Ann. Inst. Statist. Math.
34 (1982), Part A, 579-589

# EFFICIENCIES OF TESTS AND ESTIMATORS FOR $p$-ORDER AUTOREGRESSIVE PROCESSES WHEN THE ERROR DISTRIBUTION IS NONNORMAL*

MICHAEL G. AKRITAS AND RICHARD A. JOHNSON

(Received Mar. 30, 1981; revised Feb. 25, 1982)

## Summary

We consider $p$th order autoregressive time series where the shocks need not be normal. By employing the concept of contiguity, we obtain the asymptotic power for tests of hypothesis concerning the autoregressive parameters. Our approach allows consideration of the double exponential and other thicker-tailed distributions for the shocks. We derive a new result in the contiguity framework that leads directly to an expression for the Pitman efficiencies of tests as well as estimators.

The numerical values of the efficiencies suggest a lack of robustness for the normal theory least squares estimators when the shock distribution is thick tailed or an outlier prone mixed normal. An important alternative test statistic is proposed that competes with the normal theory tests.

## 1. Introduction

We consider the $p$th order autoregressive (AR($p$)) model

$$X_t - \theta_1 X_{t-1} - \cdots - \theta_p X_{t-p} = E_t\,, \qquad t=1,2,\cdots \tag{1.1}$$

where the errors $E_t$'s are independent and identically distributed (iid) random variables, $E_t$ being independent of $X_{t-k}$, $k \geqq 1$, with $\mathrm{E}(E_t)=0$ and $\mathrm{Var}(E_t)=\sigma^2$, and we examine the performance of commonly employed tests and estimators under various classes of error distributions. By employing contiguity techniques we derive the limiting distributions of statistics under local alternatives. These lead directly to expressions

* This research was supported by the Office of Naval Research under Grant No. N00014-78-C-0722 and by the Army Research Office.

AMS 1980 subject classifications: Primary 62M10; Secondary 62F05, 62F12.

Key words and phrases: Robustness of autoregressive parameter estimates, efficiencies, nonnormal shocks.

for both Pitman efficiencies of tests and asymptotic efficiencies of point estimators. With the exception of the very interesting and quite general work of Gastwirth and Rubin [4], [5] and the recent work of Martin [10] and Denby and Martin [3] little attention has been given to distributional robustness in time series models.

Motivated by the typical nonparametric approach we consider classes of thick tailed distributions as alternatives to normal shocks. Besides the double exponential, whose treatment requires our weak regularity conditions, we consider $t$-distributions as well as mixtures of normals as special cases in our numerical calculations. The latter model reflects the possibility of outliers. We avoid the Cauchy and other infinite variance distributions because of the well known results on the inefficiency of least squares estimators (see Kanter and Steiger [9]). Generally our numerical results in Section 4 establish low asymptotic efficiencies for normal theory tests and estimators when the error distributions are of the thick tailed type.

When the $E_t$'s are distributed as $(2\beta)^{-1}\exp[-|x|/\beta]$, the normal theory tests and estimators are shown to have efficiency 1/2. The test statistic for $H_0\colon \theta_1=\theta_{10},\cdots,\theta_p=\theta_{p0}$ depends on a scaled version of

$$n^{-1/2}\sum_{t=p+1}^{n}\operatorname{sign}(X_t-\theta_{10}X_{t-1}-\cdots-\theta_{p0}X_{t-p})(X_{t-1},\cdots,X_{t-p})' . \tag{1.2}$$

Previously, the normal theory MLE's (Least Squares Estimators) were the only commonly employed statistics with the property that their asymptotic distribution depended only on the parameters $\theta_1,\cdots,\theta_p$ and not on the form of the distribution. The double exponential statistic (1.2) also has this property. Moreover, this lack of dependence on the form of the error distribution can be extended to our whole class of test statistics which are optimal under a particular error distribution, provided the error distribution is sufficiently smooth. These modified statistics retain the optimal properties of the original (unmodified) statistics and should prove to be interesting competitors to least squares estimators when the error distributions have thicker tails than the normal.

In the next section we state our assumptions and establish the quadratic mean (q.m.) differentiability of certain random functions. In Section 3 we prove the main contiguity results and Section 4 contains the applications.

## 2. Assumptions and the condition of q.m. differentiability

Let $X_t$, $t\geqq 1$ be an AR$(p)$ process so that $X_t-\theta' X(t,p)=E_t$, where $\theta'=(\theta_1,\cdots,\theta_p)$, $E_t$ is described in (1.1) and

$$X(t, p)=(X_{t-1}, \cdots, X_{t-p})', \qquad p<t . \tag{2.1}$$

Let $f$ denote the density of each $E_t$ and $g(X_1, \cdots, X_t; \theta)$ the joint density of $X_1, \cdots, X_t$ when the parameter $\theta$ obtains. Then the joint density of $X_1, \cdots, X_p, E_{p+1}, \cdots, E_t$ is $g(X_1, \cdots, X_p; \theta) \prod_{j=1}^{t-p} f(E_{p+j})$. Transforming from $(X_1, \cdots, X_p, E_{p+1}, \cdots, E_t)$ to $(X_1, \cdots, X_t)$ and noting that the Jacobian is unity, we obtain $g(X_1, \cdots, X_t; \theta)=g(X_1, \cdots, X_p; \theta) \prod_{j=1}^{t-p} f(X_{p+j}-\theta' X(p+j, p))$. It follows that, under $[g(x_1, \cdots, x_p; \theta_0) \times f(x)]$, the logarithm of the likelihood ratio is

$$\Lambda_t(\theta_0, \theta)=\log \frac{g(X_1, \cdots, X_p; \theta)}{g(X_1, \cdots, X_p; \theta_0)}+\sum_{j=1}^{t-p} \log \frac{f(E_{p+j}-(\theta-\theta_0)' X(p+j, p))}{f(E_{p+j})} \tag{2.2}$$

where, of course, $E_{p+j}=X_{p+j}-\theta_0' X(p+j, p)$, when $\theta_0$ obtains, in accordance with (1.1).

Consider the parametric family of probability measures $\{f(x+\xi), \xi \in R\}$ and set

$$\rho(x, \xi)=[f(x-\xi)/f(x)]^{1/2}=s(x-\xi)/s(x) \tag{2.3}$$

where $s(x)=[f(x)]^{1/2}$. For the rest of the paper we make the following assumptions.

ASSUMPTIONS. (A1) Considered as a stochastic process in $\theta$, $g(X_1, \cdots, X_p; \theta)$ is continuous in probability.
(A2) The random function $\rho(x, \xi)$ defined in (2.3) is differentiable in quadratic mean (q.m.) $[f]$ with respect to $\xi$ at $\xi=0$.
(A3) The roots of $1-\theta_1 B-\theta_2 B^2-\cdots-\theta_p B^p=0$ lie outside the unit circle, so that (see Box and Jenkins [2])

$$X_t=\sum_{k=0}^{\infty} \psi_k E_{t-k} . \tag{2.4}$$

*Remark* 2.1. Assumption (A3) implies that the process $X_t$, $t \geqq 1$ is ergodic (cf. Hannan [7], p. 204 where $\{|\psi_k|\}$ has a geometric bound by (A3)).

THEOREM 2.1. *Suppose that assumption* (A2) *holds and let* $\dot{\rho}(X)$ *denote the q.m. derivative. Then the random function* $\phi_j(\theta, \theta^*)$ *defined by*

$$\phi_j(\theta, \theta^*)=\rho(E_{p+j}, (\theta^*-\theta)' X(p+j, p)) \tag{2.5}$$

*is differentiable in q.m.* $[f(x) \times g(x_1, \cdots, x_p; \theta)]$ *with respect to* $\theta^*$ *at* $\theta^*=\theta$ *with q.m. derivative*

$$\dot{\phi}_j(\theta)=X(p+j, p) \dot{\rho}(E_{p+j}) . \tag{2.6}$$

PROOF. By assumption $E_{g\times f}[h'\dot{\phi}_1(\theta)]^2 = E_g[h'X(p+1, p)]^2 E_f[\dot{\rho}(X)]^2 < \infty$. Thus, by Vitali's theorem we have to show that

$$\frac{1}{\lambda^2} E_{g\times f}[\phi_1(\theta, \theta+\lambda h)-1]^2 \to E_f[\dot{\rho}(X)]^2 E_g[h'X(p+1, p)]^2 .$$

According to a well known theorem (Hájek and Šidák [6], p. 64) the above follows from

$$\frac{1}{\lambda^2} E_{g\times f}[\phi_1(\theta, \theta+\lambda h)-1]^2 \leqq E_f[\dot{\rho}(X)]^2 E_g[h'X(p+1, p)]^2$$

which, in turn, is implied by

$$\frac{1}{\lambda^2} E_f\{[\rho(E_{p+1}, \lambda\xi)-1]^2 \mid \xi\} \leqq \xi^2 E_f[\dot{\rho}(X)]^2$$

where we set $\xi = h'X(p+1, p)$. But the last relation follows by a sequence of routine arguments.

COROLLARY 2.1. *$t^{1/2}[\phi_j^2(\theta, \theta+t^{-1/2}h)-1] \to 2h'\dot{\phi}_j(\theta)$ as $t\to\infty$ in the first mean $[f(x)\times g(x_1, \cdots, x_p; \theta)]$.*

## 3. The property of contiguity and some consequences

Let $\Theta$ denote the set of values of $\theta$ specified by assumption (A3). For each $\theta \in \Theta$ let $P_{t,\theta}$ denote the probability measure describing the behavior of $X_1, \cdots, X_p, X_{p+1}, \cdots, X_t$, $t>p$. The dependence of $P_{t,\theta}$ on the density of the error shocks $f$ will not be indicated when there is no danger of confusion. Let

$$\theta_t = \theta + h_t t^{-1/2}, \qquad h_t \to h \in R^p . \tag{3.1}$$

In this section we establish the contiguity of the sequences $\{P_{t,\theta}\}$, $\{P_{t,\theta_t}\}$ and explore some consequences that are relevant for the applications of Section 4. The property of contiguity follows from a sequence of lemmas whose proofs are omitted since they are mainly based on the quadratic mean differentiability of the random functions $\phi_j(\theta, \theta^*)$, $j\geqq 1$ (which has been established in Theorem 2.1), and standard arguments using the stationarity and ergodicity of the process $X_t$, $t\geqq 1$ (cf. Roussas [11], p. 54–63).

Let $\Gamma_X(\theta)$, $\Gamma(\theta)$ be the covariance matrix of $2X(p+1, p)$, $2\dot{\phi}_1(\theta)$ respectively and set $\dot{\rho}_j = \dot{\rho}(E_{p+j})$. Then

$$\Gamma(\theta) = \Gamma_X(\theta) E_f[\dot{\rho}_1^2] . \tag{3.2}$$

For notational convenience set $\phi_{tj}(\theta) = \phi_j(\theta, \theta_t)$, $\Lambda_t(\theta) = \Lambda_t(\theta, \theta_t)$, and define

$$\Delta_t(\theta)=2t^{-1/2}\sum_{j=1}^{t}\dot{\phi}_j(\theta) . \tag{3.3}$$

LEMMA 3.1. $\max\{|\phi_{tj}(\theta)-1|;\ 1\leqq j\leqq t\}\rightarrow 0$, *in $P_\theta$-probability.*

LEMMA 3.2. $\sum_{j=1}^{t}\log\phi_{tj}^2(\theta)-2\left\{\sum_{j=1}^{t}[\phi_{tj}(\theta)-1]-\frac{1}{2}\sum_{j=1}^{t}[\phi_{tj}(\theta)-1]^2\right\}\rightarrow 0$ *in $P_{t,\theta}$-probability.*

LEMMA 3.3. $\sum_{j=1}^{t}[\phi_{tj}(\theta)-1]^2\rightarrow \mathrm{E}_\theta[h'\dot{\phi}_1(\theta)]^2$ *in $P_\theta$-probability.*

Recalling the form of $\phi_{tj}(\theta)$ and using the fact that $E_{p+j}$ is independent of $X_1,\cdots,X_{p+j-1}$, we have $\mathrm{E}_\theta(\phi_{tj}^2(\theta)|\mathcal{A}_{j-1})=1$. Here $\mathcal{A}_n=\sigma(X_1,\cdots,X_{p+n})$, $n\geqq 0$. The last relation, Corollary 2.1 and standard properties of conditional expectation imply

LEMMA 3.4. $\mathrm{E}\,(\dot{\phi}_j(\theta)|\mathcal{A}_{j-1})=0$, *a.s.* $[P_\theta]$.

Finally, it may be shown that

LEMMA 3.5. $\sum_{j=1}^{t}(\phi_{tj}(\theta)-1)-t^{-1/2}\sum_{j=1}^{t}h'\dot{\phi}_j(\theta)\rightarrow -\frac{1}{2}\mathrm{E}_\theta[h'\dot{\phi}_1(\theta)]^2$ *in $P_{t,\theta}$-probability.*

THEOREM 3.1. $\Lambda_t(\theta)-h'\Delta_t(\theta)\rightarrow -\frac{1}{2}h'\Gamma(\theta)h$, *in $P_{t,\theta}$-probability.*

PROOF. The result follows by a straightforward combination of assumption (A1), and Lemmas 3.2, 3.3, 3.5.

THEOREM 3.2. $\mathcal{L}[\Delta_t(\theta)|P_{t,\theta}]\Rightarrow N(0,\Gamma(\theta))$.

PROOF. Claim follows from Lemma 3.4 and the central limit theorem for martingales.

An implication of Theorems 3.1, 3.2 is

THEOREM 3.3. $\mathcal{L}[\Lambda_t(\theta)|P_{t,\theta}]\Rightarrow N\left(-\frac{1}{2}h'\Gamma(\theta)h,\ h'\Gamma(\theta)h\right)$.

COROLLARY 3.1. *The sequences of probability measures $\{P_{t,\theta}\}$, $\{P_{t,\theta_t}\}$ are contiguous.*

As a consequence of Corollary 3.1 and Theorems 3.1–3.3 we have

THEOREM 3.4. (i) $\Lambda_t(\theta)-h'\Delta_t(\theta)\rightarrow -\frac{1}{2}h'\Gamma(\theta)h$, *in $P_{t,\theta_t}$-probability.*

(ii) $\mathcal{L}[\Lambda_t(\theta)|P_{t,\theta_t}]\Rightarrow N\left(\frac{1}{2}h'\Gamma(\theta)h,\ h'\Gamma(\theta)h\right)$.

(iii) $\mathcal{L}[\Delta_t(\theta)|P_{t,\theta_t}]\Rightarrow N(\Gamma(\theta)h,\Gamma(\theta))$.

In view of the applications of next section we need some further results. To formulate them, let $f^{(1)}$, $f^{(2)}$ be two different densities satisfying the assumptions (A1), (A2), each of which can be thought as generating the innovations process $E_t$, $t\geq 1$. Denote by $P_{t,\theta}^{(1)}$, $P_{t,\theta}^{(2)}$ the corresponding probability measures and by $\Delta_t^{(1)}(\theta)$, $\Delta_t^{(2)}(\theta)$, $\dot{\rho}_j^{(1)}$, $\dot{\rho}_j^{(2)}$ etc. the corresponding quantities. Also $\mathrm{E}^{(i)}$ will denote expectation under $f^{(i)}$ while $\mathrm{E}_\theta^{(i)}$ will denote expectation under $P_\theta^{(i)}$, $i=1, 2$.

THEOREM 3.5. *$\mathcal{L}[(\Delta_t^{(2)}(\theta), \Delta_t^{(1)}(\theta))' \mid P_{t,\theta}^{(2)}] \Rightarrow N\left(\begin{pmatrix}\mu\\0\end{pmatrix}, D\right)$ where $\mu=-\frac{1}{2}\cdot h'\Gamma^{(2)}(\theta)h$ and $D=(D_{ij})$, $i, j=1, 2$ with $D_{11}=\mathrm{E}^{(2)}[(\dot{\rho}_1^{(2)})^2]h'\Gamma_X^{(2)}(\theta)h$, $D_{12}=\mathrm{E}^{(2)}[\dot{\rho}_1^{(1)}\dot{\rho}_1^{(2)}]h'\Gamma_X^{(2)}(\theta)+8\sum_{k=1}^{\infty} h'\,\mathrm{E}_\theta^{(2)}[\dot{\rho}_1^{(2)}X(p+1, p)X'(p+k+1, p)\dot{\rho}_{k+1}^{(1)}]$ and $D_{22}=\mathrm{E}_\theta^{(2)}[(\dot{\rho}_1^{(1)})^2]\Gamma_X^{(2)}(\theta)+8\sum_{k=1}^{\infty}\mathrm{E}_\theta^{(2)}[\dot{\rho}_1^{(1)}X(p+1, p)X'(p+k+1, p)\dot{\rho}_{k+1}^{(1)}]$.*

PROOF. By Theorem 3.1, it suffices to show

$$\mathcal{L}[(h'\Delta_t^{(2)}, \Delta_t^{(1)})' \mid P_{t,\theta}^{(2)}] \Rightarrow N\left(\begin{pmatrix}0\\0\end{pmatrix}, D\right) \tag{3.4}$$

where $\theta$ is dropped in our notation. A linear combination of the vector $(h'\Delta_t^{(2)}, \Delta_t^{(1)})$ is of the form $2t^{-1/2}\sum_{j=1}^{t}(a_2'\dot{\phi}_j^{(2)}+a_1'\dot{\phi}_j^{(1)})=2t^{-1/2}\sum_{j=1}^{t}Y_j$, where $a_2=ah$ and (see (2.6)),

$$Y_j=(\dot{\rho}_j^{(1)}a_1+\dot{\rho}_j^{(2)}a_2)'X(p+j, p)\ . \tag{3.5}$$

Consider the representation $X_t=\sum_{k=0}^{\infty}\psi_k E_{t-k}$, set $X_{l,t}=\sum_{k=0}^{l}\psi_k E_{t-k}$, $X_l(p+j, p)=(X_{l,p+j-1},\cdots, X_{l,j})'$, and define

$$Y_{l,j}=(\dot{\rho}_j^{(1)}a_1+\dot{\rho}_j^{(2)}a_2)'X_l(p+j, p)\ . \tag{3.6}$$

By a well known theorem (Billingsley [1], p. 184) if we can show that

$$\sum_{l=0}^{\infty}\nu(l)^{1/2}<\infty\ , \qquad \text{where } \nu(l)=\mathrm{E}_\theta^{(2)}(Y_1-Y_{l,1})^2 \tag{3.7}$$

holds true, then $2t^{-1/2}\sum_{j=1}^{t}Y_j$ has an asymptotically normal distribution with variance $4\,\mathrm{E}_\theta^{(2)}Y_1^2+8\sum_{k=1}^{\infty}\mathrm{E}_\theta^{(2)}Y_1Y_{k+1}$. But $4\,\mathrm{E}_\theta^{(2)}Y_1^2=a\,\mathrm{E}^{(2)}[\dot{\rho}_1^{(2)}]^2h'\Gamma_X^{(2)}h+\mathrm{E}^{(2)}[\dot{\rho}_1^{(1)}]^2a_1'\Gamma_X^{(2)}a_1+2a\,\mathrm{E}^{(2)}[\dot{\rho}_1^{(1)}\dot{\rho}_1^{(2)}]a_1'\Gamma_X^{(2)}h$, and $8\,\mathrm{E}_\theta^{(2)}Y_1Y_{k+1}=8a_1'\,\mathrm{E}_\theta^{(2)}[\dot{\rho}_1^{(1)}X(p+1, p)X'(p+k+1, p)\dot{\rho}_{k+1}^{(1)}]a_1+8a_2'\,\mathrm{E}_\theta^{(2)}[\dot{\rho}_1^{(2)}X(p+1, p)X'(p+k+1, p)\dot{\rho}_{k+1}^{(1)}]a_1$ which verifies the covariance matrix of the theorem. It remains to verify that (3.7) is true. It is easy to see that this amounts to showing that $\sum_{l=0}^{\infty}\left(\sum_{k=l+1}^{\infty}|\psi_k\psi_{k+i}|\right)^{1/2}<\infty$. But this is shown in Johnson and Bagshaw [8].

*Remark* 3.1. If $E^{(2)}(\dot{\rho}_1^{(1)})=0$, then $D_{12}=E^{(2)}[\dot{\rho}_1^{(1)}\dot{\rho}_1^{(2)}]h'\Gamma_X^{(2)}(\theta)$, $D_{22}=E^{(2)}[(\dot{\rho}_1^{(1)})^2]\Gamma_X^{(2)}(\theta)$, while $D_{11}$ remains the same. This happens, for instance, if $f^{(1)}(x)=(2\pi)^{-1/2}\exp(-x^2/2)$ and $f^{(2)}(x)$ is any density with zero mean value.

Theorem 3.5 and well known properties of contiguous probability measures imply

THEOREM 3.6. *$\mathcal{L}[\Delta_t^{(1)}(\theta)\,|\,P_{t,\theta_t}^{(2)}]\Rightarrow N(D_{21}, D_{22})$ where $D_{21}$, $D_{22}$ are defined in Theorem* 3.5.

## 4. Evaluation of asymptotic efficiencies

I) *Pitman efficiency of tests*

First we will consider testing hypotheses for AR(1) and then for AR($p$).

a) Consider testing the hypothesis $H_0$: $\theta=\theta_0$ vs. $H_1$: $\theta\neq\theta_0$ in AR(1), and let $f^{(1)}$, $f^{(2)}$ be two candidate densities for generating the innovations process. It is well known that under $f^{(i)}$ the asymptotically most powerful test for $H_0$ is based on $\Delta_t^{(i)}$. From Theorems 3.4, 3.5, 3.6 it follows that under $f^{(2)}$ the Pitman efficiency of $\Delta_t^{(1)}(\theta_0)$ with respect to $\Delta_t^{(2)}(\theta_0)$ is

$$e_{1,2}=\frac{(D_{21})^2}{D_{22}\,E^{(2)}[(\dot{\rho}_1^{(2)})^2]\Gamma_X^{(2)}h} \tag{4.1}$$

where $D_{21}$, $D_{22}$ are defined in Theorem 3.5. According to Remark 3.1, if $E^{(2)}(\dot{\rho}_1^{(1)})=0$ relation (4.1) simplifies to

$$e_{1,2}=\frac{[E^{(2)}(\dot{\rho}_1^{(1)}\dot{\rho}_1^{(2)})]^2}{E^{(2)}[(\dot{\rho}_1^{(2)})^2]\,E^{(2)}[(\dot{\rho}_1^{(1)})^2]}\,. \tag{4.2}$$

Table 1 contains the Pitman efficiencies of $\Delta_t^{(1)}$ with respect to $\Delta_t^{(2)}$ when $f^{(1)}(x)=(2\pi)^{-1/2}\exp(-x^2/2)$ and $f^{(2)}$ as indicated. When $f^{(1)}(x)=(1/2)\exp(-|x|)$, the Pitman efficiency of $\Delta_t^{(1)}$ with respect to $\Delta_t^{(2)}$ for various $f^{(2)}$ is given in Table 2.

*Remark* 4.1. (i) In practice it is not realistic to assume that we know the variance of the innovations distribution, and hence it is desirable to have a test statistic independent of this unknown parameter. To obtain this statistic we note that, according to the notation introduced in Theorem 2.1,

$$\dot{\rho}(x)=\frac{1}{\sigma}\dot{\rho}\left(\frac{x}{\sigma};\,1\right) \tag{4.3}$$

Table 1. Efficiencies of $\Delta_t^{(1)}$ based on the normal error distribution, $f^{(1)}(x)=(2\pi)^{-1/2}\exp(-x^2/2)$

| true p.d.f.: $f^{(2)}(x)$ | efficiency $e_{1,2}$ |
|---|---|
| $(2\beta)^{-1}\exp(-\|x\|/\beta)$ | .5 |
| $t_\nu$ | $\dfrac{(\nu-2)(\nu+2)^3(\nu+3)}{\nu^{5/2}(\nu+1)(\nu+4)^{3/2}}$ |
| $t_3, t_4, t_6, t_{10}$ | .6495, .8353, .9443, .9863 |
| $(2\pi\sigma^2)^{-1/2}\exp(-x^2/2\sigma^2)$ | 1 |
| $\dfrac{p}{(2\pi)^{1/2}}e^{-x^2/2}+\dfrac{(1-p)}{(2\pi\sigma^2)^{1/2}}e^{-x^2/2\sigma^2}$ | $\left\{[p+(1-p)\sigma^2]\int_{-\infty}^{\infty}\dfrac{\left[\frac{px}{\sqrt{2\pi}}e^{-x^2/2}+\frac{(1-p)x\sigma^{-3}}{\sqrt{2\pi}}e^{-x^2/2\sigma^2}\right]^2}{\frac{p}{\sqrt{2\pi}}e^{-x^2/2}+\frac{(1-p)}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}}dx\right\}^{-1}$ |
| $\sigma=3$: $p=.9, .7, .5, .3, .1$ | .70, .55, .58, .72, .96 |

Table 2. Efficiencies of double exponential statistics based on the error distribution, $f^{(1)}(x)=(1/2)\exp(-\|x\|)$

| true p.d.f. $f^{(2)}(x)$ | efficiency $e_{1,2}$ |
|---|---|
| $(2\pi\sigma^2)^{-1/2}\exp(-x^2/2\sigma^2)$ | $2/\pi$ |
| $(2\beta)^{-1}\exp(-\|x\|/\beta)$ | 1 |
| $t_\nu$ | $\dfrac{4}{\pi}\left[\dfrac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\right]^2\dfrac{\nu^{1/2}(\nu+3)}{(\nu+1)(\nu+4)^{1.5}}$ |
| $t_3, t_{10}, t_{20}, t_{40}$ | .2274, .4321, .5173, .5715 |
| $\dfrac{p}{(2\pi)^{1/2}}e^{-x^2/2}+\dfrac{(1-p)}{(2\pi\sigma^2)^{1/2}}e^{-x^2/2\sigma^2}$ | $\dfrac{2}{\pi}\dfrac{[p+(1-p)\sigma^{-1}]^2}{\int_{-\infty}^{\infty}\dfrac{\left[\frac{px}{\sqrt{2\pi}}e^{-x^2/2}+\frac{(1-p)x\sigma^{-3}}{\sqrt{2\pi}}e^{-x^2/2\sigma^2}\right]^2}{\frac{p}{\sqrt{2\pi}}e^{-x^2/2}+\frac{(1-p)}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}}dx}$ |
| $\sigma=3$: $p=.9, .7, .5, .3, .1$ | .70, .77, .82, .86, .80 |

where $\dot\rho(y)$, $\dot\rho(y;1)$ correspond to the densities $f(y)$ and $f_{st}(y)=\sigma f(\sigma y)$, respectively. Provided that $\dot\rho$ is sufficiently smooth and $\hat\sigma^2$ is any $n^{1/2}$ consistent estimator,

$$\Delta_t^*(\theta_0)=2t^{-1/2}\sum_{j=1}^{t}\frac{1}{\hat\sigma}\dot\rho\left(\frac{E_{p+j}}{\hat\sigma};1\right)X(p+j,p) \tag{4.4}$$

will have the same asymptotic properties as $\Delta_t(\theta_0)$. A prime example is

$$\Delta_t^*(\theta_0)=2t^{-1/2}\sum_{j=1}^{t}\frac{\operatorname{sign}(E_{p+j})}{\hat\beta}X(p+1,p) \tag{4.5}$$

derived from double exponential errors. Here $(2t)^{-1}\sum_{j=p+1}^{t}(X_t-\theta_{10}X_{t-1}-\cdots-\theta_{p0}X_{t-p})^2$ is one choice for $\hat\beta^2$. The asymptotic distribution of $\Delta_t^*$ is clearly the same as $\Delta_t$ and the Pitman efficiencies, which depend only on Theorems 3.2 and 3.4–3.6, remain in force. These efficiencies show (4.5) to be more robust than the normal theory test against the outlier prone mixed normal.

(ii) From a practical standpoint, we would like asymptotic probability statements concerning the statistic (4.4) (such as level of significance of test) to be free of the form of the error distribution. Note that according to Theorem 3.2, the asymptotic variance of $\Delta_t^*(\theta_0)$ is $\mathrm{E}\left[\dot\rho_1\left(\frac{X}{\sigma};1\right)\right]^2\left(\frac{1}{\sigma^2}\Gamma_X\right)$. Note also that $\frac{1}{\sigma^2}\Gamma_X$ is independent of $\sigma^2$ and of the form of the distribution. Proceeding heuristically we consider the test statistic

$$\hat\Delta_t(\theta_0)=\left\{t^{-1}\sum_{j=1}^{t}\left[\dot\rho\left(\frac{E_{p+j}}{\hat\sigma};1\right)\right]^2\right\}^{-1/2}\Delta_t^*(\theta_0)\,. \tag{4.6}$$

For double exponential errors, $\dot\rho^2=1$ almost everywhere, and the resulting test statistic $\hat\Delta_t(\theta_0)$ remains the same as (4.5). Thus the limiting distribution of (4.5) is the same, whatever the underlying symmetric error distribution.

b) Consider now the multiparameter hypothesis $H_0:\theta=\theta_0$ vs $H_1:\theta\neq\theta_0$ in AR$(p)$. Let $f^{(1)}, f^{(2)}$ be two candidate densities for the innovations distribution. It is well known that, under $f^{(i)}$, the test based on $\Delta_t^{(i)}(\theta_0)$ has certain asymptotic optimality properties such as asymptotically most stringent and asymptotically highest average power over certain ellipsoids. We consider the smooth cases where $\Delta_t^{(i)}(\theta_0)$ and $\hat\Delta_t^{(i)}(\theta_0)$ are asymptotically equivalent. Then, the test based on $\hat\Delta_t(\theta_0)$ rejects $H_0$ when $\hat\Delta_t^{(i)}(\theta_0)'[\Gamma^{(i)}(\theta_0)]^{-1}\hat\Delta_t^{(i)}(\theta_0)\geq\chi_p^2(\alpha)$. It follows easily from Theorems 3.4, 3.5, 3.6 that (dropping $\theta_0$ from our notation),

$$\mathcal{L}[\hat\Delta_t^{(1)\prime}[\Gamma^{(1)}]^{-1}\hat\Delta_t^{(1)}\,|\,P_{t,\theta_0}^{(2)}]\Rightarrow\chi_p^2$$

and

$$\mathcal{L}[\hat\Delta_t^{(1)\prime}[\Gamma^{(1)}]^{-1}\hat\Delta_t^{(1)}\,|\,P_{t,\theta_t}^{(2)}]\Rightarrow\chi_p^2\left(\frac{\mathrm{E}^{(2)}(\dot\rho^{(1)}(X/\sigma;1)\dot\rho^{(2)}(X/\sigma;1))}{\sqrt{\mathrm{E}^{(2)}(\dot\rho^{(1)}(X/\sigma;1))^2}\,\sigma}h'\Gamma_X h\right)\,.$$

Also,

$$\mathcal{L}[\hat{\Delta}_t^{(2)\prime}[\Gamma^{(2)}]^{-1}\hat{\Delta}_t^{(2)} \mid P_{t,\theta_0}^{(2)}] \Rightarrow \chi_p^2$$

and

$$\mathcal{L}[\hat{\Delta}_t^{(2)\prime}[\Gamma^{(2)}]^{-1}\hat{\Delta}_t^{(2)} \mid P_{t,\theta_t}^{(2)}] \Rightarrow \chi_p^2\left(\frac{\mathrm{E}^{(2)}\,(\dot{\rho}^{(2)}(X/\sigma\,;1))^2}{\sigma^2}\,h'\Gamma_X h\right).$$

The Pitman efficiency thus is

$$e_{\hat{\Delta}_t^{(1)},\hat{\Delta}_t^{(2)}} = \left[\frac{\mathrm{E}^{(2)}[\dot{\rho}_1^{(1)}\dot{\rho}_1^{(2)}]}{\sqrt{\mathrm{E}^{(2)}(\dot{\rho}_1^{(1)})^2}\,\sqrt{\mathrm{E}^{(2)}(\dot{\rho}_1^{(2)})^2}}\right]^2 \tag{4.7}$$

which is the same as the formula given in (4.2), and thus the efficiency calculations in Tables 1 and 2 apply for all order autoregressions. Although these efficiencies pertain to testing $H_0\colon \theta=\theta_0$, the important hypothesis $H_0\colon X_t$ are independent, is included. Moreover the statistic $\hat{\Delta}_t(\theta_0)$ provides an important alternative to the $\chi^2$ test based on least squares estimators.

II) *Efficiency of L.S. estimators*

It follows from Theorems 3.1–3.4 that, within a certain wide class of estimators, the best attainable asymptotic covariance matrix, under the innovations distribution $f$, is $\mathrm{E}_f[\dot{\rho}_1^2]^{-1}\Gamma_X^{-1}$. It is also well known (see Whittle [12], [13]) that the asymptotic covariance matrix of the L.S. estimators is independent of the innovations distribution. Thus, the efficiency of the L.S. estimator with respect to the most efficient estimator under $f$ (although we have not checked if such an estimator exists in all cases considered) is $1\Big/\mathrm{E}_f\left[2\dot{\rho}\left(\frac{X}{\sigma};1\right)\right]^2$, or an appropriate power. Table 3 give some values. Note that for the double exponen-

Table 3. Efficiency of the L.S. estimator

| $f$ | $\left\{\mathrm{E}_f\left[2\dot{\rho}\left(\frac{X}{\sigma};1\right)\right]^2\right\}^{-1}$ |
|---|---|
| $(2\beta)^{-1}\exp(-\lvert x\rvert/\beta)$ | 1/2 |
| $t_\nu$ | $\frac{(\nu-2)(\nu+3)}{\nu(\nu+1)}$ |
| $t_3, t_4, t_6, t_{10}$ | .5, .7, .857, .945 |
| $\frac{p}{\sqrt{2\pi}}e^{-x^2/2} + \frac{(1-p)}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}$ | $\left\{[p+(1-p)\sigma^2]\int_{-\infty}^{\infty}\frac{\left[\frac{px}{\sqrt{2\pi}}e^{-x^2/2}+\frac{(1-p)x\sigma^3}{\sqrt{2\pi}}e^{-x^2/2\sigma^2}\right]^2}{\frac{p}{\sqrt{2\pi}}e^{-x^2/2}+\frac{(1-p)}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}}dx\right\}^{-1}$ |

tial and mixed normals models, the Pitman efficiency of the normal theory test coincides with the efficiency of the L.S. estimator and the corresponding numbers in Table 1 apply.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY*
UNIVERSITY OF WISCONSIN-MADISON

## REFERENCES

[1] Billingsley, P. (1968). *Convergence of Probability Measures*, John Wiley, New York.

[2] Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Models for Forecasting and Control*, Holden Day, San Francisco.

[3] Denby, L. and Martin, R. D. (1979). Robust estimation of the first order autoregressive parameter, *J. Amer. Statist., Ass.*, **74**, 140-146.

[4] Gastwirth, J. L. and Rubin, H. (1975a). The asymptotic distribution theory of the empiric cdf for mixing stochastic processes, *Ann. Statist.*, **3**, 809-924.

[5] Gastwirth, J. L. and Rubin, H. (1975b). The behavior of robust estimators on dependent data, *Ann. Statist.*, **3**, 1070-1100.

[6] Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*, Academic Press, New York.

[7] Hannan, E. J. (1970). *Multiple Time Series*, John Wiley, New York.

[8] Johnson, R. A. and Bagshaw, M. (1974). The effect of serial correlation on the performance of CUSUM tests, *Technometrics*, **16**, 103-112.

[9] Kanter, M. and Steiger, W. L. (1974). Regression and autoregression with infinite variance, *Advances Appl. Prob.*, **6**, 768-783.

[10] Martin, R. D. (1979). Robust estimation for time series autoregressions, *Robustness in Statistics* (eds. R. L. Launer and G. Wilkinson), Academic Press, New York.

[11] Roussas, G. G. (1972). *Contiguity of Probability Measures: Some Applications in Statistics*, Cambridge University Press.

[12] Whittle, P. (1953). Estimation and information in stationary time series, *Arch. Math.*, **2**, 423-434.

[13] Whittle, P. (1962). Gaussian estimation in stationary time series, *Bull. Int. Statist. Inst.*, **39**, 105-129.

* Now at University of Rochester.