LEAST-SQUARES ESTIMATION OF AUTOREGRESSIONS WITH SOME UNIT ROOTS

by

Christopher A. Sims

Discussion Paper No. 78-95, March 1978

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, Minnesota 55455

LEAST-SQUARES ESTIMATION OF AUTOREGRESSIONS

WITH SOME UNIT ROOTS

by Christopher A. Sims, February 1978

It is a widespread practice in applied time series analysis to difference series one or more times before modeling them when it appears that without differencing a model to fit the series would have to be non-stationary. In some applications this procedure can be justified as based on prior knowledge of a physical or behavioral mechanism [1/]; but it is my impression that a more important reason for the procedure's popularity is the widespread impression that if the true model is non-stationary, the usual asymptotic distribution theory for its estimated coefficients breaks down.

Now the usual asymptotic distribution theory does break down in this case, as has been known since the work of T. W. Anderson (1959) and J. S. White (1958), who studied the case of a first-order autoregression with a unit coefficient on the lagged variable. One cannot estimate a first-order autoregression and use the usual distribution theory to test the null hypothesis that the coefficient is one. However this case is not relevant to the issue of preliminary differencing of the data, since once the unit

---

1/ E.g. in modeling price time series in economics where rational market behavior can be argued to imply random walk behavior for the price process.

coefficient is assumed and the data differenced, there is no further estimation to be done. Where differencing might seem more useful is in a model with <u>some</u> unstable roots, e.g.

$$1) \qquad y(t) = ay(t-1) + by(t-2) + u(t) ,$$

where $a + b = 1$. Even with stationary $u(t)$, $y(t)$ in this model is non-stationary, and OLS estimates of a and b do not have a limiting nonsingular joint normal distribution. If we first-difference the data, we obtain

$$2) \qquad y(t) - y(t-1) = -b(y(t-1) - y(t-2)) + u(t) ,$$

wherein both the differenced y's and u are stationary; we have avoided non-stationarity by using (2).

The point of this paper is that the estimation of (1) by OLS, when u is serially uncorrelated and stationary, yields estimates for which the usual asymptotic distribution theory breaks down in a convenient way. In particular, if $\hat{a}$ and $\hat{b}$ are OLS estimates of the parameters of (1) with $a + b = 1$, every linear combination of $\hat{a}$ and $\hat{b}$ <u>except</u> $\hat{a} + \hat{b}$ has an asymptotic normal distribution which is correctly estimated by the usual formulas. Thus, e.g., whether or not $a + b = 1$, we are justified in testing the hypothesis that $b = 0$ by the usual t-statistic. Further, leaving the unit root unconstrained has no effect on the asymptotic distribution of other linear combinations of $\hat{a}$ and $\hat{b}$.

One way to explain this apparently paradoxical result is to consider the sample covariance matrix of the right-hand-side variables in (1), which is proportional to the usual estimate of the variance-covariance matrix of $\hat{a}$ and $\hat{b}$. As sample size increases, the ratio of the larger to the smaller eigenvalue of this matrix increases without bound. If we scale the matrix by a scalar so that its inverse converges to a limit, that limit turns out to be singular. We can instead scale the matrix by a matrix-valued function of sample-size so that the "limit" is non-singular (and not a constant). Correspondingly, it turns out, we can scale the vector $(\hat{a}, \hat{b})$ by a scalar so that it converges to a limiting singular normal distribution, or instead by a matrix so that it converges to a limiting non-normal joint distribution. Unless we are interested in that linear combination of $\hat{a}$ and $\hat{b}$ which has zero variance in the limiting singular distribution, we can legitimately use the limiting singular distribution for forming confidence regions and hypothesis tests.

Some Lemmas

To demonstrate these results we employ repeatedly the following two lemmas. The notation "$E_{t-1}[u(t)]$" means "$E[u(t)|u(0),\ldots,u(t-1)]$".

Lemma 1: Suppose:

i) $z(t) = \sum_{s=0}^{t} u(t)$ and $x(t) = \sum_{s=0}^{t} c(t-s)u(s)$;

ii) $E_{t-1}(u(t)) = 0$; $E_{t-1}(u(t)^2) = \sigma^2$; $u(t)$ an ergodic process;

iii) $E(u(t)^4) = \mu_4$;

iv) $|a(t)| < B(t)$; $B(t)$ monotone decreasing;

$$\sum_{t=0}^{\infty} B(t) < \infty;$$

v) $$V_T = \sum_{s=0}^{T} z(t)x(t).$$

Then $\mathrm{Var}(T^{-1}V_T)$ is bounded.

Lemma 2: Under assumptions (i) - (iii) of Lemma 1, $T^{-2}\sum_{s=0}^{T} z(t)^2$ converges in distribution to the distribution of $\int_0^1 W(s)^2 ds$, where $W(s)$ is a Weiner process with $W(0) = 0$.

The proof of the first lemma is postponed to an appendix, as it is fairly long. The second lemma is a corollary of the martingale central limit theorem for processes with ergodic increments, as stated, e.g., in Scott (1973) or Billingsley (1968).[2/] The quantity whose convergence in distribution is asserted is a fixed, well-behaved functional of the functions whose distribution is asserted by the martingale CLT to converge to that of a Weiner process on (0,1). We need to be able to apply these lemmas to a slightly more general class of processes:

Corollary: Lemmas 1 and 2 continue to hold it z is replaced by $z^*(t) = z(t) + z_0(t)$ and x by $x^*(t) = x(t) + x_0(t)$, where $E(z_o(t)^2)$ is bounded, $E(x_0(t)^2)t$ is bounded, and $x_0(s)$, $z_0(s)$, for all s, are included in the conditioning set for the conditional

[2/] Recent references in this area are Rootzen (1977) and Hall (1977).

expectations in (ii) of Lemma 1.

The corollary is easily verified and the proof omitted. The processes x_0 and z_0 represent the effects of initial conditions.

Now if (1) has held from time $t = 0$ onward, and if u in (1) satisfies the conditions of the lemmas, then it easily is seen that $y(t)$ can be written as the sum of two components, one non-stationary and behaving as z^* in the corollary, the other stationary and behaving as x^*. Using this fact and the lemmas, it is easy to prove the following result, which is really a digression from the path toward our main result:

Proposition 1: If (1) has held from time $t = 0$ onward, and if u in (1) satisfies the conditions of Lemma 1, then C_T, the matrix of cross products of the right-hand-side variables in (1), defined by

$$C_T = \begin{bmatrix} \Sigma y(t-1)^2 & \Sigma y(t-1)y(t-2) \\ \Sigma y(t-1)y(t-2) & \Sigma y(t-2)^2 \end{bmatrix},$$

where the limits of summation are $t = 1$ to $t = T$, satisfies

$$(\Sigma y(t-1)^2)^{-1} \det(C_T)\, C_T^{-1} \overset{P}{\to} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} .$$

The proof is straightforward and is omitted.

The Main Result

Theorem: If $u(t)$ in (1) satisfies (ii) and (iii) of Lemma 1 and if $S_T = c_1\hat{a}_T + c_2\hat{b}_T - (c_1a + c_2b)$, where $\hat{a}_T$ and $\hat{b}_T$ are OLS estimates of a and b from (1) from a sample of size T, then $T^{-\frac{1}{2}}S_T$ has a limiting normal distribution with mean 0 and variance $\sigma^2(c_1 - c_2)^2$.

The proof of this result, though not difficult, is long, so it is postponed to the appendix.

Note that, this result, together with the earlier Proposition, implies that the usual formulas for confidence intervals and tests on $c_1a + c_2b$ will be asymptotically correct, unless $c_1 = c_2$. Furthermore, test statistics found from the regression (1), with levels data, have the same asymptotic distribution as those formed from (2), using differenced data.

Conclusions

This paper deals explicitly only with the second-order univariate autoregression of equation (1). My main motivation for exploring this problem is applied work with vector autoregressions involving many variables and many lags. It remains to be shown, of course, that high-order lags and multivariate systems would yield the same qualitative conclusions as this simple system. Nonetheless it appears to me that the main complication to be expected is that the linear combinations of estimated parameters which have singular distributions become harder to list. The practical conclusion, that the conservative procedure when roots on the unit circle are suspected is to estimate a model which leaves those roots unconstrained, seems to me likely to generalize. It is likely to generalize also, I would guess, to the nonlinear autoregressions generated by "ARIMA" models discussed, e.g., in Box and Jenkins (1970).

Appendix

Proof of Lemma 1

We can write $\mathrm{Var}(V_T) = \sum_{s=1}^{T} \sum_{r=1}^{T} E[x(s)z(s)x(r)z(r)]$. Consider the element of this sum with index s, r. Using the expressions in (i) for z and x in terms of u let us write this s, r element as

$$E[\sum_{q_1=0}^{s} \sum_{q_2=0}^{s} a(q_1)u(s-q_1)u(s-q_2) \sum_{p_1=0}^{r} \sum_{p_2=0}^{r} a(p_1)u(r-p_1)u(r-p_2)].$$

The expected value of the product of the u's in this sum is, by (ii), 0 unless two pairs of u-arguments are equal. This can occur in the following three ways: (a) $q_1 = q_2$, $p_1 = p_2$; (b) $s-q_1 = r-p_1$, $s-q_2 = r-p_2$; and (c) $s-q_1 = r-p_2$, $s-q_2 = r-p_1$. The three cases overlap in that all three include the case where all four u-arguments are identical. The sum of all case - (a) terms is bounded by

$$\text{a)} \qquad \sum_{q=0}^{s} \sum_{p=0}^{s} a(q)a(p)\mu_4 \leq \mu_4(\sum_{q=0}^{\infty} |a(q)|)^2 = K_a$$

For case (b) the corresponding bound is, for $s \geq r$,

$$\text{b)} \qquad \sum_{q_1=s-r}^{s} \sum_{q_2=s-r}^{s} a(q_1)a(q_1 + r-s)\mu_4 = r\mu_4 \sum_{q_1=s-r}^{s} a(q_1)a(q_1 + r-s)$$

$$\leq B(s-r)r\mu_4(\sum_{s=0}^{\infty} B(s)) = rB(s-r)K_b.$$

And for case (c) the bound is

c) $$\sum_{q_1=0}^{s} \sum_{q_2=r-s}^{s} a(q_1)a(q_2 + r-s)\mu_4 \leq \mu_4 \sum_{s=0}^{\infty} |a(q)|^2 = K_c.$$

Thus

$$\mathrm{Var}(T^{-1}V_T) \leq T^{-2} \sum_{s=1}^{T} \sum_{r=1}^{T} K_a + K_b \min(s,r)B(|s-r|) + K_c$$

$$= K_a + K_c + T^{-2}K_b \sum_{s=1}^{T} \sum_{r=1}^{T} \min(s,r)B(|s-r|)$$

$$= K_a + K_c + T^{-2}K_b \sum_{s=1}^{T} \left(\sum_{r=1}^{s} rB(s-r) + \sum_{r=s+1}^{T} s\, B(r-s)\right)$$

$$\leq K_a + K_c + T^{-2}K_b \sum_{s=1}^{T} 2T\left(\sum_{s=0}^{\infty} B(s)\right) = K_a + K_c + 2K_b\left(\sum_{s=0}^{\infty} B(s)\right).$$

Q.E.D.

<u>Proof of the Theorem</u>.

$$S_T = (\det(C_T))^{-1}[\{c_1\Sigma y(t-2)^2 - c_2\Sigma y(t-1)y(t-2)\}\ \Sigma y(t-1)u(t)$$

$$+ \{c_2\Sigma y(t-1)^2 - c_1\Sigma y(t-1)y(t-2)\}\ \Sigma y(t-2)u(t)].$$

$$= [\det(C_T)]^{-1}[\{c_1\Sigma y(t-2)^2 - c_2\Sigma y(t-1)^2\}\ \Sigma\{y(t-1)-y(t-2)\}u(t)$$

$$+ c_1\Sigma y(t-2)\{y(t-2)-y(t-1)\}\ \Sigma y(t-2)u(t)$$

$$+ c_2\Sigma y(t-1)\{y(t-1)-y(t-2)\}\ \Sigma y(t-1)u(t)].$$

Now observe that $T\Sigma y(t-2)^2/\det(C_T) \xrightarrow{P} [\mathrm{Var}(y(t)-y(t-1))]^{-1}$.

To see this, write

$$T\Sigma y(t-2)^2/\det(C_T) = T[\Sigma y(t-1)^2 - (\frac{\Sigma y(t-1)y(t-2)}{\Sigma y(t-2)^2})\ \Sigma y(t-1)y(t-2)]^{-1}$$

$$= T[\Sigma y(t-1)\{y(t-1)-y(t-2)\} + \frac{\Sigma\{y(t-2)-y(t-1)\}y(t-2)}{\Sigma y(t-2)^2}\ \Sigma y(t-1)y(t-2)]^{-1}$$

$$= T[\Sigma\{y(t-1)-y(t-2)\}^2 + \frac{\Sigma\{y(t-1)-y(t-2)\}y(t-2)\Sigma y(t-2)\{y(t-2)-y(t-1)\}}{\Sigma y(t-2)^2}]^{-1}$$

$$= [\frac{1}{T}\ \Sigma\{y(t-1)-y(t-2)\}^2 - \frac{1}{T}\ \frac{[\Sigma\{y(t-1)-y(t-2)\}y(t-2)/T]^2}{\Sigma y(t-2)^2/T^2}\ ]^{-1} .$$

The first term inside the bracket in the last line above converges a.s. to $\mathrm{Var}(y(t-1)-y(t-2))$ by the ergodicity of $u(t)$, while the second converges in probability to zero by Lemmas 1 and 2.

This suggests rewriting S_T once again to get

$$\begin{aligned}\sqrt{T}\ S_T = {} & [T\Sigma y(t-2)^2/\det(C_T)] \\ & \cdot [(c_1 - \{\Sigma y(t-1)^2/\Sigma y(t-2)^2\}c_2)\ \Sigma\{y(t-1)-y(t-2)\}u(t) \\ & + T^{-\frac{1}{2}}\ c_1\Sigma y(t-2)\{y(t-2)-y(t-1)\}\ \Sigma y(t-2)u(t)/\Sigma y(t-2)^2 \\ & + T^{-\frac{1}{2}}\ c_2\Sigma y(t-1)\{y(t-1)-y(t-2)\}\ \Sigma y(t-1)u(t)/\Sigma y(t-2)^2].\end{aligned}$$

Now the first line above we have just found to converge in probability to $1/\mathrm{Var}(y(t)-y(t-1))$. The second line, first term obviously converges in probability to $c_1 - c_2$, while the second term converges in distribution by the martingale CLT to a

$N(0, \sigma^2 Var(y(t)-y(t-1))$ variable. The last two lines converge in probability to zero, as can be seen because Lemma 1 applies to both factors in their numerators and Lemma 2 to their denominators. Putting these conclusions together yields $\sqrt{T}\ S_T$ converges in distributions to $N(0, \sigma^2(c_1 - c_2)^2)$.

Q.E.D.

References

1. Anderson, T. W. (1959), "On Asymptotic Distributions of Estimates of Parameters of Difference Equations", Annals of Math. Stat., 30, 676-687.

2. Billingsley, P. (1968), Convergence of Probability Measures, Wiley, New York.

3. Box, G. E. P., and G. M. Jenkins (1970), Time Series Analysis: Forecasting and Control (San Francisco, Cambridge, London, Amsterdam: Holden-Day.

4. Hall, Peter (1977), "Martingale Invariance Principles", Annals of Probability, 5, 875-887.

5. Rootzen, H. (1977), "On the Functional Central Limit Theorem For Martingales", Z. Warscheinlichkeitstheorie verw. Gebiete, 38, 199-210.

6. Scott, D. J. (1973), "Central Limit Theorems for Martingales and for Processes with Stationary Increments Using a Skorokhod Representation Approach", Advances in Applied Probability, 5, 119-137.

7. White, J. S. (1958), "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case", Annals of Math. Stat., 29, 1188-1197.