On Asymptotic properties of the Least Squares Estimators for Autoregressive Time Series With a Unit Root.

Sastry G. Pantula

Summary

Let n observations $Y_1, Y_2, \ldots, Y_n$ be generated by the model $Y_t = \rho Y_{t-1} + e_t$, where Y_0 is a fixed constant and $\{e_t\}$ is a sequence of martingale differences with a constant conditional variance and bounded $(2+\delta)$-th moment. Properties of the regression estimator of ρ are obtained under the assumption $\rho = 1$. Extensions for the p-th order stochastic difference equation with a unit root are given.

1. **Introduction.** Consider the autoregressive model

$$(1) \quad Y_t = \rho Y_{t-1} + e_t , \quad t = 1,2,\ldots,$$

where $Y_0 = 0$, ρ is a real number, and the e_t are uncorrelated random variables. Given n observations $Y_1, Y_2,\ldots,Y_n$, the least squares estimator of ρ is

$$(2) \quad \hat{\rho} = \left(\sum_{t=1}^{n} Y_{t-1}^2\right)^{-1} \left(\sum_{t=1}^{n} Y_t Y_{t-1}\right) .$$

Mann and Wald (1943) established that the limiting distribution of the least squares estimator is normal, for $|\rho| < 1$ under the assumption that $\{e_t\}$ is a sequence of independent normal random variables with mean zero and variance σ^2 [i.e., e_t NID$(0,\sigma^2)$] . Anderson (1959) extended Mann and Wald's result to the case where e_t are assumed to be independent $(0,\sigma^2)$ random variables with bounded

---

AMS 1970 subject classifications Primary 62M10; secondary 62J05.

Key words and phrases. Autoregressive process, asymptotic properties, martingale differences.

$(2 + \delta)$-th moments, for some $\delta > 0$. White (1958) obtained a representation for the limiting distribution of $n(\hat{\rho}-1)$ under the assumptions that the e_t are $NID(0,\sigma^2)$ and $\rho = 1$.

Hannan and Heyde (1972), Hannan and Nicholls (1972), Reinsel (1976), Fuller (1976), Anderson and Taylor (1979), and Fuller, Hasza, and Goebel (1981) considered the estimation of a p-th order stochastic difference equation

$$(3) \quad Y_t = \sum_{i=1}^{n} \psi_{ti}\beta_i + \sum_{j=1}^{p} \alpha_j Y_{t-j} + e_t \quad ,$$

where $\{\psi_{ti}\}$ are fixed sequences and the roots of the characteristic equation

$$(4) \quad m^p - \sum_{j=1}^{p} \alpha_j m^{p-j} = 0$$

are less than unity in absolute value. Crowder (1980) obtained the asymptotic distribution of the least squares estimators of the parameters of (3) for the case where it is assumed that $\{e_t\}$ is a sequence of martingale differences and the roots of (4) are less than unity in absolute value.

Dickey (1976), Fuller (1976), and Dicky and Fuller (1979) considered the estimation of equation (3) assuming one of the roots of the equation (4) to be one, $\{e_t\}$ to be a sequence of independent normal $(0,\sigma^2)$ random variables, and the set $\{\psi_{ti}\}$ to include the constant function and time. Fuller, Hasza and Goebel (1981) extended the results of Dickey and Fuller to the case where $\{e_t\}$ is assumed to be a sequence of independent $(0,\sigma^2)$ random variables with bounded $(2+\delta)$-th moments and permitted the set $\{\psi_{ti}\}$ to include functions other than the constant function and time trend.

In this note we extend the results of Dickey and Fuller (1979), and of Fuller, Hasza and Goebel (1981) to the case where $\{e_t\}$ is assumed to satisfy

$$(5)\quad \begin{aligned} &E(e_t \mid F_{t-1}) = 0 \qquad \text{a.s.} \\ &E(e_t^2 \mid F_{t-1}) = \sigma^2 \qquad \text{a.s.} \end{aligned}$$

and

$$E(e_t^{2+\delta}) < L < \infty \quad \text{for some } L \text{ , and } \delta > 0 \text{ ,}$$

where F_t is the σ-field generated by $\{e_1, e_2, \ldots, e_t\}$.

2. Representations for the limit distributions. Let Y_t satisfy (1) with $\rho = 1$ and let e_t satisfy (5). The statistic analogous to the regression t-statistic for the test of the hypothesis that $\rho = 1$ is

$$(6)\quad \hat{\tau} = (\hat{\rho}-1)\left(\sum_{t=1}^{n} Y_{t-1}^2\right)^{1/2} S^{-1}$$

where $\hat{\rho}$ is defined by (2) and

$$(7) \quad S^2 = (n-2)^{-1} \sum_{t=2}^{n} (Y_t - \hat{\rho}\, Y_{t-1})^2$$

Because the statistics $\hat{\rho}$ and $\hat{\tau}$ are ratios of quadratic forms, we assume without loss of generality that $\sigma^2 = 1$. Following Dickey (1976), we have

$$n(\hat{\rho}-1) = \Big(\sum_{t=1}^{n} Y_{t-1}^2\Big)^{-1} \sum_{t=1}^{n} Y_{t-1} e_t$$

$$= \tfrac{1}{2} \Big(\sum_{t=1}^{n} Y_{t-1}^2\Big)^{-1}\Big(Y_n^2 - \sum_{t=1}^{n} e_t^2\Big) .$$

Since the conditional variance of e_t is assumed to be a constant,

$$\mathrm{Cov}(e_t^2, e_j^2) = 0 \quad , \text{ for } t \neq j .$$

and it follows from Theorem 2.13 of Hall and Heyde (1980, p.29), that

$$n^{-1} \sum_{t=1}^{n} e_t^2 \xrightarrow{P} 1 .$$

Therefore,

$$(8) \quad n(\hat{\rho} - 1) = (2\Gamma_n)^{-1}(T_n^2 - 1) + o_p(1) ,$$

where

$$T_n = n^{1/2} Y_{n-1}$$

$$= \sum_{i=1}^{n-1} a_{in} Z_{in} + O_p(n^{-1/2})$$

$$\Gamma_n = n^{-2} \sum_{t=2}^{n} Y_{t-1}^2$$

$$= n^{-2} \sum_{i=1}^{n-1} \lambda_{in} Z_{in}^2 ,$$

$$\lambda_{in} = (\tfrac{1}{4})\sec^2\big((n-i)\pi/(2n-1)\big) , \quad i = 1,2,\ldots, n-1$$

$$\underset{\sim}{Z} = \underset{\sim}{M}\, \underset{\sim}{e}_n$$

$$= (Z_{1n}, Z_{2n},\ldots, Z_{n-1,n})' ,$$

$$\underset{\sim}{e}_n = (e_1, e_2,\ldots, e_{n-1})' ,$$

$$a_{in} = \mathrm{Cov}(T_n, Z_{in})$$

and $\underset{\sim}{M}$ is an orthonormal matrix whose (i,t)-th element is

$$m_{it} = 2(2n-1)^{-1/2}\cos\left((4n-2)^{-1}(2t-1)(2i-1)\pi\right) .$$

We now give two results that are used in the derivation of the limiting distribution of (T_n, Γ_n) . The following Lemma is an extension of Lemma 1 of Dickey and Fuller (1979) and the proof is omitted.

<u>Lemma 1</u>. Suppose $\{Z_{in}: i = 1,2,\ldots, n-1;\ n = 1,2,\ldots\}$ is a triangular array of random variables. Suppose $E(Z_{in}) = 0$, $Var(Z_{in}) = \sigma^2$, and $Cov(Z_{in}, Z_{jn}) = 0$ for $i \neq j$. Let $\{w_i: i = 1,2,\ldots\}$ be a sequence of real numbers and let $\{w_{in}: i = 1,2,\ldots, n-1;\ n = 1,2,\ldots\}$ be a triangular array of real numbers. If

$$\sum_{i=1}^{\infty} w_i^2 < \infty ,$$

$$\lim_{n\to\infty} \sum_{i=1}^{n-1} w_{in}^2 = \sum_{i=1}^{\infty} w_i^2 ,$$

and

$$\lim_{n\to\infty} w_{in} = w_i ,$$

then

$$\sum_{i=1}^{n-1} w_{in} Z_{in} = \sum_{i=1}^{n-1} w_i Z_{in} + o_p(1) .$$

The limiting distribution of $\{Z_{in}\}$ is obtained in the following lemma.

Lemma 2. Let $\{e_t\}$ be a sequence of random variables satisfying conditions (5). Let $\underset{\sim}{M}_n$ be an $(n-1)$ by $(n-1)$ orthonormal matrix with

$$\lim_{n\to\infty} \sup_{1<t<n-1} |m_{it}(n)| = 0$$

for each fixed i , where $m_{it}(n)$ is the (i,t)-th element of $\underset{\sim}{M}_n$. Then for a fixed k ,

$$\underset{\sim}{Z}_n(k) = (Z_{1n}, Z_{2n},\ldots, Z_{kn})' \xrightarrow{L} \underset{\sim}{N}(\underset{\sim}{0}, \sigma^2 \underset{\sim}{I}_k) ,$$

where

$$\underset{\sim}{Z}_n = (Z_{1n}, Z_{2n},\ldots, Z_{n-1,n})' = \underset{\sim}{M}_n \underset{\sim}{e}_n ,$$

and

$$\underset{\sim}{e}_n = (e_1, e_2,\ldots, e_{n-1})' .$$

Proof. Let η be an arbitrary vector such that $\eta'\eta = 1$. Consider

$$S_n = \eta' Z_n(k)$$

$$= \sum_{t=1}^{n-1} X_{tn} ,$$

where

$$X_{tn} = e_t d_{tn}$$

and

$$d_{tn} = \sum_{i=1}^{k} \eta_i m_{it}(n) .$$

We apply the results of Scott (1973) to obtain the result. Note that

$$E[X_{tn} | F_{t-1}] = 0 \quad \text{a.e.}$$

$$E[X_{tn}^2 | F_{t-1}] = d_{tn}^2 \sigma^2 \quad \text{a.e.}$$

and

$$s_{nn}^2 = \sum_{t=1}^{n-1} d_{tn}^2 \sigma^2 = \sigma^2$$

because $\underset{\sim}{M}_n$ is orthogonal. Therefore,

$$V_{nn}^2 = \sum_{t=1}^{n-1} E[X_{tn}^2 | F_{t-1}] \quad \text{a.e.}$$

$$= s_{nn}^2 \quad \text{a.e.}$$

and the first condition of Scott's theorem is satisfied. To verify the second condition, consider

$$s_{nn}^{-2} \sum_{t=1}^{n-1} E[X_{tn}^2 I(|X_{tn}| > \varepsilon\ \sigma)]$$

$$< \sigma^{-2} \sup_{1<t<n-1} E[e_t^2\ I(|e_t d_{tn}| > \varepsilon\ \sigma)]$$

because $\Sigma_{t=1}^{n-1} d_{tn}^2 = 1$. Now

$$\sup_{1<t<n-1} E[e_t^2\ I(|e_t d_{tn}| > \varepsilon\ \sigma)]$$

$$< \sup_{1<t<n-1} \{E(e_t^{2+\delta})\}^{\frac{2}{2+\delta}} \cdot \{P[|e_t d_{tn}| > \varepsilon\sigma\}^{\frac{\delta}{2+\delta}}$$

$$< L^{\gamma}\ \varepsilon^{-\gamma\delta} \sup_{1<t<n-1} d_{tn}^{\gamma\delta}$$

$\to 0$ as $n \to \infty$,

where $\gamma = 2(2+\delta)^{-1}$. Therefore, the second condition of Scott's theorem is satisfied and

$$s_{nn}^{-1} S_n \xrightarrow{L} N(0,1) . \qquad []$$

Now we obtain the limiting distribution of (T_n, Γ_n) .

Theorem 1. Let $\{Z_i\}_{i=1}^{\infty}$ be a sequence of independent normal (0,1) random variables. Let Y_t satisfy (1) with $\rho = 1$. Let e_t satisfy (5). Let $\zeta_n' = (T_n, \Gamma_n)$, where T_n and Γ_n are defined in (8). Let $\zeta' = (T, \Gamma)$, where

$$T = \sum_{i=1}^{\infty} 2^{1/2} \gamma_i Z_i ,$$

$$\Gamma = \sum_{i=1}^{\infty} \gamma_i^2 Z_i^2 ,$$

$$\gamma_i = (-1)^{i+1} \sqrt{\gamma_i^2}$$

and

$$\gamma_i^2 = 4[(2i-1)\pi]^{-2} .$$

Then ζ_n converges in distribution to ζ .

Proof. From (8) and Lemma 1, it follows that

$$T_n = \sum_{i=1}^{n-1} a_i Z_{in} + o_p(1) ,$$

where

$$a_i = 2^{1/2} \gamma_i .$$

Also, since

$$n^{-2} \lambda_{in} - \gamma_i^2 = O(n^{-2})$$

we obtain

$$\Gamma_n = \sum_{i=1}^{n-1} \gamma_i^2 Z_{in}^2 + o_p(1) .$$

Note that for a fixed k ,

$$T_n = \sum_{i=1}^{k} a_i Z_{in} + \sum_{i=k+1}^{n-1} a_i Z_{in} + o_p(1) .$$

For a fixed k, from Lemma 2 it follows that

$$\sum_{i=1}^{k} a_i Z_{in} \xrightarrow{L} \sum_{i=1}^{k} a_i Z_i \quad \text{as} \quad n \to \infty$$

and as $k \to \infty$

$$\sum_{i=1}^{k} a_i Z_i \xrightarrow{L} T \ .$$

Also,

$$\text{Var}\left(\sum_{i=k+1}^{n-1} a_i Z_{in}\right) < \sum_{i=k+1}^{\infty} a_i^2$$

which converges to zero as k tends to infinity, uniformly in n. Therefore, using Lemma 6.3 of Fuller (1976),

$$T_n \xrightarrow{L} T$$

and

$$\zeta_n \xrightarrow{L} \zeta \qquad [\,]$$

<u>Corollary 1</u>. Let Y_t satisfy (1) with $\rho = 1$. Let e_t satisfy (5). Then

$$n(\hat{\rho} - 1) \xrightarrow{L} (\tfrac{1}{2})\Gamma^{-1}(T^2 - 1)$$

and

$$\hat{\tau} \xrightarrow{L} \frac{1}{2} \Gamma^{-1/2}(T^2 - 1) ,$$

where $\hat{\rho}$ and $\hat{\tau}$ are defined by (2) and (6), respectively.

Proof. The proof is an immediate consequence of Theorem 1 because the denominator of $\hat{\rho}$ is a continuous function of η that has probability one of being positive. Note under the model (1),

$$S^2 = (n-2)^{-1} \sum_{t=2}^{n} (Y_t - \hat{\rho}\, Y_{t-1})^2$$

$$= (n-2)^{-1} \sum_{t=1}^{n} e_t^2 + O_p(n^{-1}) . \qquad []$$

The remaining results of Dickey and Fuller (1979), extend to the case where e_t is assumed to satisfy (5) and are given in Pantula (1982).

We now extend the results of Fuller, Hasza and Goebel (1981), (FHG for short), to the case where $\{e_t\}$ is a sequence of martingale differences. Consider the model given in (3). We consider two cases of practical interest: (a) $\psi_{t1} \equiv 1$ and (b) $\psi_{t1} \equiv 1, \psi_{t2} = t$. We use the notation of FHG. An extension of Theorem 2 of FHG is given in the following theorem.

Theorem 2. Let model (3) hold with $m_1 = 1$ and $m_2, m_3, \ldots, m_p$ less than one in absolute value, where the m_i are the roots of $m^p - \alpha_1 m^{p-1} - \ldots - \alpha_p = 0$. Let $\{e_t\}$ be a sequence of martingale differences satisfying the conditions (5). Assume that $E[e_t^{4+2\delta}] < M < \infty$ for some M and $\delta > 0$. Let

$$\hat{\underset{\sim}{\theta}}_{(u)n} = [\Sigma_{t=1}^{n} \underset{\sim}{X}'_{(u)tn}\underset{\sim}{X}_{(u)tn}]^{-1}[\Sigma_{t=1}^{n} \underset{\sim}{X}'_{(u)tn}Y_t] ,$$

where $\underset{\sim}{X}_{(u)tn}$ is defined in (16) of FHG. Let $\underset{\sim}{D}_{(u)n}$ be the diagonal matrix whose elements are the square roots of the diagonal elements of $\Sigma_{t=1}^{n} \underset{\sim}{X}'_{(u)tn}\underset{\sim}{X}_{(u)tn}$. Let

$$\underset{\sim}{G}_{(u)n} = \underset{\sim}{D}_{(u)n}^{-1} \Sigma_{t=1}^{n} \underset{\sim}{X}'_{(u)tn}\underset{\sim}{X}_{(u)tn}\underset{\sim}{D}_{(u)n}^{-1} ,$$

$$\hat{\underset{\sim}{\delta}}_n = \sigma^{-1} \underset{\sim}{G}_{(u)n}^{1/2}\underset{\sim}{D}_{(u)n}(\hat{\underset{\sim}{\theta}}_{(u)n} - \underset{\sim}{\theta}_{(u)n}) ,$$

where $\underset{\sim}{G}_{(u)n}^{1/2}$ is the positive definite square root of $\underset{\sim}{G}_{(u)n}$ and $\underset{\sim}{\theta}_{(u)n}$ is defined in (16) of FHG. Assume that the conditions (10), (11), (17), (18) and (19) of FHG hold. then the last element of $\hat{\underset{\sim}{\delta}}_n$ converges in distribution to the statistic $\hat{\tau}_\mu$ for case (a) and to $\hat{\tau}_\tau$ for case (b), where $\hat{\tau}_\mu$ and $\hat{\tau}_\tau$ are characterized in Dickey and Fuller (1979). The limiting distribution of the remaining $q + p - 1$ elements of $\hat{\underset{\sim}{\delta}}_n$ is normal with zero mean and identity covariance matrix for both cases.

Proof. Using the arguments of FHG, it follows that the first q elements of the last row of $\underset{\sim}{G}_{(u)n}$ converge in probability to zero. To show that the next $(p-1)$ elements of the last row of $\underset{\sim}{G}_{(u)n}$ converge in probability to zero, we need to establish that

$$\Sigma_{t=1}^{n}(\Sigma_{k=0}^{t-i} v_k^* e_{t-k})(\Sigma_{s=1}^{t} e_s) = O_p(n)$$

for $i = 1,2,\ldots, j$ and $j = 1,2,\ldots, p-1$, where v_k^* are defined in FHG. Note that

$$\Sigma_{t=1}^{n}(\Sigma_{k=0}^{t-i} v_k^* e_{t-k})(\Sigma_{s=1}^{t} e_s)$$

$$= \Sigma_{t=1}^{n} \Sigma_{\ell=1}^{t} v_{t-\ell}^* e_\ell^2 + \Sigma_{t=1}^{n} \Sigma_{\ell=i}^{t} v_{t-\ell}^* \Sigma_{\substack{k=1 \\ \neq \ell}}^{t} e_k e_\ell .$$

Since $\Sigma_{\ell=0}^{\infty} |v_\ell^*| < \infty$, the first term is of $O_p(n)$. Because

$$E(e_k e_\ell e_s e_r) = 0 ,$$

if at most two subscripts are equal and because the e_t^2 are uncorrelated, it follows that

$$\mathrm{Var}(\Sigma_{t=1}^{n} \Sigma_{\ell=i}^{t} v_{t-\ell}^* e_\ell \Sigma_{\substack{k=1 \\ \neq \ell}}^{t} e_k) = O(n^2) .$$

Therefore, the first $q + p - 1$ elements of the elements of $\underset{\sim}{G}_{(u)n}$

converge in probability to zero. The last element of $\underset{\sim}{D}^{-1}_{(u)n} \Sigma^n_{t=1} \underset{\sim}{X}'_{(u)tn} e_t$ is

$$(9) \quad \frac{\Sigma^n_{t=1} W^+_{tpn} e_t}{[\Sigma^n_{t=1}(W^+_{tpn})^2]^{1/2}} = \frac{n^{-1} \Sigma^n_{t=1} u^+_{t-1} e_t}{[n^{-2} \Sigma^n_{t=1}(u^+_{t-1})^2]^{1/2}} + o_p(1) ,$$

where $u^+_t = A^+_{tn} + B^+_{tn}$ and A^+_{tn} and B^+_{tn} are defined in FHG. From the results of Theorem 1 the limiting distribution of the first quantity on the right of (9) is that of the $\hat{\tau}_\mu$-statistic for case (a) and that of $\hat{\tau}_\tau$ for case (b).

From the results of Crowder (1980), it follows that the limiting distribution of the remaining elements of $\underset{\sim}{D}^{-1}_{(u)n} \Sigma^n_{t=1} \underset{\sim}{X}'_{(u)tn} e_t$ is multivariate normal. The condition $E[e_t^{4+2\delta}] < M < \infty$ is sufficient to establish the Lindeberg condition of Crowder (1980). []

Similar arguments can be used to extend Theorem 3 of FHG to the case where $\{e_t\}$ is a sequence of martingale differences with constant conditional variance and bounded $(4+2\delta)$-th moments.

Acknowledgements

I wish to express my thanks to Dr. Wayne Fuller for helpful suggestions.

REFERENCES

Anderson, T. W. (1959). On asymptotic distributions of estimates of parameters in stochastic difference equations. Ann. Math. Statist. **30**: 676-687.

Anderson, T. W., and Taylor, J. B. (1979). Strong consistency of least squares estimates in dynamic models. The Annals of Statistics **7**: 484-489.

Crowder, M. J. (1980). On the asymptotic properties of least squares estimators in autoregression. The Annals of Statistics **8**: 132-146.

Dickey, D. A. (1976). Hypothesis testing for nonstationary time series. Unpublished Ph.D. thesis. Iowa State University, Ames, Iowa.

Dickey, D. A., and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. J. Amer. Statist. Assoc. **74**: 427-431.

Fuller, W. A. (1976). Introduction to Statistical Time Series. Wiley, New York.

Fuller, W. A., Hasza, D. P., and Goebel, J. J. (1981). Estimation of the parameters of stochastic difference equations. The Annals of Statistics **9**: 531-543.

Hall, P., and Heyde, C. C. (1980). Martingale Limit Theory And Its Application. Academic Press, New York.

Hannan, E. J., and Heyde, C. C. (1972). On limit theorems for quadratic functions of discrete time series. Ann. Math. Statist. **43**: 2058-2066.

Hannan, E. J., and Nicholls, D. F. (1972). The estimation of mixed regression, autoregression, moving average and distributed lag models. Econometrica **40**: 529-548.

Mann, H. B., and Wald, A. (1943). On the statistical treatment of linear stochastic difference equations. Econometrica **11**: 173-220.

Pantula, S. G. (1982). Properties of estimators of the parameters of autoregressive time series. Unpublished Ph.D. Thesis. Iowa State University, Ames, Iowa.

Reinsel, G. (1976). Maximum likelihood estimation of stochastic linear difference equations with autoregressive moving average errors. Technical Report 112, Dept. Statist., Carnegie-Mellon Univ., Pittsburgh.

Scott, D. J. (1973). Central limit theorems for martringales and for proceses with stationary increments using a Skorokhod representation approach. Advances Appl. Probability 5: 119-137.

White, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. Ann. Math. Statist. 29: 1188-1197.