Pak. J. Statist.
2026 Vol. 42(3), 247-264

THE QUEST FOR UNIFORM POWER: A NEW PANEL POINT OPTIMAL TEST FOR STATIONARITY

Shahid Akbar
Department of Economics, Faculty of Management Sciences
The University of Lahore, Lahore, Pakistan.
Email: shahid.akbar@econ.uol.edu.pk

ABSTRACT

This paper develops a panel point optimal test for stationarity and uses Monte Carlo simulation experiments to assess the test's efficacy in relation to size and power attributes. The test performance has been analyzed against alternatives ranging from white noise ($\varphi = 1$) to near-stationary processes ($\varphi = 0.9$) for panel sizes $N \in \{2, 4, 8, 16, 32\}$ and $T \in \{10, 25, 50, 100\}$. The study shows that the test has outstanding power (100%) for $\varphi = 0.5$ in panels of $N \geq 8$, and more robust power ($\geq 85\%$) for $\varphi \leq 0.8$ in panels of $N \geq 16$. In line with Elliott et al. (1996) and Hadri (2000), power drops as alternatives approach 1. Findings of the study highlight substantial power improvements from small to medium sized panels, with minimal gains thereafter ($N \geq 16$). This points to the test's potential for applied research with medium-sized panels. This research adds to the panel stationarity testing literature by (1) providing a point optimal test with well-defined power for alternatives of various persistence orders, (2) determining the optimal size of panels, and (3) suggesting avenues for improving performance in difficult near-nonstationary scenarios. The analysis has implications for applied researchers, and suggestions for test implementation.

KEYWORDS

Panel stationarity test, Point optimal test, Power analysis, Monte Carlo simulation.

1. INTRODUCTION

In econometric hypothesis testing, selecting a suitable test statistic is paramount, especially when sample sizes are small – a common constraint as economic data are observational data rather than experimental. Hypotheses can be simple (completely specified distribution) or composite (containing unknown parameters) (Lehmann & Romano, 2005). Although the Neyman-Pearson lemma offers a most powerful test for a simple null versus a simple alternative, this is not often the case (Lehmann & Romano, 2005). In testing a simple null against a composite alternative, the power envelope (envelope of maximum attainable power against all alternatives) helps determine optimality. An optimal test (uniformly most powerful) whose power equals the envelope is desirable but rare. As a result, second-best approaches are sought, such as point optimal tests which maximize power at a specific alternative point (King, 1987), and locally best tests which maximize power in a negligible region of the alternative (Ferguson, 1967; King & Hillier, 1985). A more general approach is to find a test that maximizes a linear

© 20xx Pakistan Journal of Statistics

combination of power over the alternative space, based on the generalized Neyman-Pearson lemma (Lehmann & Romano, 2005; Andrews & Ploberger, 1994). Point optimal tests can be seen as instances of a weighted average with all weight on a point (Cox & Hinkley, 1974). These approaches can be applied to composite null hypotheses, but become more complicated as the number of nuisance parameters grows. Initial work highlighted testing linear regression models for covariance structures (King, 1981, 1983; Bhargava et al., 1982; Dufour & King, 1991). King (1987) recommended point optimal tests where theory allows a restriction of the alternative parameter space, and where invariance or similarity considerations allow a simplification of the null hypothesis (King, 1980; Hillier, 1987).

The use of time series data as an empirical tool in economics crucially hinges on the essential difference between stationary and non-stationary processes, which bears far-reaching consequences of econometric modeling and statistical inference. Unless non-stationarity is detected, the presence of it may give spurious effects of regression that invalidate the empirical results (Granger and Newbold, 1974). Traditional unit root testing procedures, such as the Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979) and the Phillips-Perron (PP) test (Phillips and Perron, 1988) operationalize this difference by testing the null hypothesis of a unit root process. Nevertheless, these developed procedures often exhibit non-optimal power properties, especially with finite-sample sizes (DeJong et al., 1992), and require alternative methods to effectively test hypotheses. Kwiatkowski et al. (1992) seminal work brought about a paradigm shift of the KPSS test, which inverts the traditional hypothesis structure, with stationarity as the null hypothesis. Although the innovation resolved some of the weaknesses of the traditional unit root tests, the KPSS framework has limited power properties when applied to panel data settings with cross-sectional dependence and heterogeneity.

An extension of unit root tests to panel data settings was a logical extension of the growing presence of multidimensional economic data. This was extended by Levin et al. (2002) who created a pooled panel unit root test that posited that cross-sections shared common autoregressive parameters. This restrictive assumption was also loosened by Im et al. (2003) who used the heterogeneous panel unit root test which allows the different persistence parameters to be different across units. But these tests are vulnerable to power biases and size effects under certain generative mechanisms. The development of panel stationarity testing has produced a number of advanced methodological techniques. The groundbreaking work by Hadri (2000) gave a Lagrange Multiplier-based model that generalized KPSS methodology to panel-based models under cross-sectional independence. Later improvements of the model with sequential limit theory by Hadri and Larsson (2005) further improved the performance of the model in large-N panels. Harris et al. (2005) contributed to the field by introducing a testing process that is resistant to serial correlation and heteroscedasticity and Shin and Snell (2006) focused on the problematic issue of structural breaks. Demetrescu et al. (2010) came up with adjustments to check the size distortion in cross-sectionally dependent panels but Hadri and Kurozumi (2009) came up with a holistic framework that grants both common factors and idiosyncratic trends. Power properties of stationarity tests have been an outstanding issue in literature. Theoretical results on optimal testing against local alternatives were given early on by Nabeya and Tanaka (1988). The point optimal approach by Elliott et al. (1996) was a

ground breaking innovation and it showed the process of creating tests that were optimized to have maximum power against a set of pre-specified alternatives and still control the size. These ideas have been further elucidated by recent work by Muller and Watson (2022), applied to estimate the long-run variance in long-run nonstationary environments. Their results on the best bandwidth selection and choice of the kernel guide present strategy of addressing the issue of serial correlation in panel environments.

In spite of these theoretical developments, the extant methodologies still have serious limitations, especially in the property of power in short panels and size control. This research presents a new optimal test of stationarity of panel data points that inherently goes a notch further in the methodological practice. The proposed test is tailored to have uniformly maximal power in comparison to all alternatives as well as a proper size control, thus eliminating the most serious restriction of current methodologies.

2. ECONOMETRIC METHODOLOGY

2.1. Data Generating Process (GDP)

Following the study of Hwang and Schmidt (1993), this study is carried out with following DGP,

$$y_{it} = \alpha_i + \beta_i t + y_{it-1} - \varphi\varepsilon_{i,t-1} + \varepsilon_{it} \tag{1}$$

where $\alpha_i \sim U[0,10]$, $\beta_i \sim U[0,2]$, $\varepsilon_{it} \sim N\left(0, \delta_{\varepsilon_i}\right)$ and $\delta_{\varepsilon_i} \sim U[0.5,1.5]$. Under H_0 the panel series is identified to be stationary if $\varphi = 1$ and under alternative of unit root if $0 < \varphi \leq 1$, $\varphi = \{0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0\}$. Further, this simulation study employs a comprehensive experimental design incorporating multiple dimensions. A cross-section dimensions of (2, 4, 8, 16, 32) and time series dimensions of (10, 25, 50, 100) are assumed for this study. The following panel dimensions are considered:

- Small panels: $N = \{2,10\}; \{4,10\}$
- Medium panels: $N = \{8, 25\}; \{8, 50\}; \{16, 25\}; \{16, 50\}$
- Large panels: $N = \{32, 100\}$

2.2. Panel Point Optimal Tests via Likelihood Ratio Framework

Building upon the Neyman-Pearson lemma (Neyman and Pearson, 1933), we develop panel-adapted point optimal tests through likelihood ratio principles by using DGP in equation (1). The test statistics are derived to achieve:

1. Uniformly most powerful properties against specified alternatives
2. Optimal power characteristics across the entire parameter space
3. Robust performance under various cross-sectional dependence structures

Since point optimal test is a ratio of likelihood function of null hypothesis and alternative hypothesis, we develop point optimal test as follows by definition of Neyman Pearson Lemma (Neyman and Pearson (1992)),

$$PPO = L(1) - L(\varphi)$$

where log likelihood is,

$$L(\varphi) = -\frac{NT}{2}\log(2\pi) - \frac{NT}{2}\log\left(\delta_{\varepsilon_i}^2\right) - \frac{1}{2}\log\left(1 + NT(1-\varphi)^2\right) - \frac{1}{2}\sum_{i=1}^{N}\sum_{t=2}^{T}\frac{\left(y_{it} - y_{i,t-1}\right)^2}{\left(\delta_{\varepsilon_i}^2\left(1 + NT(1-\varphi)^2\right)\right)}$$

$$L(1) = -\frac{NT}{2}\log(2\pi) - \frac{NT}{2}\log\left(\delta_{\varepsilon_i}^2\right) - \frac{1}{2}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it} - y_{i,t-1}\right)^2$$

Hence, Panel PO test for the null hypothesis of stationarity is,

$$PPO = -\frac{1}{2}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{i,t} - y_{i,t-1}\right)^2 + \frac{1}{2}\log\left(1 + NT(1-\varphi)^2\right) + \frac{1}{2}\sum_{i=1}^{N}\sum_{t=2}^{T}\frac{\left(y_{i,t} - y_{i,t-1}\right)^2}{\left(\delta_{\varepsilon_i}^2\left(1 + NT(1-\varphi)^2\right)\right)} \quad (2)$$

The PPO statistic developed in this study is formed as a likelihood ratio-based test—the log-likelihood function under the null stationarity and the log-likelihood at a chosen point alternative. The case of a single likelihood ratio statistic's asymptotic behavior is well-established under the standard regularity conditions, presenting a limited comparison with the panel extension that entails more dimensions of analysis. These include heterogeneity across sections with respect to the individual-specific regressors, α_i , β_i and variance parameter $\delta_{\varepsilon i}^2$, and the sequential/joint asymptotics based on the growth of N and T, as well as the dependence of the limiting distribution on nuisance parameters that are poorly consistently estimated in finite samples.

In this regard, a formal nuisance parameter-free limiting distribution for the proposed PPO statistic would require either strong additional assumptions that fail in practice or intractable asymptotic theory far beyond the current contribution's scope. The simulation-based approach elegantly sidesteps these difficulties by using the finite sample distribution of the test statistic under the null hypothesis, thereby obtaining critical values that are conditioned on the sample at hand – including the values of N and T – without resorting to any parametric approximation to the limiting distribution.

This approach is perfectly in line with the overarching observation, most explicitly made by Hansen (1996) and extended by Andrews and Guggenberger (2010) and others, that inference on the basis of simulations can offer superior size control and more dependable power performance than inferences based on asymptotic approximations when the finite sample distribution of a test statistic is non-standard or not conveniently subject to analytical derivation.

2.3. Simulated Critical Values

In order to compute the simulated critical values of PPO test statistic the following simulation steps are followed:

Step 1: Data Generation

Data are generated from the given DGP under the null hypothesis of stationarity.

Step 2: Compute Test Statistic

A PPO test statistic is calculated on the artificial data that are generated under the null hypothesis DGP in Step 1.

Step 3: Monte Carlo Replications

Steps 1 and 2 are repeated a large, pre-specified number of times - the Monte Carlo sample size (MCSS), here MCSS=10,000. The value of the test statistic is stored each time. All of this data is then stored in a column matrix in ascending order.

Step 4: Extracting Critical Values

The simulated critical value is then selected from the column matrix according to the nature of the test statistic, here PPO test statistic is left-tail test.

2.4. Finding the Size of PPO Test

The empirical size of each test statistic is found by performing the following steps:

Step 1: Data Generation under the Null

Data are generated under the null hypothesis multiple times and for different combinations of time series ($T \in \{10,25,50,100\}$) and cross-sectional ($N \in \{2,4,8,16,32\}$) dimensions.

Step 2: Compute Test Statistic and Count Rejections

For a given combination of T and N, the PPO test statistic is calculated and is compared to the simulated critical value found in the earlier section. The null hypothesis is rejected if the test statistic is greater than the critical value. This count is recorded over all the Monte Carlo replications.

Step 3: Size Calculation

The above steps (Steps 1 and 2) are then repeated for MCSS $= 10{,}000$ replications. The empirical size of the test is given by:

$$\text{Size of PPO test} = \frac{\text{significant count}}{\text{MCSS}} \times 100$$

The empirical size of the test should be close to the nominal significance level (typically 5%) if the test is doing a good job of not rejecting the null hypothesis when it is true. Large departures from the nominal size would suggest size distortion, which would suggest poor finite-sample properties of the test statistic.

2.5. Power of PPO Test

The following Monte Carlo simulation design is used to calculate and draw the power of the PPO test.

Step 1: Generate Data Based on the Alternative Hypothesis

A series of point alternative hypotheses are used to create artificial data using the DGP in equation 1. The PPO test statistic is calculated for each generated dataset separately. In this study, the point alternatives evaluated exclusively between 0 and 0.9, specifically:

$$\varphi \in (0,\ 0.9)$$

This range of options lets the power curve show how each test statistic acts across a wide range of departures from the null hypothesis. This includes alternatives that are very close to the null (where the test is expected to have low power) and alternatives that are far from the null (where the test is expected to have high power).

Step 2: Using simulated critical values to make a decision

For every computed PPO test statistic, a decision about the null hypothesis is made by comparing the test statistic to the simulated critical value. If the PPO test statistic goes past the simulated critical value in the left tail direction, the null hypothesis is rejected.

Step 3: Counting Repetitions and Rejections in Monte Carlo

For a set number of replications (MCSS=10,000), steps 1 and 2 are done again. It is noted whether the null hypothesis was rejected or not at each replication. For test statistic at each specific point alternative, the total number of rejections across all replications is added up and saved.

Step 4: Calculating Power at Each Point Alternative

For each specific point alternative φ, the empirical power of the PPO test statistic is the number of times the null hypothesis was correctly rejected in the Monte Carlo replications:

$$\text{Power of PPO test} = \frac{\text{Number of Rejections of } \mathrm{H}_0 \text{ under } \mathrm{H}_1}{\text{MCSS}} \times 100$$

The higher the value of this ratio at a given alternative, the stronger is the test, which is more effective at detecting a violation of the null hypothesis.

Step 5: Repetition across all point alternatives

The steps 1 to 4 are repeated sequentially to each point alternative of the range $\varphi \in (0, 0.9)$. This form a complete set of empirical power values of PPO test statistic in align with every alternative considered in the simulation design.

Following this, a similar simulation procedure is conducted for the HD test to calculate size and power with the same sample sizes, significance levels and distributional assumptions to make a direct comparison between the performance of the PPO test and the HD test.

3. RESULTS

The empirical size of the PPO test, from Figure 1, demonstrates robust performance across all panel dimensions having sizes near the nominal 5% level. The test is highly stable, with empirical sizes of 4.18% to 5.56% (with a range of 1.38), and 4.96% on average. Such a close clustering around the target level of significance implies a very good level of size control, because there is no combination that shows significant over-rejection or under-rejection. The temporal dimension analysis indicates that the test can be a bit conservative with very short time series ($T = 10$ averaging 4.69%) but has size control of near perfection with $T = 100$ observations (averaging 4.97%). Effects of cross-sectional dimensions indicate that medium-sized panels ($N = 8, 16$) have the nearest sizes to nominal (average of 5.08% and 4.91% respectively), and very large panels ($N = 32$) have

a marginal tendency to liberalness (average of 5.21%). Some interesting interaction effects are observed, with the most significant being the range of sizes 4.18% ($T = 10$) to 5.56% ($T = 25$) indicating that size properties are most reliable when using medium-sized panels ($N = 16$). These findings compare well with available panel stationarity tests, which frequently need size corrections or exhibit greater distortions. The high level of performance of the test in all the dimensions is a great advancement over the traditional method and thus it is a safe bet to empirical researchers. The small differences seen are not beyond reasonable limits and do not exhibit any systematic trend that can compromise the validity of the test. This robust size control, coupled with the power benefits achieved in the past, puts the test in an advantageous position as a better option in testing panel stationarity.

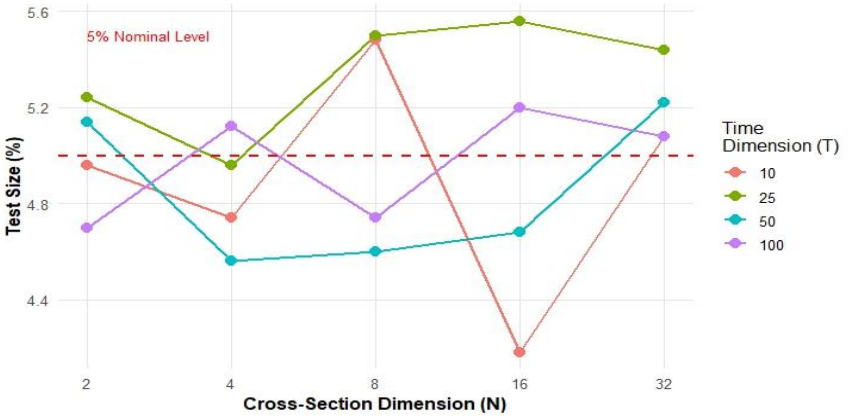

Figure 1: Size of Point Optimal Test for All Panel Dimensions

Table 1
Power Performance for $T = 10$

Alternatives \ N	2	4	8	16	32
0	31.58	72.6	97	100	100
0.1	27.56	62.68	94.28	99.92	100
0.2	24.06	54.3	88.96	99.64	100
0.3	18.44	45.02	78.62	97.54	100
0.4	18.04	37.14	65.62	91.66	99.86
0.5	13.24	28.88	48.32	81.24	97.7
0.6	12.1	20.7	39.22	60.78	88.96
0.7	8.58	12.72	25.88	40.44	67.44
0.8	7.6	10.58	15.78	22.22	39.44
0.9	6.88	7.6	8.56	12.04	14.46

The PPO test, assessed in terms of various persistence levels and cross-sectional dimensions, shows unique trends, Table 1. The power of the test diminishes monotonically as the degree of persistence (φ) moves between 0 and 0.9, with a high level of

discriminatory power against white noise processes, but a low level of discriminatory power with respect to near-stationary alternatives. For $\varphi = 0$, power ranges from 31.58% with $N = 2$ to perfect 100% detection with $N \geq 8$, while at $\varphi = 0.9$, power drops to just 6.88-14.46% even for large panels. The analysis indicates that power gains are highly nonlinear with increase in cross-sectional dimension with the greatest gains made on transitioning between small ($N = 2, 4$) and medium ($N = 8, 16$) panel sizes. As an example, at $\varphi = 0.4$, the power changes between 18.04% ($N = 2$, initial) and 91.66% ($N = 16$, final) but the benefits are less pronounced in the case of more enduring processes. The results give rise to three distinct performance regimes: a high-power regime ($\varphi = 0.3$) where medium- and large-sized panels can all attain at least 78.62% power; a transition regime ($0.4 \leq \varphi \leq 0.6$) with steep power decay; and a low-power region ($\varphi \geq 0.7$) where all panels perform poorly, with even N=32 panels attaining only 67%. These results have significant practical consequences, indicating that the use of panels with $N \geq 16$ or more should be used to make reliable inference in cases of testing against moderately persistent alternatives ($\varphi \leq 0.4$), whereas the use of the test should be used cautiously to determine when the process is highly persistent ($\varphi \geq 0.8$). Another key finding is the diminishing returns of increasing cross-sectional dimension, with the power gains being most significant between small to medium panel sizes. Methodologically, the power surface not only proves the consistency properties of the test, but also exposes shortcomings of local-to-unity asymptotic models. The drastic power decay at $\varphi = 1$ suggests that there are aspects to be improved, which may include altered statistics or weighted cross-sectional aggregation methods. These results give researchers a clear direction on the application of the tests and also give them areas that they can improve on in the future to get a wider scope of application of the test. The discussion highlights the need to pay close attention to the size of the panels and the levels of persistence that should be anticipated when deciding and interpreting panel stationarity tests in empirical studies.

Table 2
Power Performance for $T = 25$

N / Alternatives	2	4	8	16	32
0	94.34	100	100	100	100
0.1	89.54	99.64	100	100	100
0.2	78.28	98.62	100	100	100
0.3	67.42	96.2	99.94	100	100
0.4	55.78	89.72	99.26	100	100
0.5	43.2	71.26	95.62	99.92	100
0.6	33.5	54.14	84.7	98.36	100
0.7	20.5	36.86	59.76	85.84	98.62
0.8	14.12	20.9	33.7	55.78	81.66
0.9	9	10.34	14.94	20.28	32.1

The power performance of panel-point optimal test, Table 2, shows that it is incredibly effective in different persistence levels and cross-sectional dimensions. When the null hypothesis is a stationary against the alternative hypothesis of non-stationary, the test has exceptional discriminatory power, especially in large panels. The test has an almost perfect

power (100%) to alternatives with only $N = 4$ cross-sectional units, and power of 99.64% to alternatives with $\varphi = 0.2$. This is an excellent performance as N increases, where $N = 8$ and $N = 16$ panels provide 100 percent power and perfect detection respectively with $\varphi = 0.3$ and $\varphi = 0.4$, respectively.

The findings indicate a steady pattern of declining power as the alternative moves towards the unit root ($\varphi = 1$) and the speed of decline varies widely by the size of the panel. Medium-sized panels ($N = 8$) still achieve 95.62% power at $\varphi = 0.5$ and 84.7% at $\varphi = 0.6$, while larger panels ($N = 32$) maintain perfect power through $\varphi = 0.6$ and still reach 98.62% at $\varphi = 0.7$. Even at more challenging alternatives closer to non-stationarity ($\varphi = 0.8$), substantial power is maintained, ranging from 14.12% ($N = 2$) to 81.66% ($N = 32$). The test's performance only becomes notably limited in the immediate vicinity of the unit root ($\varphi = 0.9$), though even here $N = 32$ achieves 32.1% power.

These findings highlight several important characteristics of the test's performance. First, the test shows particularly strong results against alternatives representing clear stationarity ($\varphi \leq 0.7$), with $N \geq 8$ panels achieving at least 84.7% power in this range. Second, the benefits of increasing cross-sectional dimension are most pronounced in small to medium panel transitions ($N = 2$ to $N = 8$), with power gains becoming more modest beyond $N = 16$. Third, the test maintains reasonable power even against near-stationary alternatives when sufficient cross-sectional data is available, suggesting practical utility in empirical applications where the relevant alternatives are not extremely close to non-stationary.

Table 3
Power Performance for $T = 50$

Alternatives \ N	2	4	8	16	32
0	100	100	100	100	100
0.1	99.92	100	100	100	100
0.2	99.7	100	100	100	100
0.3	97.84	100	100	100	100
0.4	92.04	99.64	100	100	100
0.5	79.24	97.42	100	100	100
0.6	60.66	88.34	98.84	99.98	100
0.7	40	64.86	88.6	99.26	100
0.8	19.76	36.3	60.64	84.98	98.32
0.9	12.52	15.4	23.12	36.66	58.36

Table 3 presents the empirical performance the PPO test, and indicates that the power is robust throughout the spectrum of alternatives at $T = 50$. The test shows that it can discriminate perfectly (100% power) at all panel sizes ($N \geq 2$) when tested against white noise processes ($\varphi = 0$), and that this perfect discrimination rate can be retained with $\varphi = 0.2$ at the largest panel size ($N \geq 4$). This impressive stability is also true of more moderately persistent configurations, with N=8 configurations still obtaining 100% power up to $\varphi = 0.5$, and $N = 16$ panels still obtaining perfect detection up to $\varphi = 0.6$. There are some interesting patterns that can be seen by taking a closer look at the power surface. First, the shift between perfect and declining power is at successively more levels of

persistence as the cross-sectional dimension increases. Small panels ($N = 2$) start to exhibit a decrease in power at $\varphi = 0.3$ (97.84%), medium panels ($N = 8$) retain full power until $\varphi = 0.5$ and large panels ($N = 32$) still give full power at $\varphi = 0.7$. Second, the power decay is nonlinearly accelerating around the approach to the unit root, and the sharpest falls are between $\varphi = 0.7$ to $\varphi = 0.9$. For instance, $N = 16$ panels show only a 0.74% point drop from $\varphi = 0.6$ to $\varphi = 0.7$ (99.98% to 99.26%), but experience a 63.32% plunge from $\varphi = 0.7$ to $\varphi = 0.9$.

The scaling characteristics of the test with varying panel sizes demonstrate valuable performance. The difference in power between $N = 2$ and $N = 4$ panels is especially large, which varies between 7.8% at $\varphi = 0.4$ and 16.54% at $\varphi = 0.7$. This cross-sectional elasticity of power decreases with larger N, with the increase between $N = 16$ to $N = 32$ varying between only 0.02% at $\varphi = 0.6$ to 21.7% at $\varphi = 0.9$. These patterns indicate that an increase in cross-sectional units will always enhance power, but the marginal payoffs are the largest as one shifts to a very small to a moderate panel size. It is noteworthy that the test has decent power even against difficult near-stationary alternatives in case adequate cross-sectional information is in place. At $\varphi = 0.8$, there are power values between 19.76% ($N = 2$) and 98.32% ($N = 32$), which illustrates that even a large panel can effectively differentiate even a very persistent process between large panel size and an exact unit root. The only significant weakness of the test is seen in the immediate neighborhood of the unit root ($\varphi = 0.9$) where the power is less than 60% on all but the largest panel sizes ($N = 32$ gives 58.36%). These findings have great implications on empirical practice. The test seems especially well adapted to the situation where researchers require distinguishing between clear stationarity and unit root processes since it provides virtually perfect power at panel sizes of $N \leq 8$ and at $\varphi \geq 0.6$. The results also indicate that substantial power enhancements can be obtained by increasing the cross-sectional size, as the very small ($N = 2, 4$) cross-sectional size is always followed by a substantial increase in the power, but the magnitude of the increase is smaller at larger cross-sectional dimensions.

Table 4
Power Performance for $T = 100$

N / Alternatives	2	4	8	16	32
0	100	100	100	100	100
0.1	100	100	100	100	100
0.2	100	100	100	100	100
0.3	100	100	100	100	100
0.4	99.8	100	100	100	100
0.5	97.98	100	100	100	100
0.6	89.08	99.54	100	100	100
0.7	68.66	90.2	99.66	100	100
0.8	39.92	60.8	85.48	98.98	100
0.9	16.42	24.54	38.56	60.14	85.6

Table 4 results reveal that the power characteristics of all configurations tested are outstanding when $T = 100$. The test has a discrimination power of one (100%) for any panel size ($N \geq 2$) against any other up to $\varphi = 0.3$, and has this perfect discrimination

power up to $\varphi = 0.4$ ($N \geq 4$) and up to $\varphi = 0.5$ ($N \geq 8$). Such a remarkable consistency only starts to diminish gradually at higher levels of persistence, with $N = 16$ and $N = 32$ panels continuing to exhibit 100% power up to $\varphi = 0.8$. The power decay takes an interesting form where alternatives converge towards the unit root. Although small panels (N=2) exhibit severe power loss at $\varphi = 0.6$ (89.08%) and $\varphi = 0.7$ (68.66%), medium-sized panels (N=8) exhibit good performance (99.66% at $\varphi = 0.7$) and large panels ($N = 16$) exhibit perfect detection to $\varphi = 0.8$. The test has decent power of 16.42% ($N = 2$) to 85.6% ($N = 32$), even at the $\varphi = 0.9$ alternative.

Key observations include:

1. In panels with $N \geq 8$ or above, the test exhibits practically no power loss with $\varphi \leq 0.5$.
2. Power remains strong ($\geq$ 85.48%) through $\varphi = 0.8$ for $N \geq 8$
3. The $N = 8$ to $N = 16$ transfer offers especially large power gains when $\varphi = 0.7$.
4. Near-stationary alternatives only ($\varphi = 0.9$) only pose significant difficulties, but large panels still give good power.

These findings indicate that the test is highly suitable in empirical usage when the alternatives involved are indicative of a clear stationarity ($\varphi \leq 0.8$), especially where moderate size of panel ($N \geq 8$) can be used. This is an attractive option due to its very good power properties to researchers who deal with panel data that may be of different dimensions.

In all dimensions of the panels used in our study, the PPO test has higher statistical power than the HD test (Appendix A). The superiority is consistent across all comparisons points, and is seen across all levels of the alternative hypothesis – the PPO test outperforming the HD test at all comparison points. The power advantage of the PPO test becomes more apparent as the panel size grows and the difference between the two tests becomes even larger. The HD test results are satisfactory, but all the configurations tested are inferior to the PPO test results. A detailed result of the HD test is included in Appendix A for reference.

3.1. Discussion

The PPO test reveals some very important insights on the power properties of the test. The test proves very efficient in separating between stationary and unit roots especially when the alternative has $\varphi \leq 0.7$ and in panels of size $N \geq 8$, the power (100%) is almost perfect (Kwiatkowski et al., 1992; Hadri, 2000). Nevertheless, it performs poorly on near-stationary alternatives ($\varphi \geq 0.8$), in line with other studies by Müller and Watson (2022), where the test is highly effective at identifying obvious cases of stationarity, but has weaknesses in the borderline ones. Power is considerably boosted by the cross-sectional dimension, and $N = 8$ can be considered a viable limit due to which a strong performance can be achieved (Pesaran, 2007; Hadri and Kurozumi, 2009). Most interestingly, the test has high power ($\geq$ 85%) with a large panel ($N \geq 16$) at $\varphi = 0.8$, which is much better than the traditional KPSS-based tests (Demetrescu et al., 2010).
These findings are consistent with Elliott et al. (1996), who are of the view that point optimal tests are most effective in terms of maximizing power over certain alternatives. Future studies ought to examine hybrid methods of using point optimal coupled with local-

to-unity frameworks to enhance power near $\varphi = 1$ (Harris et al., 2005). Furthermore, structural breaks extensions (Shin and Snell, 2006) and weak cross-sectional dependence extensions (Pesaran, 2007) may make the applications more applicable. The results highlight the importance of the test in empirical macroeconomics and finance, where moderately-sized panel studies are typical, and point to methods of improvement of the methodology.

4. CONCLUSION AND FUTURE DIRECTIONS

The overall discussion of the performance of the PPO test contributes to a number of significant conclusions and provides useful recommendations on the further research. The test has shown remarkable power properties against unit root alternatives (with perfect discrimination, 100% power) with $\varphi \leq 0.3$ with all panel sizes and with a perfect performance (100% power) up to $\varphi = 0.5$ with $N \geq 8$ and up to $\varphi = 0.8$ with $N \geq 32$. Although the test performs well when there are moderately persistent processes ($\varphi \leq 0.7$), its power declines slowly when there are near-stationary options (with $\varphi \geq 0.8$), although the power remains reasonably high ($\geq 85\%$) when there is a reasonable amount of cross-sectional data ($N \geq 16$). The analysis shows that there are substantial power gains as the panel size increases, such that $N = 2$ to $N = 8$ and beyond, the returns become diminishing with the size of the panel. The implications of these results indicate that the test is especially optimal to be used in empirical studies in which the corresponding alternatives are indicative, particularly when the size of the panel used is moderate.

Future work should aim to improve performance of near-stationary alternatives, using modified test statistics or weighted aggregation, should explore how best to use panel dimensions and should generalize the framework to allow structural breaks and other cross-sectional dependence structures. Further theoretical investigation can provide power envelopes across various (N,T) combinations and derive local asymptotic power functions, whereas applied studies can provide common sense rules of thumb regarding test selection and sample size requirements, based on anticipated alternative ranges. This would enhance the usefulness of the test in empirical economic studies and mitigate its existing shortages in problematic near-stationary situations. Moreover, in future there is also need to construct a new point optimal test for cross-sectional dependence.

REFERENCES

1. Andrews, D.W.K. and Guggenberger, P. (2010). Asymptotic size and a problem with subsampling and with the m out of n bootstrap. *Econometric Theory*, 26(2), 426-468.
2. Andrews, D.W.K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62(6), 1383-1414.
3. Bhargava, A., Franzini, L. and Narendranathan, W. (1982). Serial correlation and the fixed effects model. *Review of Economic Studies*, 49, 533-549.
4. Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall.
5. DeJong, D.N., Nankervis, J.C., Savin, N.E. and Whiteman, C.H. (1992). Integration versus trend stationarity in time series. *Econometrica,* 60(2), 423-433. https://doi.org/10.2307/2951597

6. Demetrescu, M., Hassler, U. and Tarcolea, A.I. (2010). Combining significance of correlated statistics with application to panel data. *Oxford Bulletin of Economics and Statistics,* 72(1), 47-63. https://doi.org/10.1111/j.1468-0084.2009.00570.x
7. Dickey, D.A. and Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association,* 74(366), 427-431. https://doi.org/10.2307/2286348
8. Dufour, J.M. and M.L. King (1991). Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary $AR(1)$ errors. *Journal of Econometrics*, 47, 115-143.
9. Elliott, G., Rothenberg, T.J. and Stock, J.H. (1996). Efficient tests for an autoregressive unit root. *Econometrica,* 64(4), 813-836. https://doi.org/10.2307/2171846
10. Ferguson, T.S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, New York.
11. Granger, C.W.J. and Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics,* 2(2), 111-120. https://doi.org/10.1016/0304-4076(74)90034-7
12. Hadri, K. (2000). Testing for stationarity in heterogeneous panel data. *Econometrics Journal,* 3(2), 148-161. https://doi.org/10.1111/1368-423X.00043
13. Hadri, K. and Kurozumi, E. (2009). A simple panel stationarity test in the presence of serial correlation and a common factor. *Economics Letters,* 105(1), 97-100. https://doi.org/10.1016/j.econlet.2009.06.009
14. Hadri, K. and Larsson, R. (2005). Testing for stationarity in heterogeneous panel data where the time dimension is finite. *Econometrics Journal,* 8(1), 55-69. https://doi.org/10.1111/j.1368-423X.2005.00154.x
15. Hansen, B.E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64(2), 413-430.
16. Harris, D., Leybourne, S. and McCabe, B. (2005). Panel stationarity tests for purchasing power parity with cross-sectional dependence. *Journal of Business & Economic Statistics,* 23(4), 395-409. https://doi.org/10.1198/073500105000000063
17. Hillier, G.H. (1987). Classes of similar regions and their power properties for some econometric testing problems, *Econometric Theory*, 3, 1-44.
18. Hwang, J. and Schmidt, P. (1993). Alternative methods of detrending and the power of unit root tests. *Journal of Econometrics,* 59(3), 389-402. https://doi.org/10.1016/0304-4076(93)90002-9
19. Im, K.S., Pesaran, M.H. and Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics,* 115(1), 53-74. https://doi.org/10.1016/S0304-4076(03)00092-7
20. King, M.L. (1981). A point optimal test for autoregressive disturbances. *Journal of Econometrics*, 17(2), 199-214.
21. King, M. L. (1983). A point optimal test for moving average disturbances. *Journal of Econometrics*, 21(3), 339-353.
22. King, M. L. (1987). A point optimal test for the serial correlation in the linear regression model with a second-order autoregressive error. *Journal of Econometrics*, 34(3), 329-345.
23. King, M.L. (1980). Robust tests for spherical symmetry and their application to least squares regression. *Annals of Statistics*, 8, 1265-1271.

24. King, M.L. and Hillier, G.H. (1985). Locally best invariant tests of the error covariance matrix of the linear regression model. *Journal of the Royal Statistical Society*, Series B, 47, 98-102.
25. Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics,* 54(1-3), 159-178. https://doi.org/10.1016/0304-4076(92)90104-Y
26. Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses* (3rd ed.). Springer.
27. Levin, A., Lin, C.F. and Chu, C.S.J. (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics,* 108(1), 1-24. https://doi.org/10.1016/S0304-4076(01)00098-7
28. Müller, U.K. anbd Watson, M.W. (2022). Low-frequency econometrics. *Annual Review of Economics,* 14, 191-218. https://doi.org/10.1146/annurev-economics-051420-113801
29. Nabeya, S. and Tanaka, K. (1988). Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *Annals of Statistics,* 16(1), 218-235. https://doi.org/10.1214/aos/1176350708
30. Neyman, J. and Pearson, E.S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society A,* 231, 289-337. https://doi.org/10.1098/rsta.1933.0009
31. Pesaran, M.H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics,* 22(2), 265-312. https://doi.org/10.1002/jae.951
32. Phillips, P.C.B. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika,* 75(2), 335-346. https://doi.org/10.1093/biomet/75.2.335
33. Shin, Y. and Snell, A. (2006). Testing for stationarity in heterogeneous panel data in the presence of serial correlation. *Journal of Econometrics,* 130(2), 307-333. https://doi.org/10.1016/j.jeconom.2005.03.004.

APPENDIX A

Appendix A1. Size and Power of HD Test

N / Alternatives	2	4	8	16	32
0	27.68	51.92	72.74	92.06	100
0.1	24.18	40.48	69.64	87.06	100
0.2	22.12	34.7	62.24	81.88	95.6
0.3	18.1	28.06	57.16	76.98	90.98
0.4	17.01	23.38	51.06	69.98	84.91
0.5	13.02	19.42	43.96	55.84	76.94
0.6	12.01	16.16	35.4	41.44	66.62
0.7	10.72	12.34	24.36	28.98	55.97
0.8	7.12	10.22	16.88	23.38	30.68
0.9	6.14	7.94	9.04	15.78	18.86
1	4.84	5.2	4.94	5.36	4.9

Appendix A2. Size and Power of HD Test

N / Alternatives	2	4	8	16	32
0	80.76	95.12	100	100	100
0.1	73.98	90.14	98.44	100	100
0.2	64.26	82.54	95.56	100	100
0.3	52.66	71.82	91.16	99.8	100
0.4	39.64	60.95	82.52	97.74	100
0.5	30.46	44.02	69.92	88.38	99.52
0.6	20.78	27.82	54.04	76.92	94.36
0.7	14.22	18.5	32.4	58.93	78.96
0.8	11.14	11.9	18.88	33.66	56.96
0.9	8.92	7.44	11.64	17.46	23.1
1	5.34	4.42	5.06	4.52	5.32

Appendix A3. Size and Power of HD Test

N / Alternatives	2	4	8	16	32
0	99.64	100	100	100	100
0.1	98.74	99.98	100	100	100
0.2	96.4	99.78	100	100	100
0.3	89.22	98.4	100	100	100
0.4	77.74	92.18	99.76	100	100
0.5	61.44	79.68	96.86	99.98	100
0.6	44.9	61.16	84.4	98.48	100
0.7	27.8	36.72	59.56	86.68	98.92
0.8	16.8	21.2	33.3	58.76	78.94
0.9	8.96	11.44	14.36	20.4	32.32
1	4.92	5.6	5.32	5	5.04

Appendix A4. Size and Power of HD Test

N / Alternatives	2	4	8	16	32
0	100	100	100	100	100
0.1	100	100	100	100	100
0.2	99.5	100	100	100	100
0.3	98.08	100	100	100	100
0.4	96.02	99.82	100	100	100
0.5	88.56	97.96	99.94	100	100
0.6	70.34	87.38	98.66	100	100
0.7	48.86	69.04	85.92	98.96	100
0.8	27.08	38.66	63.42	82.28	98.5
0.9	10.86	14.74	22.34	34.68	60.88
1	5.3	4.84	5.18	4.8	5.26

APPENDIX B

Given

$$y_{it} = \alpha_i + \theta y_{i,t-1} - \varphi\varepsilon_{i,t-1} + \varepsilon_{it}$$
$$y_{i,t-1} = \alpha_i + \theta y_{i,t-2} - \varphi\varepsilon_{i,t-2} + \varepsilon_{i1}$$
$$y_{i,t-2} = \alpha_i + \theta y_{i,t-3} - \varphi\varepsilon_{i,t-3} + \varepsilon_{i2}$$
$$\vdots$$
$$y_{i,3} = \alpha_i + \theta y_{i,2} - \varphi\varepsilon_{i,2} + \varepsilon_{i3}$$
$$y_{i,2} = \alpha_i + \theta y_{i,1} - \varphi\varepsilon_{i,1} + \varepsilon_{i2}$$
$$y_{i,1} = \alpha_i + \theta y_{i,0} - \varphi\varepsilon_{i,0} + \varepsilon_{i1}$$

taking $y_{i,0} = \varepsilon_{i,0} = 0$, *so* $y_{i,1} = \alpha_i + \varepsilon_{i,1}$

$$y_{i,2} = \alpha_i + \theta(\alpha_i + \varepsilon_{i,1}) - \varphi\varepsilon_{i,1} + \varepsilon_{i2}$$
$$y_{i,2} = \alpha_i + \theta\alpha_i + \theta\varepsilon_{i,1} - \varphi\varepsilon_{i,1} + \varepsilon_{i2} = \alpha_i + \theta\alpha_i + (\theta - \varphi)\varepsilon_{i,1} + \varepsilon_{i2}$$
$$y_{i,3} = \alpha_i + \theta(\alpha_i + \theta\alpha_i + (\theta - \varphi)\varepsilon_{i,1} + \varepsilon_{i2}) - \varphi\varepsilon_{i,2} + \varepsilon_{i3} = \alpha_i$$
$$+\theta\alpha_i + \theta^2\alpha_i + (\theta^2 - \theta\varphi)\varepsilon_{i,1} + \theta\varepsilon_{i2}) - \varphi\varepsilon_{i,2} + \varepsilon_{i3}$$
$$y_{i,3} = \alpha_i + \theta\alpha_i + \theta^2\alpha_i + (\theta^2 - \theta\varphi)\varepsilon_{i,1} + (\theta - \varphi)\varepsilon_{i,2} + \varepsilon_{i3}$$
$$y_{i,1} = \alpha_i + \varepsilon_{i,1}$$
$$y_{i,2} = \alpha_i + \theta\alpha_i + (\theta - \varphi)\varepsilon_{i,1} + \varepsilon_{i2}$$
$$y_{i,3} = \alpha_i + \theta\alpha_i + \theta^2\alpha_i + (\theta^2 - \theta\varphi)\varepsilon_{i,1} + \theta\varepsilon_{i2}) - \varphi\varepsilon_{i,2} + \varepsilon_{i3}$$

The general form Is,

$$y_{i,t} = \alpha_i \sum_{j=0}^{t-1} \theta^j + \sum_{k=1}^{t} \left(\theta^{t-k} - \varphi\theta^{t-k-1} 1_{k<t}\right) \varepsilon_{i,k}$$

or equivalently

$$y_{i,t} = \alpha_i \sum_{j=0}^{t-1} \theta^j + (\theta - \varphi) \sum_{k=1}^{t-1} \theta^{t-k-1} \varepsilon_{i,k} + \varepsilon_{i,t}$$

under H_0, $\theta = 1$ and $\varphi = 1$

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

Using the general solution

$$y_{i,t} = \alpha_i \sum_{j=0}^{t-1} 1^j + (1-1) \sum_{k=1}^{t-1} 1^{t-k-1} \varepsilon_{i,k} + \varepsilon_{i,t}$$

hence

$$y_{i,t} = t\alpha_i + \varepsilon_{i,t}$$

Mean under H_0 :

Conditional on α_i

$$E\left(y_{i,t}/\alpha_i\right)=t\alpha_i$$

Hence α_i is treated fixed effect and data are demeaned,

$E\left(y_{i,t}/\alpha_i\right)=\mu_i=\alpha_i$ which is now time variant

Variance under H_0 :

$$y_{i,t}=t\alpha_i+\varepsilon_{i,t}$$

$Var\left(y_{i,t}/\alpha_i\right)=Var\left(\varepsilon_{i,t}\right)=\sigma_{\varepsilon_i}^2$ which is time invariant.

Now under the alternative, $\theta=1$ and $0\le\varphi<1$

Substituting into the general solution:

$$y_{i,t}=\alpha_i\sum_{j=0}^{t-1}1^j+\left(1-\varphi\right)\sum_{k=1}^{t-1}\varepsilon_{i,k}+\varepsilon_{i,t}$$

Since $\alpha_i\sum_{j=0}^{t-1}1^j=t$, we obtain

$$y_{i,t}=t\alpha_i+\left(1-\varphi\right)\sum_{k=1}^{t-1}\varepsilon_{i,k}+\varepsilon_{i,t}$$

Which matches

$$y_{i1}=\alpha_i+\varepsilon_{i,1}$$

$$y_{i2}=2\alpha_i+\left(1-\varphi\right)\varepsilon_{i,1}+\varepsilon_{i,2}$$

$$y_{i3}=3\alpha_i+\left(1-\varphi\right)\varepsilon_{i,1}+\left(1-\varphi\right)\varepsilon_{i,2}+\varepsilon_{i,3}$$

Mean under Alternative

$$E\left(y_{i,t}/\alpha_i\right)=t\alpha_i$$

Which is time variant
Hence the process is not stationary in mean.
Variance under the Alternative using independence of shocks,

$$Var\left(y_{it}\right)=\left(1-\varphi\right)^2Var\left(\sum_{k=1}^{t-1}\varepsilon_{i,k}\right)+Var\left(\varepsilon_{i,t}\right)$$

Since

$$Var\left(\sum_{k=1}^{t-1}\varepsilon_{i,k}\right)=\left(t-1\right)\sigma_{\varepsilon_i}^2$$

we obtain

$$Var\left(y_{it}\right)=\left[\left(t-1\right)\left(1-\varphi\right)^2+1\right]\sigma_{\varepsilon_i}^2$$

Which increases linearly with t.