iff

$$\Omega\mathfrak{M} = \mathfrak{M}.$$

Comment on these results.

REFERENCES

1. Kruskal, W. When are Gauss Markov and least squares estimators identical: A coordinate free approach. *Annals of Mathematical Statistics*, 39 (1968): 70–75.
2. Malinvaud, E. *Statistical methods of econometrics*. Amsterdam: North Holland, 1980.

90.4.8. *A Best Linear Unbiased Estimator of $R\beta$ with a Scalar Variance Matrix*, proposed by R.W. Farebrother. Consider the model

$$y = X\beta + \epsilon \qquad \epsilon \sim N(0, \sigma^2 I_n)$$

where y is an $n \times 1$ matrix of observations on the dependent variable, X is an $n \times k$ matrix of observation on the k regressors, β is a $k \times 1$ matrix of parameters, and ϵ is an $n \times 1$ matrix of disturbances, and where X has full column rank k.

Let R be an $m \times k$ matrix of fixed numbers of rank $p \leq m$; then the reader is asked to derive (and suggest possible applications for) a linear unbiased estimator of $R\beta$ with variance proportional to $\sigma^2 I_m$ and such that the constant of proportionality is as small as possible.

SOLUTIONS

89.3.1. *The Asymptotic Distribution of the Iterated Gauss-Newton Estimators of an ARIMA Process* — Solution,[1] proposed by Juan J. Dolado and Javier Hidalgo-Moreno. Let $\{y_t\}$ have the following DGP:

$$y_t = \rho y_{t-1} + e_t + \beta e_{t-1} \tag{1}$$

where $\rho = 1$, $|\beta| < 1$, $e_0 = y_0 = 0$, $\{e_t\} \sim$ n.i.d.(0,1). Let $\theta' = (\rho, \beta)$ and define

$$e_t(\theta) = y_t - \rho y_{t-1} - \beta e_{t-1}(\theta) = [1 - \rho L/1 + \beta L]\, y_t, \qquad t \geq 1. \tag{2}$$

Then, expanding $e_t(\theta)$ around an initial estimator $\hat{\theta}$ we get

$$\begin{aligned} e_t(\theta) = {} & e_t(\hat{\theta}) + (\partial e_t(\hat{\theta})/\partial\rho)(\rho - \hat{\rho}) + (\partial e_t(\hat{\theta})/\partial\beta)(\beta - \hat{\beta}) \\ & + 1/2(\partial^2 e_t(\dot{\theta})/\partial\rho^2)(\rho - \hat{\rho})^2 + (\partial^2 e_t(\dot{\theta})/\partial\rho\,\partial\beta)(\rho - \hat{\rho})(\beta - \hat{\beta}) \\ & + 1/2(\partial^2 e_t(\dot{\theta})/\partial\beta^2)(\beta - \hat{\beta})^2. \end{aligned} \tag{3}$$

The first step in a Gauss–Newton iteration procedure is based upon the following regression model obtained as a rearrangement of (3):

$$e_t(\hat{\theta}) = v_{1t}(\hat{\theta})\gamma_1 + v_{2t}(\hat{\theta})\gamma_2 + u_t(\theta, \dot{\theta}) \tag{4}$$

where, by differentiation of (2), we get

$$v_{1t}(\hat{\theta}) = -(\partial e_t(\hat{\theta})/\partial\rho) = (1/1 + \hat{\beta}L)y_{t-1}$$

$$v_{2t}(\hat{\theta}) = -(\partial e_t(\hat{\theta})/\partial\beta) = [(1 - \hat{\rho}L)/(1 + \hat{\beta}L)^2]y_{t-1}$$

$$u_t(\theta,\dot{\theta}) = e_t(\theta) + r_{1t}(\dot{\theta}) + r_{2t}(\dot{\theta}) + r_{3t}(\dot{\theta})$$

$$r_{1t}(\dot{\theta}) = -1/2(\partial^2 e_t(\dot{\theta})/\partial\rho^2)(\rho - \hat{\rho})^2 \equiv 0$$

$$r_{2t}(\dot{\theta}) = -(\partial^2 e_t(\dot{\theta})/\partial\rho\,\partial\beta)(\rho - \hat{\rho})(\beta - \hat{\beta})$$
$$= -(1/(1 + \dot{\beta}L)^2)y_{t-2}(\rho - \hat{\rho})(\beta - \hat{\beta})$$

$$r_{3t}(\dot{\theta}) = -1/2(\partial^2 e_t(\dot{\theta})/\partial\beta^2)(\beta - \hat{\beta})^2$$
$$= -[(1 - \dot{\rho}L)/(1 + \dot{\beta}L)^3]y_{t-2}(\beta - \hat{\beta})^2$$

$$\gamma_1 = (\rho - \hat{\rho})$$

$$\gamma_2 = (\beta - \hat{\beta}).$$

The remainder terms r_{it} ($i = 1,2,3$) are evaluated at a point $\dot{\theta}$ between θ and $\hat{\theta}$ and the initial estimators are taken as $\hat{\rho} = 1$ and $\hat{\beta} = \beta^* = \beta + O_p(T^{-1/4})$. Note that, under the null, by choosing $\hat{\rho} = \rho = 1$, we obtain $r_{2t} = 0$ and, hence, we are only left with the term r_{3t}.

Let us write (4) in vector notation as

$$e_t(\hat{\theta}) = v_t'(\hat{\theta})\gamma + u_t(\theta,\dot{\theta}) \tag{5}$$

where $v_t' = [v_{1t}v_{2t}]$, $\gamma' = [\gamma_1\gamma_2]$, and choose the scaling matrix $\Upsilon_T = \text{diag}(T, T^{1/2})$. Then

$$\Upsilon_T(\hat{\gamma} - \gamma) = [\Upsilon_T^{-1}\Sigma v_t v_t' \Upsilon_T^{-1}]^{-1}\Upsilon_T^{-1}\Sigma v_t u_t. \tag{6}$$

To obtain the limiting distribution of (6), it is convenient to start by showing that the limit of the matrix in brackets is diagonal matrix. The off-diagonal term is

$$T^{-3/2}\Sigma v_{1t}(\hat{\theta})v_{2t}(\hat{\theta}) = T^{-1/2}[T^{-1}\Sigma v_{1t}(\hat{\theta})v_{2t}(\hat{\theta})] = o_p(1) \tag{7}$$

since the term in brackets is $O_p(1)$ (see [1] for the proof that the sum of the products of an $I(1)$ and an $I(0)$ is $O_p(T)$). This argument lies on ρ and β being consistent estimators under the null to be able to apply the continuous mapping theorem [2].

By similar arguments it is easy to show that

$$T^{-2}\Sigma v_{1t}^2(\hat{\theta}) \to \int_0^1 B_e^2(r)\,dr$$

$$T^{-1}\Sigma v_{1t}(\hat{\theta})e_t(\theta) \to \frac{1}{2}\,[B_e^2(1) - 1]$$

where "$\rightarrow$" denotes weak convergence in distribution and $B_e(r)$ is a unit variance Brownian motion generated from e_t (see [1]).

Note that

$$T^{-1}\Sigma v_{1t}(\hat{\theta})r_{3t}(\dot{\theta}) = T^{-1}\Sigma(y_{t-1}/1 + \beta^* L)(\Delta y_{t-2}/(1 + \dot{\beta}L)^3)(\beta - \hat{\beta})^2$$
$$= o_p(1)$$

since $(\beta - \hat{\beta})^2$ is $O_p(T^{-1/2})$ and the product term is $O_p(T)$.

Therefore

$$T\hat{\gamma}_1 \rightarrow \frac{1}{2}\,[B_e^2(1) - 1]\bigg/\int_0^1 B_e^2(r)\,dr$$

and

$$t_\gamma \rightarrow \frac{1}{2}\,[B_e^2(1) - 1]\bigg/\int_0^1 B_e^2(r)\,dr$$

that is, the distributions tabulated in [3].

Using similar arguments it is easy to prove that $T^{1/2}(\hat{\gamma}_2 - \gamma_2)$ is $O_p(1)$ since

$$T^{-1}\Sigma v_{2t}^2(\hat{\theta}) = T^{-1}\Sigma(e_{t-1}/1 + \beta^* L)^2 \rightarrow 1/1 - \beta^2$$

$$T^{-1/2}\Sigma v_{2t}(\hat{\theta})e_t(\theta) \rightarrow N(0, 1/1 - \beta^2)$$

$$T^{-1/2}\Sigma v_{2t}(\hat{\theta})r_{3t}(\dot{\theta})$$
$$= -T^{-1/2}\Sigma(\Delta y_{t-1}/(1 + \beta^* L)^2)(\Delta y_{t-2}/(1 + \dot{\beta}L)^3)(\beta - \hat{\beta})^2 = o_p(1)$$

if $T^{1/2}(\beta - \beta^*)^2$ is $o_p(1)$.

Remark 1. Under the assumption $\beta - \beta^* = O_p(T^{-1/4})$, this last term does not vanish, given that the cross-product term is $O_p(T)$ contrary to the argument given in [4] (expression (2) in p. 374). It tends to a nonzero limit (η) that changes the limiting distribution presented in [4] to $N(\eta(1 - \beta^2), 1 - \beta^2)$.

After this prolegomenon we are now able to answer the posed questions.

1. The second iteration in Gauss-Newton (θ^2) is obtained from (4) now evaluating in the improved estimator $\theta^1 = \hat{\theta} + \hat{\gamma}$. That is,

$$\theta^2 = \theta^1 + (\Sigma v_t(\theta^1)v_t'(\theta^1))^{-1}\Sigma v_t(\theta^1)e_t(\theta^1). \tag{8}$$

Now, expanding $e_t(\theta)$ around θ^1 as in (3) and substituting in (8) we get

$$\theta^2 = \theta + (\Sigma v_t(\theta^1)v_t'(\theta^1))^{-1}\Sigma v_t(\theta^1)\,[e_t(\theta) + r_{2t}(\ddot{\theta}) + r_{3t}(\ddot{\theta})]$$

where $r_t' = [r_{2t}, r_{3t}]$ evaluated at θ, a point between θ^1 and θ.

Since θ^1 is consistent by the arguments given in the prolegomenon, the limiting distributions of $\Upsilon_T(\theta^2 - \theta)$ and $\Upsilon_T(\theta^1 - \theta)$ are identical if we can prove that

$$(\Upsilon_T^{-1}\Sigma v_t(\theta^1)v_t'(\theta^1)\Upsilon_T^{-1})^{-1}\Upsilon_T^{-1}\Sigma v_t(\theta^1)(r_{2t}(\ddot{\theta}) + r_{3t}(\ddot{\theta})) = o_p(1).$$

The limiting distribution of the matrix is a diagonal matrix, as shown before since θ^1 is consistent, and hence is $o_p(1)$. The cross product $\Sigma v_t(\theta^1)r_{3t}(\theta)$ is $o_p(1)$ by the same arguments as before if $T^{1/2}(\beta - \beta^*)^2$ is $o_p(1)$. If this last term is $O_p(1)$, then the limiting distribution corresponds to that given in the Remark 1. Hence, we are only left with the following cross products:

$$T^{-1}\Sigma v_{1t}(\theta^1)r_{2t}(\ddot{\theta})$$
$$= -T^{-1}\Sigma(y_{t-1}/1 + \beta^1 L)(y_{t-2}/(1 + \ddot{\beta}L)^2)(\rho - \rho^1)(\beta - \beta^1)$$

which is $o_p(1)$ since the cross product is $O_p(T^2)$, $(\rho - \rho^1)$ is $O_p(T^{-1})$, and $(\beta - \beta^1)$ is either $o_p(T^{-1/4})$ or $O_p(T^{-1/4})$, and

$$T^{-1/2}\Sigma v_{2t}(\theta^1)r_{2t}(\ddot{\theta})$$
$$= -T^{-1/2}\Sigma(e_{t-1}/1 + \beta^1 L)(y_{t-2}/(1 + \ddot{\beta}L)^2)(\rho - \rho^1)(\beta - \beta^1)$$

which is $o_p(1)$ since the cross product is $O_p(T)$, $(\rho - \rho^1)$ is $O_p(T^{-1})$, and $(\beta - \beta^1)$ is either $O_p(T^{-1/4})$ or $o_p(T^{-1/4})$.

If we interchange θ^1 by θ^n and θ^2 by θ^{n+1}, the argument is identical and therefore the asymptotic distribution of the iterated estimators and their corresponding t-ratios are the same.

2. It is not necessary to choose $\hat{\rho} = 1$. Choose instead $\hat{\rho} = 1 + o_p(T^{-3/4})$, then the limiting distributions of $(\Upsilon_T^{-1}\Sigma v_t(\hat{\theta})v_t'(\hat{\theta})\Upsilon_T^{-1})$, $(\Upsilon_T^{-1}\Sigma v_t(\hat{\theta})e_t(\theta))$, and $(\Upsilon_T^{-1}\Sigma v_t(\hat{\theta})r_{3t}(\dot{\theta}))$ are the same as before, since $\hat{\rho}$ is consistent under the null.

The different terms are now

$$T^{-1}\Sigma v_{1t}(\hat{\theta})r_{2t}(\dot{\theta})$$
$$= -T^{-1}\Sigma(y_{t-1}/1 + \beta^* L)(y_{t-2}/(1 + \dot{\beta}L)^2)(1 - \hat{\rho})(\beta - \beta^*)$$

which is $o_p(1)$ since the cross product is $O_p(T^2)$, $(\beta - \beta^*)$ is $O_p(T^{-1/4})$, and $1 - \hat{\rho}$ is $o_p(T^{-3/4})$ and

$$T^{-1/2}\Sigma v_{2t}(\hat{\theta})r_{2t}(\dot{\theta})$$
$$= -T^{-1/2}\Sigma((1 - \rho L)/(1 + \beta^* L))y_{t-1}(y_{t-2}/(1 + \dot{\beta}L)^2)(1 - \hat{\rho})(\beta - \beta^*)$$

which is also $o_p(1)$ since the cross product is $O_p(T)$, $(\beta - \beta^*)$ is $O_p(T^{-1/4})$, and $(1 - \hat{\rho})$ is $o_p(T^{-3/4})$.

Remark 2. If β^* is such that $\beta^* - \beta$ is $o_p(T^{-1/4})$, then $\hat{\rho}$ can be chosen such that $\hat{\rho} = 1 + O_p(T^{-3/4})$.

3. The previous results extend to ARIMA$(p,1,q)$ processes. The regression model (2) will now contain $p + q + 1$ regressors, reflecting the inclusion

of the additional derivatives. One of the regressors will behave again as an $I(1)$ series and the remaining will be $I(0)$. The previous arguments apply identically to this case.

NOTE

1. A very good solution has been proposed by Sastry G. Pantula (the poser of the problem).

REFERENCES

1. Phillips, P.C.B. Time series regression with a unit root. *Econometrica* 55 (1987): 277–301.
2. Billingsley, P. *Convergence of probability measures.* New York: Wiley, 1968.
3. Fuller, W. *Introduction to statistical time-series.* New York: Wiley, 1976.
4. Said, S.E. & D.A. Dickey. Hypothesis testing in ARIMA$(p,1,q)$ models. *Journal of the American Statistical Association* 80 (1985): 369–374.

89.3.5. *The Limit Distribution of the Generalized Inverse of a Singular Covariance Matrix Estimate*—Solution,[1] proposed by Pietro Balestra.

1. Using Phillips' notation, we have

$$M_{xx} - \Sigma = \frac{1}{n}\sum_i^n (X_i X_i' - \Sigma)$$

or, equivalently,

$$\operatorname{vec}(M_{xx} - \Sigma) = \frac{1}{n}\sum_i^n Z_i, \qquad Z_i = X_i \otimes X_i - \operatorname{vec}\Sigma.$$

Now

$$E(Z_i) = \operatorname{vec}\Sigma - \operatorname{vec}\Sigma = 0 \qquad V(Z_i) = E(Z_i Z_i')$$

$$= E(X_i X_i' \otimes X_i X_i') - (\operatorname{vec}\Sigma)(\operatorname{vec}\Sigma)' = (I + K)(\Sigma \otimes \Sigma)$$

where K is the commutation matrix. Therefore

$$Z_i \sim \text{i.i.d.}(0, (I + K)(\Sigma \otimes \Sigma))$$

and consequently

$$\sqrt{n}\operatorname{vec}(M_{xx} - \Sigma) = \frac{1}{\sqrt{n}}\sum_i^n Z_i \xrightarrow{\mathfrak{D}} N(0, (I + K)(\Sigma \otimes \Sigma)).$$

2. If Σ is positive definite,

$$M_{xx}^{-1} - \Sigma^{-1} = -M_{xx}^{-1}(M_{xx} - \Sigma)\Sigma^{-1}$$

$$\sqrt{n}\operatorname{vec}(M_{xx}^{-1} - \Sigma^{-1}) = -(\Sigma^{-1} \otimes M_{xx}^{-1})\sqrt{n}\operatorname{vec}(M_{xx} - \Sigma)$$

and its limiting distribution is normal with zero mean and covariance matrix

$$(\Sigma^{-1} \otimes \Sigma^{-1})(I + K)(\Sigma \otimes \Sigma)(\Sigma^{-1} \otimes \Sigma^{-1}) = (I + K)(\Sigma^{-1} \otimes \Sigma^{-1}).$$