ORTHOGONALITY TESTS WITH DE-TRENDED DATA
Interpreting Monte-Carlo Results Using Nagar Expansions *

Anindya BANERJEE

University of Florida, Gainesville, FL 32611, USA

Juan DOLADO

Research Department, Bank of Spain, 28014 Madrid, Spain

John W. GALBRAITH

McGill University, Montreal, Que. H3A 2T7, Canada

Received 1 February 1989
Accepted 19 June 1989

We extend previous results concerning the behaviour of a finite-sample approximation to the distribution of the t-statistic used in testing orthogonality of a variable to a given information set. In particular, we look at the case in which the data are de-trended, innovations in the explanatory variable are correlated with the regressand, and the explanatory variable is substantially autocorrelated.

1. Introduction

This paper considers the behaviour of a Nagar-type expansion of the expected t-statistic for the slope coefficient in a simple model used to test the orthogonality of a variable to an element of a particular information set, the interesting case is that in which the latter quantity is substantially autocorrelated and its innovations are correlated with the regressand. This approximation was employed by Banerjee and Dolado (1987, 1988) for the case where the only deterministic component of the model is a constant. It offered a satisfactory analytical explanation of the Monte-Carlo results of Mankiw and Shapiro (1985, 1986) where it was shown that, under the previous assumptions, there may be substantial over-rejection of the null hypothesis. Our purpose here is to extend the analysis to the case in which a linear time trend is also included in the model, under the erroneous belief that the data are stationary around this trend. We also extend the analysis to the case in which the correlation is perfect and the explanatory series has a unit root, which corresponds to a test for a unit root where a linear trend is present under the alternative hypothesis [see Fuller (1976)].

We show that the Nagar expansion provides a fairly good approximation to the true bias, especially for values of the autoregressive parameter up to 0.95. As well, using the continuous normalisation of the bias and applying a simple rule, we derive the empirical percentiles of the

* We thank David Hendry and Manuel Arellano for their help and suggestions. A previous version of this paper was presented at the 1988 European Meeting of the Econometric Society, Bologna.

distribution of the t-statistic; the crucial difference between this distribution and the t-distribution seems to be a shift in the lower tail of the asymptotic t-distribution by the amount of the normalised bias. The percentiles in the lower tail (those normally used in applied work) correspond closely to those obtained by Monte-Carlo simulation.

2. The orthogonality test

The standard test consists in regression of a variable Y_t on another variable X_t, lagged once, which is presumed to belong to the information set at $t-1$. In the classical example of the 'rational expectations' version of consumption behaviour, Y_t is the change in consumption and X_t is income; the test asks whether or not consumption is excessively sensitive to income, once the role of current income in signalling changes in permanent income has been taken into account [see Flavin (1981) and Mankiw and Shapiro (1985)].

The canonical model characterising the null hypothesis can be described by the following data generation process (henceforth DGP):

$$Y_t = \epsilon_t; \quad X_t = \lambda X_{t-1} + \nu_t; \quad \mathrm{E}(\epsilon_t \nu_t) = \delta_{ts}\rho, \tag{1}$$

where ϵ_t and ν_t are taken to be IN(0, 1) without loss of generality [see Mankiw and Shapiro (1986)]; δ is the Kronecker delta and ρ is the correlation coefficient. Hence under the hypothesis that the data are stationary around a linear trend, the standard procedure would be to test the null hypothesis H_0: $\pi = 0$ in an OLS regression using the model [1]

$$Y_t = \beta_0 + \beta_1 t + \pi X_{t-1} + u_t. \tag{2}$$

If we consider (for example) the process

$$X_t = \gamma_1 + \gamma_2 t + v_t \quad \text{and} \quad v_t = \rho v_{t-1} + \epsilon_t, \tag{3}$$

then, where $\lambda = \rho = 1$, we have $\Delta X_t = \gamma_2 + \epsilon_t$; using (2) corresponds to testing whether or not X_t follows a random walk, when the relevant alternative hypothesis is stationarity around a linear trend [see Fuller (1976)].

The case on which Mankiw and Shapiro focus is that in which X_t follows a borderline stationary process (i.e., $|\lambda|$ is less than, but close to, unity). Their Monte-Carlo study shows that the inferences drawn are incorrect if the standard t-distribution is applied, leading to over-rejection [2] of H_0. Given that the t-statistic does not follow the standard distribution when $\lambda = 1$ [see Phillips (1987a)], it is not surprising that in borderline cases, by continuity in finite samples, [Phillips (1987b)], a similar argument applies.

The usual explanation of the over-rejection of H_0 depends upon the distribution of the 't-statistic' on $\hat{\pi}$ in (2) being skewed and downwardly biased; simulation studies show this bias to be especially pronounced in the case where a linear trend is present. Hence we wish to check the performance of the analytical approximation in that case.

[1] In fact the observed series is postulated to be $Z_t = \phi + \xi t + X_t$. However, the de-trended series derived from Z_t is independent of ϕ and ξ; hence we set $\phi = \xi = 0$ without loss of generality.

[2] See however Galbraith et al. (1987).

3. A Nagar-type expansion for the borderline-stationary case

As explained in Banerjee and Dolado (1987), the analytical approximation to the correct percentiles of the distribution of the t-ratio of $\hat{\pi}$ in (2) is based upon three steps: (a) computation of a Nagar expansion for the bias; (b) approximation of the t-ratio by the continuous normalisation of the bias, and (c) shifting of the distribution using the standard critical values of the t-distribution.

To implement step (a) consider the DGP given in (1), and define the disturbance ω_t as another IN(0, 1) process independent of ν_t, such that by construction

$$\epsilon_t = (1-\rho^2)^{1/2}\omega_t + \rho\nu_t. \tag{4}$$

From (3), we have that

$$\hat{\pi} = (X'_{-1}MX_{-1})^{-1}(X'_{-1}MY), \tag{5}$$

with $Y=(Y_2, \dots Y_T)'$, $X_{-1}=(X_1, \dots X_{T-1})'$, $i=(1, 1, \dots 1)'$, $t=(2, 3, \dots T)'$, $Z=[i, t]$ and $M = I - Z(Z'Z)^{-1}Z'$, and where I is the identity matrix.

Since we can write $Y_t \equiv \epsilon_t = (1-\rho^2)^{\frac{1}{2}}\omega_t + \rho\nu_t$, we have that

$$\hat{\pi} = (X'_{-1}MX_{-1})^{-1}\left[(1-\rho^2)^{1/2}(X'_{-1}M\omega) + \rho X'_{-1}M\nu\right], \quad \text{and} \tag{6}$$

$$E(\hat{\pi}) = \rho \cdot E\left[(X'_{-1}MX_{-1})^{-1}X'_{-1}M\nu\right] = \rho \cdot E\left[\frac{\nu' N\nu}{\nu' D\nu}\right], \tag{7}$$

given that X_{-1} can be expressed in terms of ν, and that ν is independent of ω. The second term in the equality has been derived by Grubb and Symons (1987); the exact expressions for N and D for this particular case, along with the method of evaluating them, are contained in an appendix available upon request from the authors.

It is readily shown using a simple Taylor expansion that

$$E(\hat{\pi}) = \rho \cdot E\left[\frac{2E(A)}{E(B)} - \frac{E(AB)}{[E(B)]^2}\right] + O(T^{-1}) \equiv \rho \cdot E_N(\hat{\pi}) + O(T^{-1}), \tag{8}$$

where $E(A) = \text{tr}(N)$, $E(B) = \text{tr}(D)$ and $E(AB) = \text{tr}(N)\cdot\text{tr}(D) + 2\text{tr}(ND)$. Hence $E(\hat{\pi})$ can be approximate by [3] $\rho E_N(\hat{\pi})$.

To implement step (b), we obtain the continuous normalisation of the bias, denoted by $E_N(t_\pi)$ [see Evans and Savin (1984)]. The normalisation factor is derived from the information matrix on the assumption that X_0 follows the marginal distribution $N(0, (1-\lambda^2)^{-1})$. The expected t-statistic is therefore approximated by

$$E(t_\pi) = \rho E_N(t_\pi) + O(T^{-3/2}), \quad \text{with} \tag{9}$$

$$E_N(t_\pi) = E(B)^{1/2}E_N(\hat{\pi}). \tag{10}$$

Note from (8) and (9) that the computed expressions (the central values) are proportional to ρ; using this fact reduces much of the computational burden.

[3] The approximation can be related to the standard Hurwicz bias by assuming that var(B) = 0.

Table 1
Bias and continuous normalization of the bias. ($\rho = 1$) [a]

λ	$T = 50$		$T = 200$	
	$E_N(\hat{\pi})$	$E_N(t_{\pi=0})$	$E_N(\hat{\pi})$	$E_N(t_{\pi=0})$
0.999	−0.1376 (−0.1722)	−1.7762 (−2.1086)	−0.0361 (−0.0460)	−1.8687 (−2.0802)
0.990	−0.1332 (−0.1638)	−1.7201 (−1.9981)	−0.0336 (−0.0383)	−1.6692 (−1.7046)
0.980	−0.1296 (−0.1559)	−1.6659 (−1.8867)	−0.0325 (−0.0360)	−1.5019 (−1.5217)
0.950	−0.1232 (−0.1386)	−1.5283 (−1.6315)	−0.0308 (−0.0330)	−1.1542 (−1.1712)
0.900	−0.1166 (−0.1202)	−1.3362 (−1.3399)	−0.0290 (−0.0297)	−0.8551 (−0.8749)

[a] The approximations are based on the formulae derived in (8). The figures in brackets are based on a Monte-Carlo study with 5000 replications, and are provided for comparison with the unbracketed quantities calculated from the Nagar expansion. $E_N(\hat{\pi})$ is the estimate of the mean of the estimated coefficient $\hat{\pi}$ in (2) for a sample of size N; $E_N(t_{\pi=0})$ is the estimate of the mean of the t-statistic for the hypothesis H_0: $\pi = 0$ in (2) again for a sample of size N.

Finally, step (c) is based on the observation that the critical values reported by Mankiw and Shapiro are in fact one-tailed tests at the five per cent level, although reported as two-tailed tests [see Banerjee and Dolado (1987) for details]. With this in mind we have computed approximations to the pseudo-two-tailed critical values by adding the one-tailed five per cent critical values of the t-distribution to the central values computed in step (b). Hence, denoting by C_N the approximate critical value and by C the corresponding percentile of the t-distribution at significance level α, we have

$$C_N(\alpha) = C(\alpha) + E_N(t_\pi). \qquad (11)$$

To implement the approach, we evaluate (8), (10) and (11) for the values in the parameter space $(T \times \lambda \times \rho)$ considered by Mankiw and Shapiro (1986): that is, $T = \{50, 200\}$; $\lambda = \{0.999, 0.99, 0.95, 0.90\}$; $\rho = \{1.0, 0.9, 0.8, 0.7, 0.5\}$.

Table 1 reports both the Nagar approximation to the bias and the continuous normalisation of the bias, for the two sample sizes, when $\rho = 1$. In all cases we observe both that both statistics are centred (as expected) around negative values, indicating that deviations from a standard central t-distribution can be substantial. For comparison, the means of the distribution obtained from a Monte-Carlo simulation using 5000 replications are also included; these appear in brackets beneath the analytical results. In general, results for the bias and the approximate t-ratio are reasonably good for values of λ as large as 0.95.

Next we compute approximations to the pseudo-two-tailed critical values given by (11), using the central values given in table 1 (multiplied by the corresponding values of ρ). [4] The results are tabulated in the bottom entries of table 2 and when compared with the critical values obtained by Mankiw and Shapiro, shown in the topmost entries, can be seen to be very similar. Hence we have an explanation of the Mankiw–Shapiro Monte-Carlo results; it seems that the positive bias involved in

[4] The critical values of the Student t-distribution are $t(47) = 1.680$ and $t(197) = 1.645$.

Table 2
Monte-Carlo and approximate critical values in the borderline case. [a]

λ	$T=50$					$T=200$				
	$\rho=1.0$	0.9	0.8	0.7	0.5	1.0	0.9	0.8	0.7	0.5
0.999	−3.5	−3.2	−3.0	−3.0	2.7	−3.3	−3.3	−3.1	−3.0	−2.7
	−3.5	−3.3	−3.1	−2.9	2.6	−3.5	−3.3	−3.1	−3.0	−2.6
0.990	−3.4	−3.2	−3.0	−2.9	−2.6	−3.1	−3.0	−2.9	−2.7	−2.5
	−3.4	−3.2	−3.0	−2.9	−2.6	−3.1	−3.0	−2.9	−2.7	−2.5
0.980	−3.3	−3.1	−3.0	−2.8	−2.6	−2.9	−2.8	−2.7	−2.6	−2.4
	−3.3	−3.2	−3.0	−2.8	−2.5	−3.1	−3.0	−2.8	−2.7	−2.4
0.950	−3.1	−3.0	−2.8	−2.7	−2.5	−2.6	−2.5	−2.4	−2.3	−2.2
	−3.2	−3.0	−2.9	−2.7	−2.4	−2.8	−2.7	−2.5	−2.4	−2.2
0.900	−2.9	−2.8	−2.8	−2.6	−2.3	−2.3	−2.3	−2.2	−2.2	−2.1
	−3.0	−2.9	−2.8	−2.6	−2.3	−2.5	−2.4	−2.3	−2.2	−2.1

[a] The entries in the first row of each box show the five percent critical values reproduced from Mankiw and Shapiro (1986). The entries in the second row show the approximations calculated using the formulae given in (11).

the use of the standard five per cent one-tailed critical value is offset by the negative bias in the computed central values of the t-ratio.

It is important to point out that the complete finite-sample distribution cannot be recovered simply by shifting the (asymptotically valid) standard distribution in the way described in this section. The method performs poorly in the upper tails; for instance, the empirical 0.95 percentiles for $\lambda = 0.99$ and $\rho = 1$ are -0.82 and -0.88 for $T = 50$ and $T = 200$ respectively, whereas the corresponding approximations are -0.08 and -0.19. Rather our observation is that for the lower

Table 3
Monte-Carlo and approximate critical values in the unit root case. [c]

α	$T=25$	$T=50$	$T=200$
0.010	−4.38	−4.15	−4.02
	−4.14	−4.19	−4.22
0.025	−3.95	−3.80	−3.71
	−3.71	−3.80	−3.86
0.050	−3.60	−3.50	−3.44
	−3.35	−3.46	−3.54
0.100	−3.24	−3.18	−3.14
	−2.98	−3.08	−3.18

[c] The symbol α denotes size. The entries in the first row of each box show the critical values reproduced from Fuller (1976); the entries in the second row show the approximations calculated using the formulae given in (11). To compute these second-row entries, note the following: the central normalizations of the bias are -1.632, -1.783 and -1.898. The critical values of the t-distribution for the different sizes (0.01, 0.025, 0.050 and 0.100) are:
2.508 2.074 1.717 1.321 ($t(22)$),
2.411 2.014 1.680 1.301 ($t(47)$),
2.326 1.960 1.645 1.282 ($t(197)$).

tail of the distribution (up to roughly the 25% level), as typically used in empirical work, the approximation is quite accurate.

4. Approximation in the unit root case

When $\lambda = 1$, we can implement the same approximation, now taking X_0 equal to zero [again see Evans and Savin (1984)]. Table 3 contains the approximate values for $\rho = 1$ and $T = \{25, 50, 200\}$. The topmost entries are Fuller's (1976, table 8.5.2) critical values, corresponding to the Nagar approximations shown as the lower entries. In addition to five per cent critical values, we have included a range of other critical values corresponding to the lower tail of the distribution. Again the similarity is clear, especially for the larger sample sizes.

References

Banerjee, A. and J. Dolado, 1987, Do we reject rational expectations models too often? Interpreting evidence using Nagar expansions, Economics Letters 24, 27–32.

Banerjee, A. and J. Dolado, 1988, Tests of the life cycle-permanent income hypothesis in the presence of random walks: Asymptotic theory and small sample interpretations, Oxford Economic Papers 40, 610–633.

Evans, G.B.A. and N.E. Savin, 1984, Testing for unit roots: 2, Econometrica 52, 1241–1269.

Flavin, M., 1981, The adjustment of consumption to change in expectations about future incomes, Journal of Political Economy 89, 974–1009.

Fuller, W.A., 1976, Introduction to statistical time series (Wiley, New York).

Galbraith, J.W., J. Dolado and A. Banerjee, 1987, Rejections of orthogonality in rational expectations models: Further Monte-Carlo results for an extended set of regressors. Economics Letters 25, 243–247.

Grubb, D. and J. Symons, 1987, Bias in regressions with a lagged dependent variable, Econometric Theory 3, 371–386.

Mankiw, N.G. and M.D. Shapiro, 1985, Trends, random walks and tests of the permanent income hypothesis, Journal of Monetary Economics 16, 165–174.

Mankiw, N.G. and M.D. Shapiro, 1986, Do we reject too often? Small sample properties of tests of rational expectations models. Economics Letters 20, 139–145.

Phillips, P.C.B., 1987a, Time series regression with a unit root, Econometrica 55, 277–301.

Phillips, P.C.B., 1987a, Towards a unified asymptotic theory of autoregression, Biometrika 74, 535–548.