90.3.5. *Optimal Structural Estimation of Triangular Systems: II. The Nonstationary Case*—Solutions. Three solutions have been proposed independently by Peter C.B. Phillips (the poser of the problem), Juan J. Dolado, and H. Peter Boswijk. These solutions provide a correct answer to the problem. In addition, they each contain interesting derivations and provide additional insights into the results obtained. Faced with the very difficult task to chose among these solutions, the Editor solved this dilemma by publishing all three solutions.

1. Solution—proposed by Peter C.B. Phillips, Yale University.

PART (a). Setting $x_t = t$, we start with the 2SLS estimator $\hat{\beta} = (y_2' P_t y_2)^{-1}(y_2' P_t y_1)$. Since

$$n^{-3} y_2' P_t y_2 \underset{p}{\rightarrow} (\tfrac{1}{3})\gamma^2$$

and

$$n^{-3/2} y_2' P_t u_1 \underset{d}{\rightarrow} N(0, (\tfrac{2}{3})\sigma^2\gamma^2)$$

we find

$$n^{3/2}(\hat{\beta} - \beta) \underset{d}{\rightarrow} N(0, 6\sigma^2/\gamma^2). \tag{6}$$

Next, OLS on (1)′ yields $\tilde{\beta} = (y_2' Q_t y_2)^{-1}(y_2' Q_t (y_1 - y_2))$. Now

$$n^{-1} y_2' Q_t y_2 = n^{-1} u_2' Q_t u_2 \underset{p}{\to} \sigma^2$$

and

$$n^{-1/2} y_2' Q_t v = n^{-1/2} u_2' Q_t v \underset{d}{\to} N(0, \sigma^4).$$

Thus

$$n^{1/2}(\tilde{\beta} - \beta) \underset{d}{\to} N(0,1), \tag{7}$$

giving the same limit distribution as in the stationary case. Finally, for the MLE we have, as before,

$$\beta^+ - \beta = [y_2'(2I - P_t)y_2]^{-1}[2y_2'u_1 - y_2'P_t u_1 - 2u_2'(I - P_t)u_2]$$

and

$$n^{-3} y_2'(2I - P_t)y_2 = n^{-3} y_2' y_2 + n^{-3} y_2'(I - P_t)y_2 \underset{p}{\to} (\tfrac{1}{3})\gamma^2$$

$$\begin{aligned} n^{-3/2}&[2y_2'u_1 - y_2'P_t u_1 - 2u_2'(I - P_t)u_2] \\ &= \gamma(n^{-3/2}\Sigma_1^n t u_{1t}) + o_p(1) \underset{d}{\to} N(0, (\tfrac{2}{3})\sigma^2\gamma^2). \end{aligned}$$

Thus

$$n^{3/2}(\beta^+ - \beta) \underset{d}{\to} N(0, 6\sigma^2/\gamma^2). \tag{8}$$

REMARKS

(i) Note that the limit distributions (6) and (8) are the same, so that econometrician A is right in asserting that 2SLS is optimal in this case. This outcome relies critically on $\gamma \neq 0$, in which case both variables y_{2t} and y_{1t} carry a deterministic trend. The trend in y_{2t}, in particular, dominates its asymptotic behavior and thereby overrules the effect of the covariance between y_{2t} and u_{1t}. In fact, OLS and 2SLS on equation (1) are asymptotically equivalent, since

$$n^{-3} y_2' y_2 \underset{p}{\to} (\tfrac{1}{3})\gamma^2$$

and

$$n^{-3/2} y_2' u_1 = \gamma n^{-3/2} \Sigma_1^n t u_{1t} + o_p(1) \underset{d}{\to} N(0, (\tfrac{2}{3})\sigma^2\gamma^2),$$

as for the components of 2SLS.

(ii) We observe that knowledge of Σ_0 has no effect on the asymptotic distribution of the MLE, since (8) and (6) are equivalent. This contradicts traditional theory (e.g., Rothenberg [1], p. 74) whereby restrictions on the error covariance matrix in a simultaneous equations model generally lead to improved efficiency when the information is used in maximum likelihood estimation. This is due to the fact that the coefficient β is estimated at a higher rate of consistency than Σ and the information matrix turns out to be block diagonal when partitioned conformably with these parameters.

(iii) The augmented regression estimator $\tilde{\beta}$ is consistent for β but at a slower rate than $\hat{\beta}$ and β^+. Thus, econometrician B would be wrong in this case, at least unless $\gamma = 0$. Note that the reason for the slower rate of convergence of $\tilde{\beta}$ is that the presence of the regressor $x_t = t$ in (1)′ ensures that the variables y_{1t} and y_{2t} are effectively detrended, thereby reducing the stochastic order of magnitude of their sample second moments.

PART (b). The 2SLS estimator is $\hat{\beta} = (y_2' P_{-1} y_2)^{-1}(y_2' P_{-1} y_1)$, with P_{-1} representing the projection matrix on the space spanned by the observations of the instrument y_{2t-1}. We have

$$n(\hat{\beta} - \beta) = (n^{-2} y_2' P_{-1} y_2)^{-1}(n^{-1} y_2' P_{-1} u_1),$$

$$n^{-2} y_2' P_{-1} y_2 = (n^{-2} y_2' y_{2-1})(n^{-2} y_{2-1}' y_{2-1})^{-1}(n^{-2} y_2' y_{2-1}) \underset{d}{\rightarrow} \int_0^1 B_2^2,$$

and

$$n^{-2} y_2' P_{-1} u_2 = (n^{-2} y_2' y_{2-1})(n^{-2} y_{2-1}' y_{2-1})^{-1}(n^{-1} y_{2-1}' u_1) \underset{d}{\rightarrow} \int_0^1 B_2\, dB_1,$$

where $(B_1, B_2) \equiv \mathrm{BM}(\Sigma)$, i.e. vector Brownian motion with covariance matrix Σ (see Phillips [2] for a review of the weak convergence methods by which these limit results are obtained). It follows that

$$n(\hat{\beta} - \beta) \underset{d}{\rightarrow} \left(\int_0^1 B_2^2\right)^{-1} \int_0^1 B_2\, dB_1 = \left(\int_0^1 B_2^2\right)^{-1} \left(\int_0^1 B_2\, dB_{1\cdot 2}\right) + \left(\int_0^1 B_2^2\right)^{-1} \left(\int_0^1 B_2\, dB_2\right) \tag{9}$$

where we use the decomposition (see Lemma 3.1 of Phillips [3]) $B_1 = \sigma_{21}' \Sigma_{22}^{-1} B_2 + B_{1\cdot 2} = B_2 + B_{1\cdot 2}$ in which $B_{1\cdot 2} \equiv \mathrm{BM}(\sigma_{11\cdot 2}) = \mathrm{BM}(\sigma^2)$ is independent of B_2. (Observe that $\sigma_{21} = \Sigma_{22} = \sigma^2$ here, so that $\sigma_{11\cdot 2} = \sigma_{11} - \sigma_{21}' \Sigma_{22}^{-1} \sigma_{21} = \sigma^2$ also.) The first term on the far right of (9) is a mixture of normals and the second term is a unit root distribution as in Phillips [4].

From the OLS regression on (1)″ we set $\tilde{\beta} = (y_2' y_2)^{-1}(y_2'(y_1 - \Delta y_2))$. Thus

$$n(\tilde{\beta} - \beta) = (n^{-2} y_2' y_2)^{-1}(n^{-1} y_2'(u_1 - u_2)) \underset{d}{\rightarrow} \left(\int_0^1 B_2^2\right)^{-1} \left(\int_0^1 B_2\, dB_{1\cdot 2}\right). \tag{10}$$

Thus the limit distribution of $\tilde{\beta}$ is mixed normal and avoids the unit root component that is present in (9).

REMARKS

(i) Both $\tilde{\beta}$ and $\hat{\beta}$ are consistent for β at the rate $1/n$. However the limit distribution for $\hat{\beta}$ involves a second-order bias effect, from the presence of the unit root component in (9), and this leads to a mislocation and asymmetry of the limit distribution. These effects are systematically studied in Phillips [5] and the distributional differences between $\tilde{\beta}$ and $\hat{\beta}$ in finite samples are explored by simulation exercises in Phillips and Loretan [6].

(ii) The estimator $\tilde{\beta}$ is the maximum likelihood estimator of β. This is because the likelihood function factors into a component based on the joint density of $(v_{1t})_1^n$ and a component based on the joint density of $(u_{2t})_1^n$. Since the latter component carries no information about β, the

MLE is based on maximizing the first component which in turn reduces minimizing the residual sum of squares $\Sigma_1^n v_{1t}^2$, that is, OLS on (1)″. Thus, econometrician B's estimator $\tilde{\beta}$ is the preferred choice in this case.

PART (c). As in Part (a), econometrician A recommends $\hat{\beta} = (y_2'P_t y_2)^{-1} \times (y_2'P_t y_1)$. Set $\underline{t} = (1, \ldots, n)$ and then

$$n^{-2}y_2'P_t y_2 = (n^{-5/2}y_2'\underline{t})(n^{-3}\underline{t}'\underline{t})^{-1}(n^{-5/2}y_2'\underline{t})$$
$$\underset{d}{\rightarrow} \left(\int_0^1 rB_2\right)\left(\frac{1}{3}\right)^{-1}\left(\int_0^1 rB_2\right) = 3\left(\int_0^1 rB_2\right)^2,$$
$$n^{-1}y_2'P_t u_1 = (n^{-5/2}y_2'\underline{t})(n^{-3}\underline{t}'\underline{t})^{-1}(n^{-3/2}\underline{t}'u_1)$$
$$\underset{d}{\rightarrow} \left(\int_0^1 rB_2\right)\left(\frac{1}{3}\right)^{-1}\left(\int_0^1 r\,dB_1\right),$$

and

$$n(\hat{\beta} - \beta) \underset{d}{\rightarrow} \left(\int_0^1 rB_2\right)^{-1}\left(\int_0^1 r\,dB_1\right). \tag{11}$$

Next, OLS on (1)′ gives $\tilde{\beta} = (y_2'Q_t y_2)^{-1}(y_2'Q_t(y_1 - y_2))$ as before, but $y_1 = \beta y_2 + u_t$ so that

$$\tilde{\beta} - \beta = (y_2'Q_t y_2)^{-1}(y_2'Q_t u_1) - 1 \underset{p}{\rightarrow} -1$$

since

$$n^{-2}y_2'Q_t y_2 = n^{-2}y_2'y_2 - (n^{-5/2}y_2'\underline{t})(n^{-3}\underline{t}'\underline{t})^{-1}(n^{-5/2}\underline{t}'y_2) \underset{d}{\rightarrow} \int_0^1 \underline{B}_2^2$$

and

$$n^{-2}y_2'Q_t u_1 = n^{-1}y_2'u_1 - (n^{-5/2}y_2'\underline{t})(n^{-3}\underline{t}'\underline{t})^{-1}(n^{-3/2}\underline{t}'u_1) \underset{d}{\rightarrow} \int \underline{B}_2\,dB_1$$

where

$$\underline{B}_2(r) = B_2(r) - \left(\int_0^1 rB_2\right)\left[\left(\frac{1}{3}\right)\right]^{-1} r = B_2(r) - 3r\left(\int_0^1 rB_2\right)$$

is detrended Brownian motion. We deduce that

$$\tilde{\beta} \underset{p}{\rightarrow} \beta - 1. \tag{12}$$

REMARKS

(i) The limit distribution (11) is a ratio of dependent normal variates, each of which has zero mean. Its distribution is Fieller [6] and it has Cauchy-type tails. Its variance is undefined.

(ii) The estimator $\tilde{\beta}$ is inconsistent so we would prefer the estimation suggested by econometrician A in this case.

(iii) However, as shown in Part (b) the optimal estimator is the maximum likelihood estimator which utilizes the information that the generating mechanism is the random walk (2)′. This estimator is simply OLS on (1)″ and its limit distribution was shown earlier to be (10).

(iv) We conclude that, when stochastic trends are taken to be deterministic trends, econometrician A's recommended estimator is preferred because it is consistent whereas B's is not. However, A's is inferior to the optimal estimator obtained by maximum likelihood under the correct information about the generating mechanism. In particular, the limit distribution of A's estimator is seen from (11) to be asymmetric and to involve bias (induced by the correlation of the normal variates in the numerator and denominator of (11)). Thus, neither A's nor B's procedure provides an adequate bias for inference about β when the trend is misspecified as deterministic.

REFERENCES (FOR SOLUTIONS 1 AND 2)

1. Rothenberg, T. *Efficient Estimation with a priori Information.* New Haven: Yale University Press, 1973.
2. Phillips, P.C.B. Multiple regression with integrated processes. In N.U. Prabhu (ed.), *Statistical Inference from Stochastic Processes, Contemporary Mathematics, Vol. 80.* Providence: American Mathematical Society, 1988.
3. Phillips, P.C.B. Partially identified econometric models. *Econometric Theory* 5 (1989): 181–240.
4. Phillips, P.C.B. Time series regression with a unit root. *Econometrica* 55 (1987): 277–301.
5. Phillips, P.C.B. Optimal inference in cointegrated systems, Cowles Foundation Discussion Paper No. 866, 1988.
6. Phillips, P.C.B. & M. Loretan. Estimating long-run economic equilibria, Cowles Foundation Discussion Paper No. 928, 1989.
7. Fieller, E.C. The distribution of the index in a normal bivariate population. *Biometrika* 24 (1932): 428–440.
8. Park, J.Y. & P.C.B. Phillips. Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory* 4 (1988): 468–498.