}^{S}$ is not. The asymptotic relative optimality of these estimators is discussed in Section 4 in terms of ADR.

4. ADR and ADRE. In order to obtain the expressions for the ADR of the above estimators, we follow Saleh and Sen (1986) and Sen and Saleh (1987) and use the sequence of alternatives $K_{(n)}$: $\boldsymbol{\rho}_{2(n)} = n^{-1/2}\boldsymbol{\xi}$, $\boldsymbol{\xi} \in R^{p_2}$. In addition, we need the following assumption:

(a)*(ii) The error d.f. F has an absolutely continuous density f with its a.e. derivative f' satisfying $0 < I(f) = \int (f'/f)^2\, dF < \infty$.

In what follows, $H_m(\cdot; \delta)$ stands for the cdf of a noncentral chi-square r.v. with m DF and the noncentrality parameter δ and $E(\chi_m^{-2r}(\delta)) = \int_0^\infty x^{-r}\, dH_m(x, \delta)$, $r \geq 1$. Finally, for any estimator $\boldsymbol{\rho}_{1n}^{*}$ of $\boldsymbol{\rho}_1$, for which $\sqrt{n}\,(\boldsymbol{\rho}_{1n}^{*} - \boldsymbol{\rho}_1)$ converges in distribution under $K_{(n)}$ to a p_1-variate-normal with mean $\mathbf{0}$ and a positive definite dispersion matrix $\mathbf{V}^*$, define the ADR with respect to the loss $n(\boldsymbol{\rho}_{1n}^{*} - \boldsymbol{\rho}_1)'\mathbf{W}(\boldsymbol{\rho}_{1n}^{*} - \boldsymbol{\rho}_1)$ to be

$$(4.1)\qquad R(\boldsymbol{\rho}_{1n}^{*}: \mathbf{W}) \equiv \operatorname{tr}(\mathbf{W}\mathbf{V}^*),$$

where $\mathbf{W}$ is a positive definite matrix and tr stands for the trace operator.

Akritas and Johnson (1982) have shown that under (a)(i), (b) and (a)*(ii), $K_{(n)}$ is contiguous to H_0. Using this and an argument similar to the one appearing in the proof of Theorem 3.2 of Sen and Saleh (1987), one obtains the following.

THEOREM 4.1. *Under* (1.1), (1.2), (a)(i), (b) *and* (a)*(ii) *the ADR's of the four estimators are given by*

$$R(\tilde{\boldsymbol{\rho}}_{1n}:\mathbf{W}) = d^2\operatorname{tr}(\mathbf{W}\boldsymbol{\Sigma}_{11\cdot 2}^{-1}), \tag{4.2}$$

$$R(\hat{\boldsymbol{\rho}}_{1n}:\mathbf{W}) = d^2\operatorname{tr}(\mathbf{W}\boldsymbol{\Sigma}_{11}^{-1}) + \boldsymbol{\xi}'\mathbf{M}\boldsymbol{\xi}, \tag{4.3}$$

$$\begin{aligned} R(\hat{\boldsymbol{\rho}}_{1n}^{PT}:\mathbf{W}) = d^2\Big\{&\operatorname{tr}(\mathbf{W}\boldsymbol{\Sigma}_{11\cdot 2}^{-1})\big[1 - H(\chi^2_{p_2,\alpha};\delta)\big] \\ &+\operatorname{tr}(\mathbf{W}\boldsymbol{\Sigma}_{11}^{-1})H_{p_2+2}(\chi^2_{p_2,\alpha};\delta)\Big\} \\ &+(\boldsymbol{\xi}'\mathbf{M}\boldsymbol{\xi})\big[2H_{p_2+2}(\chi^2_{p_2,\alpha};\delta) - H_{p_2+4}(\chi^2_{p_2,\alpha};\delta)\big], \end{aligned} \tag{4.4}$$

$$\begin{aligned} R(\hat{\boldsymbol{\rho}}_{1n}^{S}:\mathbf{W}) = d^2\Big\{&\operatorname{tr}(\mathbf{W}\boldsymbol{\Sigma}_{11\cdot 2}^{-1}) - c\operatorname{tr}(\mathbf{M}\boldsymbol{\Sigma}_{22\cdot 1}^{-1})\big[2E(\chi^{-2}_{p_2+2}(\delta)) \\ &-cE(\chi^{-4}_{p_2+2}(\delta))\big]\Big\} \\ &+ c(c+4)(\boldsymbol{\xi}'\mathbf{M}\boldsymbol{\xi})E(\chi^{-4}_{p_2+4}(\delta)), \end{aligned} \tag{4.5}$$

where

$$\mathbf{M} = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{W}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}. \tag{4.6}$$

We shall now discuss the *asymptotic distribution risk efficiency* (*ADRE*) results.

First note that the risk of $\tilde{\boldsymbol{\rho}}_{1n}$ does not depend on $\boldsymbol{\xi}$ and if we consider the Mahalanobis distance (loss) then $\mathbf{W} = d^{-2}\boldsymbol{\Sigma}_{11\cdot 2}$, and $R(\tilde{\boldsymbol{\rho}}_{1n}:\mathbf{W})$ reduces to p_1. Now, let

$$\mathbf{M}^* = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{11\cdot 2}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \quad\text{and}\quad \mathbf{M}^0 = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}. \tag{4.7}$$

THEOREM 4.2. *Under the conditions of Theorem* 4.1, *we have*:

(4.8) (i) $R(\tilde{\boldsymbol{\rho}}_{1n}:\mathbf{W}) \gtreqless R(\hat{\boldsymbol{\rho}}_{1n}:\mathbf{W})$ *according as* $d^{-2}(\boldsymbol{\xi}'\mathbf{M}^*\boldsymbol{\xi}) \lesseqgtr \operatorname{tr}(\mathbf{M}^0)$.

(ii) $R(\tilde{\boldsymbol{\rho}}_{1n}:\mathbf{W}) \gtreqless R(\hat{\boldsymbol{\rho}}_{1n}^{PT};\mathbf{W})$ *according as*

$$(\boldsymbol{\xi}'\mathbf{M}^*\boldsymbol{\xi}) \lesseqgtr \frac{d^2\operatorname{tr}(M^0)H_{p_2+2}(\chi^2_{p_2,\alpha};\delta)}{2H_{p_2+2}(\chi^2_{p_2,\alpha};\delta) - H_{p_2+4}(\chi^2_{p_{21},\alpha};\delta)}.$$

(iii) *If* $c = (p_2 - 2)$ *and* $\mathbf{W}$ *is an arbitrary positive definite matrix, then*

$$R(\hat{\boldsymbol{\rho}}_{1n}^{S};\mathbf{W}) \le R(\tilde{\boldsymbol{\rho}}_{1n},\mathbf{W}) \qquad \forall\,\mathbf{W}\in\mathscr{W} \text{ and } \boldsymbol{\xi}\in\mathscr{R}^{p_2},$$

where

$$\mathscr{W} = \left\{\mathbf{W}\text{ p.d.}: \frac{\operatorname{tr}(\mathbf{M}\boldsymbol{\Sigma}_{22\cdot 1}^{-1})}{ch_{\max}(\mathbf{M}\boldsymbol{\Sigma}_{22\cdot 1}^{-1})} \ge \frac{p_2+2}{2}\right\} \tag{4.9}$$

with $\mathbf{M}$ *given by* (4.6) *and* $ch_{\max}(\mathbf{A})$ = *maximum of the characteristic roots of* $\mathbf{A}$.

The proof of (i) and (ii) is straightforward. We prove (iii). Use the identity

$$E\big(\chi_{p_2+2}^{-2}(\delta)\big) - (p_2 - 2)E\big(\chi_{p_2+2}^{-4}(\delta)\big) = \delta E\big(\chi_{p_2+4}^{-4}(\delta)\big) \tag{4.10}$$

in Theorem 4.1 to obtain

$$\begin{aligned} R(\hat{\boldsymbol{\rho}}_{1n}^S; \mathbf{W}) = R(\tilde{\boldsymbol{\rho}}_{1n}: \mathbf{W}) & \\ -(p_2 - 2)d^2 \operatorname{tr}\big(\mathbf{M}\boldsymbol{\Sigma}_{22\cdot 1}^{-1}\big) & \\ \times\bigg\{(p_2 - 2)E\big(\chi_{p_2+2}^{-4}(\delta)\big) & \\ +\left[1 - \frac{(p_2+2)(\boldsymbol{\xi}'\mathbf{M}\boldsymbol{\xi})\sigma_\varphi^{-2}\gamma^2}{2\operatorname{tr}\big(\mathbf{M}\boldsymbol{\Sigma}_{22\cdot 1}^{-1}\big)\delta}\right] 2\delta E\big(\chi_{p_2+4}^{-4}(\delta)\big)\bigg\}. & \end{aligned} \tag{4.11}$$

From (4.11), it follows that $R(\hat{\boldsymbol{\rho}}_{1n}^S; \mathbf{W}) \le R(\tilde{\boldsymbol{\rho}}_{1n}; \mathbf{W})$ for all $\boldsymbol{\xi} \in R^{p_2}$ if $\operatorname{tr}(\mathbf{M}\boldsymbol{\Sigma}_{22\cdot 1}^{-1})/ch_{\max}(\mathbf{M}\boldsymbol{\Sigma}_{22\cdot 1}^{-1}) \ge (p_2 + 2)/2$, that is, if $\mathbf{W} \in \mathscr{W}$.

In particular, if $\mathbf{W} = d^{-2}\boldsymbol{\Sigma}_{11\cdot 2}$, then the risk expression (4.11) is the same with $\mathbf{M}$ replaced by $\mathbf{M}^*$ of (4.7). In this case the condition (4.9) reduces to

$$\mathscr{W} = \left\{\mathbf{W}\ \text{p.d.}: \frac{\operatorname{tr}(\mathbf{M}^0)}{Ch_{\max}(\mathbf{M}^0)} \ge \frac{p_2+2}{2}\right\} \quad \text{since } \operatorname{tr}\big(\mathbf{M}^*\boldsymbol{\Sigma}_{22\cdot 1}^{-1}\big) = \operatorname{tr}(\mathbf{M}^0). \tag{4.12}$$

We note that $\operatorname{tr}(\mathbf{M}^0) = \operatorname{tr}(I_{p_1} - \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}) \le \min(p_1, p_2) = p^* > 3$ since $p_2 \ge 3$. Also, if the rank of $\mathbf{M}^0 \le 2$, then (4.12) does not hold with $\mathbf{M} = \mathbf{M}^*$. However, if $p_2 \ge 3$ and $2\operatorname{tr}(\mathbf{M}^0) \ge Ch_{\max}(\mathbf{M}^0)$, then (4.12) holds. Thus, for the Mahalanobis loss, (iii) holds. Fortunately, the condition (4.12) is verifiable for a given $\boldsymbol{\Sigma}_n$ for any AR(p) model. For an arbitrary $\mathbf{W}$ not belonging to $\mathscr{W}$, (iii) may not hold for all $\mathbf{M}$. In that case one may use the SRE given in (3.7).

Now, we assume c is an arbitrary positive integer and that $\mathbf{W} = d^{-2}\boldsymbol{\Sigma}_{11\cdot 2}$. The following theorem gives a sufficient condition on c for which SRE dominates URE under ADR.

THEOREM 4.3. *A sufficient condition for the asymptotic dominance of* SRE *over* URE [*i.e.*, $R(\tilde{\boldsymbol{\rho}}_{1n}: \mathbf{W}) \ge R(\hat{\boldsymbol{\rho}}_{1n}^S: \mathbf{W})\ \forall\ \boldsymbol{\xi} \in R^{p_2}$] *is that the shrinkage factor* c *is positive and it satisfies the inequality*

$$2E\big(\chi_{p_2+2}^{-2}(\delta)\big) - cE\big(\chi_{p_2+2}^{-4}(\delta)\big) - (c+4)h\delta E\big(\chi_{p_2+2}^{-4}(\delta)\big) \ge 0 \qquad \forall\ \delta, \tag{4.13}$$

which in turn, requires that

$$p_2 \ge 3, \qquad 0 < c < 2(p_2 - 2) \quad \textit{and} \quad h(c+4) \le 2,$$

where $h = ch_{\max}(\mathbf{M}^0)/\operatorname{tr}(\mathbf{M}^0)$ *so that* $0 < h \le 1$.

PROOF. Details of the proof are similar to Sen and Saleh (1987) and are left out. □

Similarly, one has the following theorem.

THEOREM 4.4. *Under the conditions of Theorem* 4.1, *the PTRE fails to dominate SRE. Also, if for an* $\alpha(0 < \alpha < 1)$, *the level of significance of the PT, we have*

$$H_{p_2+2}\left(\chi^2_{p_2,\alpha};0\right) \geq q\{2(p_2-2)-q\}/p_2(p_2-2), \tag{4.14}$$
$$q = (p_2-2) \wedge (2/h-4),$$

then, the SRE fails to dominate the PTRE.

However, one may verify that under H_0 the ordering of the ADR of the four estimators is as follows:

$$R(\hat{\boldsymbol{\rho}}_{1n};W) \leq R\left(\hat{\boldsymbol{\rho}}_{1n}^{PT};\mathbf{W}\right) \leq R\left(\hat{\boldsymbol{\rho}}_{1n}^{S};\mathbf{W}\right) \leq R(\tilde{\boldsymbol{\rho}}_{1n};\mathbf{W}). \tag{4.15}$$

The general behavior of these risks is similar to those discussed in Sen and Saleh (1987) in connection with linear regression models.

5. Concluding remarks. It follows from the discussions of Section 4 that both PTRE and SRE are robust from the risk-efficiency point of view. Of the two SRE possesses asymptotic minimax character in terms of ADR characteristics, while PTRE does not have this character. However, PTRE is attractive for any $p_2 \geq 1$, while SRE needs $p_2 \geq 3$ in such a case, that is, for $p_2 \leq 2$, PTRE is the only course left. However, for $p_2 \geq 3$, SRE in general is the choice while PTRE needs the appropriate choice of α, the level of significance of the PT. However, one can develop a maximum rule to select appropriate α to get the optimum PTRE for $\boldsymbol{\rho}_1$ but still SRE is preferable since, the interval $(0, \delta_0)$ in which PTRE is at best not known generally.

APPENDIX

PROOF OF LEMMA 2.1. The proof of Lemma 2.1 will follow at the end of the following two lemmas.

LEMMA A.1. *In addition to* (1.1), (1.2) *and* (a)(i) *of Theorem* 2.1, *assume that the following hold*:

(b1) *The d.f. F is continuous and strictly increasing.*
(b2) *The function g on* $[0,1]$ *to* $\mathscr{R}$ *satisfies* $\int|g(t)|^4\,dt < \infty$.

Then the following holds:
For any $0 \leq u < v < w \leq 1$ *and for all* $1 \leq j \leq p$,

$$\limsup_{n\to\infty} E\{Z_j(v)-Z_j(u)\}^2\{Z_j(w)-Z_j(v)\}^2 \leq Cm_1m_2, \tag{1}$$

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

PROOF. Since u, v, w are fixed, we shall suppress these entities in the notation. Further, to simplify writing let $x = F^{-1}(u)$, $y = F^{-1}(v)$ and $z = F^{-1}(w)$ and define

$$\begin{aligned} &p_1 = G(v) - G(u), p_2 = G(w) - G(v), \qquad q_j = 1 - p_j, j = 1, 2, \\ &\alpha_i = g(F(\varepsilon_i))\mathbf{I}(x < \varepsilon_i \le y) - p_1, \qquad \beta_i = g(F(\varepsilon_i))\mathbf{I}(y < \varepsilon_i < z) - p_2. \end{aligned} \tag{2}$$

Then

$$\{Z_j(v) - Z_j(u)\}^2\{Z_j(w) - Z_j(v)\}^2 = n^{-2}(\Sigma_i X_{i-j}\alpha_i)^2 \cdot (\Sigma_l X_{l-j}\beta_l)^2. \tag{3}$$

Let $\mathscr{F}_i := \sigma\text{-field }\{Y_0, \varepsilon_1, \dots, \varepsilon_i\}$, $i \ge 1$; $\mathscr{F}_0 := \sigma\text{-field }\{Y_0\}$. In carrying out the computations that follow we have repeatedly used the following facts: α_i, β_l are centered; α_i, β_i are $\mathscr{F}_{k-1}$ measurable for all $i < k$ and X_{i-j} is $\mathscr{F}_{i-1}$ measurable and independent of ε_i, $i \ge 1$. Thus, one has

$$E\alpha_i = 0 = E\beta_k \quad \text{for all } i, k.$$

$$\begin{aligned} EX_{i-j}X_{k-j}\alpha_i\beta_k &= E\big[X_{i-j}X_{k-j}\alpha_i E(\beta_k|\mathscr{F}_{k-1})\big] \\ &= E\big[X_{i-j}X_{k-j}\alpha_i\big]E(\beta_k) = 0, \qquad i < k. \end{aligned} \tag{4}$$

$$EX_{i-j}X_{k-j}X_{l-j}^2\alpha_i\alpha_k\beta_l^2 = E\big[X_{i-j}X_{k-j}X_{l-j}^2\beta_l^2\alpha_i E(\alpha_k|\mathscr{F}_{k-1})\big] = 0, \qquad i, l < k.$$

Using facts like these one can write

$$\begin{aligned} &E(\Sigma_i X_{i-j}\alpha_i)^2(\Sigma_l X_{l-j}\beta_l)^2 \\ &\quad = \Sigma_i EX_{i-j}^4\alpha_i^2\beta_i^2 + \sum_{i \ne l}\Sigma EX_{i-j}^2 X_{l-j}^2\alpha_i^2\beta_l^2 \\ &\qquad + 4\sum_{i<k}\Sigma EX_{i-j}^2X_{k-j}^2\alpha_i\beta_i\alpha_k\beta_k \\ &\qquad + 2\sum_{i<k<l}\Sigma\Sigma EX_{i-j}X_{k-j}X_{l-j}^2(\alpha_i\alpha_k\beta_l^2 + \beta_i\beta_k\alpha_l^2) \\ &\qquad + 4\Big(\sum_{i<l<k}\Sigma\Sigma + \sum_{l<i<k}\Sigma\Sigma\Big)\big(EX_{i-j}X_{l-j}X_{k-j}^2\alpha_i\beta_l\alpha_k\beta_k\big) \\ &\quad = T_1 + T_2 + 4T_3 + 2(T_4 + T_5) + 4(T_6 + T_7), \quad \text{say.} \end{aligned} \tag{5}$$

We shall now show that $n^{-2}T_j \to 0$, for $j = 1, 4, 5, 6, 7$, and that $\limsup n^{-2}(T_2 + 4T_3) \le Cm_1m_2$. The basic idea of the proof is to exploit the iterative nature of the process.

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 we shall give the details only for $n^{-2}T_7 \to 0$. The assumption that $\{\varepsilon_i\}$ are i.i.d. mean 0 r.v.'s will be used repeatedly and without mention. Observe that

$$E\big(X_{i-j}X_{l-j}X_{k-j}^2\alpha_i\beta_l\alpha_k\beta_k\big) = EX_{i-j}X_{l-j}X_{k-j}^2\alpha_i\beta_l E(\alpha_k\beta_k|\mathscr{F}_{k-1}),$$

$$EX_{i-j}X_{l-j}X_{k-j}^2\alpha_i \equiv 0, \qquad l < i, k - j \le i - 1$$

and

$$E(\alpha_k \beta_k | \mathscr{F}_{k-1}) = -p_1 p_2, \qquad k \geq 1.$$

Therefore,

$$\begin{aligned} n^{-2} T_7 = n^{-2} p_1 p_2 \bigg\{ & \sum_{l<i} \Sigma E X_{i-j} X_{l-j} X_i^2 \alpha_i \beta_l \\ (6) \qquad & + \sum_{l<i} \Sigma \sum_{k-j \geq i+1} E X_{i-j} X_{l-j} X_{k-j}^2 \alpha_i \beta_l \bigg\} \\ = & -p_1 p_2 n^{-2} \{ T_{71} + T_{72} \}, \quad \text{say}. \end{aligned}$$

Now, it is well known that the AR(p) process given by (1.1) and (1.2) satisfy

$$X_i = \sum_{k=-\infty}^{i} \delta_{i-k} \varepsilon_k, \qquad i \geq 1, \tag{7}$$

where $\{\delta_k\}$ are real numbers satisfying $\delta_0 = 1$, $d_1 < \infty$, with $d_r = \sum_{k=0}^{\infty} |\delta_k|^r$, $r \geq 1$. See, for example, Ash and Gardner [(1976), Sections 2.3.4–2.3.7] and Lukacs [(1968), Theorem 4.2.1]. Note that $\sup_k |\delta_k| \leq d_1$ and hence $d_1 < \infty$ implies that

$$d_r \leq d_1^r < \infty, \quad \text{for all } r \geq 1. \tag{8}$$

Next, define

$$A_{m,n} = \sum_{r=-\infty}^{n} \delta_{m-r} \varepsilon_r, \qquad H_{m,k}^n = \sum_{r=k}^{n} \delta_{m-r} \varepsilon_r, \qquad 0 \leq n \leq m < \infty, k \leq n.$$

(9) $a_r = E(\alpha \varepsilon^r)$, $b_r = E(\beta \varepsilon^r)$, $1 \leq r \leq 4$, where α, β are copies of α_1, β_1.

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

Observe that

$$H_{m,k}^n = A_{m,n} - A_{m,k-1}, \qquad ki \leq n \leq m,$$

$$X_i = A_{i,i} = A_{i,i-j} + H_{i,i-j+1}^{i-1} + \varepsilon_i = A_{i,i-1} + \varepsilon_i, \quad \text{for all } i, \tag{10}$$

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

Moreover, (8) implies that for all $n \leq m < \infty$,

$$EA_{m,n}^2 = \sum_{r=m-n}^{\infty} \delta_r^2 \mu_2 \leq \mu_2 d_1^2 < \infty, \tag{11}$$

$$EA_{m,n}^4 = \sum_{k=m-n}^{\infty} \delta_k^4 \mu_4 + 3 \sum_{k=m-n}^{\infty} \sum_{\substack{l=m-n \\ l \neq k}}^{\infty} \delta_k^2 \delta_l^2 \mu_2 \leq (\mu_4 + \mu_2) d_1^4 < \infty, \tag{12}$$

where $\mu_r = E\varepsilon^r$, $1 \leq r \leq 4$.

From (10), we obtain $E\{X_i^2\alpha_i|\mathscr{F}_{i-1}\} = 2A_{i,\iota-1}a_1 + a_2$, for all i. Use this and argue as for (6) to obtain

$$T_{71} = \Sigma_l EX_{l-j}\beta_l\{2a_1L_{l+j} + a_2X_l\} + \sum_{l<i;\, i-j\geq l+1} \Sigma EX_{l-j}\beta_l\{2a_1L_{i+j} + a_2X_{i-j}\}$$

$$= T_{711} + T_{712}, \quad \text{say, where } L_i = X_{i-j}A_{i,i-1}.$$

Now, the Cauchy–Schwarz (C–S) inequality, the stationarity of the process $\{X_i\}$ and (11) imply that $E|X_{l-j}\beta_lL_{l+j}| \leq \{E(X_{l-j}\beta_l)^2EL_{l+j}^2\}^{1/2} \leq D_4C_4 < \infty$, $E|X_{l-j}\beta_lX_l| \leq D_4C_4 < \infty$, where D_k is a constant depending on the kth moment of $g(U)$ and C_k is a constant depending on the kth moment of ε and d_k, $1 \leq k \leq 4$. These facts imply that

$$n^{-2}|T_{711}| = O(n^{-1}) = o(1). \tag{13}$$

Next, to handle T_{712}, use (11) to get that

$$L_i = \left\{A_{i-j,l-1} + \delta_{i-j-l}\varepsilon_l + H_{i-j,l+1}^{i-j}\right\}\left\{A_{i,l-1} + \delta_{i-l}\varepsilon_l + H_{i,l+1}^{i-1}\right\},$$
$$i-j \geq l+1.$$

The above type of conditioning argument yields that for $i - j \geq l + 1$,

$$EX_{l-j}\beta_lL_i = EX_{l-j}\left\{\left[\delta_{i-l}A_{i-j,l-1} + \delta_{i-l-j}A_{i,l-1}\right]b_1 + \delta_{i-j-l}\delta_{i-l}b_2\right\},$$

where b_r as in (9), $r = 1, 2, 3, 4$. Similarly,

$$EX_{l-j}\beta_lX_{i-j} = EX_{l-j}\left\{\delta_{i-j-l}^2b_2 + 2\delta_{i-j-l}A_{i-j,l-1}b_1\right\}, \qquad i-j \geq l+1.$$

Use these facts together with (11) and an argument like the one that led to (13) to conclude that $n^{-2}|T_{712}| = O(n^{-2})$. This and (13) yield that

$$n^{-2}|T_{71}| = O(n^{-1}) = o(1). \tag{14}$$

Now we turn to T_{72}. Use (10) to write $X_{k-j} = A_{k-j,i-1} + \delta_{k-j-i}\varepsilon_i + H_{k-j,i+1}^{k-j}$, $k - j \geq i + 1$, and use arguments like those above to get that $E\{X_{k-j}^2\alpha_i|\mathscr{F}_{i-1}\} = \delta_{k-j-i}^2a_2 + 2A_{k-j,i-1}\delta_{k-j-i}a_1$, $k - j \geq i + 1$, so that

$$\begin{aligned} T_{72} &= \sum\sum\sum_{l<i<k;\, k-j\geq i+1} EX_{l-j}\beta_lX_{\iota-j}\left\{\delta_{k-j-i}^2a_2 + 2A_{k-j,i-1}\delta_{k-j-i}a_1\right\} \\ &= a_2T_{721} + 2a_1T_{722}, \quad \text{say.} \end{aligned} \tag{15}$$

Arguing as above obtains

$$EX_{l-j}\beta_lX_{i-j} = EX_{l-j}\beta_l\left\{A_{i-j,l-1} + \delta_{i-j-l}\varepsilon_l + H_{i-j,l+1}^{i-j}\right\} = 0, \qquad i-j \geq l+1.$$

Thus,

$$|T_{721}| \leq a_2d_2\sum_{l<i}\sum|EX_{l-j}\beta_lX_{i-j}| = 0.$$

Similar arguments show that $|n^{-2}T_{722}| = o(1)$ thereby completing the proof of $n^{-2}|T_{72}| = o(1)$. This together with (14) shows that $n^{-2}|T_7| = o(1)$.

We shall now consider T_2: Rewrite

$$T_2 = \left(\sum\sum_{i<l} + \sum\sum_{i>l}\right)\left(EX_{i-j}^2 X_{l-j}^2 \alpha_i^2 \beta_l^2\right) = T_{21} + T_{22}, \quad \text{say}.$$

Again, by a conditioning argument,

$$\begin{aligned} T_{21} &= \sigma_2^2 \cdot \sum\sum_{i<l} EX_{i-j}^2 X_{l-j}^2 \alpha_i^2 \\ &= \sigma_2^2 \cdot \left(\sum\sum_{i<l;\, l-i\le j} + \sum\sum_{i<l;\, l-i\ge j+1}\right)\left(EX_{i-j}^2 X_{l-j}^2 \alpha_i^2\right) \qquad (16) \\ &= \sigma_2^2 \cdot \{T_{211} + T_{212}\}, \quad \text{say}. \end{aligned}$$

Again, the C–S inequality, the stationarity of the process $\{X_t\}$, the assumptions (a)(i) and (b2) imply that $0 < T_{211} \le j \cdot n \cdot EX_0^4 = O(n)$, by (8) and (10), so that

$$n^{-2}T_{211} = O(n^{-1}) = o(1). \qquad (17)$$

Next, arguing as for (15) we obtain

$$\begin{aligned} T_{212} &= \sum_{i<l;\, l-i\ge j+1} \sum EX_{i-j}^2 X_{l-j}^2 \alpha_i^2 \\ &= \sum_{i<l;\, l-i\ge j+1} EX_{i-j}^2 \alpha_i^2 \left\{A_{l-j,i-1} + \delta_{l-j-i}\varepsilon_i + H_{l-j,i+1}^{l-j}\right\}^2 \\ &= \sigma_1^2 \cdot \sum_{i<l;\, l-i\ge j+1} \sum EX_{i-j}^2 \left\{A_{l-j,i-1}^2 + \sum_{m=i+1}^{l-j} \delta_{l-j-m}^2 E\varepsilon^2\right\} \qquad (18) \\ &\quad + 2c \sum_{i<l;\, l-i\ge j+1} \sum EX_{i-j}^2 A_{l-j,i-1}\delta_{l-j-i} \\ &= \sigma_1^2 \cdot B_1 + 2c \cdot B_2, \quad \text{say, where } c = E(\varepsilon\alpha)^2. \end{aligned}$$

The C–S inequality and (12) yield that

$$n^{-2}B_1 \le C < \infty, \quad \text{for all } n \ge 1, \qquad (19)$$

where C is a constant depending on μ_4 and d_1. A similar argument shows that $|B_2| = O(n^{-1})$. This together with (19) and (18) yield that

$$\limsup n^{-2}|T_{212}| \le C\sigma_1^2. \qquad (20)$$

Hence, from (18)–(20) one readily sees that $\limsup n^{-2}|T_{21}| \le C\sigma_1^2 \cdot \sigma_2^2 \le Cm_1 m_2$, by (10).

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

$$\limsup n^{-2}|T_2| \le Cm_1m_2, \quad \text{where } C \text{ is as in (19)}. \tag{21}$$

Finally consider $n^{-2}T_3$. An exactly similar argument gives the inequality

$$\limsup|n^{-2}T_3| \le C(p_1p_2)^2 \le Cm_1m_2, \quad \text{where } C \text{ is as in (19)},$$

because $p_1^2 = \{\int_u^v g(t)\,dt\}^2 \le (v-u)\int_u^v g^2(t)\,dt \le m_1$ and because a similar inequality holds for p_2. The proof is now terminated. □

LEMMA A.2. *In addition to* (1.1) *and* (1.2) *assume that* $E\varepsilon = 0$, $E\varepsilon^2 < \infty$ *and* $Eg^2(U) < \infty$. *Then the finite dimensional distribution of* Z_j, *for every* $1 \le j \le p$, *converges weakly to that of* $\{E(X_0)^2\}^{1/2}B(\cdot)$, *where B is the Brownian motion in* $C[0,1]$ *with the covariance function* $G(min(u,v)) - G(u)G(v)$, $0 \le u, v \le 1$.

PROOF. The proof uses Corollary 3.1 of Hall and Heyde [(1980), page 58] and the Cramér–Wold device. Details are left out for the sake of brevity. □

THE PROOF OF LEMMA 2.1. Use the indirect method of Billingsley [(1968), Section 12], the right continuity of $\{Z_j\}$ and Lemmas A.1 and A.2 to conclude Lemma 2.1. For the nature of details see Billingsley [(1986), page 107]. The details are left out for the sake of brevity. □

Acknowledgments. The authors wish to thank the referees for their careful reading and comments.

REFERENCES

ASH, R. B. and GARDNER, M. F. (1975). *Topics in Stochastic Processes*. Academic, New York.

AKRITAS, M. G. and JOHNSON, R. A. (1982). Efficiencies of tests and estimators of p-order autoregressive processes when the error distribution is nonnormal. *Ann. Inst. Statist. Math.* **34** 579–589.

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

BERGER, J. O., BOCK, M. E., BROWN, L. D., CASELLA, G. and GLESER, L. (1977). Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Ann. Statist.* **5** 763–771.

BILLINGSLEY, P. (1986). *Convergence of Probability Measures*. Wiley, New York.

HALL, P. and HEYDE, C. (1980). *Martingale Limit Theory and Its Applications*. Academic, New York.

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic, New York.

HALLIN, M. and PURI, M. L. (1988). Optimal rank-based procedures for time series analysis: Testing an ARMA model against other ARMA models. *Ann. Statist.* **16** 402–432.

JAECKEL, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.* **43** 1449–1458.

KOUL, H. L. (1985). Minimum distance estimation in multiple linear regression. *Sankhyā Ser. A.* **47** 57–74.

KOUL, H. L. (1991). A weak convergence result useful in robust autoregression. *J. Statist. Plann. Inference* **29** 291–308.

LUKACS, E. (1968). *Stochastic Convergence*. Heath, Lexington, MA.

SALEH, A. K. MD. E. and SEN, P. K. (1986). On shrinkage R-estimation in a multiple regression model. *Comm. Statist. Theory Methods* **15** 2222–2244.

SCLOVE, S. L., MORRIS, C. and RADHAKRISHNAN (1972). Non-optimality of preliminary test estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* **43** 1481–1490.

SEN, P. K. and SALEH, A. K. MD. E. (1987). On preliminary test and shrinkage M-estimation in linear models. *Ann. Statist.* **15** 1580–1591.

DEPARTMENT OF STATISTICS
AND PROBABILITY
WELLS HALL
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824-1027

DEPARTMENT OF MATHEMATICS
AND STATISTICS
DUNTON TOWERS
CARLETON UNIVERSITY
OTTAWA
CANADA K1S 5B6