Simple Tests for Cointegration in Dependent Panels with Structural Breaks*

Joakim Westerlund† and David Edgerton

July 11, 2006

Abstract

This paper develops two very simple tests for the null hypothesis of no cointegration in panel data. The tests are general enough to allow for heteroskedastic and serially correlated errors, unit specific time trends, cross-sectional dependence and an unknown structural break in both the intercept and slope of the cointegrated regression, which may be located at different dates for different units. The limiting distributions of the tests are derived, and are found to be normal and free of nuisance parameters under the null. A small simulation study is also conducted to investigate the small-sample properties of the tests. In our empirical application, we provide new evidence concerning the purchasing power parity hypothesis.

JEL Classification: C12; C32; C33; F40.

Keywords: Panel Cointegration; Cointegration Test; Structural Break; Cross-Sectional Dependence; Common Factor; Purchasing Power Parity.

1 Introduction

In recent years, there has been an upsurge in the availability and use of panel data sets where economic variables are observed over extended periods of time across a large number of cross-sectional units, such as countries, industries and households. This development has in turn given rise to a great amount of interest in econometric techniques for dealing with potentially nonstationary panel data variables. In particular, there has been a rapid growth in the use of cointegration methods with large panel data sets to empirically test important economic theories. See Breitung and Pesaran (2006) for a recent survey of the econometric research in this area.

*This project is financed by the Jan Wallander and Tom Hedelius Foundation, research grant numbers P2005-0117:1 and W2006-0068:1.

†Corresponding author: 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.

An important motivation for using such panel cointegration tests is that we wish to gain statistical power through the pooling of information across units, and to thereby improve upon the poor precision of univariate tests. So far, this has mainly been achieved by the application of what might be called a first generation of panel cointegration tests, which rely on many simplifying assumptions. This is problematic in the sense that violations of these assumptions can significantly distort the size of the tests. The increase in power therefore usually comes with nonnegligible costs.

One problem is that most panel cointegration tests are not equipped to handle the kind of structural change that is usually present when analyzing data that covers long periods of time. Although breaks of this kind can manifest themselves in many ways, the most recognized ones are those that affect the cointegration vector itself, which seems directly at odds with the conventional wisdom that there should be a stable long-run cointegrating relationship. They also complicates the interpretation of the test outcome, since structural breaks and unit roots share many qualitative features. In fact, most tests that attempt to distinguish between spurious and cointegrated processes will tend to favor the spurious model when the true process is subject to structural breaks but is otherwise cointegrated within the break regimes.

Another important problem that the first generation of tests has been unable handle is cross-sectional dependence. When studying macroeconomic and financial data for example, cross-sectional dependencies are likely to be the rule rather than the exception, due to strong inter-economy linkages. The tests of Larsson *et al.* (2001), McCoskey and Kao (1998), Pedroni (1999, 2004) and Westerlund (2005a, 2005b) all require independence among the cross-sectional units, and their size properties become suspect when this assumption does not hold.

Although these problems have been more or less ignored in the first generation of studies, some attempts have recently been made to obtain at least a partial solution. Westerlund (2006a, 2006b) relaxes the assumption of a stable cointegrating relationship, and allows for breaks in the deterministic component. By contrast, Groen and Kleibergen (2003) make an attempt to relax the cross-sectional independence assumption, and develop tests based on seemingly unrelated regressions.

While potentially very appealing by themselves, these solutions will, however, be inadequate in any application that is characterized by both structural change and cross-sectional dependence. What one would like to have is not two tests, one for each problem, but one test that takes care of both problems simultaneously. In a recent working paper, Banerjee and Carrion-i-Silvestre (2006) have proposed a number of tests that are appropriate under various degrees of structural breaks and cross-sectional dependence. Although much more general than the first generation of tests, they do not provide practitioners with an uniform solution that can be applied in all situations. In particular, it is not possible to test for cointegration while entertaining the possibility of both cross-sectional dependence and heterogeneous breaks, which is likely to be the case in practice.

Another drawback that applies to all these approaches is that the limiting test distributions depend critically on the nuisance parameters associated with both the number of regressors and deterministic specification of the cointegrated regression. Thus, there is not just one set of critical values, but one for each combination of regressors and deterministic specification. Moreover, since the asymptotic distribution is often a poor approximation in small samples, a new set of critical values is usually needed for each sample size, see for example Banerjee and Carrion-i-Silvestre (2006) and Westerlund (2006b).

In this paper, we propose two simple tests for the null hypothesis of no cointegration that can be used under very general conditions. The tests are derived from the popular Lagrange multiplier (LM) based unit root tests developed by Schmidt and Phillips (1992), Ahn (1993) and Amsler and Lee (1996), and they are able to accommodate heteroskedastic and serially correlated errors, individual specific intercepts and time trends, cross-sectional dependence and an unknown break in both the intercept and slope of the cointegrated regression, which may be located different dates for different units. In spite of this generality, both tests are very straightforward and easy to implement.

The asymptotic analysis reveals that the tests have limiting normal distributions that are free of nuisance parameters under the null hypothesis. In particular, it is shown that the asymptotic null distributions are independent of both the structural break and the common factors, which are used here to model the cross-sectional dependence. Moreover, since the null distributions are also independent of the regressors, there is only one set of critical values for all testing situations considered. Another convenient property of the new tests is that the null distributions are unaffected by erroneous omission of structural breaks in the intercept, although this will naturally affect their power. A small simulation study is also undertaken to evaluate the small-sample properties of the new tests, and the results show that the asymptotic properties are borne out well in small samples.

In our empirical application, we reevaluate the purchasing power parity (PPP) hypothesis, using data on nominal dollar exchange rates and relative price levels for 17 industrialized countries between the first quarter of 1973 and the fourth quarter of 1998. The results suggest that the two variables are nonstationary but not cointegrated, thus implying that PPP should be rejected for all the members of the panel. This result holds even when we extend our tests to allow for multiple structural breaks, representing the episodic behavior of the dollar in the 1980s.

The rest of this paper is organized as follows. The model is specified Section 2, and in Section 3, the new cointegration tests are developed. The asymptotic properties of the tests are analyzed in Section 4, while Section 5 is devoted to the simulation study. Section 6 holds the empirical application to PPP and Section 7 concludes. Proofs of important results are given in the appendix.

2 Model and assumptions

We consider the scalar variate y_{it} given by

$$y_{it} = \alpha_i + \eta_i t + \delta_i D_{it} + x_{it}'\beta_i + (D_{it}x_{it})'\gamma_i + z_{it}, \tag{1}$$

$$x_{it} = x_{it-1} + w_{it}, \tag{2}$$

where $t = 1, ..., T$ and $i = 1, ..., N$ indexes the time series and cross-sectional units, respectively. The vector x_{it} has dimension K and contains the regressors. This variable is modeled as a pure random walk process. The variable D_{it} is a break dummy such that $D_{it} = 1$ if $t > T_i^b$ and zero otherwise. Thus, in this setup, α_i and β_i represent the intercept and slope before the break, while δ_i and γ_i represent the change in these parameters at the time of the shift.

The disturbance z_{it} is assumed to have the following data generating process that permits for cross-sectional dependence through the use of unobserved common factors

$$z_{it} = \lambda_i' F_t + v_{it}, \tag{3}$$

$$F_{jt} = \rho_j F_{jt-1} + u_{jt},\ j = 1, ..., k, \tag{4}$$

$$\phi_i(L)\Delta v_{it} = \phi_i v_{it-1} + e_{it}, \tag{5}$$

where $\phi_i(L) = 1 - \sum_{j=1}^{p_i} \phi_{ij} L^j$ is a scalar polynomial in the lag operator L, F_t is an k dimensional vector of unobservable common factors and λ_i is a vector of loading parameters. By assuming that $\rho_j < 1$ for all j, we ensure that F_t is strictly stationary, which implies that the order of integration of the composite regression error z_{it} depend only on the integratedness of the idiosyncratic disturbance v_{it}. Thus, in this data generating process, the relationship in (1) is cointegrated if $\phi_i < 0$ and it is spurious if $\phi_i = 0$.

Next, we lay out the key assumptions needed for developing the new tests.

Assumption 1. (Error process.)

(a) $E(e_{it}e_{kj}) = 0$ and $E(w_{it}w_{kj}) = 0$ for all $i \neq k$, t and j;

(b) The groups e_{it} and w_{it} are mutually independent across both i and t;

(c) $\text{var}(e_{it}) = \sigma_i^2 > 0$ and $\text{var}(w_{it}) = \Omega_i$ is positive definite.

For the asymptotic theory, the following condition is also required.

Assumption 2. (Invariance principle.) The partial sum processes of e_{it} and w_{it} satisfy an invariance principle. In particular, $T^{-1/2} \sum_{t=1}^{[rT]} e_{it} \Rightarrow \sigma_i W_i(r)$ as $T \to \infty$ for each i, where $\Rightarrow$ denotes weak convergence and $W_i(r)$ is a standard Brownian motion defined on the unit interval $r \in [0, 1]$.

Finally, to be able to handle the factors and structural breaks, the following conditions are assumed to hold.

Assumption 3. (Common factors.)

(a) $E(u_t) = 0$ and $\text{var}(u_t) < \infty$;

(b) u_t is independent of e_{it} and w_{it} for all i and t;

(c) $N^{-1} \sum_{i=1}^{N} \lambda_i \lambda_i' \to \Sigma$, where Σ is positive definite.

Assumption 4. (Structural breaks.)

(a) The location of the break is an integer T_i such that $T_i^b = \lambda_i^b T$, where $\lambda_i^b \in [n, 1-n]$ and $n \in (0,1)$;

(b) The magnitude of the slope break is such that $\gamma_i = O(T^{-a})$, where $a > 0$.

Assumptions 1 through 4 form the basic conditions needed for developing the new cointegration tests given in the next section. Consider first Assumption 1 (a). This assumption implies that the cross-sectional dependence is restricted to the common factors, and is used in the derivation of the asymptotic distribution of our tests. Note that while e_{it} and w_{it} are assumed to be cross-sectionally independent, $E(z_{it} z_{jt}) = \lambda_i' E(F_t F_t') \lambda_j$ for $i \neq j$ so z_{it} is permitted to be cross-sectionally dependent. The extent of this dependence is determined by λ_i.

Assumption 1 (b) states that e_{it} and w_{it} are independent as groups, which is satisfied if the regressors are strictly exogenous. Of course, in many applications, imposing strict exogeneity can be quite restrictive, and in Section 4 we therefore discuss some possibilities for dealing with the case when this assumption is violated. Assumption 1 (c) states that Ω_i is positive definite, which means that x_{it} cannot be cointegrated in case we have multiple regressors.

Assumption 2 states that an invariance principle applies to the partial sum processes constructed from e_{it} and w_{it} as T grows for a given i. Apart from some mild regulatory conditions, this assumption places little restriction on the temporal dependencies of e_{it} and w_{it}, and encompasses for example the broad class of all stationary autoregressive moving average (ARMA) models. It also allows the covariance structure absorbed in σ_i^2 and Ω_i to differ between the cross-sectional units. Serial correlation in Δv_{it} is permitted through $\phi_i(L)$ giving the long-run variance $\omega_i^2 = \sigma_i^2 / \phi_i(1)^2$.

Assumptions 3 (a) and (b) ensure the consistency of the principal components estimates of the common factors. Moreover, since $\rho_j < 1$, (a) and (b) ensure that an invariance principle holds for the partial sum process of F_t, and that it is independent of the idiosyncratic disturbances. These assumptions are standard in common factor analysis. Assumption 1 (c) ensures that the common factors have a nontrivial contribution to the variation of z_{it}, which in turn ensures that the factor model is identified.

Finally, consider Assumption 4. Assumption 4 (a) ensures that the breakpoint is identified and that it does not lie too close to the beginning or end of the sample. It also ensures that each regime contains a positive fraction of the sample even in the limit as T grows. Assumption 4 (b) embodies the idea that the post-break parameters of the regressors must return to their pre-break values as T increases. This assumption is necessary to obtain the limiting distribution of the our new statistics in the case when there is a break in the slope.

Equations (1) through (5) are general enough to encompass three interesting break cases. In Case 1, $\delta_i = \gamma_i = 0$ so there is no break, whereas, in Case 2, $\gamma_i = 0$ but δ_i is unrestricted suggesting a model with break in the intercept. In Case 3, both γ_i and δ_i are unrestricted so both the intercept and slope are permitted to shift.

3 Test construction

The hypothesis to be tested is that all the members of the panel are not cointegrated against the alternative that at least one unit is cointegrated, which is equivalent to testing $H_0 : \phi_i = 0$ for all i against $H_1 : \phi_i < 0$ for some i. This hypothesis can be tested using the LM principle that the score vector has zero mean when evaluated at the vector of true parameters under the null.

Consider therefore the following pooled log-likelihood function

$$\log L \;=\; \text{constant} - \frac{1}{2}\sum_{i=1}^{N}\left(T\log(\sigma_i^2) - \frac{1}{\sigma_i^2}\sum_{t=1}^{T} e_{it}^2\right).$$

The tests that we propose can be derived as in Westerlund and Edgerton (2006) by first concentrating the log-likelihood function with respect to σ_i^2 and then evaluating the resulting score at the restricted maximum likelihood estimates. If we let $\widehat{\sigma}_i^2 = T^{-1}\sum_{t=1}^{T} e_{it}^2$, then the score contribution for unit i is given by

$$\frac{\partial \log L}{\partial \phi_i} \;=\; \frac{1}{\widehat{\sigma}_i^2}\sum_{t=2}^{T}\Delta\widetilde{S}_{it}\widetilde{S}_{it-1},$$

where $\widehat{S}_{it}$ is a residual from (1) to be defined in detail below, and where $\Delta\widetilde{S}_{it}$ and $\widetilde{S}_{it-1}$ are the mean deviations of $\Delta\widehat{S}_{it}$ and $\widehat{S}_{it-1}$. The point here is that the score vector is proportional to the numerator of the least squares estimate of ϕ_i in the regression

$$\Delta\widehat{S}_{it} \;=\; \text{constant} + \phi_i\widehat{S}_{it-1} + \text{error}. \tag{6}$$

It follows that a test of the hypothesis of the null of no cointegration for cross-section unit i can be formulated equivalently as a zero slope restriction in (6), which can be tested using either the least squares estimate of ϕ_i or its t-ratio. Thus, by considering the form of the log-likelihood function, it is not difficult to see that a panel test of H_0 versus H_1 can be constructed by using the cross-sectional sum of these statistics for each i.

If there is no cross-sectional dependence, and thus no common factors, then the variable $\widehat{S}_{it}$ is defined as

$$\widehat{S}_{it} = y_{it} - \widehat{\alpha}_i - \widehat{\delta}_i D_{it} - \widehat{\eta}_i t - x'_{it}\widehat{\beta}_i - (D_{it}x_{it})'\widehat{\gamma}_i, \tag{7}$$

where $\widehat{\alpha}_i = y_{i1} - \widehat{\eta}_i - \widehat{\delta}_i D_{i1} - x'_{i1}\widehat{\beta}_i - (D_{i1}x_{i1})'\widehat{\gamma}_i$ is the restricted maximum likelihood estimate of α_i. The remaining parameter estimates $\widehat{\eta}_i$, $\widehat{\delta}_i$, $\widehat{\beta}_i$ and $\widehat{\gamma}_i$ can readily be obtained by running least squares on the first differenced version of (1). That is, we have

$$\Delta y_{it} = \widehat{\eta}_i + \widehat{\delta}_i \Delta D_{it} + \Delta x'_{it}\widehat{\beta}_i + \Delta(D_{it}x_{it})'\widehat{\gamma}_i + \Delta\widehat{z}_{it}. \tag{8}$$

The construction of the test is therefore very simple in this case.

When there is cross-sectional dependence, and thus unobserved common factors to be taken into account, the procedure becomes a little more involved. We begin by taking first differences, in which case (3) becomes

$$\Delta z_{it} = \lambda'_i \Delta F_t + \Delta v_{it}.$$

Thus, had Δz_{it} been known, we could have estimated λ_i and ΔF_t directly by the method of principal components. However, Δz_{it} is not known, and we must therefore apply principal components to $\Delta\widehat{z}_{it}$, the residual from (8), instead. In so doing, note that the above expression can be rewritten in terms of $\Delta\widehat{z}_{it}$ as

$$\Delta\widehat{z}_{it} = \lambda'_i \Delta F_t + \Delta\widetilde{v}_{it}, \tag{9}$$

where $\Delta\widetilde{v}_{it}$ can be thought of as a composite error term, comprised of Δv_{it} and the projections of ΔF_t and Δv_{it} onto all the explanatory variables of (8).[1] Let λ, ΔF and $\Delta\widehat{z}$ be $K \times N$, $(T-1) \times K$ and $(T-1) \times N$ matrices of stacked observations on λ_i, ΔF_t and $\Delta\widehat{z}_{it}$, respectively. The principal components estimator $\Delta\widehat{F}$ of ΔF can be obtained by computing $\sqrt{T-1}$ times the eigenvectors corresponding to the K largest eigenvalues of the $(T-1)\times(T-1)$ matrix $\Delta\widehat{z}\Delta\widehat{z}'$. The corresponding matrix of estimated factor loadings is given by $\widehat{\lambda} = \Delta\widehat{F}'\Delta\widehat{z}/(T-1)$. The estimated common factor can be recovered as

$$\widehat{F}_t = \sum_{j=2}^{t} \Delta\widehat{F}_j.$$

Given $\widehat{\lambda}_i$ and $\widehat{F}_t$, the variable $\widehat{S}_{it}$ can be recovered by simply subtracting $\widehat{\lambda}'_i\widehat{F}_t$ from the right-hand side of (7). That is, we compute

$$\widehat{S}_{it} = y_{it} - \widehat{\alpha}_i - \widehat{\delta}_i D_{it} - \widehat{\eta}_i t - x'_{it}\widehat{\beta}_i - (D_{it}x_{it})'\widehat{\gamma}_i - \widehat{\lambda}'_i\widehat{F}_t,$$

[1] In the appendix we show that $\widetilde{v}_{it}$ is consistent for v_{it}. Thus, the presence of the explanatory variables has no asymptotic effect on the estimation of the common factors, which can therefore proceed exactly as in Bai and Ng (2004).

This defactoring makes the test robust against cross-sectional dependence generated by common factors. To also make it robust against serial correlation, we suggest augmenting the test regression in the following fashion

$$\Delta \widehat{S}_{it} \;=\; \text{constant} + \phi_i \widehat{S}_{it-1} + \sum_{j=1}^{p_i} \phi_{ij} \Delta \widehat{S}_{it-j} + \text{error}. \tag{10}$$

To obtain the new tests, define the variance ratio $S_i = \widehat{\omega}_i / \widehat{\sigma}_i$, where $\widehat{\sigma}_i$ is the estimated regression standard error in (10) and $\widehat{\omega}_i^2$ is an semiparametric consistent estimator of ω_i^2, the long run variance of Δv_{it}, which is defined as

$$\widehat{\omega}_i^2 \;=\; \frac{1}{T-1} \sum_{j=-M_i}^{M_i} \left(1 - \frac{j}{M_i + 1}\right) \sum_{t=j+1}^{T} \Delta \widehat{S}_{it} \Delta \widehat{S}_{it-j},$$

where M_i is a kernel bandwidth parameter that determines how many lagged covariances of $\Delta \widehat{S}_{it}$ to estimate in the kernel. If $\widehat{\phi}_i$ is the least squares estimate of ϕ_i in (9) and τ_i its t-ratio, then the panel LM based statistics of H_0 versus H_1 are defined as follows

$$\phi_N \;=\; \frac{1}{N} \sum_{i=1}^{N} T \widehat{\phi}_i S_i \quad \text{and} \quad \tau_N \;=\; \frac{1}{N} \sum_{i=1}^{N} \tau_i.$$

Some remarks are in order. First, note the adjustment factor S_i is only needed to construct ϕ_N. As demonstrated in the next section, this is necessary to make the test invariant with respect to the serial correlation properties on the data.

Second, as in the unit root case studied by Schmidt and Phillips (1992), the parameters used to compute $\widehat{S}_{it}$ are estimated from the first differentiated data. By contrast, the residual-based tests studied by Banerjee and Carrion-i-Silvestre (2006) and Westerlund (2006a) are based on estimating the parameters from the data in levels. Since y_{it} and x_{it} are nonstationary, this type of level regression is spurious, which implies that the estimated regression parameters do not converge to constants, but remain random even asymptotically. As we shall see, this lower degree of randomness makes the asymptotic properties of ϕ_N and τ_N very simple.

Third, to handle the case when the location of the break is unknown, we follow the strategy of Bai and Perron (1998), and estimate the breakpoint by minimizing the sum of squared residuals from the first difference regression in (8). Thus, the breakpoint estimator is defined as follows

$$\widehat{\lambda}_i^b \;=\; \underset{n \le \lambda_i^b \le 1-n}{\arg\min}\; T^{-1} \sum_{t=1}^{T} \Delta \widehat{z}_{it}^2.$$

In estimating the number of common factors M_i, we follow the recommendation of Bai and Ng (2004) and minimize the following information criterion

$$\widehat{k} = \underset{0\le k\le k_{max}}{\arg\min} \log(\widehat{\sigma}^2) + k \log\left(\frac{NT}{N+T}\right)\frac{N+T}{NT},$$

where $\widehat{\sigma}^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\widehat{v}_{it}^2$, $\Delta\widehat{v}_{it} = \Delta\widehat{z}_{it} - \widehat{\lambda}_i'\widehat{F}_t$ is an consistent estimator of Δv_{it} in (9) and k_{max} is an bounded integer such that $k \le k_{max}$.

Fourth, in applied work, the lag length p_i in (9) must be estimated before the testing can begin. This is done preferably by using a data dependent rule. For example, we may use the Campbell and Perron (1991) rule, which is a sequential test rule based on the significance of the individual lag parameters ϕ_{ij} in (10). Another possibility is to use an information criterion, such as the Schwarz Bayesian criterion. Alternatively, the number of lags can be determined independent of the data. Two examples of such deterministic rules involves choosing p_i arbitrarily or as a fixed function of T, which ensures that the estimated test regression provides an increasingly good approximation of the true autoregressive process in (5).[2]

4 Asymptotic distribution

In this section, we derive the asymptotic null distributions of the new statistics under Assumptions 1 through 4. The asymptotic properties of $T\widehat{\phi_i}$ and τ_i were analyzed by Westerlund and Edgerton (2006) in the case of a single time series with serially uncorrelated innovations. In our case, the data generating process is much more general.

As shown in the appendix, under the null and Assumptions 1 to 4, then

$$T\widehat{\phi_i}S_i \Rightarrow B_i = -\left(2\int_0^1 U_i(r)^2 \mathrm{d}r\right)^{-1} \quad \text{as } T\to\infty,$$

where $V_i(r) = W_i(r) - rW_i(1)$ is a standard Brownian bridge and $U_i(r) = V_i(r) - \int_0^1 V_i(s)\mathrm{d}s$ is a demeaned standard Brownian bridge. The corresponding time series limit for the individual t-statistic is shown to be

$$\tau_i \Rightarrow C_i = -\left(4\int_0^1 U_i(r)^2 \mathrm{d}r\right)^{-1/2}.$$

Thus, it is clear that under the assumption of no cross-sectional dependence, $\sqrt{N}$ times cross-sectional average of these statistics converge in distribution to a normal variate by standard Lindberg-Lévy central limit theorem arguments. The following theorem shows that the effect of the common factor is asymptotically negligible, and that ϕ_N and τ_N are indeed asymptotically normal.

[2]In this paper, we consider both types of lag selection rule. In particular, while Section 5 considers a fixed lag length, Section 6 considers a data dependent choice.

Theorem 1. (Asymptotic distributions.) Under the null hypothesis H_0 and Assumptions 1 through 4 (a), as $T \to \infty$ and then $N \to \infty$

$$\frac{\sqrt{N}(\phi_N - E(B_i))}{\sqrt{\text{var}(B_i)}}, \quad \frac{\sqrt{N}(\tau_N - E(C_i))}{\sqrt{\text{var}(C_i)}} \;\Rightarrow\; N(0,1).$$

Remark 1. The proof is given in the appendix but it is instructive to consider briefly why it holds. It begins by showing that

$$T^{-1/2}\widehat{S}_{it} \;=\; T^{-1/2}\sum_{j=2}^{t}(v_{ij} - v_i) + o_p(1),$$

where v_i is the time series mean of v_{it}. This implies that $T^{-1/2}\widehat{S}_{it} \Rightarrow \omega_i V_i(r)$ as T grows, which in turn implies that T^{-2} times the denominator of $\widehat{\phi}_i$ converges to $\omega_i^2 \int_0^1 U_i(r)^2 \mathrm{d}r$ while T^{-1} times the numerator converges to $-\sigma_i\omega_i/2$. The asymptotic distributions of $T\widehat{\phi}_i S_i$ and τ_i now follow directly from the fact that S_i is consistent for ω_i/σ_i. Note that these distributions are unaffected by the presence of the common factors, which derives from the fact that $\widehat{F}_t$ converges to F_t at rate $\sqrt{T}$ suggesting that the estimated factors can basically be treated as known. Thus, the defactoring essentially removes the common component from the limiting distribution of $T^{-1/2}\widehat{S}_{it}$. Because of this invariance, asymptotic normality of ϕ_N and τ_N follows by the Lindberg-Lévy central limit theorem.

Remark 2. While unbiased in the absence of deterministic intercept and trend terms, $\widehat{\phi}_i$ is downwards biased in (1) with individual specific intercepts and trends, which implies that $T\widehat{\phi}_i$ diverges towards negative infinity. To account for this, ϕ_N and τ_N are adjusted by the mean and standard deviation of their respective limiting distributions as T grows. Numerical values of these adjustment terms can be obtained via simulation methods. For this purpose, $10,000$ random walks of length $T = 1,000$ were generated. By using these random walks as a simulated Brownian motions, it is possible to evaluate the functionals B_i and C_i, and then to compute the moments. The resulting numerical values of $E(B_i)$ and $\text{var}(B_i)$ are given by -1.9675 and 0.3301 with -8.4376 and 25.8964 being the corresponding values of $E(C_i)$ and $\text{var}(C_i)$. Because the normalized statistics diverge towards negative infinity under the alternative, the computed test value should be compared with the left tail of the normal distribution. If the value is smaller than the appropriate left tail critical value, the null hypothesis is rejected.

Remark 3. Since the limit of $T^{-1/2}\widehat{S}_{it}$ does not depend on x_{it}, ϕ_N and τ_N have the unusual property that their limiting distributions are unaffected by the presence of the regressors. This is a great operational advantage as the same set of moments can be applied regardless of the dimension of x_{it}. Similarly, the asymptotic null distributions are unaffected by the presence of structural breaks.

In fact, as shown by Westerlund and Edgerton (2006), since $T^{-1/2}D_{it}$ vanishes asymptotically, a disregarded break in the intercept has no effect on $T^{-1/2}\widehat{S}_{it}$, which leaves the asymptotic null distributions of the tests unaffected. The problem is that incorrect placement, or exclusion, of the break make the tests biased towards accepting the null hypothesis. Thus, although the break does not affect the asymptotic properties of the tests under the null, it does reduce their power, which is why accounting for the break is important. Also, although $T^{-1/2}D_{it}$ vanishes asymptotically, this is not the case for $T^{-1/2}(D_{it}x_{it})'$, and thus the invariance property does not extend to the case with a shift in the slope.

Remark 4. Assumption 1 (b) requires that the regressors are strictly exogenous. This is not necessary. To see this, note that if we redefine $z_{it} = \lambda_i' F_t + g_{it}$ where $g_{it} = w_{it}'\gamma_i(L) + v_{it}$ and $\gamma_i(L) = \sum_{j=-q_i}^{p_i} \gamma_{ij} L^j$ is a K dimensional lag polynomial, then (1) can be rewritten as

$$y_{it} \;=\; \alpha_i + \eta_i t + \delta_i D_{it} + x_{it}'\beta_i + (D_{it}x_{it})'\gamma_i + \sum_{j=-q_i}^{p_i} \Delta x_{it}'\gamma_{ij} + z_{it}.$$

Thus, since z_{it} is orthogonal to w_{it}, we can allow for weakly exogenous regressors by simply augmenting (1) with the lags and leads of Δx_{it}, which is similar to the conventional dynamic least squares arguments. The testing can now be carried out as before, after modifying (7) and (8) accordingly.

In the appendix, we only provide the proof for the case when the tests are constructed based on the true breakpoint. To prove that the same results hold if λ_i^b is replaced by $\widehat{\lambda}_i^b$, it is sufficient to show that the asymptotic null distributions of the tests are unaffected when computed for an estimated breakpoint different from λ_i^b. This is done in Westerlund and Edgerton (2006), where it is shown that Assumption 4 (b) suffices to get rid of the bias caused by using $\widehat{\lambda}_i^b$ rather than λ_i^b. The proof in our case is almost identical, and is therefore omitted.

5 Monte Carlo simulations

In this section, we investigate the small-sample properties of the new tests through a small simulation study using the following data generating process

$$\begin{aligned} y_{it} &= \delta D_{it} + (D_{it}x_{it})\gamma + x_{it}\beta + z_{it}, \\ z_{it} &= \lambda_i F_t + v_{it}, \\ \Delta v_{it} &= \phi v_{it} + e_{it}, \end{aligned}$$

where x_{it} and F_t are scalars, $\lambda_i \sim N(0,1)$ and $\beta = 1$. For the disturbance e_{it}, we have two scenarios. In the first, $e_{it} = u_{it} + \theta u_{it-1}$ so e_{it} follows a first order moving average process, while, in the second, $e_{it} = \rho e_{it-1} + u_{it}$, and e_{it} therefore follows a first order autoregressive process. In both cases, we have $u_{it} \sim N(0,1)$, $F_t \sim N(0,1)$ and $\Delta x_{it} \sim N(0,1)$, and we use the value zero for initiation. The data is generated for $3,000$ panels with $T+50$ time series observations, where the first 50 is disregarded to reduce the effect of the initial values. For brevity, the number of cross-sectional units is kept fixed at 20, which seemed sufficient to obtain reasonably good precision in the estimated factors.

For the structural breaks, we have three different configurations, each of which correspond to one of our three deterministic cases. For Case 1 with no breaks, we have $\gamma = \delta = 0$, whereas, for Case 2 with an intercept shift, $\gamma = 0$ and $\delta = 5$. In Case 3, $\gamma = \delta = 5$ so there is a break in both intercept and slope. For convenience, we assume a common breakpoint in the middle of the sample.

In constructing the test statistics, we need to determine the appropriate lag order p_i to handle the serial correlation. Although this can in principle be done using a data dependent rule, as we do in Section 6, in the simulations we choose p_i as a function of T, which is computationally less demanding. In particular, since there is no obvious choice, we choose p_i to the largest integer less than $4(T/100)^{2/9}$. The same rule is used for choosing the bandwidth M_i. In order to eliminate the sample endpoints, we use a truncation of 0.1. The maximum number of factors is 3. All computational work was performed in GAUSS.

We begin by presenting the results on the size and power of a nominal 5% level test, and then we go on to discuss some results on the estimation of λ_i^b and k. All results that are referred to but not reported are available from the corresponding author upon request.

5.1 Size and power

Tables 1 through 3 contains the results of the size and power for the new tests. Due to the different size properties of the tests, all powers are adjusted for size.

Consider first the results reported in Table 1 on the size of the tests when $\phi = 0$. As expected, we see that τ_N and ϕ_N generally perform well with good size accuracy in most experiments, especially when there is no serial correlation. Consistent with other simulation studies, such as Haug (1996), however, we see that the adjustments to accommodate serial correlation are not completely effective, and that some distortions seem to remain. In particular, while τ_N tends to perform well in almost all the experiments, there is a tendency for ϕ_N to become oversized when θ and ρ are negative. However, in most cases there is a significant improvement as T increases. Indeed, with $T = 200$, size is almost perfect for both tests.

The overall best size accuracy is obtained by using τ_N, which seems very robust against the different forms of serial correlation considered here. Among the different versions of the tests considered, the best size is not surprisingly obtained by using the true breakpoint and number of factors. The tests based on an estimated breakpoint and number of factors are, however, almost as accurate, and perform only slightly less well. At the other end of the scale, we have the tests that incorrectly ignore the presence of the common factors, which suffer from severe distortions. This result clearly demonstrates the importance of accounting for the cross-sectional dependence, and the potential effects of ignoring it.

Note also the relatively good performance of the no break tests in Case 2 when there is a shift in the intercept. This is to be expected since these tests are asymptotically independent of the intercept break under the null. The ranking of the tests is reversed in the slope shift model, in which case the tests based on estimated and known breakpoints perform best. As explained in Section 4, this is also well in line with what might be expected based on theory.

The results of the power of the tests reported in Tables 2 and 3 can be summarized as follows. Firstly, the power increases quickly as T increases, and as ϕ departs form zero, which is a reflection of the consistency of the tests. Secondly, the tests based on the assumption of no break can have very poor power in Case 3 when there is a shift in the slope, thus corroborating the result that erroneous omission of breaks should affect the tests by lowering their power. In fact, since there is rarely any power beyond the size of the tests, this type of misspecification is almost certain to result in a failure to reject the null.

Thirdly, with exception of the no break tests in Case 3, there are no big differences in terms of power among the various tests considered. In particular, it is seen that there is generally no loss of power involved in estimating the break and the number of common factors rather than treating them as known. This is of course very good news in applications, when there is usually little or no *a priori* knowledge about the location of the break and number of factors.

5.2 Break and factors

Table 4 presents some results of the estimated breakpoint and number of factors under the null when $\phi = 0$.

The results suggest that the new tests perform well in terms of accuracy in the estimated breakpoint and number of common factors. In fact, the correct selection frequencies are nearly 100% in most cases, with averages that are very close to their corresponding true values. This explains in part the good performance of the tests in terms of size and power. Note also how the precision in the estimated number of factors drops markedly when a true break is omitted, from almost 100% with breaks to almost 0% without. This is also to be expected as the principal components estimator is based on the regression error in (8), which then suffers from an omitted variables bias. This illustrates the importance of considering both break and cross-section dependence in the cointegration testing.

This good precision in the estimated break and factors is important not only to the extent that it ensures good performance of the ensuing cointegration tests, but also in its own right, because researchers often seek to draw conclusions regarding these estimated parameters, especially the location of the break.

6 An application to the PPP hypothesis

In this section we illustrate the empirical implementation of our new panel cointegration tests using as an example the PPP hypothesis, which states that real exchange rates across countries should be long-run mean reverting. We investigate the United States dollar real exchange rate, which is defined as

$$q_{it} = e_{it} + p_t^* - p_{it},$$

where q_{it} is the dollar real exchange rate for country i at the end of time period t, e_{it} is the corresponding dollar nominal exchange rate, p_{it} is the domestic consumer price index and p_t^* is the United States consumer price index. All the variables are expressed in logarithms.

The most common way of testing PPP is to use a conventional univariate unit root test on the real exchange rate while allowing for a nonzero mean. A time trend is not usually included, since this is deemed inconsistent with PPP theory. The results obtained from this testing strategy have, however, been very mixed and far from convincing. As a response to this, several authors including Papell (2002) and Hegwood and Papell (1998) have argued that much of the evidence against PPP may be due to misspecification of the estimated test regression. Other, most recently Murray and Papell (2005) and Harris *et al.* (2005), have argued that a failure to reject the null hypothesis of a unit root in the real exchange rate can be at least partially explained by the low power of univariate tests.

In this section, we attempt to address both of these problems simultaneously by using the LM based tests we have developed. This approach has several advantages in comparison to conventional testing approaches. Firstly, by not imposing a unit slope on the relative price level, the estimated PPP relationship is not restricted to be proportional. This is likely to be of relevance in the presence of trade barriers and measurement errors. Secondly, structural breaks in both the level and slope of the PPP relationship can be taken into account. Thirdly, by using panel data, we can improve upon the power relative to conventional univariate testing approaches. Fourthly, as first pointed out by O'Connell (1998), by permitting for common factors the cross-sectional dependence induced by the numeraire country is effectively eliminated.

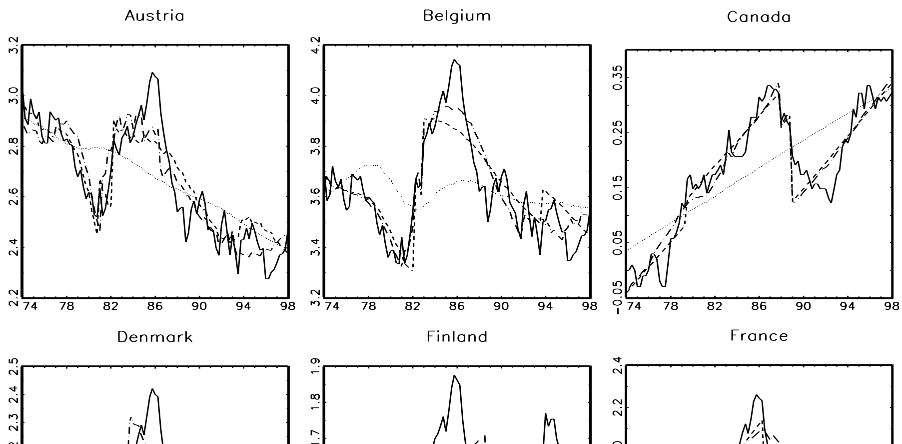
The data that we use is taken from the International Financial Statistics database of the International Monetary Fund, and covers the recent float period from the first quarter of 1973 to the fourth quarter of 1998. It is comprised of 17 countries, namely Austria, Belgium, Canada, Denmark, Finland, France, Greece, Italy, Japan, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom.[3]

[3]Apart from the frequency, this is the same sample as considered by Harris *et al.* (2005), who implemented their test of the PPP hypothesis as a unit root test on the real exchange rate. It is therefore interesting to examine whether their conclusions are robust with respect

The main purpose of this section is to illustrate the practical implementation of the new tests, and to show how they can be extended to accommodate multiple breaks. The rationale for this extension is provided by Papell (2002) who points out that most dollar based exchange rates have exhibited three large structural breaks during the 1980s, when there were large appreciations of the dollar, followed by equally large offsetting depreciations.

Before testing for cointegration, however, we must first check that both the nominal dollar exchange rate and the relative price level are nonstationary. To do this, we use the panel stationarity test recently proposed by Carrion-i-Silvestre *et al.* (2005), which allows for both unknown structural breaks and cross-sectional dependency. The results reported in Table 5 show that the null hypothesis of stationarity can be rejected at the 5% level for all specifications of the deterministic component, including the break models.[4] We can thus conclude that both the nominal exchange rate and relative price level appear to be nonstationary, and can therefore proceed to test for cointegration between these variables.

Figure 1: Nominal dollar exchange rates with fitted trend functions.

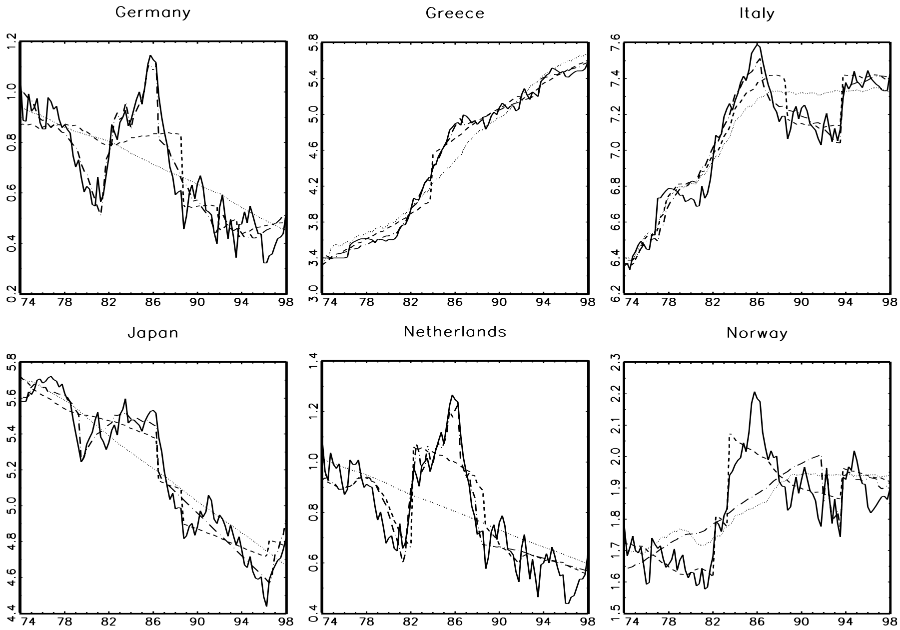


When applying the LM based tests developed in this paper, we follow Papell (2002) and Harris *et al.* (2005), and allow for three breaks. The positions of

to the assumed proportionality between nominal exchange rates and relative price levels, and to the frequency of the data being used.

[4]In constructing the Carrion-i-Silvestre *et al.* (2005) test, we use a maximum of five breaks.

Figure 2: Nominal dollar exchange rates with fitted trend functions.



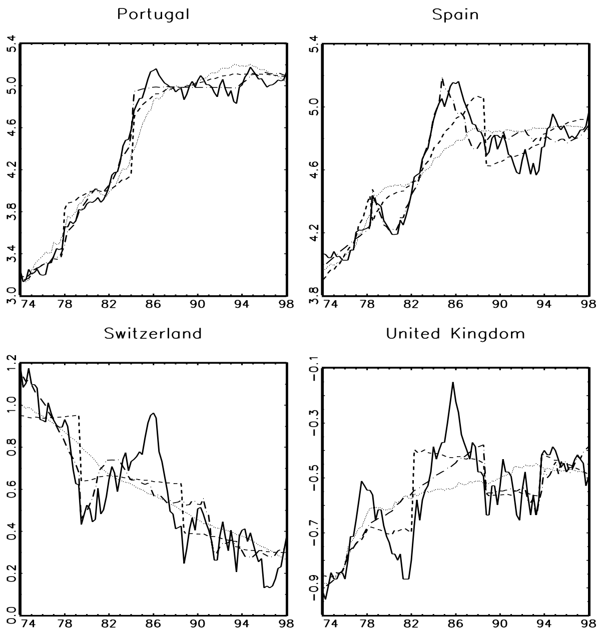
these breaks are estimated using a grid search to minimize the sum of squared residuals from the first difference regression in (8).[5] To determine the number of lags to use in the augmented test regression in (10), we follow Campbell and Perron (1991) who suggest using a sequential procedure based on the significance of the lag parameters. The maximum number of lags is set to 10 and we use a significance level of 5%. The maximum number of common factors is set to five.

The results of the cointegration tests are reported in Table 6, and the fitted trend functions are shown graphically in Figures 1 to 3.[6] In the figures, the solid line represents the nominal exchange rate, the semisolid line represents the regime shift model, the dashed line represents the level break model and the dotted line represents the model without any break. It can easily be seen that allowing for breaks substantially improves the fit of the estimated trend

[5]We also implemented the test using the sequential break estimation procedure of Bai (1997). Although computationally much more convenient, this algorithm did not succeed in finding the grid search minimum. In fact, in agreement with the results of Hegwood and Papell (1998), this procedure generally produced unsatisfactory results with a preponderance of breaks occurring at the sample endpoint.

[6]For space considerations, the results of the estimated breaks are not reported here. There can be obtained upon request from the corresponding author. The median break dates are the first quarter of 1981, the fourth quarter of 1987 and the fourth quarter of 1992 in the level break model, and the first quarter of 1980, the second quarter of 1985 and the fourth quarter of 1992 in the regime shift model.

Figure 3: Nominal dollar exchange rates with fitted trend functions.



functions, and that the regime shift model generally lies closest to the nominal exchange rate. In particular, it is interesting to see how the break models, although very simple, can provide a good fit even to this type of volatile data. In most cases, the shifts in the 1980s are well accounted for.

The cointegration results in Table 6 suggest that the null of no cointegration cannot be rejected at any conventional significance level for any of the models being considered. In fact, with the exception of ϕ_N in the level break model, all the test statistics are positive, and thus lie to the right of the center of the asymptotic normal distribution. The most extreme observation is for the τ_N test in the no break model, which lies way out in the right tail. Thus, the PPP hypothesis cannot be supported by the data.

Although this is consistent with the finding of Harris *et al.* (2005), our result is actually much stronger. In their case, the authors only reject the null of stationarity for the real exchange rate for the whole panel, which merely says that the strong form of PPP fails for at least one country. By contrast, we do not reject the null that nominal exchange rates and relative price levels are not cointegrated for the entire panel. The latter conclusion is stronger in the sense that it implies the failure of PPP for the panel as a whole, and it is robust against nonproportionalities that may well have affected the Harris *et al.* (2005) stationarity test. In either case, these findings cast doubts on the claim by Papell (2002) that PPP emerges once the breaks of the 1980s have been taken into account.

7 Conclusions

This paper proposes two new tests for the null hypothesis of no cointegration in panel data. The tests are based on the univariate unit root tests developed by Schmidt and Phillips (1992) and Amsler and Lee (1995), and they are derived by applying the LM principle to an unobserved common factor representation of the data. Allowable features include heteroskedastic and serially correlated errors, individual specific intercepts and time trends, cross-sectional dependence and an unknown break in both the intercept and slope of the cointegrated regression, which may be located different dates for different units.

By using sequential limit arguments, we show that the tests have limiting normal distributions that are free of nuisance parameters under the null hypothesis. In particular, it is shown that the asymptotic null distributions are independent of the nuisance parameters associated with both the structural break and common factors used to model the cross-sectional dependence. We also show that the asymptotic null distributions are independent of the regressors.

The small-sample performance of the tests is evaluated via simulation methods. The results, which generally accords well with the asymptotic theory, suggest that the tests generally perform well with small size distortions and good power even in small samples. The t-test seems to have better size properties then the coefficient test, and simultaneously maintain relatively good power, and is probably the test to recommend.

In our application, we provide new evidence about the familiar PPP hypothesis. In particular, drawing upon a panel of quarterly data covering 17 countries between 1973 and 1998, we cannot find any evidence in favor of cointegration between nominal exchange rates and relative price levels, which leads us to the conclusion that PPP fails for all the members of the panel. This is true even when we generalize our tests to accommodate multiple structural breaks in both the intercept and slope of the estimated PPP relationship.

Appendix: Mathematical proofs

In this appendix, we prove asymptotic null distributions of ϕ_N and τ_N. For convenience, we will introduce the notation $X_p(x_{it})$ to indicate the error from projecting x_{it} onto the generic vector X_{it} of projection variables. Thus, $X_p(x_{it}) = x_{it} - a_i X_{it}$, where a_i is a vector of projection parameters.

Lemma A.1. (Preliminaries for Lemma A.2.) Let $X_{it} = (1, \Delta\widehat{S}_{it-1}, ..., \Delta\widehat{S}_{it-p_i})'$ and $S_{it} = v_{it} - v_{i1} - \Delta v_i(t-1)$. Under H_0 and Assumptions 1 to 4, as $T \to \infty$

(a) $T^{-1/2}\widetilde{v}_{it} \;=\; T^{-1/2}v_{it} + o_p(1)$;

(b) $T^{-1/2}\widehat{F}_t \;=\; T^{-1/2}HF_t + o_p(1)$;

(c) $T^{-1/2}\widehat{S}_{it} \;=\; T^{-1/2}S_{it} + o_p(1)$;

(d) $T^{-1}\sum_{t=2}^{T} X_p(\Delta\widehat{S}_{it})^2 \;=\; T^{-1}\sum_{t=2}^{T} e_{it}^2 + o_p(1)$;

(e) $T^{-1}\sum_{t=2}^{T} X_p(\widehat{S}_{it-1})X_p(\Delta\widehat{S}_{it}) \;=\; T^{-1}\sum_{t=2}^{T} X_p(S_{it-1})(e_{it} - e_i) + o_p(1)$.

Proof of Lemma A.1.

We begin with (a). If $X_{it} = (1, \Delta D_{it}, \Delta b_{it}', \Delta x_{it}')'$, then $\Delta\widetilde{v}_{it}$ becomes

$$\begin{aligned}
T^{-1/2}\widetilde{v}_{it} &= T^{-1/2}\sum_{j=2}^{t}\Delta\widetilde{v}_{ij} \\
&= T^{-1/2}v_{it} - T^{-1}\sum_{t=2}^{T}\Delta v_{it}X_{it}'\left(T^{-1}\sum_{t=2}^{T}X_{it}X_{it}'\right)^{-1}T^{-1/2}\sum_{j=2}^{t}X_{ij} \\
&- \lambda_i' T^{-1}\sum_{t=2}^{T}\Delta F_t X_{it}'\left(T^{-1}\sum_{t=2}^{T}X_{it}X_{it}'\right)^{-1}T^{-1/2}\sum_{j=2}^{t}X_{ij}. \qquad \text{(A1)}
\end{aligned}$$

The highest ordered term in X_{it} is Δx_{it}. Thus, by using Assumption 2, we get $T^{-1/2}\sum_{j=2}^{t} X_{ij} = O_p(1)$. The term $T^{-1}\sum_{t=2}^{T} X_{it}X_{it}'$ can be written as

$$T^{-1}\sum_{t=2}^{T}X_{it}X_{it}' \;=\; T^{-1}\sum_{t=2}^{T}\begin{bmatrix} 1 & \Delta D_{it} & \Delta b_{it}' & \Delta x_{it}' \\ \Delta D_{it} & \Delta D_{it}^2 & \Delta D_{it}\Delta b_{it}' & \Delta D_{it}\Delta x_{it}' \\ \Delta b_{it} & \Delta b_{it}\Delta D_{it} & \Delta b_{it}\Delta b_{it}' & \Delta b_{it}\Delta x_{it}' \\ \Delta x_{it} & \Delta x_{it}\Delta D_{it} & \Delta x_{it}\Delta b_{it}' & \Delta x_{it}\Delta x_{it}' \end{bmatrix}.$$

Now, since ΔD_{it} equals one only at one point, all sums involving ΔD_{it} are $o_p(1)$ when normalized by T^{-1}. Also, since Δx_{it} is mean zero by Assumption 2, it is clear that $T^{-1}\sum_{t=2}^{T}\Delta b_{it},\ T^{-1}\sum_{t=2}^{T}\Delta x_{it} \to_p 0$ as $T \to \infty$, where $\to_p$ denotes convergence in probability. The remaining terms involve cross-products of Δx_{it}

and Δb_{it}, which, by Assumption 1 (c), converges in probability to Ω_i if $t > T_i$ and zero otherwise. Hence, we can show that $T^{-1}\sum_{t=2}^{T} X_{it}X_{it}' = O_p(1)$.

Next, consider $T^{-1}\sum_{t=2}^{T} \Delta F_t X_{it}'$, which can be written as

$$T^{-1}\sum_{t=2}^{T} \Delta F_t X_{it}' \;=\; T^{-1}\sum_{t=2}^{T} \Big(\Delta F_t,\ \Delta F_t \Delta D_{it},\ \Delta F_t \Delta b_{it}',\ \Delta F_t \Delta x_{it}'\Big),$$

where $T^{-1}\sum_{t=2}^{T} \Delta F_t,\ T^{-1}\sum_{t=2}^{T} \Delta F_t \Delta D_{it} \to_p 0$ as $T \to \infty$. Hence, since Δx_{it} and ΔF_t are orthogonal by Assumption 3 (b), $T^{-1}\sum_{t=2}^{T} \Delta F_t X_{it}' = o_p(1)$. By the same arguments, $T^{-1}\sum_{t=2}^{T} \Delta v_{it} X_{it}' = o_p(1)$. This implies that (A1) can be written as

$$T^{-1/2}\widetilde{v}_{it} \;=\; T^{-1/2} v_{it} + o_p(1), \tag{A2}$$

which establishes (a).

Consider (b). Because F_t can only be identified up to a scale matrix H, say, we considered the rotation HF_t of F_t. If we let $\Delta\widehat{F}_t = H\Delta F_t + v_t$, then $\widehat{F}_t$ becomes

$$\begin{aligned} T^{-1/2}\widehat{F}_t &= T^{-1/2}\sum_{j=2}^{t} \Delta\widehat{F}_j \\ &= T^{-1/2}\sum_{j=2}^{t} H\Delta F_j + T^{-1/2}\sum_{j=2}^{t} v_j, \end{aligned} \tag{A3}$$

where the second term is $O_p(\min(T^{-1/2}, N^{-1/2}))$ by equation (A.3) of Bai and Ng (2004). Here we also need (a) to ensure that v_{it} is asymptotically unaffected by the projections of ΔF_t and Δv_{it} onto X_{it}. Thus, we can show that

$$T^{-1/2}\widehat{F}_t \;=\; T^{-1/2}H(F_t - F_1) + o_p(1) \;=\; T^{-1/2}HF_t + o_p(1), \tag{A4}$$

where $F_1 = O_p(1)$. This establishes (b).

Consider (c). If we let $b_{it} = x_{it}D_{it}$, then $\widehat{S}_{it}$ can be written as

$$\begin{aligned} \widehat{S}_{it} &= y_{it} - \widehat{\alpha}_i - \widehat{\delta}_i D_{it} - \widehat{\eta}_i t - x_{it}'\widehat{\beta}_i - b_{it}'\widehat{\gamma}_i - \widehat{\lambda}_i'\widehat{F}_t \\ &= y_{it} - y_{i1} - \widehat{\delta}_i(D_{it} - D_{i1}) - \widehat{\eta}_i(t-1) - (x_{it} - x_{i1})'\widehat{\beta}_i \\ &- (b_{it} - b_{i1})'\widehat{\gamma}_i - \widehat{\lambda}_i'\widehat{F}_t \\ &= (v_{it} - v_{i1}) - (\widehat{\delta}_i - \delta_i)(D_{it} - D_{i1}) - (\widehat{\eta}_i - \eta_i)(t-1) \\ &- (x_{it} - x_{i1})'(\widehat{\beta}_i - \beta_i) - (b_{it} - b_{i1})'(\widehat{\gamma}_i - \gamma_i) - \widehat{\lambda}_i'\widehat{F}_t + \lambda_i'(F_t - F_1). \end{aligned} \tag{A5}$$

For the third term, we have

$$\begin{aligned} \widehat{\eta}_i - \eta_i &= \Delta y_i - \widehat{\delta}_i \Delta D_i - \Delta x_i'\widehat{\beta}_i - \Delta b_i'\widehat{\gamma}_i - \eta_i \\ &= \Delta v_i - (\widehat{\delta}_i - \delta_i)\Delta D_i - \Delta x_i'(\widehat{\beta}_i - \beta_i) - \Delta b_i'(\widehat{\gamma}_i - \gamma_i) + \lambda_i'\Delta F \\ &= \Delta v_i - (\widehat{\delta}_i - \delta_i)/(T-1) - w_i'(\widehat{\beta}_i - \beta_i) \\ &- \Delta b_i'(\widehat{\gamma}_i - \gamma_i) + \lambda_i'\Delta F, \end{aligned} \tag{A6}$$

where we have suppressed the index t to denote that the time series average has been taken. For the third equality, we used that $\Delta x_{it} = w_{it}$ and $\Delta D_i = 1/(T-1)$.

Now, by using (A6), the expression in (A5) can be written as

$$\begin{aligned}
T^{-1/2}\widehat{S}_{it} &= T^{-1/2}(v_{it} - v_{i1} - \Delta v_i(t-1)) - T^{-1/2}\sum_{j=2}^{t}(w_{ij} - w_i)'(\widehat{\beta}_i - \beta_i) \\
&- T^{-1/2}\left(D_{it} - D_{i1} - \left(\frac{t-1}{T-1}\right)\right)(\widehat{\delta}_i - \delta_i) \\
&- T^{-1/2}\left(b_{it} - b_{i1} - \left(\frac{t-1}{T-1}\right)(b_{iT} - b_{i1})\right)'(\widehat{\gamma}_i - \gamma_i) \\
&- T^{-1/2}(\widehat{\lambda}_i - \lambda_i)'\widehat{F}_t \\
&+ T^{-1/2}\lambda_i'\left((F_t - \widehat{F}_t) - F_1 - \left(\frac{t-1}{T-1}\right)(F_T - F_1)\right) \\
&= T^{-1/2}(v_{it} - v_{i1} - \Delta v_i(t-1)) - I - II - III \\
&- IV + V, \text{ say.} \qquad \text{(A7)}
\end{aligned}$$

To prove (c), we show that I, II, III, IV and V are $o_p(1)$.

First, consider I. Let $X_{it} = (1, \Delta D_{it}, \Delta b_{it}')'$, then $\sqrt{T}(\widehat{\beta}_i - \beta_i)$ is given as

$$\begin{aligned}
\sqrt{T}(\widehat{\beta}_i - \beta_i) &= \left(T^{-1}\sum_{t=2}^{T} X_p(\Delta x_{it})X_p(\Delta x_{it})'\right)^{-1} \\
&\cdot\ T^{-1/2}\sum_{t=2}^{T} X_p(\Delta x_{it})X_p(\Delta z_{it}), \qquad \text{(A8)}
\end{aligned}$$

where the denominator can be expanded as

$$\begin{aligned}
T^{-1}\sum_{t=2}^{T} X_p(\Delta x_{it})X_p(\Delta x_{it})' &= T^{-1}\sum_{t=2}^{T}\Delta x_{it}\Delta x_{it}' - T^{-1}\sum_{t=2}^{T}\Delta x_{it}X_{it}' \\
&\cdot \left(T^{-1}\sum_{t=2}^{T} X_{it}X_{it}'\right)^{-1} T^{-1}\sum_{t=2}^{T} X_{it}\Delta x_{it}'.
\end{aligned}$$

By the same arguments used in the proof of part (a) of this lemma, it is clear that $T^{-1}\sum_{t=2}^{T} X_{it}\Delta x_{it}'$ and $T^{-1}\sum_{t=2}^{T} X_{it}X_{it}'$ are $O_p(1)$. It follows that

$$T^{-1}\sum_{t=2}^{T} X_p(\Delta x_{it})X_p(\Delta x_{it})' = T^{-1}\sum_{t=2}^{T}\Delta x_{it}\Delta x_{it}' + O_p(1) = O_p(1). \qquad \text{(A9)}$$

For the numerator of (A8), we have

$$T^{-1/2}\sum_{t=2}^{T} X_p(\Delta x_{it})X_p(\Delta z_{it}) = T^{-1/2}\sum_{t=2}^{T}\Delta x_{it}\Delta v_{it} - T^{-1}\sum_{t=2}^{T}\Delta x_{it}X_{it}' \cdot \left(T^{-1}\sum_{t=2}^{T}X_{it}X_{it}'\right)^{-1} T^{-1/2}\sum_{t=2}^{T}X_{it}\Delta z_{it},$$

where $T^{-1}\sum_{t=2}^{T} X_{it}\Delta z_{it} \to_p 0$ as $T \to \infty$ because Δv_{it}, Δx_{it} and ΔF_t are independent by Assumptions 1 and 2 (b). Also, $T^{-1/2}\sum_{t=2}^{T}\Delta x_{it}\Delta z_{it} = O_p(1)$ by standard arguments for stationary processes. Hence, by using (A9), we get

$$\sqrt{T}(\widehat{\beta}_i - \beta_i) = O_p(1)\cdot O_p(1) = O_p(1).$$

By using this result, and that $T^{-1/2}\sum_{j=2}^{t}(w_{ij} - w_i) = O_p(1)$ by Assumption 2, we can show that

$$I = T^{-1}\sum_{j=2}^{t}(w_{ij} - w_i)'\sqrt{T}(\widehat{\beta}_i - \beta_i) = O_p(T^{-1/2}). \tag{A10}$$

Next, consider II. If $X_{it} = (1, \Delta x_{it}', \Delta b_{it}')'$, then $\sqrt{T}(\widehat{\delta}_i - \delta_i)$ becomes

$$\sqrt{T}(\widehat{\delta}_i - \delta_i) = \left(T^{-1}\sum_{t=2}^{T}X_p(\Delta D_{it})^2\right)^{-1} T^{-1/2}\sum_{t=2}^{T}X_p(\Delta D_{it})X_p(\Delta z_{it}), \tag{A11}$$

where the denominator is given by

$$T^{-1}\sum_{t=2}^{T}X_p(\Delta D_{it})^2 = T^{-1}\sum_{t=2}^{T}(\Delta D_{it})^2 - T^{-1}\sum_{t=2}^{T}\Delta D_{it}X_{it}' \cdot \left(T^{-1}\sum_{t=2}^{T}X_{it}X_{it}'\right)^{-1} T^{-1}\sum_{t=2}^{T}X_{it}\Delta D_{it}.$$

It is clear that $T^{-1}\sum_{t=2}^{T}(\Delta D_{it})^2$ and $T^{-1}\sum_{t=2}^{T}X_{it}\Delta D_{it}$ are $o_p(1)$. The remaining term involves cross-products of the elements in X_{it}, among which the highest ordered term is $O_p(1)$. Hence, $T^{-1}\sum_{t=2}^{T}X_{it}X_{it}'$ is $O_p(1)$, which implies that $T^{-1}\sum_{t=2}^{T}X_p(\Delta D_{it})^2 = o_p(1)$.

For the numerator of (A11), we have

$$T^{-1/2}\sum_{t=2}^{T}X_p(\Delta D_{it})X_p(\Delta z_{it}) = T^{-1/2}\sum_{t=2}^{T}\Delta D_{it}\Delta z_{it} - T^{-1/2}\sum_{t=2}^{T}\Delta D_{it}X_{it}' \cdot \left(T^{-1}\sum_{t=2}^{T}X_{it}X_{it}'\right)^{-1} T^{-1}\sum_{t=2}^{T}X_{it}\Delta z_{it}.$$

This expression is clearly $o_p(1)$, which implies that $\sqrt{T}(\widehat{\delta}_i - \delta_i)$ is $o_p(1)$ too. From this result, we deduce that

$$II = T^{-1}\left((D_{it} - D_{i1}) - \left(\frac{t-1}{T-1}\right)\right)\sqrt{T}(\widehat{\delta}_i - \delta_i) = o_p(T^{-1}). \quad \text{(A12)}$$

For part III, we have

$$\begin{aligned}\sqrt{T}(\widehat{\gamma}_i - \gamma_i) &= \left(T^{-1}\sum_{t=2}^{T} X_p(\Delta b_{it})X_p(\Delta b_{it})'\right)^{-1} \\ &\cdot\ T^{-1/2}\sum_{t=2}^{T} X_p(\Delta b_{it})X_p(\Delta z_{it}), \end{aligned} \quad \text{(A13)}$$

where $X_{it} = (1, \Delta x_{it}', \Delta D_{it})'$ defines the projection errors. The denominator of this expression is given by

$$\begin{aligned}T^{-1}\sum_{t=2}^{T} X_p(\Delta b_{it})X_p(\Delta b_{it})' &= T^{-1}\sum_{t=2}^{T} \Delta b_{it}\Delta b_{it}' - T^{-1}\sum_{t=2}^{T} \Delta b_{it}X_{it}' \\ &\cdot\ \left(T^{-1}\sum_{t=2}^{T} X_{it}X_{it}'\right)^{-1} T^{-1}\sum_{t=2}^{T} X_{it}\Delta b_{it}'.\end{aligned}$$

By using the results derived earlier, it is clear that both terms on the right-hand side of this expression are $O_p(1)$ so that $T^{-1}\sum_{t=2}^{T} X_p(\Delta b_{it})X_p(\Delta b_{it})'$ is also $O_p(1)$.

The numerator of (A13) is qual to

$$\begin{aligned}T^{-1/2}\sum_{t=2}^{T} X_p(\Delta b_{it})X_p(\Delta z_{it}) &= T^{-1/2}\sum_{t=2}^{T} \Delta b_{it}\Delta z_{it} - T^{-1}\sum_{t=2}^{T} \Delta b_{it}X_{it}' \\ &\cdot\ \left(T^{-1}\sum_{t=2}^{T} X_{it}X_{it}'\right)^{-1} T^{-1/2}\sum_{t=2}^{T} X_{it}\Delta z_{it},\end{aligned}$$

where all terms with normalizing order $T^{-1/2}$ are $O_p(1)$ by standard arguments for stationary processes. Hence, since $T^{-1}\sum_{t=2}^{T} \Delta b_{it}X_{it}'$ and $T^{-1}\sum_{t=2}^{T} X_{it}X_{it}'$ are $O_p(1)$, we can deduce that $T^{-1/2}\sum_{t=2}^{T} X_p(\Delta b_{it})X_p(\Delta z_{it}) = O_p(1)$, from which it follows that $\sqrt{T}(\widehat{\gamma}_i - \gamma_i) = O_p(1)$.

This shows that IV must be $o_p(1)$ as seen by writing

$$\begin{aligned}III &= T^{-1/2}\left(T^{-1/2}(b_{it} - b_{i1}) - \left(\frac{t-1}{T-1}\right)T^{-1/2}\left(b_{iT} - b_{i1}\right)\right)' \\ &\cdot\ \sqrt{T}(\widehat{\gamma}_i - \gamma_i) = O_p(T^{-1/2}). \end{aligned} \quad \text{(A14)}$$

Now, consider IV. By using (b), this term can be written as

$$\begin{aligned}IV &= T^{-1/2}(\widehat{\lambda}_i - \lambda_i)'\widehat{F}_t = \sqrt{T}(\widehat{\lambda}_i - \lambda_i)'T^{-1}HF_t + o_p(1) \\ &= O_p(1)\cdot o_p(1) + o_p(1), \end{aligned} \quad \text{(A15)}$$

where $\sqrt{T}(\widehat{\lambda}_i - \lambda_i) = O_p(1)$ follows from Lemma 1 of Bai and Ng (2004).

Similarly, (b) implies that V reduces to

$$\begin{aligned} V &= T^{-1/2}\lambda_i' \left((F_t - \widehat{F}_t) - F_1 - \left(\frac{t-1}{T-1} \right) (F_T - F_1) \right) \\ &= T^{-1/2}\lambda_i' \left(\frac{t-T}{T-1} \right) F_1 + o_p(1) \\ &= O_p(T^{-1/2}) + o_p(1). \end{aligned} \quad \text{(A16)}$$

Putting everything together, (A7) can be written as

$$\begin{aligned} T^{-1/2}\widehat{S}_{it} &= T^{-1/2}(v_{it} - v_{i1} - \Delta v_i(t-1)) + o_p(1) \\ &= T^{-1/2}S_{it} + o_p(1). \end{aligned} \quad \text{(A17)}$$

This proves part (c).

Next, consider (d). Define

$$\Delta\widehat{v}_{it} = \Delta\widehat{z}_{it} - \widehat{\lambda}_i'\Delta\widehat{F}_t. \quad \text{(A18)}$$

Note that $\Delta\widehat{z}_{it}$ can be written as

$$\begin{aligned} \Delta\widehat{z}_{it} &= \Delta y_{it} - \widehat{\eta}_i - \Delta x_{it}'\widehat{\beta}_i - \Delta D_{it}\widehat{\delta}_i - \Delta b_{it}'\widehat{\gamma}_i \\ &= \Delta z_{it} - (\widehat{\eta}_i - \eta_i) - \Delta x_{it}'(\widehat{\beta}_i - \beta_i) - \Delta D_{it}(\widehat{\delta}_i - \delta_i) - \Delta b_{it}'(\widehat{\gamma}_i - \gamma_i). \end{aligned}$$

Thus, by using (A6), and some algebra, we get

$$\begin{aligned} \Delta\widehat{z}_{it} &= \Delta z_{it} + (\widehat{\eta}_i - \eta_i) - T^{-1/2}\Delta x_{it}'\sqrt{T}(\widehat{\beta}_i - \beta_i) - T^{-1/2}\Delta D_{it}\sqrt{T}(\widehat{\delta}_i - \delta_i) \\ &- T^{-1/2}\Delta b_{it}'\sqrt{T}(\widehat{\gamma}_i - \gamma_i) \\ &= \Delta z_{it} - \Delta z_i - T^{-1/2}(\Delta x_{it} - \Delta x_i)'\sqrt{T}(\widehat{\beta}_i - \beta_i) \\ &- T^{-1/2}(\Delta D_{it} - \Delta D_i)\sqrt{T}(\widehat{\delta}_i - \delta_i) - T^{-1/2}(\Delta b_{it} - \Delta b_i)'\sqrt{T}(\widehat{\gamma}_i - \gamma_i) \\ &= \Delta z_{it} - d_{it}, \text{ say.} \end{aligned}$$

Note that, by (c) and $\Delta z_i \to_p 0$, $d_{it} = o_p(1)$. Moreover, from (A5),

$$\begin{aligned} \widehat{S}_{it} &= y_{it} - y_{i1} - \widehat{\delta}_i(D_{it} - D_{i1}) - \widehat{\eta}_i(t-1) - (x_{it} - x_{i1})'\widehat{\beta}_i \\ &- (b_{it} - b_{i1})'\widehat{\gamma}_i - \widehat{\lambda}_i'\widehat{F}_t \\ &= \sum_{j=2}^{t}(\Delta y_{it} - \widehat{\delta}_i\Delta D_{it} - \widehat{\eta}_i - \Delta x_{it}'\widehat{\beta}_i - \Delta b_{it}'\widehat{\gamma}_i - \widehat{\lambda}_i'\Delta\widehat{F}_t), \end{aligned}$$

where we have used the fact that $\widehat{F}_1 = 0$. It follows that

$$\begin{aligned} \Delta\widehat{S}_{it} &= \Delta y_{it} - \widehat{\delta}_i\Delta D_{it} - \widehat{\eta}_i - \Delta x_{it}'\widehat{\beta}_i - \Delta b_{it}'\widehat{\gamma}_i - \widehat{\lambda}_i'\Delta\widehat{F}_t \\ &= \Delta\widehat{z}_{it} - \widehat{\lambda}_i'\Delta\widehat{F}_t \\ &= \Delta\widehat{v}_{it}. \end{aligned}$$

By substituting $\Delta\widehat{S}_{it}$ and $\Delta\widehat{z}_{it}$ into (A18), we get

$$
\begin{aligned}
\Delta\widehat{S}_{it} &= \Delta\widehat{z}_{it} - \widehat{\lambda}_i'\Delta\widehat{F}_t \\
&= \Delta z_{it} - d_{it} - \widehat{\lambda}_i'\Delta\widehat{F}_t \\
&= \Delta v_{it} - d_{it} - (\widehat{\lambda}_i - \lambda_i)'\Delta F_t - (\widehat{\lambda}_i - \lambda_i)'(\Delta\widehat{F}_t - \Delta F_t) \\
&- \lambda_i'(\Delta\widehat{F}_t - \Delta F_t) \\
&= \Delta v_{it} - d_{it} - c_{it}, \text{ say,}
\end{aligned}
$$

where $c_{it} = o_p(1)$ by Lemma 1 of Bai and Ng (2004).

Now, let $X_{it} = (1, \Delta\widehat{S}_{it-1}, ..., \Delta\widehat{S}_{it-p_i})'$, then

$$
\begin{aligned}
T^{-1}\sum_{t=2}^{T} X_p(\Delta\widehat{S}_{it})^2 &= T^{-1}\sum_{t=2}^{T}(X_p(\Delta v_{it}) - X_p(d_{it}) - X_p(c_{it}))^2 \\
&= T^{-1}\sum_{t=2}^{T} X_p(\Delta v_{it})^2 + T^{-1}\sum_{t=2}^{T} X_p(d_{it})^2 + T^{-1}\sum_{t=2}^{T} X_p(c_{it})^2 \\
&- 2T^{-1}\sum_{t=2}^{T} X_p(\Delta v_{it})X_p(d_{it}) - 2T^{-1}\sum_{t=2}^{T} X_p(\Delta v_{it})X_p(c_{it}) \\
&- 2T^{-1}\sum_{t=2}^{T} X_p(c_{it})X_p(d_{it}).
\end{aligned}
$$

The second and third terms are obviously $o_p(1)$. For the fourth term,

$$
\begin{aligned}
T^{-1}\sum_{t=2}^{T} X_p(\Delta v_{it})X_p(d_{it}) &\le \left(T^{-1}\sum_{t=2}^{T} X_p(\Delta v_{it})^2\right)^{1/2}\left(T^{-1}\sum_{t=2}^{T} X_p(d_{it})^2\right)^{1/2} \\
&= O_p(1)\cdot o_p(1),
\end{aligned}
$$

where the inequality follows from the Cauchy-Schwarz inequality. By the same arguments, it can be shown that the fifth and sixth terms are also $o_p(1)$. Since $e_{it} = \phi_i(L)\Delta v_{it}$ under the null, it follows that

$$
\begin{aligned}
T^{-1}\sum_{t=2}^{T} X_p(\Delta\widehat{S}_{it})^2 &= T^{-1}\sum_{t=2}^{T} X_p(\Delta v_{it})^2 + o_p(1) \\
&= T^{-1}\sum_{t=2}^{T} e_{it}^2 + o_p(1).
\end{aligned}
$$

This proves (d).

Finally, consider (e). Note that $X_p(\widehat{S}_{it})^2$ can be written as

$$
\begin{aligned}
X_p(\widehat{S}_{it})^2 &= (X_p(\widehat{S}_{it-1}) - X_p(\Delta\widehat{S}_{it}))^2 \\
&= X_p(\widehat{S}_{it-1})^2 + X_p(\Delta\widehat{S}_{it})^2 + 2X_p(\widehat{S}_{it-1})X_p(\Delta\widehat{S}_{it}),
\end{aligned}
$$

so that

$$T^{-1}\sum_{t=2}^{T} X_p(\widehat{S}_{it-1})X_p(\Delta\widehat{S}_{it}) = \frac{1}{2}T^{-1}(X_p(\widehat{S}_{iT})^2 - X_p(\widehat{S}_{i1})^2) - \frac{1}{2}T^{-1}\sum_{t=2}^{T} X_p(\Delta\widehat{S}_{it})^2. \tag{A19}$$

Similarly, for $T^{-1}\sum_{t=2}^{T} X_p(S_{it-1})X_p(\Delta S_{it})$, we have

$$T^{-1}\sum_{t=2}^{T} X_p(S_{it-1})X_p(\Delta S_{it}) = \frac{1}{2}T^{-1}(X_p(S_{iT})^2 - X_p(S_{i1})^2) - \frac{1}{2}T^{-1}\sum_{t=2}^{T} X_p(\Delta S_{it})^2. \tag{A20}$$

Comparing the right hand side of the two expressions, since $X_p(\widehat{S}_{it}) = \widehat{S}_{it} - X_{it}'a_i$ with $X_{it}'a_i = O_p(1)$, we have $X_p(\widehat{S}_{iT})^2/T - X_p(S_{iT})^2/T = o_p(1)$ by part (c) with $t = T$.

For the third term, since $X_p(\Delta S_{it}) = X_p(\Delta v_{it} - \Delta v_i) = e_{it} - e_i$,

$$\begin{aligned} T^{-1}\sum_{t=2}^{T} X_p(\Delta S_{it})^2 &= T^{-1}\sum_{t=2}^{T}(e_{it} - e_i)^2 = T^{-1}\sum_{t=2}^{T}(e_{it}^2 - e_i^2) \\ &= T^{-1}\sum_{t=2}^{T} e_{it}^2 + o_p(1), \end{aligned}$$

which, together with $T^{-1}\sum_{t=2}^{T} X_p(\Delta\widehat{S}_{it})^2 = T^{-1}\sum_{t=2}^{T} e_{it}^2 + o_p(1)$ from (d), leads to

$$T^{-1}\sum_{t=2}^{T} X_p(\Delta\widehat{S}_{it})^2 = T^{-1}\sum_{t=2}^{T} X_p(\Delta S_{it})^2 + o_p(1),$$

These results imply that (A19) and (A20) can be written as

$$T^{-1}\sum_{t=2}^{T} X_p(\widehat{S}_{it-1})X_p(\Delta\widehat{S}_{it}) = T^{-1}\sum_{t=2}^{T} X_p(S_{it-1})(e_{it} - e_i) + o_p(1),$$

which proves (e). ■

Lemma A.2. (Preliminaries for Theorem 1.) Under the conditions of Lemma A.1, as $T \to \infty$

(a) $T^{-1/2}v_{it} \Rightarrow \dfrac{1}{\phi_i(1)}\sigma_i W_i(r)$;

(b) $T^{-1/2}\widehat{S}_{it} \;\Rightarrow\; \frac{1}{\phi_i(1)}\sigma_i V_i(r);$

(c) $T^{-2}\sum_{t=2}^{T} X_p(\widehat{S}_{it-1})^2 \;\Rightarrow\; \frac{1}{\phi_i(1)^2}\sigma_i^2\int_0^1 U_i(r)^2\mathrm{d}r;$

(d) $T^{-1}\sum_{t=2}^{T} X_p(\widehat{S}_{it-1})X_p(\Delta\widehat{S}_{it}) \;\to_p\; -\frac{1}{2\phi_i(1)}\sigma_i^2.$

Proof of Lemma A.2.

Consider (a). From (3), under the null, we have that

$$\phi_i(L)\Delta v_{it} \;=\; e_{it}.$$

Using the Beveridge-Nelson decomposition of $\phi_i(L)$, we can write $\phi_i(L) = \phi_i(1) + \phi_i^*(L)(1-L)$, and obtain

$$\phi_i(L)\Delta v_{it} \;=\; \phi_i(1)\Delta v_{it} + \phi_i^*(L)\Delta^2 v_{it} \;=\; e_{it},$$

or, equivalently

$$\Delta v_{it} \;=\; -\frac{\phi_i^*(L)}{\phi_i(1)}\Delta^2 v_{it} + \frac{1}{\phi_i(1)}e_{it}.$$

This implies that

$$\begin{aligned} T^{-1/2}v_{it} &= -\frac{\phi_i^*(L)}{\phi_i(1)}T^{-1/2}\Delta v_{it} + \frac{1}{\phi_i(1)}T^{-1/2}\sum_{j=2}^{t} e_{ij} \\ &= \frac{1}{\phi_i(1)}T^{-1/2}\sum_{j=2}^{t} e_{ij} + o_p(1) \\ &\Rightarrow \frac{1}{\phi_i(1)}\sigma_i W_i(r), \end{aligned} \tag{A21}$$

where the last result follows from Assumption 2, thus proving (a).

Part (b) is an immediate consequence of part (a) and Lemma A.1 (c), as can be seen by writing

$$\begin{aligned} T^{-1/2}\widehat{S}_{it} &= T^{-1/2}\sum_{j=2}^{t}(v_{it} - v_i) + o_p(1) \\ &= \frac{1}{\phi_i(1)}T^{-1/2}\sum_{j=2}^{t}(e_{it} - e_i) + o_p(1) \\ &\Rightarrow \frac{1}{\phi_i(1)}\sigma_i V_i(r). \end{aligned} \tag{A22}$$

Consider (c). By using (b) and the rules for projections, we get

$$\begin{aligned}
T^{-2}\sum_{t=2}^{T} X_p(\widehat{S}_{it-1})^2 &= T^{-2}\sum_{t=2}^{T}\widehat{S}_{it-1}^2 - T^{-2}\left(T^{-1}\sum_{t=2}^{T}\widehat{S}_{it-1}X_{it}'\right) \\
&\cdot \left(T^{-2}\sum_{t=2}^{T}X_{it}X_{it}'\right)^{-1} T^{-1}\sum_{t=2}^{T}X_{it}\widehat{S}_{it-1} \\
&= T^{-2}\sum_{t=2}^{T}\widehat{S}_{it-1}^2 - \left(T^{-3/2}\sum_{t=2}^{T}\widehat{S}_{it-1}\right)^2 + o_p(1) \\
&\Rightarrow \frac{1}{\phi_i(1)^2}\sigma_i^2\int_0^1 V_i(r)^2 \mathrm{d}r - \frac{1}{\phi_i(1)^2}\sigma_i^2\left(\int_0^1 V_i(r)\mathrm{d}r\right)^2 \\
&= \frac{1}{\phi_i(1)^2}\sigma_i^2\int_0^1 U_i(r)^2\mathrm{d}r,
\end{aligned}$$

where the second equality follows from the fact that all terms in X_{it} but the constant are stationary and can be discarded asymptotically.

Finally, consider (d). From Lemma A.1, we have

$$S_{it-1} = v_{it} - v_{i1} - \Delta v_i(t-1) = \frac{1}{\phi_i(1)}\sum_{t=2}^{t}(e_{ij}-e_i),$$

which, together with (a), implies

$$\begin{aligned}
T^{-1}\sum_{t=2}^{T}X_p(\widehat{S}_{it-1})X_p(\Delta\widehat{S}_{it}) &= T^{-1}\sum_{t=2}^{T}X_p(S_{it-1})(e_{it}-e_i) + o_p(1) \\
&= \frac{1}{\phi_i(1)}T^{-1}\sum_{t=2}^{T}\sum_{j=2}^{t}(e_{ij}-e_i)(e_{it}-e_i) + o_p(1) \\
&= \frac{1}{\phi_i(1)}T^{-1}\sum_{t=2}^{T}g_{it-1}\Delta g_{it} + o_p(1), \qquad \text{(A23)}
\end{aligned}$$

where $g_{it} = \sum_{j=2}^{t}(e_{ij}-e_i)$. To obtain the limit of the sum in (A23), we use the formula in (A19), which suggests that

$$\begin{aligned}
T^{-1}\sum_{t=2}^{T}g_{it-1}\Delta g_{it} &= \frac{1}{2}T^{-1}(g_{iT}^2 - g_{i1}^2) - \frac{1}{2}T^{-1}\sum_{t=2}^{T}(\Delta g_{it})^2 \\
&= -\frac{1}{2}T^{-1}\sum_{t=2}^{T}(\Delta g_{it})^2 + o_p(1) \\
&\to_p -\frac{1}{2}\sigma_i^2.
\end{aligned}$$

By inserting this into (A23), we get the required result, thus proving (d). ∎

Proof of Theorem 1.

The least squares estimator of ϕ_i can be written as

$$T\widehat{\phi}_i = \left(T^{-2}\sum_{t=2}^{T} X_p(\widehat{S}_{it-1})^2\right)^{-1} T^{-1}\sum_{t=2}^{T} X_p(\widehat{S}_{it-1})X_p(\Delta\widehat{S}_{it}).$$

Hence, by Lemma A.2, $T\widehat{\phi}_i$ has the following limit as $T \to \infty$

$$T\widehat{\phi}_i \Rightarrow -\left(2\int_0^1 U_i(r)^2 \mathrm{d}r\right)^{-1}.$$

Now, consider the variance ratio $S_i = \widehat{\omega}_i/\widehat{\sigma}_i$. Using the results of Lemma A.2 (c), it follows that $\widehat{\sigma}_i^2$ can be written

$$\begin{aligned}\widehat{\sigma}_i^2 &= T^{-1}\sum_{t=2}^{T}(X_p(\Delta\widehat{S}_{it}) - \widehat{\phi}_i X_p(\widehat{S}_{it-1}))^2 \\ &= T^{-1}\sum_{t=2}^{T} X_p(\Delta\widehat{S}_{it})^2 - 2(T\widehat{\phi})T^{-2}\sum_{t=2}^{T} X_p(\Delta\widehat{S}_{it})X_p(\widehat{S}_{it-1}) \\ &+ (T\widehat{\phi})^2 T^{-3}\sum_{t=2}^{T} X_p(\widehat{S}_{it-1})^2 \\ &= T^{-1}\sum_{t=2}^{T} e_{it}^2 + o_p(1) \to_p \sigma_i^2. \qquad \text{(A24)}\end{aligned}$$

Thus, $\widehat{\sigma}_i^2$ is consistent. Moreover, suppose that $T,\ M_i \to \infty$ with $M_i/T \to 0$, then the consistency of $\widehat{\omega}_i^2$, the estimated long-run variance of Δv_{it} based on $\Delta\widehat{S}_{it}$, is an immediate consequence of the fact that $\Delta\widehat{S}_{it} = \Delta\widehat{v}_{it}$, and $\Delta\widehat{v}_{it}$ is consistent for Δv_{it} as shown in Lemma A.1 (d).

Since $S_i \to_p \omega_i/\sigma_i = \phi_i(1)$, we can therefore show that

$$\phi_N = \frac{1}{N}\sum_{i=1}^{N} T\widehat{\phi}_i S_i \Rightarrow \frac{1}{N}\sum_{i=1}^{N} B_i \to_p E(B_i),$$

where the limit arguments are taken first as $T \to \infty$ and then as $N \to \infty$. The sequential limit distribution of ϕ_N is readily obtained by writing

$$N^{-1/2}\phi_N - \sqrt{N}E(B_i) = \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} T\widehat{\phi}_i S_i - E(B_i)\right).$$

Assume that $\mathrm{var}(B_i) < \infty$ exists. By Assumption 1 through 4, and the Lindberg-Lévy central limit theorem, then as $T \to \infty$ and then $N \to \infty$

$$N^{-1/2}\phi_N - \sqrt{N}E(B_i) \Rightarrow N(0, \mathrm{var}(B_i)).$$

This establishes the first part of the proof.

To obtain the corresponding sequential limit for τ_N, we need to evaluate $SE(\widehat{\phi}_i)$, the standard error of $\widehat{\phi}_i$, which may be written as

$$SE(\widehat{\phi}_i) \; = \; \left(\widehat{\sigma}_i^{-2} \sum_{t=2}^{T} X_p (\widehat{S}_{it-1})^2 \right)^{-1/2} .$$

By using (A24) and Lemma A.2 (b), as $T \to \infty$, we get

$$T^2 SE(\widehat{\phi}_i)^2 \; = \; \left(\widehat{\sigma}_i^{-2} T^{-2} \sum_{t=2}^{T} X_p (\widehat{S}_{it-1})^2 \right)^{-1} \; \Rightarrow \; \phi_i(1)^2 \left(\int_0^1 U_i(r)^2 \mathrm{d}r \right)^{-1} ,$$

which implies that

$$\tau_N \; = \; \frac{1}{N} \sum_{i=1}^{N} \frac{\widehat{\phi}_i}{SE(\widehat{\phi}_i)} \; = \; \frac{1}{N} \sum_{i=1}^{N} \frac{T\widehat{\phi}_i}{\sqrt{T^2 SE(\widehat{\phi}_i)^2}} \; \Rightarrow \; \frac{1}{N} \sum_{i=1}^{N} C_i \; \to_p \; E(C_i).$$

Next, similar to the analysis of ϕ_N, we expand the statistic as

$$\sqrt{N}\tau_N - \sqrt{N} E(C_i) \; = \; \sqrt{N} \left(N^{-1} \sum_{i=1}^{N} \frac{\widehat{\phi}_i}{SE(\widehat{\phi}_i)} - E(C_i) \right) .$$

Again, by assuming that $\mathrm{var}(C_i) < \infty$ exists, we obtain

$$N^{-1/2} \tau_N - \sqrt{N} E(C_i) \; \Rightarrow \; N(0, \mathrm{var}(C_)),$$

as $T \to \infty$ and then $N \to \infty$ by the Lindeberg-Lévy central limit theorem. This completes the second part of the proof. ■

Table 1: Size at the 5% level.

Case	T	θ	ρ	τ_N $(\widehat{\lambda},\widehat{k})$	$(\widehat{\lambda},0)$	$(0,\widehat{k})$	(λ,k)	ϕ_N $(\widehat{\lambda},\widehat{k})$	$(\widehat{\lambda},0)$	$(0,\widehat{k})$	(λ,k)
1	100	0	0	6.7	12.0	6.7	6.7	5.8	34.5	5.8	5.8
		0.3	0	7.0	8.5	7.0	7.0	4.9	19.7	4.9	4.9
		−0.3	0	5.9	15.7	5.9	5.9	12.4	56.9	12.4	12.4
		0	0.3	7.8	10.8	7.8	7.8	3.9	19.3	3.9	3.9
		0	−0.3	6.8	14.2	6.8	6.8	9.8	54.7	9.8	9.8
	200	0	0	6.9	10.5	6.9	6.9	5.9	35.8	5.9	5.9
		0.3	0	7.3	10.2	7.3	7.3	4.5	21.0	4.5	4.5
		−0.3	0	6.1	16.2	6.1	6.1	11.5	65.0	11.5	11.5
		0	0.3	7.0	8.6	7.0	7.0	3.8	17.5	3.8	3.8
		0	−0.3	4.9	15.7	4.9	4.9	8.1	54.2	8.1	8.1
2	100	0	0	7.5	10.3	9.3	8.9	7.0	24.4	7.6	6.8
		0.3	0	8.0	9.0	8.7	8.9	4.6	16.2	5.4	5.3
		−0.3	0	5.6	10.1	6.3	6.0	9.8	39.1	10.7	9.2
		0	0.3	7.3	8.8	8.0	7.6	4.1	14.3	2.9	3.3
		0	−0.3	5.2	8.3	5.3	5.4	8.2	33.4	8.2	8.1
	200	0	0	7.1	10.6	7.4	6.6	6.3	28.4	6.4	6.1
		0.3	0	5.7	6.3	6.7	5.4	2.9	17.1	3.6	2.4
		−0.3	0	6.9	12.3	7.3	6.1	12.0	50.7	12.9	11.7
		0	0.3	6.6	7.6	6.6	6.6	3.6	15.0	3.1	2.1
		0	−0.3	5.6	8.6	5.9	5.1	8.0	42.0	7.7	7.9
3	100	0	0	7.4	11.4	20.3	7.4	6.7	31.3	18.0	6.7
		0.3	0	8.2	9.4	21.0	7.8	4.4	19.8	18.5	4.1
		−0.3	0	6.3	16.6	22.4	6.3	11.0	52.3	20.8	10.6
		0	0.3	9.0	10.9	19.2	8.4	5.6	17.6	17.5	5.6
		0	−0.3	5.7	11.7	24.8	5.4	8.6	45.6	23.7	8.2
	200	0	0	5.7	10.7	19.6	5.9	5.6	33.3	16.7	5.9
		0.3	0	7.0	9.4	19.7	6.9	4.9	20.9	15.4	5.1
		−0.3	0	4.6	13.3	23.6	4.4	10.0	58.7	22.0	9.9
		0	0.3	7.1	9.3	19.4	7.0	4.1	17.1	15.4	3.9
		0	−0.3	5.7	14.7	18.4	5.6	8.6	57.7	17.3	8.7

Notes: The notation (a,b) is used to indicate that the test has been computed using the rules a and b for selecting the breakpoint and number of common factors. The value ρ refers to the autoregressive parameter and θ refers to the moving average parameter. Case 1 refers to the no break model, Case 2 refers to the intercept break model, and Case 3 refers to the model with a break in both intercept and slope.

Table 2: Power at the 5% level when $\phi = -0.03$.

				τ_N				ϕ_N			
Case	T	θ	ρ	$(\widehat{\lambda}, \widehat{k})$	$(\widehat{\lambda}, 0)$	$(0, \widehat{k})$	(λ, k)	$(\widehat{\lambda}, \widehat{k})$	$(\widehat{\lambda}, 0)$	$(0, \widehat{k})$	(λ, k)
1	100	0	0	21.0	22.1	21.0	21.0	20.4	16.9	20.4	20.4
		0.3	0	18.4	19.6	18.4	18.4	20.4	15.9	20.4	20.4
		−0.3	0	20.4	17.0	20.4	20.4	20.0	11.0	20.0	20.0
		0	0.3	23.6	21.9	23.6	23.6	22.6	18.6	22.6	22.6
		0	−0.3	24.0	17.3	24.0	24.0	19.7	12.0	19.7	19.7
	200	0	0	80.6	80.0	80.6	80.6	79.1	72.4	79.1	79.1
		0.3	0	78.7	79.7	78.7	78.7	77.1	73.7	77.1	77.1
		−0.3	0	81.3	78.9	81.3	81.3	79.0	66.1	79.0	79.0
		0	0.3	80.4	79.9	80.4	80.4	77.6	74.4	77.6	77.6
		0	−0.3	83.6	74.3	83.6	83.6	78.4	61.9	78.4	78.4
2	100	0	0	15.6	12.9	16.1	18.6	16.4	10.6	16.3	19.0
		0.3	0	24.1	23.6	23.4	25.6	24.1	18.9	24.3	25.9
		−0.3	0	16.1	13.3	18.7	17.1	13.7	8.4	15.6	14.6
		0	0.3	19.1	18.0	22.9	26.6	21.9	14.1	23.3	26.4
		0	−0.3	23.1	20.1	21.3	25.6	18.0	14.0	19.9	20.4
	200	0	0	71.7	64.9	75.6	81.0	71.6	61.9	71.9	79.9
		0.3	0	77.1	78.7	77.9	83.6	75.7	73.0	76.1	82.0
		−0.3	0	74.1	65.7	76.3	82.7	68.9	57.6	70.3	79.0
		0	0.3	74.3	72.6	79.0	82.4	71.9	66.4	75.7	79.3
		0	−0.3	72.9	72.3	77.0	83.7	67.4	63.9	72.3	80.1
3	100	0	0	21.0	16.1	8.7	21.6	18.1	12.7	8.6	18.9
		0.3	0	20.1	20.0	8.6	20.0	18.0	17.3	7.9	17.9
		−0.3	0	23.3	15.1	4.4	24.3	21.1	10.9	4.6	20.4
		0	0.3	22.1	23.1	7.1	23.1	18.9	15.1	6.9	20.1
		0	−0.3	24.9	15.4	8.6	24.4	20.1	9.0	7.3	20.7
	200	0	0	79.9	76.9	10.9	79.9	74.0	70.6	10.1	75.0
		0.3	0	77.9	79.6	10.6	78.4	74.7	74.7	10.6	74.7
		−0.3	0	83.4	74.0	6.0	84.0	78.4	66.7	6.6	76.7
		0	0.3	80.9	78.0	9.3	81.0	76.6	69.3	8.6	77.6
		0	−0.3	78.6	68.3	9.3	78.7	76.0	58.9	9.3	77.1

Notes: See Table 1 for an explanation of the various features of the table.

Table 3: Power at the 5% level when $\phi = -0.05$.

				τ_N				ϕ_N			
Case	T	θ	ρ	$(\widehat{\lambda}, \widehat{k})$	$(\widehat{\lambda}, 0)$	$(0, \widehat{k})$	(λ, k)	$(\widehat{\lambda}, \widehat{k})$	$(\widehat{\lambda}, 0)$	$(0, \widehat{k})$	(λ, k)
1	100	0	0	55.8	50.5	55.8	55.8	58.1	46.2	58.1	58.1
		0.3	0	53.0	54.3	53.0	53.0	51.6	49.1	51.6	51.6
		−0.3	0	50.1	42.3	50.1	50.1	42.6	30.1	42.6	42.6
		0	0.3	51.3	46.3	51.3	51.3	56.6	45.5	56.6	56.6
		0	−0.3	51.4	45.1	51.4	51.4	50.2	42.3	50.2	50.2
	200	0	0	100.0	100.0	100.0	100.0	99.9	99.6	99.9	99.9
		0.3	0	99.9	99.7	99.9	99.9	99.9	99.1	99.9	99.9
		−0.3	0	99.9	99.7	99.9	99.9	99.9	99.0	99.9	99.9
		0	0.3	99.9	99.9	99.9	99.9	100.0	99.7	100.0	100.0
		0	−0.3	99.9	99.6	99.9	99.9	99.9	98.6	99.9	99.9
2	100	0	0	38.5	38.4	48.8	48.4	41.8	35.6	49.3	50.3
		0.3	0	45.1	40.9	50.8	50.5	47.8	42.5	51.9	51.4
		−0.3	0	39.5	31.3	53.1	50.6	37.2	30.1	50.6	47.1
		0	0.3	45.1	41.3	49.3	50.9	45.2	41.0	53.0	52.4
		0	−0.3	44.9	40.2	55.4	56.0	45.9	36.5	52.9	54.5
	200	0	0	99.0	97.9	100.0	100.0	99.1	97.4	99.7	100.0
		0.3	0	99.7	99.6	99.9	100.0	99.7	99.3	99.3	100.0
		−0.3	0	97.4	95.7	98.4	100.0	97.4	92.9	98.1	99.9
		0	0.3	99.7	99.4	99.7	100.0	99.7	99.6	99.7	100.0
		0	−0.3	99.1	97.0	98.7	99.9	98.7	96.4	99.0	99.6
3	100	0	0	51.7	44.2	10.0	52.4	47.0	41.3	8.5	46.4
		0.3	0	47.2	43.8	10.1	48.6	46.7	45.1	10.1	46.4
		−0.3	0	46.6	38.3	9.9	47.4	47.8	34.8	9.1	49.3
		0	0.3	45.1	46.4	11.5	45.6	43.0	45.0	11.2	42.5
		0	−0.3	51.0	41.5	6.3	51.7	47.4	35.0	6.9	48.6
	200	0	0	99.6	99.7	10.6	99.6	99.6	99.3	11.1	99.6
		0.3	0	99.9	99.7	14.1	99.9	100.0	99.7	13.9	100.0
		−0.3	0	99.4	98.6	6.1	99.7	99.6	97.9	7.3	99.7
		0	0.3	100.0	100.0	16.1	100.0	100.0	99.6	15.6	100.0
		0	−0.3	99.4	98.7	12.4	99.4	99.6	97.3	12.3	99.6

Notes: See Table 1 for an explanation of the various features of the table.

Table 4: Estimated break and number of factors when $\phi = 0$.

Case	T	θ	ρ	Frequency count			Average selection		
				$\widehat{\lambda}$	$\widehat{k}(\widehat{\lambda})$	$\widehat{k}(0)$	$\widehat{\lambda}$	$\widehat{k}(\widehat{\lambda})$	$\widehat{k}(0)$
1	100	0	0	–	100.0	100.0	–	1.0	1.0
		0.3	0	–	100.0	100.0	–	1.0	1.0
		−0.3	0	–	100.0	100.0	–	1.0	1.0
		0	0.3	–	100.0	100.0	–	1.0	1.0
		0	−0.3	–	100.0	100.0	–	1.0	1.0
	200	0	0	–	100.0	100.0	–	1.0	1.0
		0.3	0	–	100.0	100.0	–	1.0	1.0
		−0.3	0	–	100.0	100.0	–	1.0	1.0
		0	0.3	–	100.0	100.0	–	1.0	1.0
		0	−0.3	–	100.0	100.0	–	1.0	1.0
2	100	0	0	70.3	97.4	0.3	50.2	1.0	2.0
		0.3	0	69.5	98.0	0.2	50.1	1.0	2.0
		−0.3	0	68.9	97.8	0.1	50.1	1.0	2.0
		0	0.3	69.7	98.8	0.2	50.0	1.0	2.0
		0	−0.3	69.0	97.9	0.0	50.0	1.0	2.0
	200	0	0	65.6	100.0	20.9	100.2	1.0	1.8
		0.3	0	64.8	100.0	29.7	100.2	1.0	1.7
		−0.3	0	65.6	100.0	27.9	99.6	1.0	1.7
		0	0.3	64.9	99.9	29.7	100.6	1.0	1.7
		0	−0.3	65.0	100.0	30.1	100.2	1.0	1.7
3	100	0	0	93.5	99.1	3.5	50.0	1.0	2.0
		0.3	0	94.2	99.1	4.6	50.0	1.0	2.0
		−0.3	0	93.9	99.1	3.4	50.0	1.0	2.0
		0	0.3	93.8	99.6	4.4	50.0	1.0	2.0
		0	−0.3	93.4	99.7	3.1	50.0	1.0	2.0
	200	0	0	95.5	100.0	1.3	100.0	1.0	2.0
		0.3	0	95.5	99.9	2.1	100.0	1.0	2.0
		−0.3	0	95.6	99.7	2.6	100.0	1.0	2.0
		0	0.3	95.3	100.0	2.3	100.0	1.0	2.0
		0	−0.3	95.4	100.0	1.9	100.0	1.0	2.0

Notes: The value $\widehat{\lambda}$ refers to the estimated breakpoint and $\widehat{k}$ refers to the estimated number of factors. The notation $\widehat{k}(a)$ is used to indicate that the estimated number of factors is based on the rule a for selecting the breakpoint. The frequency counts refer to the correct selection frequencies. See Table 1 for an explanation of the remaining features of the table.

Table 5: Panel stationarity tests of the PPP hypothesis.

	Nominal exchange rate			Relative price level		
Specification	Value	p-value^a	p-value^b	Value	p-value^a	p-value^b
Constant	28.710	0.000	0.015	45.014	0.000	0.000
Constant and trend	16.592	0.000	0.020	30.009	0.000	0.000
Break in constant	5.013	0.000	0.005	13.219	0.000	0.000
Break in constant and trend	11.453	0.000	0.045	11.951	0.000	0.005

Notes: The test is implemented using a maximum of five breaks and the Bartlett kernel with the bandwidth of $4(T/100)^{2/9}$.

^{a}The p-values are for a one-sided test based on the normal distribution.

^{b}The p-values are for a one-sided test based on the bootstrapped distribution. We use 500 boostrap replications.

Table 6: Panel cointegration tests of the PPP hypothesis.

	τ_N		ϕ_N	
Model	Value	p-value	Value	p-value
No break	2.385	0.991	1.026	0.847
Level break	0.271	0.607	−0.267	0.395
Regime shift	0.636	0.738	0.422	0.664

Notes: The test is implemented using the Campbell and Perron (1991) automatic procedure to select the lag length. We use three breaks, which are determined by grid search at the minimum of the sum of squared residuals. The p-values are for a one-sided test based on the normal distribution.

References

Ahn, S. K. (1993). Some Tests for Unit Roots in Autoregressive-Integrated-Moving Average Models with Deterministic Trends. *Biometrica*, Vol. 80, pp. 855-868.

Amsler, C., and J. Lee (1995). An LM Test for a Unit Root in the Presence of a Structural Break. *Econometric Theory*, Vol. 11, pp. 359-368.

Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory*, Vol. 13, pp. 315-352.

Bai, J., and S. Ng (2004). A Panic Attack on Unit Roots and Cointegration. *Econometrica*, Vol. 72, pp. 1127-1177.

Bai, J., and P. Perron (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, Vol. 66, pp. 47-78.

Banerjee, A., and J. L. Carrion-i-Silvestre (2006). Cointegration in Panel Data with Breaks and Cross-Section Dependence. ECB working paper no. 591.

Breitung, J., and M. H. Pesaran (2006). Unit Roots and Cointegration in Panels. Forthcoming in L. Matyas, and P. Sevestre (Eds.), The Econometrics of Panel Data, Kluwer Academic Publishers.

Campbell, J., and P. Perron (1991). Pitfalls and Opportunities: What Macroeconomists should Know about Unit Roots. In Blanchard, O., and S. Fishers (Eds.), NBER Macroeconomics Annual, MIT Press, Cambridge, MA.

Carrion-i-Silvestre, J. L., T. Del Barrio-Castro and E. Lópes-Bazo (2005). Breaking the panels: An application to the GDP per capita. *Econometrics Journal*, Vol. 8, pp. 159-175.

Groen, J. J. J., and F. Kleibergen (2003). Likelihood-based Cointegration Analysis in Panels of Vector Error Correction Models. *Journal of Business and Economic Statistics*, Vol. 21, pp. 295-318.

Harris, D., B. P. M. McCabe and S. J. Leybourne (2005). Panel Stationarity Tests for Purchasing Power Parity with Cross-Sectional Dependence. *Journal of Business and Economic Statistics*, Vol. 23, pp. 395-408.

Haug, A. A. (1996). Tests for Cointegration: A Monte Carlo Comparison. *Journal of Econometrics*, Vol. 71, pp. 89115.

Hegwood, N. D., and D. H. Papell (1998). Quasi Purchasing Power Parity. *International Journal of Finance and Economics*, Vol. 3, pp. 27989.

Larsson, R., J. Lyhagen and M. Lötgren (2001). Likelihood-Based Cointegration Test in Heterogeneous Panels. *Econometrics Journal*, Vol. 4, pp. 109-142.

McCoskey, S., and C. Kao (1998). A Residual-Based Test of the Null of Cointegration in Panel Data. *Econometric Reviews*, Vol. 17, pp. 57-84.

Murray, C. J., and D. H. Papell (2003). Do Panels Help Solving the Purchasing Power Parity Puzzle? *Journal of Business and Economic Statistics*, Vol. 23, pp. 410-415.

O'Connell, P. G. J. (1998). The Overvaluation of Purchasing Power Parity. *Journal of International Economics*, Vol. 44, pp. 119.

Papell, D. H. (2002). The Great Appreciation, the Great Depreciation and the Purchasing Power Parity Hypothesis. *Journal of International Economics*, Vol. 57, pp. 5182.

Pedroni, P. (1999). Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors. *Oxford Bulletin of Economics and Statistics*, Vol. 61, pp. 653-670.

Pedroni, P. (2004). Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis. *Econometric Theory*, Vol. 3, pp. 579-625.

Schmidt, P. and P. C. B. Phillips (1992). LM Tests for a Unit Root in the Presence of Deterministic Trends. *Oxford Bulletin of Economics and Statistics*, Vol. 54, pp. 257-287.

Westerlund, J. (2005a). A Panel CUSUM Test of the Null of Cointegration. *Oxford Bulletin of Economics and Statistics* Vol. 67, pp. 231-262, 2005.

Westerlund, J. (2005b). New Simple Tests for Panel Cointegration. *Econometrics Reviews*, Vol. 24, pp. 297-316.

Westerlund, J. (2006a). Testing for Panel Cointegration with a Level Break. *Economics Letters*, Vol. 91, pp. 27-33.

Westerlund, J. (2006b). Testing for Panel Cointegration with Multiple Structural Breaks. *Oxford Bulletin of Economics and Statistics*, Vol. 68, pp. 101-132.

Westerlund, J., and D. Edgerton (2006). New Improved Tests for Cointegration with Structural Breaks. *Journal of Time Series Analysis*, forthcoming.