

Westerlund, Joakim

**Working Paper**

## Pooled Unit Root Tests in Panels with a Common Factor

Working Paper, No. 2005:9

**Provided in Cooperation with:**

Department of Economics, School of Economics and Management, Lund University

*Suggested Citation:* Westerlund, Joakim (2005) : Pooled Unit Root Tests in Panels with a Common Factor, Working Paper, No. 2005:9, Lund University, School of Economics and Management, Department of Economics, Lund

This Version is available at:

<https://hdl.handle.net/10419/259901>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Pooled Unit Root Tests in Panels with a Common Factor

Joakim Westerlund\*

January 26, 2005

## Abstract

This paper proposes new pooled panel unit root tests that are appropriate when the data exhibit cross-sectional dependence that is generated by a single common factor. Using sequential limit arguments, we show that the tests have a limiting normal distribution that is free of nuisance parameters and that they are unbiased against heterogenous local alternatives. Our Monte Carlo results indicate that the tests perform well in comparison to other popular tests that also presumes a common factor structure for the cross-sectional dependence.

**JEL Classification:** C12; C31; C33.

**Keywords:** Pooled Unit Root Tests; Panel Data; Common Factor; Cross-Sectional Dependence; Monte Carlo Simulation.

## 1 Introduction

During the last few years there has been an immense proliferation of research concerned with the problem of testing for unit roots in panel data. Some of the most influential contributions within this field include Choi (2001), Im *et al.* (2003), Levin *et al.* (2002), and Harris and Tzavalis (1999). A common feature of these studies is that they all assume that the individual time series of the panel are independent of each other. Under this assumption, various central limit theorems can be applied to obtain test statistics that achieve asymptotic normality. Although it has been widely recognized that cross-sectional independence may be an overly restrictive assumption, it has long been thought that subtracting the cross-sectional average from the data before application of the panel unit root test could be employed to, at least partially, deal with this problem. Recently, however, it has become increasingly clear that cross-sectional demeaning of this sort may not work in general as it cannot be used

---

\*Department of Economics, Lund University, P. O. Box 7082, S-220 07 Lund, Sweden. Telephone: +46 46 222 4970; Fax: +46 46 222 4118; E-mail address: joakim.westerlund@nek.lu.se.

to accommodate correlations that differ between pairs of individual time series, which seem like a more realistic assumption in many empirical applications such as macroeconomics and international finance. For instance, O'Connell (1998) argues that real exchange rates should be highly correlated across countries due to the strong links between financial markets and because of the use of a numeraire country in constructing real rates.

Recognizing this deficiency, several new panel data unit root tests have been proposed in the recent literature. These tests are distinct in that they make explicit assumptions regarding the structure of the cross-sectional correlation. Chang (2004), Maddala and Wu (1999), and Smith *et al.* (2003) avoid the restrictive nature of the cross-sectional demeaning by employing bootstrap techniques, which make the unit root test valid under quite general forms of cross-sectional dependence. Other tests that allow for general correlation structures include those of Chang (2002), O'Connell (1998) and Jönsson (2004). To eliminate the cross-sectional dependence, Chang (2002) proposes a nonlinear instrumental variables estimator, while O'Connell (1998) suggests estimation by generalized least squares. The test of Jönsson (2004) relies on a robust estimator of the standard errors to handle the impact of cross-sectional dependence. The tests proposed by Ng and Bai (2004), and Moon and Perron (2004) also allow for more general forms of cross-sectional dependence by assuming that the correlation can be modelled using a generalized common factor structure. These studies recognize that the common factors are likely to have differential effects on different cross-sectional units by allowing for the possibility of heterogeneous factor loadings. This avenue is potentially very fruitful as it takes an important step in the direction of reducing the dimension of the cross-sectional correlation and it shall therefore be employed in this paper.

The suggestions of Ng and Bai (2004), and Moon and Perron (2004) are very similar in that they both allow for a very general factor structure with an unknown number of factors. Moon and Perron (2004) proposes several pooled panel unit root test statistics based on defactored data and suggest estimating the unknown factor loadings using the method of principal components. Using joint limit arguments, Moon and Perron (2004) derive the asymptotic distribution of their statistics under both the null and local alternative hypotheses. Monte Carlo results, suggestive of good small-sample properties, are also provided. The setup of Ng and Bai (2004) is even more general insofar they allow for the possibility of unit roots and cointegration among the common factors. In so doing, however, they face the problem of how to estimate the integrated factors. Their solution involves first estimating the factors using the differentiated data by principal components and then to apply the unit root test onto the recumulated and defactored data. As in Moon and Perron (2004), Ng and Bai (2004) uses joint limits to derive the limiting distribution of their heterogeneous test statistics. Their Monte Carlo results suggest that the tests perform well even in very small samples.

In this paper, we follow Ng and Bai (2004), and Moon and Perron (2004) and assume a common factor structure for the cross-sectional dependence. In contrast to these authors, however, we make the simplifying assumption of a single common factor, which may have disparate effects on the different individual time series of the panel. Due to their extensive use in the empirical literature, and due to their increased availability through various econometric software packages, we propose two new tests based on the pooled Dickey-Fuller type tests developed by Levin *et al.* (2002), and Harris and Tzavalis (1999). In order to defactor the data, we propose using a version of the procedure recently developed by Phillips and Sul (2003), which is based on estimating the factor loadings by iterated method of moments. In contrast to the principal components method used by Ng and Bai (2004), and Moon and Perron (2004), the consistency of this procedure only requires passing the number of time series observations to infinity, whereby lending itself to simple sequential limit asymptotics. The asymptotic results reveal that the tests reach a limiting normal distribution under the null hypothesis and that they are unbiased against the heterogeneous local alternative hypothesis. In our Monte Carlo study, we demonstrate that the tests have good size properties and reasonable power. We also find that the proposed tests compares favorably to a number of different tests that also presumes a common factor structure for the cross-sectional dependence.

The paper proceeds as follows. Section 2 provides a brief presentation of the model that we use. Section 3 introduces the unit root test statistics, whereas Section 4 is concerned with their asymptotic properties. Sections 5 then present our Monte Carlo study. Section 6 concludes the paper. For notational convenience, the Brownian motion  $B(r)$  defined on the unit interval  $0 \leq r \leq 1$  will be written as only  $B$  and integrals such as  $\int_0^1 W(r)dr$  will be written  $\int_0^1 W$  and  $\int_0^1 W(r)dW(r)$  as  $\int_0^1 WdW$ . The symbols  $\Rightarrow$  and  $\xrightarrow{P}$  will be used to signify weak convergence and convergence in probability, respectively.

## 2 Model and assumptions

Let  $y_{it}$  be a vector of observed data on individual  $i = 1, \dots, N$  and time series  $t = 1, \dots, T$ . The data is assumed to be generated by

$$Y_t = \beta X_t + \rho Y_{t-1} + E_t, \quad (1)$$

where  $Y_t = (y_{1t}, \dots, y_{Nt})'$ ,  $E_t = (e_{1t}, \dots, e_{Nt})'$  and  $\rho = \text{diag}(\rho_1, \dots, \rho_N)'$ . For notational simplicity, we use  $X_t$  to indicate the vector of deterministic components and  $\beta$  is used to indicate the corresponding vector of parameters. We shall distinguish between three different deterministic specifications. In Model 1,  $X_t = \{\emptyset\}$ , which correspond to a model with no deterministic components. In Model 2,  $X_t = 1$  so the deterministic component comprises an individual

specific constant term. In Model 3,  $X_t = (1, t)$ , which is the most general specification with both individual specific constants and linear time trends. For the present,  $Y_t$  is assumed to evolve according to Model 1. Generalizations to Models 2 and 3 are straightforward and will be discussed when appropriate.

The focus of interest in this paper is the problem of testing for the presence of a common unit root in the error process  $Y_t$  against the following local alternative

$$\rho_i = I_N - \frac{\theta}{TN^{1/2}}. \quad (2)$$

In this paper, we model the individual local-to-unity parameters  $\theta_i$  comprised of  $\theta = \text{diag}(\theta_1, \dots, \theta_N)'$  as a sequence of non-negative i.i.d. random variables. The hypothesis tested may be stated as  $H_0 : \theta_i = 0$  for all individuals  $i$  versus  $H_1 : \theta_i > 0$  for some  $i$ . A special case of interest for the alternative hypothesis is when  $\theta_i = \theta > 0$  for all  $i$ . In this case, the local-to-unity parameters take on a common value  $\theta > 0$  for all  $i$  and each of the individual error processes  $Y_t$  is therefore locally stationary. Now, consider the following assumption regarding the local-to-unity parameters  $\theta_i$ .

**Assumption 1.** (Local-to-unity parametrization.) The random variable  $\theta_i$  is i.i.d. with expected value  $\mu_\theta$  and support on  $[0, M]$ .

Under Assumption 1, the null hypothesis of no cointegration is equivalent to the statement that  $\mu_\theta > 0$ . Thus, we may express the null and alternative hypotheses in the following equivalent fashion

$$H_0 : \mu_\theta = 0 \text{ versus } H_1 : \mu_\theta > 0.$$

To model the cross-sectional correlation, we follow Phillips and Sul (2003), and assume that the error process  $v_t$  is generated by the following single factor model

$$E_t = \lambda F_t + v_t, \quad (3)$$

where  $F_t$  is a scalar unobservable random factor,  $\lambda = (\lambda_1, \dots, \lambda_N)'$  is a nonrandom vector of factor loading parameters and  $v_t = (v_{1t}, \dots, v_{Nt})'$  is a vector of idiosyncratic disturbances. For convenience in deriving the asymptotic theory, we assume that the vector  $v_t$  follow a general linear process whose parameters satisfy the summability conditions of the following assumption.

**Assumption 2.** (Error process.) The vector  $v_t$  is i.i.d. cross-sectionally such that  $v_t = C(L)u_t$ , where  $L$  is the lag operator,  $C(L) = \sum_{j=0}^{\infty} C_j L^j$ ,  $C_j = \text{diag}(C_{1j}, \dots, C_{Nj})$ ,  $C(1) \neq 0$ ,  $\sum_{j=0}^{\infty} j^2 C_j C_j < \infty$  and  $u_t$  is a vector white noise sequence. The long-run covariance matrix of  $v_t$  is given by  $\Sigma \equiv C(1)C(1)$ .

The factor model in (3) is introduced to model the cross-sectional dependence between  $e_{it}$  and  $e_{ij}$ . To this end, we make the following assumption.

**Assumption 3.** (Factor model.) (i) The common factor  $F_t$  is such that  $F_t = D(L)z_t$ , where  $D(L) = \sum_{j=0}^{\infty} D_j L^j$ ,  $D(1) \neq 0$ ,  $\sum_{j=0}^{\infty} j^2 D_j^2 < \infty$  and  $z_t$  is a white noise sequence. (ii) We normalize  $\sigma_F^2 \equiv D(1)^2 = 1$ . (iii) The processes  $u_t$ ,  $z_t$  and  $\lambda$  are mutually independent.

Assumption 2 ensures that a functional central limit theorem holds individually for each cross-section as  $T$  increases. Specifically, Assumption 2 implies the following weak convergence result  $T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} v_t \Rightarrow B_v \equiv \Sigma^{1/2} W$  as  $T \rightarrow \infty$ , where  $B_v = (B_{1v}, \dots, B_{Nv})'$  is a vector Brownian motion that is conformable with  $v_t$ . The process  $W = (W_1, \dots, W_N)'$  is referred to as a vector standard Brownian motion. Because the covariance matrix of  $W$  equals identity by definition, this implies that  $W_i$  and  $W_j$  are mutually independent if  $i \neq j$ . The assumed independence of the individual cross-sections is tantamount to requiring that the parameter vector  $C(L)$  is block-diagonal, which implies that  $\Sigma$  too is block-diagonal. Also, because  $C(1) \neq 0$ , each element along the main diagonal of  $\Sigma$  is nonzero. Similarly, Assumption 3 (i) ensures that  $T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} F_t \Rightarrow B_F$  as  $T \rightarrow \infty$ , where  $B_F$  is a scalar Brownian motion. This assumption implies that  $F_t$  is stationary and that the only source of nonstationarity in (1) is  $Y_t$  itself. As required by Assumption 3 (ii), the covariance of  $B_F$  is normalized to unity, which entails no loss of generality. Assumption 3 (iii) requires  $u_t$ ,  $z_t$  and  $\lambda$  to be mutually independent.

The linear process assumption on  $v_t$  and  $F_t$  facilitates a straightforward analysis by application of the methods developed by Phillips and Solo (1992). Similar results could be obtained under the strong mixing conditions of Phillips and Perron (1988), which also ensure that a functional central limit theorem holds for each cross-section as  $T$  grows large. Notice that the asymptotic analysis of linear processes holds under a variety of conditions, and that the limiting result of our tests therefore can be generalized to different classes of time series innovations. In particular, with a strengthening of Assumption 2 and 3, our result can be generalized to panels with i.i.d. disturbances.

Most unit root tests in panel data requires that the process  $Y_t$  is independent cross-sectionally (see, e.g., Im *et al.*, 2003; Levin *et al.*, 2002; Harris and Tzavalis, 1999). This assumption is very strong and it is therefore unlikely to hold in many empirical applications. When the structure of the dependence is completely unknown, it is generally infeasible to deal with the unrestricted correlation structure because of the degree of freedom constraint. Therefore, in order to reduce the dimensionality of the covariance structure of the errors, it is common to make at least some simplifying assumption. The most common way to achieve this is to include a common time effect in the process driving  $Z_t$ . The justification for doing so is that certain co-movements of multidimensional time series may be due to a common factor. For instance, in international macroeconomics it might be argued that the common time effect represents some global shock, such as an oil price shock.

The model we use in this paper allows for a common time effect that may

affect the individual time series differently. The extent of the cross-sectional correlation is determined by the loading parameters  $\lambda_i$ , which are such that

$$E(e_{it}e_{jt}) = \lambda_i\lambda_j \text{ for } i \neq j.$$

It follows that there is no cross-sectional correlation when  $\lambda_i = 0$  for all  $i$  and there is identical correlation when  $\lambda_i = \lambda_j$  for all  $i \neq j$ . The model we use is the same single factor model studied by Phillips and Sul (2003), which considers the problem of dynamic panel data estimation and homogeneity testing when the data is correlated cross-sectionally. Note that this model is inherently distinct from the simple factor model in which the dependence is such that it may be removed by simply subtracting the cross-sectional average from each observation. This transformation of the data is equivalent to including a full set of time specific constants in (1), which is appropriate under the assumption of a single common factor that has an identical impact on all the individuals of the panel. Our model also presumes a single common factor but it is less restrictive since it allows the factor to impact the individuals of the panel differently.

### 3 Pooled unit root tests

In this section, we develop the pooled unit root tests. In doing so, we will make frequent use of the results given in Phillips and Sul (2003). As pointed out in the previous section, Assumption 1 and 2 are relatively weak and allow for quite general forms of error dynamics. In order to facilitate the construction of tests with simple enough structure, however, in this section we shall initially make some simplifying assumptions, which will subsequently be disregarded. Specifically, we strengthen Assumption 2 and 3 to the following set of conditions.

**Assumption 4.** The processes  $v_t$ ,  $F_t$  and  $\lambda$  are mutually uncorrelated white noise sequences.

Essentially, Assumption 4 allows us to focus on the cross-sectional properties of the data without simultaneously having to deal with any temporal dependencies. Specifically, it will allow us to write the covariance of  $E_t$  as

$$V \equiv E(E_t E_t') = \lambda\lambda' + \Sigma. \quad (4)$$

Notice that  $\Sigma = E(v_t v_t')$  under Assumption 4. Making use of (4), under the unit root hypothesis, it is possible to show that in Model 1 with no deterministic components,  $T^{-1/2}Y_{[Tr]} = T^{-1/2} \sum_{t=2}^{[Tr]} E_t \Rightarrow B \equiv LW$  as  $T \rightarrow \infty$ , where  $B = (B_1, \dots, B_N)'$  is an  $N$  dimensional vector Brownian motion and  $V = L'L$ . This implies that  $Y_t$  is unsuitable for unit root testing purposes since its limit distribution depend on the nuisance parameters associated with the cross-sectional dependencies of the data as captured by the off-diagonal elements in  $V$ . Note, however, that  $B$  may be decomposed as

$$B = \lambda B_F + B_v. \quad (5)$$

If we let  $\lambda^*$  be the  $N \times (N - 1)$  matrix that spans the orthogonal complement of the vector  $\lambda$ , then we may define

$$F_\lambda \equiv (\lambda^{*\prime} \Sigma \lambda^*)^{-1/2} \lambda^{*\prime}. \quad (6)$$

Making use of (5) and (6), since  $\lambda^{*\prime} \lambda = 0$  by definition, we may deduce that  $T^{-1/2} F_\lambda Y_{[Tr]} \Rightarrow F_\lambda B = F_\lambda B_v \equiv W^*$  as  $T \rightarrow \infty$ , where  $W^* = (W_1^*, \dots, W_{N-1}^*)'$  is a  $N - 1$  dimensional vector standard Brownian motion. Since the covariance matrix of  $B_v$  is given by  $\Sigma$ , it is easy to verify that the covariance of  $F_\lambda B_v$  is indeed the identity matrix. Note also that the pre-multiplication of  $F_\lambda$  reduces the dimension of  $W^*$  from  $N$  to  $N - 1$ .

This discussion suggests that  $Y_t$  may be employed to test the unit root hypothesis after pre-multiplying it by  $F_\lambda$ . Obviously, since  $\lambda^*$  and  $\Sigma$  generally are unknown, this means that  $F_\lambda$  is unobservable and that it needs to be consistently estimated. Denote this estimator by  $\hat{F}_\lambda$ . Making use of  $\hat{F}_\lambda$ , we shall consider the following two panel statistics that may be used to test the null hypothesis of a unit root.

**Definition 1.** (Pooled defactored panel unit root statistics.) Let  $\bar{Y}_t = Y_t - \hat{\beta} X_t$  and  $\bar{Y}_t^* = \hat{F}_\lambda \bar{Y}_t$ , where  $\hat{\beta}$  is the least squares estimate of  $\beta$ . The defactored panel unit root statistics are defined as follows

$$Z_\rho \equiv \left( \sum_{t=2}^T \bar{Y}_{t-1}^* \bar{Y}_{t-1}^* \right)^{-1} \sum_{t=2}^T \Delta \bar{Y}_t^* \bar{Y}_{t-1}^*, \quad (7)$$

$$Z_t \equiv \left( \sum_{t=2}^T \bar{Y}_{t-1}^* \bar{Y}_{t-1}^* \right)^{-1/2} \sum_{t=2}^T \Delta \bar{Y}_t^* \bar{Y}_{t-1}^*. \quad (8)$$

The above statistics are nothing but simple modifications of the pooled normalized bias and  $t$ -ratio statistics studied earlier in the literature by Harris and Tzavalis (1999), and Levin *et al.* (2002) using the defactored data. Note the particularly simple form of the  $Z_t$  statistic as it does not require any estimation of the error variance in the denominator, which is standard in the earlier literature. This is a direct consequence of the defactoring procedure that we use, which makes the asymptotic covariance matrix of  $T^{-1/2} F_\lambda Y_{[Tr]}$  equal to the identity matrix.

For the estimation of the transformation matrix  $F_\lambda$ , we propose using a version of the moment based procedure discussed in Phillips and Sul (2003). The purpose of this procedure is to retrieve  $\hat{\lambda}$ , the vector of estimated loading parameters, together with  $\hat{\Sigma}$ , the estimated covariance of  $v_t$ , as the minimizers of the sum of squared errors of  $\lambda$  and  $\Sigma$  from  $\hat{V} = T^{-1} \sum_{t=2}^T \hat{E}_t \hat{E}_t'$ , where  $\hat{E}_t = \bar{Y}_t - \hat{\beta} X_t - \hat{\rho} \bar{Y}_{t-1}$  and  $\hat{\rho}$  is the least squares estimator of  $\rho$ . This is different from the procedure employed by Phillips and Sul (2003), which uses  $\tilde{V} = T^{-1} \sum_{t=2}^T \Delta E_t \Delta E_t'$  to estimate  $V$ . This estimator is, however, only consistent as  $T \rightarrow \infty$  under the null hypothesis suggesting that there should be



some merit in using  $\hat{V}$ , which is consistent under the null as well as under the alternative hypothesis. Indeed, since  $\hat{V}$  is consistent for  $V$  under both the null and alternative, this indicates that the minimization of the sum of squared errors may be carried out with respect to only  $\lambda$  and  $\Sigma$ . The implication is that  $\hat{\lambda}$  and  $\hat{\Sigma}$  may be obtained as

$$(\hat{\lambda}, \hat{\Sigma}) = \arg \min_{\lambda, \Sigma} \text{tr} \left( (\hat{V} - \Sigma - \lambda\lambda')(\hat{V} - \Sigma - \lambda\lambda')' \right). \quad (9)$$

The solution of this problem satisfies the two equations  $\hat{\lambda} = (\hat{\lambda}'\hat{\lambda})^{-1}(\hat{V}\hat{\lambda} - \hat{\Sigma}\hat{\lambda})$  and  $\hat{\Sigma} = \text{diag}(\hat{V} - \hat{\lambda}\hat{\lambda}')$ . Obviously, since there exist no closed form solutions for  $\hat{\lambda}$  and  $\hat{\Sigma}$ , the minimization of (9) cannot be done directly but needs to be carried out iteratively. To this end, we engage in an recursive procedure using the following updating scheme

$$\begin{aligned} \lambda^r &= (\lambda^{r-1'}\lambda^{r-1})^{-1} (\hat{V}\lambda^{r-1} - \Sigma^{r-1}\lambda^{r-1}), \\ \Sigma^r &= \text{diag}(\hat{V} - \lambda^r\lambda^{r'}) . \end{aligned}$$

The procedure is initiated by choosing a vector  $\lambda^0$  of starting values for  $\lambda$ , which is used to compute  $\Sigma^0 = \hat{V} - \lambda^0\lambda^{0'}$ . The updating then continues until convergence or until the number of iterations reaches some predetermined upper boundary. Once  $\hat{\lambda}$  and  $\hat{\Sigma}$  has been obtained, a consistent estimator of  $\lambda^*$  may be constructed by taking the eigenvectors of the projection matrix  $Q_\lambda = I_N - \hat{\lambda}(\hat{\lambda}'\hat{\lambda})^{-1}\hat{\lambda}'$  that correspond to unit eigenvalues. If we denote the resulting estimator by  $\hat{\lambda}^*$ , then  $\hat{F}_\lambda$  may be constructed as  $\hat{F}_\lambda = (\hat{\lambda}^{*'}\hat{\Sigma}\hat{\lambda}^*)^{-1/2}\hat{\lambda}^{*}$ . Using  $\hat{F}_\lambda$ , we may then transform the data as in (7) and (8) giving  $\bar{Y}_t^* = \hat{F}_\lambda Y_t$ . Since this transformation asymptotically removes the cross-sectional dependence of the data,  $\bar{Y}_t^*$  is asymptotically independent cross-sectionally as  $T \rightarrow \infty$ . Hence, the transformed data may be used to test the unit root hypothesis.

As will be shown in Section 4, the limiting distributions of the above statistics are free of nuisance parameters associated with the underlying data generating process (DGP). Once we allow for the possibility of nonzero constants and time trends in (1), however, the distributions of the statistics will no longer be invariant with respect to these nuisance parameters. Therefore, in order to obtain statistics that are asymptotically similar in Model 2, the data should be demeaned prior to using the formulas in (7) and (8). For Model 3, the data should be both demeaned and detrended to account for the linear trend appearing in (1). Thus, as in the case of a single time series, if a deterministic element is present but not accounted for when constructing the test statistics, the ensuing unit root test will be inconsistent. Therefore, in order to obtain tests that are asymptotically similar, we use  $\bar{Y}_t$  and not  $Y_t$  when constructing the statistics.

Similarly, when Assumption 4 is relaxed, the statistics in (7) and (8) are no longer asymptotically similar and needs to be modified to account for the

temporal dependence in the DGP. Under Assumptions 2 and 3, this may be accomplished by simply augmenting the right-hand side of (1) with lagged values of  $\Delta Y_t$ . In so doing, it is necessary that the lag order  $K$ , say, is chosen sufficiently large to whiten the errors. This suggests that in order to obtain similar test statistics, we should replace  $\bar{Y}_t$  in (7) and (8) with the projection errors of  $\bar{Y}_t$  from  $K$  lags of  $\Delta Y_t$ . That is,  $\bar{Y}_t = Y_t - \hat{\beta}X_t$  should be replaced with  $\bar{Y}_t = Y_t - \hat{\beta}X_t - \sum_{k=1}^K \hat{\varphi}_k \Delta Y_{t-k}$ . The defactored data  $\bar{Y}_t^*$  may then be computed as before by pre-multiplying the projection errors  $\bar{Y}_t$  by  $\hat{F}_\lambda$ .

## 4 Asymptotic distribution

In this section, we characterized the asymptotic distribution of the proposed unit root test statistics. In so doing, we shall exploit the fact that the orthogonalization procedure applied here is consistent passing  $T \rightarrow \infty$  for a fixed  $N$ . This means that the limiting distribution of the tests may be derived in a relatively straightforward fashion using the simple sequential limit theory developed by Phillips and Moon (1999). It will be shown that both statistics require standardization based on the first two moments of the following vector Brownian motion functional

$$K_i = (K_{i1}, K_{i2})' \equiv \left( \int_0^1 \bar{W}_i^{*2}, \int_0^1 \bar{W}_i^* dW_i^* \right)', \quad (10)$$

where

$$\bar{W}_i^* = W_i^* - \left( \int_0^1 W_i^* X' \right) \left( \int_0^1 X X' \right)^{-1} X.$$

The vector functional  $K_i$  takes the scalar Brownian motion  $\bar{W}_i^*$  as its only argument. This scalar is the Hilbert projection of  $W_i^*$  onto the space orthogonal to the vector  $X$ , which is the limiting trend function. Specifically, let  $D_T = \text{diag}(1, T)$  denote a matrix of normalizing orders that is conformable with  $X_t = (1, t)'$ , then  $D_T^{-1} X_{[Tr]} \Rightarrow X = (1, r)'$  as  $T \rightarrow \infty$ . In Model 1,  $X_t = \{\emptyset\}$  so  $\bar{W}_i^*$  reduces to  $W_i^*$ . In Model 2,  $X_t = 1$  in which case  $\bar{W}_i^*$  represents the demeaned standard Brownian motion  $W_i^* - \int_0^1 W_i^*$ . Similarly, in Model 3,  $X_t = (1, t)'$  so  $\bar{W}_i^*$  is the demeaned and detrended standard Brownian motion  $W_i^* + (6r - 4) \int_0^1 W_i^* + (6 - 12r) \int_0^1 r W_i^*$ . All moments of  $K_i$  exist. In particular, we shall posit  $\Theta = (\Theta_1, \Theta_2)'$  and  $\Sigma$  to be, respectively, the mean and the variance of  $K_i$ . It will also be useful to define  $\phi \equiv (-\Theta_2 \Theta_1^{-2}, \Theta_1^{-1})'$  and  $\varphi \equiv (-2^{-1} \Theta_2 \Theta_1^{-3/2}, \Theta_1^{-1/2})'$ . With these definitions in hand, we are now ready to state our first main result.

**Theorem 1.** (*Asymptotic distribution.*) Under Assumption 1 through 3, as  $T \rightarrow \infty$  followed by  $N \rightarrow \infty$

$$TN^{1/2} Z_\rho - N^{1/2} \Theta_1 \Theta_2^{-1} \Rightarrow N(-\mu_\theta, \phi' \Sigma \phi), \quad (11)$$

$$Z_t - N^{1/2}\Theta_1\Theta_2^{-1/2} \Rightarrow N(-\mu_\theta\Theta_1^{1/2}, \varphi'\Sigma\varphi). \quad (12)$$

The proof of Theorem 1 is outlined in the appendix. It proceeds by showing that the intermediate limiting distributions passing  $T \rightarrow \infty$  of the unit root statistics can be written entirely in terms of the elements of the vector Brownian motion functional  $K_i$ . Therefore, by virtue of cross-sectional independence of the defactored data, the limiting distribution of the test statistic can be described in terms of differentiable functions of i.i.d. random variables to which the Delta method is applicable. Hence, by subsequently passing  $N \rightarrow \infty$ , we obtain a limiting normal distribution for the test statistic, which depend only on the first two moments of  $K_i$ .

Notice the difference between this result and the asymptotic theory for univariate time series, which typically involves moments of diffusion processes rather than standard Brownian motions. In particular, Theorem 1 shows that the statistics are asymptotically normal under both the null and alternative hypotheses. Under the null,  $\mu_\theta = 0$  in which case Theorem 1 shows that the distributions are mean zero and only depend on the moments of the Brownian motion functional  $K_i$ . When  $\mu_\theta > 0$ , however, although they are still asymptotically normal and similar with respect to the variance of the error process, the distributions depend on the nuisance parameter  $\mu_\theta$ . Theorem 2 shows that this dependency causes a miscentering of the limiting distributions of the test statistics. The expected value of these distributions are negative, which imply that they will tend to shift leftwards as we move away from the null hypothesis. This means that the tests are unbiased and that their asymptotic local powers therefore are greater than their size. The drift of the distributions depend on the average of the deviations  $\theta_i$ . Thus, the statistics will tend to diverge towards negative infinity as  $\theta_i$  grows arbitrarily large for at least some  $i$ . For a given value  $\theta > 0$ , optimal power is obtained when  $\theta_i = \theta$  for all  $i$ , which is not unexpected given our pooling approach.

Under a fixed specification of the alternative hypothesis, the value taken by the autoregressive parameters do not depend on  $N$  or  $T$ . In this case, the probability that the statistics take on a more negative value than the critical value provided by the standard normal distribution approaches one asymptotically as  $T \rightarrow \infty$  followed by  $N \rightarrow \infty$ . Apparently, in spite of the fact that the statistics are pooled, they may be used to construct consistent tests against the heterogeneous type of alternative considered here.

Theorem 1 indicate that each of the standardized statistics converges to a normal distribution whose moments depend on the underlying vector Brownian motion functional  $K_i$ . As we have seen, since it is a relatively straightforward matter to adjust the formulae in (7) and (8) to account for the effects of weakly dependent disturbances, the results of Theorem 1 are quite general and they apply regardless of the deterministic specification of (1). If the test statistics are based on (1) with no deterministic terms, then  $X$  is the empty set and the results of Theorem 1 apply directly to the standard Brownian motion  $W_i^*$ . If

(1) is fitted with a constant, then the limiting distributions of  $Z_\rho$  and  $Z_t$  still have the same form as in (11) and (12) but now the moments are based on the demeaned standard Brownian motion  $W_i^* - \int_0^1 W_i^*$ . Analogously, if (1) involves fitted constant and trend terms, then the limiting distributions in (11) and (12) retain their stated forms but involve moments of the demeaned and detrended standard Brownian motion  $W_i^* + (6r - 4) \int_0^1 W_i^* + (6 - 12r) \int_0^1 rW_i^*$ .

The appropriate moments needed for computing the standardized test statistics are presented in Table 1. They have been obtained by direct calculation using the properties of Brownian motion. The implication of this is that we can use the results of Theorem 1 to compute standardized test statistics, which only depend on the estimated values of  $Z_\rho$  and  $Z_t$ , and their moments. Since these moments are available from Table 1, the standardized statistics can be readily computed and used for statistical inference on the unit root hypothesis. As suggested by Theorem 1, it is possible to construct the tests as one-sided using only the left tail of the normal distribution to reject the null hypothesis.

## 5 Monte Carlo simulations

In this section, we compare and evaluate the small-sample properties of the proposed test statistics relative to that of some competing tests that build on the work of Levin *et al.* (2002), and Harris and Tzavalis (1999), Bai and Ng (2004), and Moon and Perron (2004). For this purpose, a small set of Monte Carlo experiment were conducted using the following DGP

$$y_{it} = \alpha_i + \beta_i t + \rho_i y_{it-1} + z_{it}, \quad (13)$$

$$z_{it} = \tau \lambda_i F_t + v_{it}, \quad (14)$$

where  $y_{i0} = 0$  and  $(\lambda_i, F_t, v_{it})' \sim N(0, I_3)$ . For each experiment, we generate 1,000 panels with  $N \in \{10, 20\}$  individual and  $T \in \{100, 200\}$  time series observations. The DGP is parameterized as follows. For the deterministic component, we have three different configurations, each of which correspond to one of our three model specifications. Specifically,  $\alpha_i = \beta_i = 0$  in Model 1,  $\alpha_i = 0$  and  $\beta_i \sim N(0, 1)$  in Model 2, and  $(\alpha_i, \beta_i)' \sim N(0, I_2)$  in Model 3. Each model is taken as a separate experiment.

For each model, we consider a small set of separate experiments, which corresponds to different parameterizations of  $\rho_i$  and  $\tau$ . For the autoregressive parameter  $\rho_i$ , we have two cases. Under the null hypothesis, we have  $\rho_i = 1$  for all  $i$ . Under the alternative hypothesis, we shall consider several different autoregressive parameterizations. In particular, to be able to infer the increased power that derives from increasing the size of the panel, we shall consider a fixed alternative hypothesis, which is independent of  $N$  and  $T$ .

The parameter  $\tau \in \{0, 1, 2\}$  controls the relative importance of the common and idiosyncratic disturbances. A larger value of  $\tau$  represents a greater weight

being attached to the common disturbances relative to the idiosyncratic ones. When  $\tau = 0$ , then there is no common factor and the individuals of the panel are therefore independent. Conversely, when  $\tau \neq 0$ , then there is a common factor present, which induces the cross-sectional correlation among the individuals. Specifically, while  $\tau = 1$  correspond to a situation in which the common and idiosyncratic disturbances have equal weight,  $\tau = 2$  represent a situation when the common disturbances are twice as important as those emanating from the idiosyncratic error term.

For the estimation of the loading parameters, we use the iterated method of moments procedure laid out in Section 3. Towards this end, we set the maximum number of iterations to 50 and the convergence criterion for the loading parameters is set to 0.0001. Moreover, to be able to evaluate the comparative merit of the proposed tests, we compute six alternative tests. The first two are constructed under the assumption of no cross-sectional dependence. They are the pooled Dickey-Fuller type  $t$ -ratio and normalized bias statistics proposed by Levin *et al.* (2002), and Harris and Tzavalis (1999). The  $t$ -ratio and normalized bias statistics will henceforth be denoted by  $DF_t$  and  $DF_\rho$ , respectively. The second pair of statistics, abbreviated  $MP_t$  and  $MP_\rho$ , correspond to employing the  $DF_t$  and  $DF_\rho$  statistics to the defactored data when the factor has been removed using the method suggested by Moon and Perron (2004). Essentially, what Moon and Perron (2004) proposes is to first estimate the factor loadings using principal components and then to defactor the data using the projection matrix spanning the orthogonal space of the loadings, which is similar to the procedure used here.

The third pair of statistics is also based on using  $DF_t$  and  $DF_\rho$  on defactored data but now with the factors being removed using the method of Bai and Ng (2004). The idea put forth by Bai and Ng (2004) is to eliminate the cross-sectional dependence under the null by projecting the data onto the non-stationary factor  $f_t = \sum_{j=1}^t F_j$ . In doing so, one needs to obtain consistent estimates of  $f_t$ . Bai and Ng (2004) show that this can be accomplished by first estimating  $F_t$  by principal components on differentiated data and then to construct  $\hat{f}_t = \sum_{j=1}^t \hat{F}_j$ . To have an intuition on this, consider the DGP described by (13) and (14) under the null when  $\tau = 1$  and  $\alpha_i = \beta_i = 0$  as in Model 1. In this case, the nonstationary data may be written as  $y_{it} = \lambda f_t + \sum_{j=1}^t v_{ij}$ , which suggests that  $\Delta y_{it} = \lambda F_t + v_{it}$  is stationary and that estimation by principal components is feasible. Thus, since  $\hat{f}_t$  is consistent for  $f_t$ ,  $\hat{E}_{it} = y_{it} - \lambda \hat{f}_t$  will be cross-sectionally independent and therefore suitable for testing the unit root hypothesis. The resulting  $t$ -ratio and normalized bias statistics will be denoted by  $BN_t$  and  $BN_\rho$ , respectively.

All statistics except  $BN_t$  and  $BN_\rho$  have the same critical values as the proposed statistics. The asymptotic distribution of the  $BN_t$  and  $BN_\rho$  statistics are not available for the model with a linear time trend but may be derived using the results of Theorem 3 in Bai and Ng (2004). The mean and the variance of the

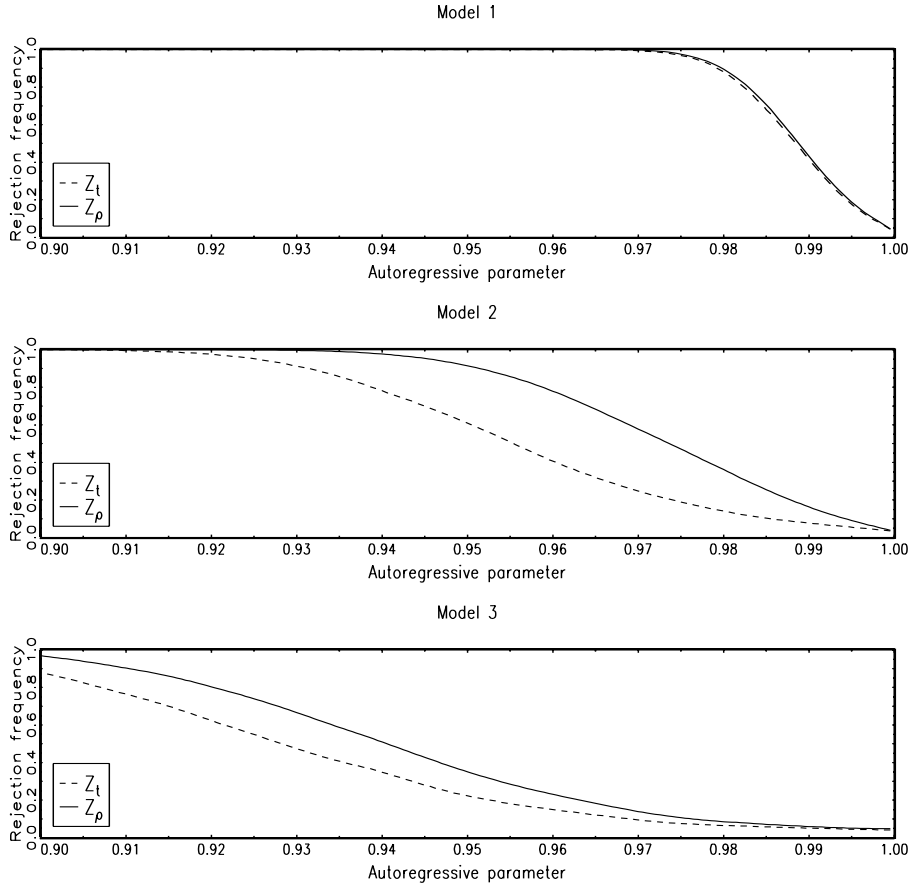
statistics are then obtained by means of Monte Carlo simulation. To this effect, we make 100,000 replications of a single random walk of length  $T = 10,000$  with standard normal innovations. The simulated mean and variance are  $-1.2255$  and  $0.3017$  for the  $BN_t$  statistic and  $-3.0039$  and  $7.2496$  for the  $BN_\rho$  statistic. All tests are carried out on the five percent level and all powers are adjusted for size so that each test has the same rejection frequency of five percent when the null hypothesis is true. All computational work is performed in GAUSS.

Consider first the empirical size of the tests presented in Table 2. All tests should have a rejection rate of five percent when  $\tau = 0$ . In agreement with this, Table 2 illustrates that the empirical size of the  $t$ -ratio tests generally lies very close to the nominal level in most panels. The corresponding results for the normalized bias statistics suggest that there is an overall tendency for these tests to over-reject the null hypothesis when it is true. When  $\tau > 0$ , we expect the  $DF_t$  and  $DF_\rho$  tests to suffer from size distortions as they do not account for the presence of cross-sectional correlation. This is confirmed by Table 1, which indicates that the size may be seriously distorted when the errors admit a common factor structure. In particular, the table suggests that the size distortions are substantial and that they tend to get larger and becomes very serious as  $N$  increases. As expected, the distortions increase as the degree of cross-sectional correlation increases. The results for the other tests are much more encouraging with only small size distortions in most cases. The results for the proposed tests are particularly good. In fact, size accuracy appear to be almost perfect in most panels and it generally improves as both  $N$  and  $T$  increases.

Next, we continue to the results on the size-adjusted power of the tests presented in Tables 3 through 5. In this case, the results suggest that the  $DF_t$  and  $DF_\rho$  tests enjoy augmented power when  $\tau = 0$  and there is no cross-sectional correlation. As expected, the power is higher the larger is the deviation from the null hypothesis. When  $\rho_i \sim U(0.9, 1)$ , we see that the power is generally very good and that it approaches one as  $N$  and  $T$  increases. Interestingly, the power of the tests in Model 1 with no deterministic components appears to be higher than that for Model 2 with a fitted intercept, which, in turn, appears to be higher than that for Model 3 with fitted intercept and trend terms. Of course, this is not unexpected given the well known incidental trends problem (see, e.g., Moon and Perron, 2004). Furthermore, as predicted by asymptotic theory, we see that the power of the tests generally falls as the support of  $\rho_i$  gets narrower.

When  $\tau > 0$  and the data is correlated cross-sectionally, the power of the  $DF_t$  and  $DF_\rho$  tests fall relative to the power of the other tests. In fact, the tables suggest that the power of  $DF_t$  and  $DF_\rho$  falls strictly below that of the other tests. In this case, the simulations suggest that the  $BN_t$  and  $BN_\rho$  tests generally perform best with the power of the proposed tests being only slightly lower. In addition, and in agreement with the simulation results presented by Moon and Perron (2004), we see that the power of all tests decline significantly

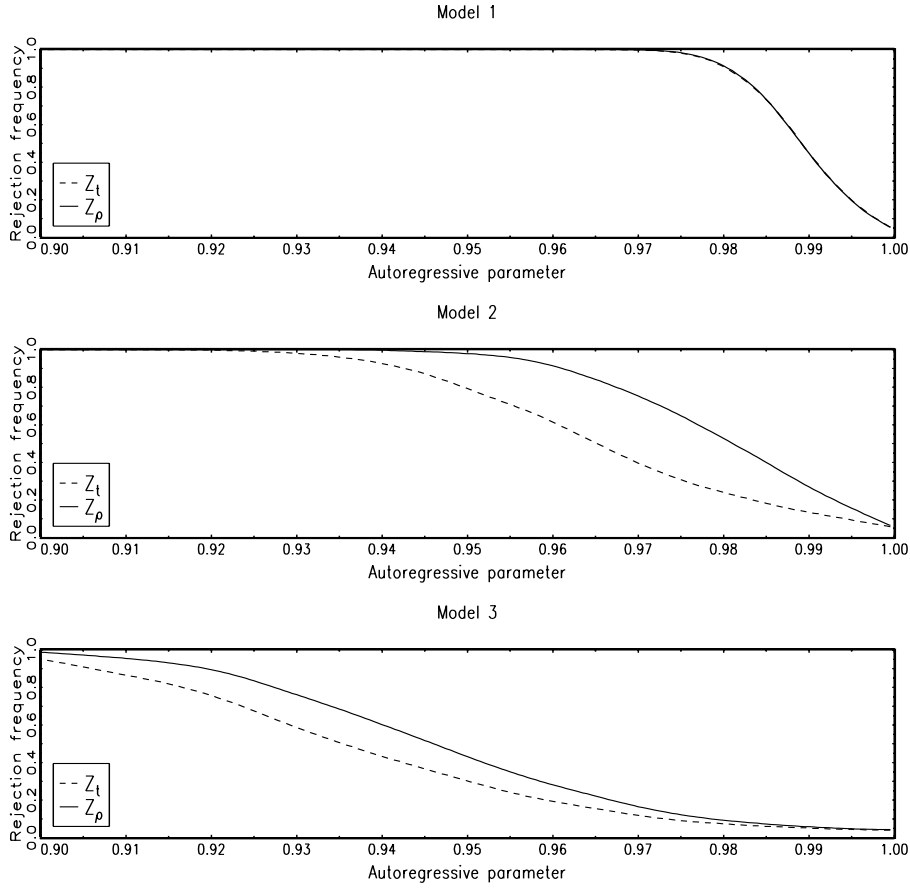
Figure 1: Local power when  $\tau = 0$ .



as the value of  $\tau$  increases. In particular, the results suggest that the power may be very poor, and practically nonexistent in some cases, when the tests are fitted with both an intercept and a linear time trend. The power usually improves, however, as the size of the panel increases, especially along the time series dimension.

Theorem 2 reveal that the tests have different drift terms under the local alternative hypothesis, which means that they have different local power functions. To study the local power of the tests, we make the simplifying assumption that the autoregressive parameters take on a common value  $\rho_i = \rho$  for all  $i$ . We then simulate size-adjusted powers for different autoregressive parameterizations as we move away from the null. The results for the leading case when  $T = 100$

Figure 2: Local power when  $\tau = 1$ .



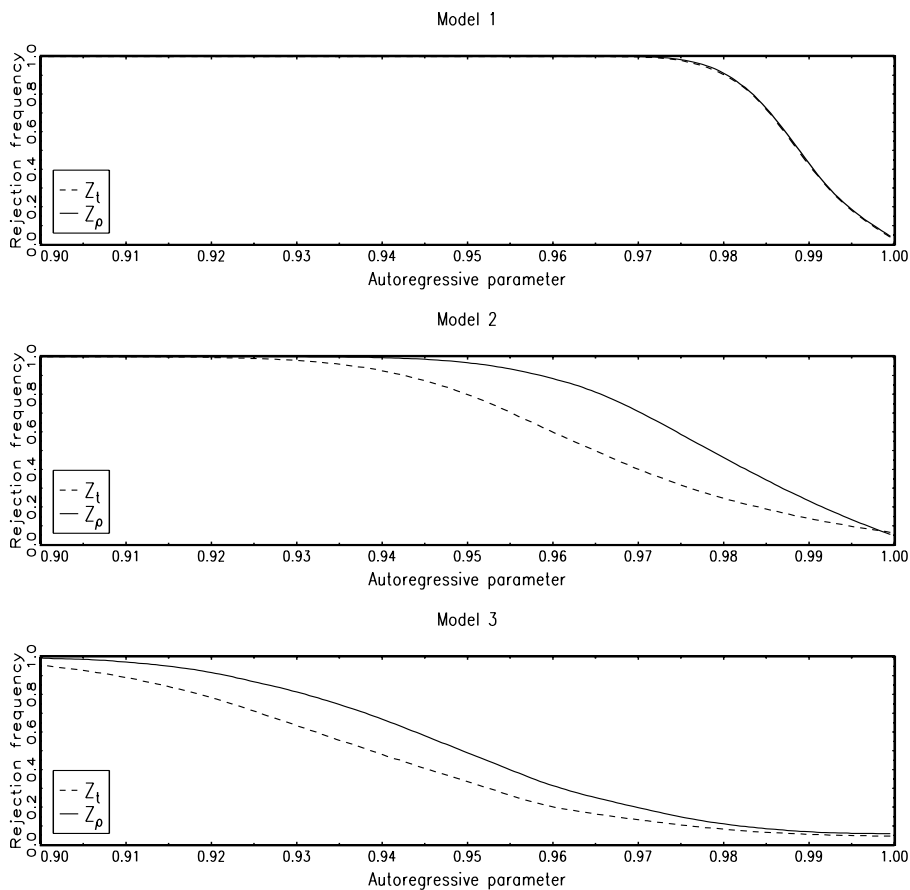
and  $N = 10$  are presented in Figures 1 through 3.<sup>1</sup> The figures suggest that the  $Z_\rho$  test is most powerful in the vicinity of the null. In addition to this, the figures suggest that the tests based on demeaned and detrended data are least powerful while those based on raw data are most powerful.

In summary, we find that the  $DF_t$  and  $DF_\rho$  tests generally suffer from serious size distortions in the presence of cross-sectional dependence. In contrast, we find that the proposed tests tend to display smaller size distortions than the other tests considered and, at the same time, maintain reasonable power in small samples.

<sup>1</sup>In Figures 1 through 3, the curves representing the local power of the test statistics have been smoothed by means of a least squares spline of neighboring points.



Figure 3: Local power when  $\tau = 2$ .



## 6 Conclusions

During the last few years there has been an immense proliferation of research concerned with the problem of testing for unit roots in panel data. Most of these studies assume that the individuals of the panel are independent. For many empirical applications, however, cross-sectional independence seem like a very restrictive assumption. Recognizing this shortcoming, this proposes two panel unit root tests that may be used when the cross-sectional dependence can be described by a single common factor that may exert disparate effects on the different individual time series of the panel. In order to defactor the data, we propose using a version of the procedure recently developed by Phillips and Sul (2003), which is based on estimating the factor loadings by iterated method of

moments. Sequential limit arguments reveal that the tests reach a limiting normal distribution under the null hypothesis and that they are unbiased against the heterogeneous local alternative hypothesis. In our Monte Carlo study, we demonstrate that the tests have good size properties and reasonable power. We also find that the proposed tests compares favorably to a number of different tests that also presumes a common factor structure for the cross-sectional dependence.

## Appendix A Mathematical proofs

In this appendix, we derive the limiting distributions of the pooled unit root test statistics. Unless otherwise stated, the limit arguments are taken passing  $T \rightarrow \infty$  with  $N$  held fixed.

**Proof of Theorem 1.** In order to derive the asymptotic distributions, we multiply the test statistics by  $T$  in which case (7) and (8) can be rewritten as

$$TZ_\rho = \left( T^{-2} \sum_{t=2}^T \bar{Y}_{it-1}^{*'} \bar{Y}_{it-1}^* \right)^{-1} T^{-1} \sum_{t=2}^T \bar{Y}_{it-1}^{*'} \Delta \bar{Y}_{it}^*, \quad (\text{A1})$$

$$Z_t = \left( T^{-2} \sum_{t=2}^T \bar{Y}_{it-1}^{*'} \bar{Y}_{it-1}^* \right)^{-1/2} T^{-1} \sum_{t=2}^T \bar{Y}_{it-1}^{*'} \Delta \bar{Y}_{it}^*. \quad (\text{A2})$$

Under Assumption 1 through 3, we have  $\rho_i = I_N - T^{-1}N^{-1/2}\theta$ , which imply that  $\Delta \bar{Y}_{it}^*$  in (A1) and (A2) can be written as

$$\begin{aligned} \Delta \bar{Y}_{it}^* &= (\rho - I_N) \bar{Y}_{it-1}^* + \bar{Z}_{it}^* \\ &= N^{-1/2} T^{-1} \theta \bar{Y}_{it-1}^* + \bar{Z}_{it}^*. \end{aligned} \quad (\text{A3})$$

Let us define  $Q_1 \equiv T^{-2} \sum_{t=2}^T \bar{Y}_{it-1}^{*'} \bar{Y}_{it-1}^*$ ,  $Q_2 \equiv T^{-1} \sum_{t=2}^T \bar{Y}_{it-1}^{*'} \bar{Z}_{it}^*$  and  $Q_3 \equiv T^{-2} \sum_{t=2}^T \bar{Y}_{it-1}^{*'} \theta \bar{Y}_{it-1}^*$ . By Theorem 4.4 of Hansen (1992), it is possible to show that  $T^{-1/2} \hat{F}_\lambda \bar{Y}_{[Tr]} \Rightarrow F_\lambda \bar{J} \equiv \bar{J}^*$  as  $T \rightarrow \infty$ , where  $\bar{J} = (\bar{J}_1, \dots, \bar{J}_{N-1})'$  is a  $N-1$  dimensional diffusion process such that

$$\bar{J}_i \equiv \int_0^r e^{\theta_i(r-s)N^{-1/2}} d\bar{W}_i = \bar{W}_i + N^{-1/2} \theta_i \int_0^r e^{\theta_i(r-s)N^{-1/2}} \bar{W}_i. \quad (\text{A4})$$

This implies that we have the following limits as  $T \rightarrow \infty$

$$Q_1 \Rightarrow \int_0^1 \bar{J}^{*'} \bar{J}^* = \sum_{i=1}^{N-1} \int_0^1 \bar{J}_i^{*2}, \quad (\text{A5})$$

$$Q_2 \Rightarrow \int_0^1 \bar{J}^{*'} dW^* = \sum_{i=1}^{N-1} \int_0^1 \bar{J}_i^* dW_i^*, \quad (\text{A6})$$

$$Q_3 \Rightarrow \int_0^1 \bar{J}^{*'} \theta \bar{J}^* = \sum_{i=1}^{N-1} \theta_i \int_0^1 \bar{J}_i^{*2}, \quad (\text{A7})$$

Now, since  $\bar{W}^*$  is i.i.d. over the cross-section, (A4) implies that  $\bar{J}^*$  must be so too. Thus, the limiting distributions of  $Q_1$ ,  $Q_2$  and  $Q_3$  passing  $T \rightarrow \infty$  is i.i.d. over the cross-section. Another implication of (A4) is that  $\bar{J}_i \Rightarrow \bar{W}_i$  as  $N \rightarrow \infty$  and that  $\bar{J}_i = \bar{W}_i$  if  $\theta_i = 0$  for a fixed  $N$ . Hence, the limiting process passing  $T \rightarrow \infty$  and then  $N \rightarrow \infty$  is the same under both

the null and the local alternative hypothesis. One implication of this is that  $\lim_{N \rightarrow \infty} N^{-1} E(Q_1, Q_2)' = \Theta = (\Theta_1, \Theta_2)'$ . Also, from Levin *et al.* (2002), the variance of  $K_i$  is given by  $\Sigma = \text{diag}(\Sigma_{11}, \Sigma_{22})$  for all  $i$ . To derive the limiting distributions of the test statistic, we shall make use of the Delta method, which provides the limiting distribution for continuously differentiable transformations of i.i.d. vector sequences. In so doing, we substitute  $\Delta \bar{Y}_{it}^*$  from (A3), which means that (A1) and (A2) can be rewritten as

$$TZ_\rho = Q_1^{-1} (Q_2 - N^{-1/2} Q_3), \quad (\text{A8})$$

$$Z_t = Q_1^{-1/2} (Q_2 - N^{-1/2} Q_3). \quad (\text{A9})$$

The statistics in (A8) and (A9) may be rewritten as

$$\begin{aligned} TN^{1/2} Z_\rho + \mu_\theta - N^{1/2} \Theta_2 \Theta_1^{-1} &= \mu_\theta + N^{1/2} (N^{-1} Q_2 - \Theta_2) (N^{-1} Q_1)^{-1} \\ &+ N^{1/2} \Theta_2 \left( (N^{-1} Q_1)^{-1} - \Theta_1^{-1} \right) \\ &- (N^{-1} Q_1)^{-1} N^{-1} Q_3, \end{aligned} \quad (\text{A10})$$

$$\begin{aligned} Z_t + \mu_\theta \Theta_1^{1/2} - N^{1/2} \Theta_2 \Theta_1^{-1/2} &= \mu_\theta \Theta_1^{1/2} + N^{1/2} (N^{-1} Q_2 - \Theta_2) (N^{-1} Q_1)^{-1/2} \\ &+ N^{1/2} \Theta_2 \left( (N^{-1} Q_1)^{-1/2} - \Theta_1^{-1/2} \right) \\ &- (N^{-1} Q_1)^{-1/2} N^{-1} Q_3. \end{aligned} \quad (\text{A11})$$

The terms appearing in (A10) and (A11) with normalizing order  $N^{-1}$  converge in probability to the means of the corresponding random variables by virtue of a law of large numbers as  $T \rightarrow \infty$  and then  $N \rightarrow \infty$ . Hence,  $N^{-1} Q_1 \xrightarrow{p} \Theta_1$  and  $N^{-1} Q_2 \xrightarrow{p} \Theta_2$ . In addition, by Corollary 1 of Phillips and Moon (1999),  $N^{-1} Q_3 \xrightarrow{p} \mu_\theta \Theta_1$  as  $T \rightarrow \infty$  prior to  $N$ . Moreover, by the Lindberg-Lévy central limit theorem,  $N^{1/2} (N^{-1} Q_2 - \Theta_2) \Rightarrow N(0, \Sigma_{22})$ . The remaining expressions involve continuously differentiable transformations of i.i.d. random variables. Thus, by the Delta method, as  $T \rightarrow \infty$  prior to  $N$

$$N^{1/2} \left( (N^{-1} Q_1)^{-1} - \Theta_1^{-1} \right) \Rightarrow N(0, \Theta_1^{-4} \Sigma_{11}), \quad (\text{A12})$$

$$N^{1/2} \left( (N^{-1} Q_1)^{-1/2} - \Theta_1^{-1/2} \right) \Rightarrow N(0, 4^{-1} \Theta_1^{-3} \Sigma_{11}). \quad (\text{A13})$$

Together, these results imply that the expressions appearing on the right hand side of equations (A10) and (A11) are mean zero with variance  $\Theta_1^{-4} \Theta_2^2 \Sigma_{11} + \Theta_1^{-2} \Sigma_{22}$  and  $4^{-1} \Theta_1^{-3} \Theta_2^2 \Sigma_{11} + \Theta_1^{-1} \Sigma_{22}$ , respectively. This establishes the required results.  $\blacksquare$

## Appendix B Tables

Table 1: Asymptotic moments

Model	$\Theta_2\Theta_1^{-1}$	$\phi'\Sigma\phi$	$\Theta_2\Theta_1^{-1/2}$	$\varphi'\Sigma\varphi$
1. No constant or trend terms	0	2	0	1
2. Individual specific constants	-3	51/5	$-\sqrt{3/2}$	8/10
3. Individual specific constants and trends	-15/2	2895/112	$-\sqrt{15/4}$	277/448

Table 2: Empirical size

Model	$\tau$	$T$	$N$	$DF_t$	$DF_\rho$	$Z_t$	$Z_\rho$	$MP_t$	$MP_\rho$	$BN_t$	$BN_\rho$
1	0	100	10	0.079	0.121	0.073	0.140	0.091	0.154	0.085	0.147
1	0	200	10	0.082	0.139	0.085	0.148	0.098	0.156	0.090	0.155
1	0	100	20	0.065	0.093	0.062	0.096	0.063	0.099	0.065	0.095
1	0	200	20	0.065	0.088	0.056	0.093	0.065	0.099	0.059	0.093
2	0	100	10	0.059	0.093	0.077	0.135	0.153	0.189	0.084	0.149
2	0	200	10	0.088	0.114	0.076	0.125	0.130	0.174	0.087	0.155
2	0	100	20	0.066	0.081	0.164	0.202	0.205	0.230	0.069	0.112
2	0	200	20	0.064	0.077	0.107	0.128	0.127	0.149	0.064	0.088
3	0	100	10	0.086	0.078	0.099	0.122	0.226	0.212	0.119	0.121
3	0	200	10	0.075	0.082	0.049	0.093	0.156	0.159	0.101	0.108
3	0	100	20	0.102	0.076	0.177	0.248	0.370	0.298	0.089	0.080
3	0	200	20	0.081	0.077	0.117	0.153	0.198	0.206	0.086	0.091
1	1	100	10	0.197	0.277	0.062	0.115	0.074	0.132	0.077	0.128
1	1	200	10	0.199	0.289	0.081	0.137	0.092	0.146	0.093	0.144
1	1	100	20	0.299	0.346	0.061	0.098	0.067	0.097	0.064	0.102
1	1	200	20	0.303	0.351	0.064	0.098	0.067	0.102	0.070	0.102
2	1	100	10	0.191	0.235	0.057	0.083	0.093	0.119	0.085	0.132
2	1	200	10	0.180	0.213	0.050	0.080	0.085	0.103	0.078	0.135
2	1	100	20	0.268	0.296	0.072	0.114	0.098	0.123	0.084	0.116
2	1	200	20	0.238	0.301	0.059	0.082	0.076	0.093	0.076	0.101
3	1	100	10	0.242	0.226	0.062	0.081	0.120	0.115	0.121	0.125
3	1	200	10	0.187	0.207	0.029	0.056	0.077	0.095	0.103	0.117
3	1	100	20	0.318	0.281	0.096	0.088	0.153	0.112	0.115	0.103
3	1	200	20	0.307	0.291	0.057	0.074	0.097	0.090	0.090	0.096
1	2	100	10	0.387	0.443	0.060	0.127	0.070	0.121	0.075	0.129
1	2	200	10	0.370	0.419	0.079	0.136	0.096	0.151	0.089	0.143
1	2	100	20	0.500	0.515	0.072	0.113	0.065	0.103	0.070	0.113
1	2	200	20	0.496	0.511	0.063	0.102	0.065	0.109	0.066	0.104
2	2	100	10	0.341	0.385	0.064	0.099	0.102	0.122	0.072	0.132
2	2	200	10	0.372	0.406	0.052	0.089	0.090	0.118	0.063	0.124
2	2	100	20	0.468	0.491	0.078	0.117	0.086	0.117	0.071	0.113
2	2	200	20	0.429	0.473	0.055	0.076	0.072	0.086	0.067	0.100
3	2	100	10	0.353	0.353	0.053	0.067	0.125	0.100	0.119	0.121
3	2	200	10	0.318	0.319	0.037	0.071	0.090	0.108	0.107	0.118
3	2	100	20	0.467	0.449	0.100	0.095	0.139	0.108	0.109	0.104
3	2	200	20	0.429	0.423	0.040	0.049	0.071	0.070	0.089	0.091

Table 3: Size-adjusted power when  $\tau = 0$

Model	T	N	$\rho_i \sim U(0.9, 1)$						$\rho_i \sim U(0.98, 1)$									
			DF <sub>t</sub>	DF <sub>p</sub>	Z <sub>t</sub>	Z <sub>p</sub>	MP <sub>t</sub>	MP <sub>p</sub>	BN <sub>t</sub>	BN <sub>p</sub>	DF <sub>t</sub>	DF <sub>p</sub>	Z <sub>t</sub>	Z <sub>p</sub>	MP <sub>t</sub>	MP <sub>p</sub>	BN <sub>t</sub>	BN <sub>p</sub>
0	100	10	0.969	0.969	0.948	0.957	0.942	0.956	0.944	0.954	0.445	0.440	0.355	0.372	0.381	0.391	0.387	0.400
0	200	10	0.991	0.990	0.985	0.987	0.984	0.984	0.985	0.987	0.753	0.761	0.701	0.714	0.689	0.699	0.691	0.711
0	100	20	0.997	0.997	0.997	0.997	0.996	0.996	0.995	0.995	0.747	0.727	0.700	0.686	0.703	0.702	0.657	0.674
0	200	20	0.998	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.939	0.943	0.911	0.913	0.914	0.919	0.918	0.919
1	100	10	0.628	0.864	0.473	0.738	0.481	0.754	0.928	0.933	0.098	0.219	0.080	0.140	0.081	0.166	0.340	0.366
1	200	10	0.948	0.978	0.903	0.969	0.889	0.959	0.977	0.978	0.301	0.506	0.191	0.339	0.179	0.330	0.710	0.713
1	100	20	0.905	0.977	0.738	0.933	0.751	0.924	0.992	0.992	0.196	0.334	0.102	0.223	0.131	0.225	0.616	0.607
1	200	20	0.994	0.999	0.982	0.999	0.977	0.998	0.999	1.000	0.463	0.678	0.337	0.614	0.909	0.908	0.043	0.059
2	100	10	0.364	0.465	0.263	0.357	0.244	0.333	0.540	0.535	0.287	0.586	0.057	0.054	0.046	0.068	0.094	0.092
2	200	10	0.825	0.899	0.736	0.823	0.729	0.827	0.882	0.882	0.112	0.149	0.095	0.114	0.088	0.113	0.149	0.149
2	100	20	0.602	0.690	0.356	0.466	0.382	0.502	0.774	0.782	0.063	0.064	0.056	0.050	0.073	0.074	0.062	0.065
2	200	20	0.945	0.980	0.916	0.953	0.905	0.957	0.988	0.988	0.126	0.147	0.105	0.137	0.099	0.133	0.205	0.204

Table 4: Size-adjusted power when  $\tau = 1$

Model	T	N	$\rho_i \sim U(0.9, 1)$						$\rho_i \sim U(0.98, 1)$									
			DF <sub>t</sub>	DF <sub>p</sub>	Z <sub>t</sub>	Z <sub>p</sub>	MP <sub>t</sub>	MP <sub>p</sub>	BN <sub>t</sub>	BN <sub>p</sub>	DF <sub>t</sub>	DF <sub>p</sub>	Z <sub>t</sub>	Z <sub>p</sub>	MP <sub>t</sub>	MP <sub>p</sub>	BN <sub>t</sub>	BN <sub>p</sub>
0	100	10	0.871	0.875	0.913	0.923	0.914	0.920	0.900	0.911	0.254	0.242	0.387	0.378	0.387	0.391	0.359	0.361
0	200	10	0.963	0.966	0.965	0.968	0.964	0.969	0.960	0.963	0.548	0.543	0.623	0.659	0.636	0.648	0.636	0.657
0	100	20	0.932	0.931	0.982	0.983	0.975	0.977	0.972	0.975	0.417	0.398	0.612	0.626	0.606	0.600	0.580	0.593
0	200	20	0.979	0.981	0.987	0.988	0.985	0.987	0.985	0.986	0.726	0.697	0.840	0.841	0.839	0.836	0.835	0.829
1	100	10	0.469	0.711	0.575	0.794	0.557	0.775	0.878	0.884	0.087	0.164	0.132	0.227	0.131	0.216	0.292	0.278
1	200	10	0.775	0.913	0.843	0.933	0.830	0.926	0.956	0.961	0.148	0.307	0.179	0.356	0.173	0.342	0.663	0.666
1	100	20	0.458	0.693	0.748	0.908	0.727	0.890	0.967	0.968	0.085	0.160	0.174	0.302	0.175	0.308	0.545	0.553
1	200	20	0.843	0.940	0.948	0.984	0.934	0.975	0.985	0.987	0.176	0.347	0.289	0.540	0.288	0.530	0.843	0.850
2	100	10	0.187	0.256	0.247	0.312	0.237	0.287	0.376	0.378	0.055	0.055	0.054	0.059	0.056	0.054	0.055	0.055
2	200	10	0.534	0.651	0.663	0.757	0.641	0.749	0.773	0.773	0.072	0.090	0.093	0.114	0.082	0.108	0.132	0.128
2	100	20	0.220	0.307	0.386	0.520	0.388	0.513	0.692	0.689	0.057	0.065	0.062	0.065	0.059	0.065	0.082	0.079
2	200	20	0.668	0.754	0.855	0.909	0.834	0.891	0.931	0.933	0.101	0.130	0.129	0.147	0.116	0.133	0.186	0.186



Table 5: Size-adjusted power when  $\tau = 2$

Model	T	N	$\rho_i \sim U(0.9, 1)$						$\rho_i \sim U(0.98, 1)$									
			DF <sub>t</sub>	DF <sub>p</sub>	Z <sub>t</sub>	Z <sub>p</sub>	MP <sub>t</sub>	MP <sub>p</sub>	BN <sub>t</sub>	BN <sub>p</sub>	DF <sub>t</sub>	DF <sub>p</sub>	Z <sub>t</sub>	Z <sub>p</sub>	MP <sub>t</sub>	MP <sub>p</sub>	BN <sub>t</sub>	BN <sub>p</sub>
0	100	10	0.697	0.685	0.812	0.815	0.782	0.788	0.769	0.762	0.217	0.197	0.274	0.275	0.266	0.270	0.279	0.275
0	200	10	0.863	0.871	0.886	0.890	0.862	0.867	0.855	0.856	0.331	0.325	0.538	0.537	0.519	0.524	0.501	0.516
0	100	20	0.688	0.705	0.867	0.877	0.844	0.854	0.832	0.831	0.205	0.202	0.489	0.521	0.481	0.510	0.467	0.495
0	200	20	0.888	0.883	0.947	0.952	0.918	0.929	0.909	0.921	0.374	0.341	0.694	0.695	0.689	0.692	0.673	0.686
1	100	10	0.252	0.405	0.397	0.558	0.364	0.523	0.720	0.723	0.085	0.128	0.117	0.170	0.099	0.157	0.259	0.264
1	200	10	0.504	0.668	0.660	0.768	0.641	0.738	0.859	0.862	0.084	0.148	0.164	0.273	0.151	0.284	0.483	0.497
1	100	20	0.205	0.422	0.557	0.732	0.507	0.670	0.844	0.845	0.046	0.116	0.121	0.213	0.120	0.200	0.502	0.491
1	200	20	0.566	0.749	0.754	0.879	0.707	0.825	0.933	0.935	0.131	0.222	0.194	0.412	0.180	0.388	0.699	0.710
2	100	10	0.102	0.151	0.186	0.243	0.183	0.240	0.300	0.304	0.052	0.054	0.067	0.062	0.069	0.063	0.068	0.072
2	200	10	0.348	0.442	0.464	0.519	0.430	0.502	0.556	0.557	0.090	0.093	0.093	0.078	0.082	0.087	0.078	0.078
2	100	20	0.151	0.204	0.287	0.354	0.248	0.326	0.458	0.458	0.057	0.069	0.057	0.053	0.053	0.054	0.062	0.061
2	200	20	0.313	0.422	0.553	0.616	0.496	0.565	0.688	0.686	0.068	0.080	0.066	0.071	0.063	0.065	0.117	0.116

## References

- [1] Bai, J., and S. Ng, 2004, A Panic Attack on Unit Roots and Cointegration, *Econometrica*, Vol. 72, No. 4, pp. 1127-1177.
- [2] Chang, Y., 2002, Nonlinear IV Unit Root Tests in Panels with Cross-Sectional Dependency, *Journal of Econometrics*, Vol. 110, pp. 261-292.
- [3] Chang, Y., 2004, Bootstrap Unit Root Test in Panels with Cross-Sectional Dependency, *Journal of Econometrics*, Vol. 120, pp. 263-293.
- [4] Choi, I., 2002, Unit Root Tests for Panel Data, *Journal of International Money and Finance*, No. 20, pp. 249-272.
- [5] Harris, R. D. F., E. Tzavalis, 1999, Inference for Unit Roots in Dynamic Panels where the Time Dimension is Fixed, *Journal of Econometrics*, No. 91, pp. 201-226.
- [6] Im, K., M. Pesaran and Y. Shin, 2003, Testing for Unit Roots in Heterogeneous Panels, *Journal of Econometrics*, No. 115, pp. 53-74.
- [7] Jönsson, K., 2004, Cross-Sectional Dependency and Size Distortion in a Small-Sample Homogenous Panel Data Unit Root Test, *Oxford Bulletin of Economics and Statistics*, forthcoming.
- [8] Levin, A., C. F. Lin and J. Chu, 2002, Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties, *Journal of Econometrics*, No. 98, pp. 1-24.
- [9] Maddala, G. S., and S. Wu, 1999, A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test, *Oxford Bulletin of Economics and Statistics*, 61, pp. 631-652.
- [10] Moon, H. R., and B. Perron, 2004, Testing for a Unit Root in Panels with Dynamic Factors, *Journal of Econometrics*, No. 122, pp. 81-126.
- [11] O'Connell, P., 1998, The Overvaluation of Purchasing Power Parity, *Journal of International Economics*, No. 44, pp. 1-19.
- [12] Phillips, P. C. B., and D. Sul, 2003, Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence, *Econometrics Journal*, Vol. 6, pp. 217-259.
- [13] Phillips, P. C. B., and H. R. Moon, 1999, Linear Regression Limit Theory of Nonstationary Panel Data, *Econometrica*, No. 67, pp. 1057-1111.
- [14] Phillips, P. C. B., and V. Solo, 1992, Asymptotics for Linear Processes, *Annals of Statistics*, Vol. 20, No. 2, pp. 971-1001.

- [15] Phillips, P. C. B., and P. Perron, 1988, Testing for a Unit Root in Time Series Regression, *Biometrika*, No. 75, pp. 335-346.
- [16] Smith, V. L., S. Leybourne, T-H. Kim and P. Newbold, 2004, More Powerful Panel Data Unit Root Tests with an Application to Mean Reversion in Real Exchange Rates, *Journal of Applied Econometrics*, No. 19, pp. 147-170.