Comput Econ (2013) 41:1–9
DOI 10.1007/s10614-011-9309-4

Response Surface Estimates of the Cross-Sectionally Augmented IPS Tests for Panel Unit Roots

Jesús Otero · Jeremy Smith

Accepted: 12 December 2011 / Published online: 21 December 2011

Abstract This paper estimates response surface coefficients for a large range of quantiles of the cross-sectionally augmented IPS (CIPS) test of Pesaran (2007), for different specifications of the deterministic components. An Excel programme is available to calculate the P value associated with a CIPS test statistic.

Keywords CIPS test · Monte Carlo · Unit roots · Response surface · Critical values · P values

JEL Classification C12 · C15

1 Introduction

During the last decade or so, the problem of testing for the presence of unit roots in panels of data has received a great deal of attention. Among the tests available in the literature, perhaps the one proposed by Im et al. (2003), commonly referred to as the IPS test, has proved to be the most popular. This panel unit root test, based on averaging individual augmented Dickey and Fuller (1979) (ADF) statistics, combines information from the time-series dimension with that from the cross-section dimension, such that fewer time observations are required for the test to have power.

J. Otero
Facultad de Economía, Universidad del Rosario, Bogota, Colombia
e-mail: jesus.otero@urosario.edu.co

J. Smith (✉)
Department of Economics, University of Warwick, Coventry, UK
e-mail: jeremy.smith@warwick.ac.uk

A critical assumption underlying the IPS test is that of cross sectional independence. However, failing to account for cross-sectional dependence leads to over-rejection of the test statistic, the magnitude of which increases as the strength of the cross-sectional dependence increases; see e.g., Strauss and Yigit (2003) and Pesaran (2007). To overcome this deficiency, Pesaran (2007) suggests augmenting the standard ADF regressions with the cross section averages of lagged levels and first-differences of the individual series in the panel; the resulting test statistic is referred to as the cross-sectionally augmented version of the IPS test, denoted as CIPS.

Both Im et al. (2003) and Pesaran (2007) tabulate critical values for the most commonly used specifications of the deterministic components in the test regressions, namely no intercept and no trend (Case I); intercept only (Case II); and intercept and trend (Case III). The tabulated critical values are based on 50,000 replications, but are only reported for a limited range of time observations, T, and a limited number of cross-sectional units, N, in the panel.

In this paper, we undertake an extensive set of Monte Carlo simulations that are summarised by means of response surface regressions, from which critical values of the Pesaran (2007) panel unit root test can be calculated for different values of N and T, and different specifications of deterministic components. Our particular interest is on the CIPS test, rather than on the IPS test, since it is applicable to the more realistic scenario in which there is cross sectional dependency among the individuals in the panel. Response surfaces have been used, among others, by MacKinnon (1991) to calculate critical values of the Dickey and Fuller (1979) and Engle and Granger (1987) unit root and cointegration tests, respectively; Cheung and Lai (1993) for the Dickey-Fuller tests allowing for the effect of lag order; Sephton (1995) for the Kwiatkowski et al. (1992) univariate stationarity tests; Mackinnon et al. (1999) for the Johansen (1988) likelihood ratio tests for cointegration; and Harvey and Van Dijk (2006) for the Hylleberg et al. (1990) seasonal unit root tests.

The plan of the paper is as follows. In Section 2 we provide a brief overview of both the IPS and CIPS tests. Section 3 then discusses the Monte Carlo simulation design and the response surface results. Section 4 concludes.

2 IPS and CIPS Panel Unit Root Tests

For Case III, that is including cross-section specific intercept and trend, the IPS test is based on individual ADF regressions:

$$\Delta y_{it} = a_i + b_i t + c_i y_{i,t-1} + \sum_{r=1}^{p_i} d_{ir} \Delta y_{i,t-r} + \varepsilon_{it}, \tag{1}$$

where $i = 1, \ldots, N$ cross section units, and $t = 1, \ldots, T$ time observations (for Case I: $a_i = b_i = 0$; for Case II: $b_i = 0$). In this setting the null hypothesis to test the presence of a unit root becomes $H_0 : c_i = 0$ for all i, against the alternative that at least one of the individual series in the panel is stationary, that is $H_1 : c_i < 0$ for at least one i. The IPS test averages the ADF statistics obtained in Eq. 1 across the N cross-sectional units of the panel, that is:

$$\text{IPS} = (N)^{-1} \sum_{i=1}^{N} ADF_i, \tag{2}$$

where ADF_i is the augmented Dickey and Fuller statistic based on the regression t statistic for $H_0 : c_i = 0$ in Eq. 1. Im et al. (2003) show that after a suitable standardisation, their statistic follows a standard normal distribution. They compute the mean and variance required to standardise the statistic in Eq. 2 via Monte Carlo simulations, for different values of T and p_i, and for different combinations of deterministic components. Im et al. further show that when the underlying error term ε_{it} follows a normal distribution, the required condition for the second moment of the statistic in Eq. 2 to exist is $T > 5$ in the model with intercept and no trend, and $T > 6$ in the model with intercept and trend.

An important assumption underlying the IPS test is that of cross sectional independence across the individual time series in the panel, as the test suffers from size distortions in the presence of cross section dependence. In order to overcome this, Pesaran (2007) augments Eq. 1 with the cross section averages of lagged level and lagged first-differences of the individual series in the panel. Thus, the test of the unit root hypothesis would be based on the following p^{th} order cross-sectionally augmented Dickey and Fuller regressions:

$$\Delta y_{it} = a_i + b_i t + c_i y_{i,t-1} + \sum_{r=1}^{p} d_{ir} \Delta y_{i,t-r} + f_i \bar{y}_{t-1} + \sum_{r=0}^{p} g_{ir} \Delta \bar{y}_{t-r} + \varepsilon_{it}, \tag{3}$$

where $\bar{y}_t$ is the cross section mean of y_{it}, defined as $\bar{y}_t = (N)^{-1} \sum_{i=1}^{N} y_{it}$. The cross-sectionally augmented version of the IPS test statistic (CIPS) is:

$$\text{CIPS} = (N)^{-1} \sum_{i=1}^{N} CADF_i, \tag{4}$$

where $CADF_i$ is the cross-sectionally augmented Dickey and Fuller t statistic for testing $H_0 : c_i = 0$ in Eq. 3. Pesaran (2007) also considers a truncated version of the $CADF_i$ statistic, denoted $CADF_i^*$, that avoids the problem of moment calculation. The truncated version of the statistic is given by:

$$\begin{aligned}
CADF_i^* &= CADF_i \quad && \text{if} \quad -K_1 < CADF_i < K_2 \\
CADF_i^* &= -K_1 && \text{if} \quad CADF_i \le -K_1 \\
CADF_i^* &= K_2 && \text{if} \quad CADF_i \ge K_2
\end{aligned}$$

where K_1 and K_2 are positive constants sufficiently large so that Pr $(-K_1 < CADF_i < K_2)$ is close to 1. Applying a normal approximation of $CADF_i$, Pesaran (2007) obtains:

Case I: No intercept, no trend: $K_1 = 6.12; \quad K_2 = 4.16$

Case II: Intercept, no trend: $K_1 = 6.19; \quad K_2 = 2.61$

Case III: Intercept and trend: $K_1 = 6.42; \quad K_2 = 1.70$

The corresponding truncated version of the CIPS statistic, denoted CIPS*, is computed as the simple average of the individual $CADF_i^*$ statistics, that is:

$$\text{CIPS}^* = (N)^{-1} \sum_{i=1}^{N} CADF_i^*. \tag{5}$$

Critical values for both the CIPS and CIPS* test statistics are tabulated by Pesaran (2007) for several values of T and N, and according to the deterministic components included in the cross-sectionally augmented Dickey and Fuller regression given in Eq. 3. Pesaran (2007) shows that the distribution of both the CIPS and CIPS* test statistics are non-standard even for sufficiently large N. This result is in sharp contrast to that obtained for the IPS test under the assumption of cross section independence which, after a suitable standardisation, was shown to be normally distributed for N sufficiently large. Pesaran also observes that it is only for very small values of T that the finite sample distributions of CIPS and CIPS* differ, being practically indistinguishable for $T > 20$. In this paper we focus on the CIPS test statistic.

3 Monte Carlo Design and Results

The data generating process (DGP) used in the Monte Carlo simulation follows closely that used by Pesaran (2007), which enables us to determine how well our results compare with those reported in that paper. Thus, we assume that y_{it} is generated by a first-order autoregressive process:

$$y_{it} = y_{i,t-1} + \varepsilon_{it}, \tag{6}$$

where $i = 1, \ldots, N$, $t = 1, \ldots, T + 1$, and $\varepsilon_{it} \sim N\left(0, \sigma_{it}^2\right)$, where $\sigma_{it}^2 = 1$ without loss of generality. We set $N = 2(1)8$, 10(2)20, 25 (5) 50, 60 (10) 100, 120 (20) 160, 200, and $T = 10$, 11, 12 (2) 20, 25 (5) 50, 60 (10) 100, 120 (20) 200, 250, 300 (100) 800, where e.g. $N = 2$ (1) 8 means that we take all samples from $N = 2$ up to $N = 8$ going up in steps of 1, and so on (the same notation is also used for the sample size, T, and later on when listing significance levels, l). Following Pesaran (2007), for Case I the initial value of y_{it} is set equal to zero, while for Cases II and III a burn-in period of 100 observations is used. Each experiment consists of 50,000 replications and then, similar to Harvey and Van Dijk (2006), we repeat each experiment 25 times to allow for sampling variability, implying that we have 25 critical values generated from each of the 50,000 replications. According to MacKinnon (1996) , conducting multiple experiments for the same sample size (and number of individuals) provides a simple way to measure experimental randomness. Although the choice for repeating each experiment (25 times in our case) is based on computational costs, it appears that numerical accuracy is adequate. For instance, our estimate of the 1% critical value for the CIPS test in Case II (III) for $N = 20$ and $T = 30$ varies from -2.381 (-2.894) to -2.371 (-2.878) with a mean of -2.377 (-2.884) and standard deviation of 0.0028 (0.0038), over the 25 different repetitions.

Turning to the specifications of the response surfaces, we follow MacKinnon (1991) by running regressions of the critical values, for a given significance level l, on an intercept and functions of $\left(\frac{1}{N}\right)$, $\left(\frac{1}{T}\right)$ and their interaction $\left(\frac{1}{NT}\right)$. The chosen functional form includes powers of order 0.5, 1, 2, and 3 of the terms $\left(\frac{1}{N}\right)$, $\left(\frac{1}{T}\right)$ and $\left(\frac{1}{NT}\right)$. It is also worth noting that including powers of order 1.5 and 2.5 and experimenting with higher powered terms, in general, yielded coefficients that did not turn out to be statistically different from zero at the 1% significance level, nor led to any increase in the $\overline{R}^2$ for these models. Therefore, the preferred functional form specification is:

$$CV^l_{N,T} = \phi^l_0 + \sum_k \phi^l_k \frac{1}{N^k} + \sum_k \theta^l_k \frac{1}{T^k} + \sum_k \varphi^l_k \frac{1}{(NT)^k} + \epsilon^l, \quad k = 0.5,\ 1,\ 2,\ 3. \tag{7}$$

where $CV^l_{N,T}$ is the critical value estimate at each of 221 significance levels ($l = 0.0001$, 0.0002, 0.0005, 0.001 (0.001) 0.01, 0.015 (0.005) 0.990, 0.991 (0.001) 0.999, 0.9995, 0.9998 and 0.9999), N denotes the number of cross sectional units, and T refers to the number of observations on Δy_{it} (which is one less than the total number of available observations). Because all terms in Eq. 7 tend to zero as $N \to \infty$ and $T \to \infty$, the intercept term (i.e., ϕ^l_0) provides an estimate for the asymptotic critical value of the test statistics.

Table 1 reports response surface regressions for 3 of the 221 significance levels, namely $l = 0.01$, 0.05 and 0.10. The coefficient of determination indicates that the fit of these response surface models is particularly good, being at least 0.997. While we only report the response surface regressions for a limited range of l, Eq. 7 has been estimated for all 221 quantiles. Across all quantiles the average R^2 in the response surface models is 0.998, with a minimum R^2 of 0.967 (which occurs in Case III for $l = 0.0001$). As to be expected, the residuals of the estimated response surfaces exhibit heteroskedasticity. Thus, to assess the robustness of the OLS results we also considered estimation using the GMM procedure described in MacKinnon (1994) and MacKinnon (1996). For the purposes of our simulation exercise, this procedure amounts to averaging the critical values across the 25 replications for each combination of T and N, and scaling all the variables in Eq. 7 by the standard error in these replications. Then, the resulting equation using the re-scaled variables can be estimated by OLS. This GMM procedure yields very similar results to those obtained when using OLS. Finally, to obtain P values of the CIPS statistic, we follow MacKinnon (1994) and MacKinnon (1996) by estimating the regression

$$\Phi^{-1}(l) = \gamma^l_0 + \gamma^l_1 \widehat{CV^l} + \gamma^l_2 \left(\widehat{CV^l}\right)^2 + \upsilon^l, \tag{8}$$

Table 1 Response surface estimates of the CIPS test

Regressors	Case I: No intercept, no trend 1%	5%	10%	Case II: Intercept, no trend 1%	5%	10%	Case III: Intercept and trend 1%	5%	10%
Intercept	−1.5539	−1.4920	−1.4350	−2.0837	−2.0467	−2.0095	−2.5352	−2.5083	−2.4803
	(0.0010)	(0.0005)	(0.0004)	(0.0017)	(0.0008)	(0.0006)	(0.0036)	(0.0013)	(0.0007)
$1/\sqrt{N}$	−0.5195	−0.0635	0.0965	−0.8072	−0.2796	−0.0952	−1.1812	−0.6187	−0.4425
$1/N$	−2.5560	−2.1712	−1.8374	−2.4299	−2.2740	−2.0520	−1.9824	−1.7685	−1.4553
$1/N^2$	3.0975	2.2762	1.6842	3.7630	3.3513	3.0305	3.7955	3.0385	2.4279
$1/N^3$	−6.5278	−4.3859	−3.2235	−7.0165	−5.3023	−4.4276	−6.9859	−5.0309	−3.8924
$1/\sqrt{T}$	0.0392	0.0676	0.0564	0.1772	0.2057	0.1707	0.5957	0.5281	0.4418
$1/T$	−1.2240	−0.7502	−0.4721	−1.7421	−0.8271	−0.3971	−7.7581	−4.6275	−3.3352
$1/T^2$	14.1154	8.6611	6.6458	54.9572	31.5314	23.3324	171.4465	99.7734	75.9939
$1/T^3$	−201.6559	−114.1294	−81.4965	−616.5336	−349.4229	−251.7362	−1635.2750	−944.4393	−702.3546
$1/\sqrt{(NT)}$	0.3376	0.0378	0.0221	1.0647	0.1170	−0.0145	3.0515	0.8114	0.3635
$1/(NT)$	−7.7465	−2.2653	−0.9743	−31.0267	−13.3402	−8.1149	−55.9265	−21.2679	−11.9148
$1/(NT)^2$	−60.0596	−20.0741	6.0607	349.6403	135.2067	97.2566	860.3642	225.6850	111.1167
$1/(NT)^3$	−2350.5005	−337.6337	−129.8883	−10819.5177	−3067.6309	−1467.0250	−21667.8348	−5030.0167	−1760.2678
Obs	21000	21000	21000	21000	21000	21000	21000	21000	21000
R^2	0.9996	0.9997	0.9997	0.9990	0.9993	0.9993	0.9967	0.9986	0.9991

Numbers in *parentheses* are White's Heteroskedastic consistent standard errors where Φ^{-1} is the inverse of the cumulative standard normal distribution at each of the 221 quantiles and $\widehat{CV^l}$ is the critical value estimate at the l quantile, which is the fitted value from Eq. 7. As in Harvey and Van Dijk (2006), Eq. 8 is estimated by OLS using seven quantile observations on either side of the desired quantile, so that Eq. 8 is effectively estimated using 15 observations in total.[1] Approximate P values of the CIPS test statistic can then be obtained as:

$$P = \Phi\left(\widehat{\gamma}_0^l + \widehat{\gamma}_1^l \mathrm{CIPS} + \widehat{\gamma}_2^l \mathrm{CIPS}^2\right). \tag{9}$$

where $\widehat{\gamma}_j^l$, $j = 0, 1, 2$ are the OLS parameter estimates from Eq. 8. An Excel spreadsheet that calculates the P value of any CIPS test statistic for each of the three specifications is available at *http://www2.warwick.ac.uk/fac/soc/economics/staff/academic/jeremysmith/research*.

The response surface models estimated in Eq. 7 are based on the DGP of Pesaran (2007), which assumes a normally distributed error term, ε_{it}. To consider the effects of skewness and leptokurtosis (thick tails), critical values are also tabulated when the error term is distributed as χ^2 with 6 degrees of freedom (χ_6^2) and Student's t with 10 degrees of freedom (t_{10}), although standardised to facilitate the comparison of the results reported earlier. In 25 different experiments, each with 50,000 replications, our results for $N = 10, 20$ and $T = 20, 30$ indicate that the critical values show a marked departure compared to those critical values reported for the case in which the errors have the normal distribution, although the extent of the departure depends upon the deterministic components included in the test regressions. Case I exhibits the greatest extent of departure, with the resulting critical values being more negative when the errors are generated as χ_6^2 or t_{10}. For instance, when $N = 10$ and $T = 20$ the 1% critical value when the errors are generated as normal was -2.004 as compared to -2.449 (-2.455) for a χ_6^2 (Student's t_{10}) distribution. In Case II there is a smaller difference in the critical values which under normality (for $N = 10$, $T = 20$ and $l = 0.01$) was -2.610 compared to -2.854 (-2.866) for a χ_6^2 (Student's t_{10}) distribution. Lastly, in Case III the critical values are very similar irrespective of the error distribution: for $N = 10$ and $T = 20$, the 1% critical values under Gaussian, χ_6^2 and Student's t_{10} errors are -3.153, -3.173, and -3.162, respectively. From a practical point of view, these findings support the inclusion of intercept and trend terms in the testing regressions when the assumption of Gaussian disturbances appears not to be valid.

Finally, Pesaran (2007) indicates that, in practice, serial correlation can be accounted for this by including p lags of both $\triangle y_{it}$ and $\triangle \bar{y}_t$; see Eq. 3. However, neither the critical values tabulated by Pesaran (2007) nor the ones tabulated in our paper formally allow for the inclusion of these additional terms in the right hand side of the test regressions. To further our understanding of the effect of inclusion of lags in Eq. 3 on the tabulated critical values, we perform a limited set of additional Monte Carlo simulations (once again, assuming that the DGP follows the setup of Eq. 6). These additional simulations reveal that the presence of lags shifts the distribution of

[1] For $l \leq 0.004$ and $l \geq 0.996$ we use the actual quantile and the 14 observations closest to the desired quantile, as there will not be seven on either side.

the CIPS test statistics to the right, but also that this displacement diminishes as T increases. For example, in Case II and for a given panel of $N = 20$ individuals and $T = 50$, 100 and 200 observations, the 5% critical values with $p = 0$ lags are -2.202, -2.203 and -2.204, respectively. With $p = 4$ lags the corresponding 5% critical values are -2.074 for $T = 50$, -2.144 for $T = 100$, and -2.176 for $T = 200$ observations, respectively. However, a more thorough analysis of the sensitivity of the CIPS test to the inclusion of lags is left for further work.

4 Conclusions

In this paper we estimate response surface models for the critical values of the CIPS panel unit root test, for the cases in which the test regression includes no intercept and no trend, intercept only, and intercept and trend. The response surface models, which are estimated for a total of 221 significance levels, are power functions of $\left(\frac{1}{N}\right)$, $\left(\frac{1}{T}\right)$ and their interaction $\left(\frac{1}{NT}\right)$. Here we report the response surface equations for 1%, 5% and 10% significance levels. The fit of the 221 response surface models is good, with all models having an R^2 in excess of 0.967. Models are then estimated to enable the calculation of finite sample probability values of the CIPS test, and an Excel spreadsheet is made available for such purpose.

Acknowledgments The author would like to thank M. Hashem Pesaran, Takashi Yamagata and an anonymous referee for constructive comments and suggestions that helped improve our paper. This paper was started while Jesús Otero was a Visiting Fellow in the Department of Economics at the University of Warwick. For this, the authors thank the ESRC (RES-000-22-1686) and Universidad del Rosario for financial assistance. The usual disclaimer applies

References

Cheung, Y. E., & Lai, K. S. (1993). Finite sample sizes of Johansen's likelihood ratio tests for cointegration. *Oxford Bulletin of Economics and Statistics, 55*, 313–328.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74*, 427–431.

Engle, R. F., & Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica, 55*, 251–276.

Harvey, D. I., & van Dijk, D. (2006). Sample size, lag order and critical values of seasonal unit root tests. *Computational Statistics and Data Analysis, 50*, 2734–2751.

Hylleberg, S., Engle, R. F., Granger, C. W. J., & Yoo, B. S. (1990). Seasonal integration and cointegration. *Journal of Econometrics, 44*, 215–238.

Im, K., Pesaran, M. H., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics, 115*, 53–74.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control, 12*, 231–254.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics, 54*, 159–178.

MacKinnon, J. G. (1991). Critical values for cointegration tests. In R. F. Engle & C. W. J. Granger (Eds.), *Long-run economic relationships: Readings in cointegration* (pp. 267–276). Oxford: Oxford University Press.

Mackinnon, J. G. (1994). Approximate asymptotic distribution functions for unit-root and cointegration tests. *Journal of Business and Economic Statistics, 12*, 167–176.

Mackinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics, 11*, 601–618.

Mackinnon, J. G., Haug, A. A., & Michelis, L. (1999). Numerical distribution functions of likelihood ratio tests for cointegration. *Journal of Applied Econometrics, 14*, 563–577.

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross section dependence. *Journal of Applied Econometrics, 22*, 265–312.

Sephton, P. S. (1995). Response surface estimates of the KPSS stationarity test. *Economics Letters, 47*, 255–261.

Strauss, J., & Yigit, T. (2003). Shortfalls of panel unit root testing. *Economics Letters, 81*, 309–313.