Optimal panel unit root testing with covariates

ARTŪRAS JUODIS† AND JOAKIM WESTERLUND‡,§

†Faculty of Economics and Business, University of Groningen, Nettelbosje 2, 9747 AE Groningen, The Netherlands. Email: a.juodis@rug.nl

‡Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden.

§Centre for Financial Econometrics, Deakin Business School, Deakin University, 221 Burwood Highway Burwood, Victoria 3125, Australia. Email: joakim.westerlund@nek.lu.se

First version received: 28 May 2018; final version accepted: 30 July 2018.

Summary: This paper provides asymptotic optimality results for panel unit root tests with covariates by deriving the Gaussian power envelope. The main conclusion is that the use of covariates holds considerable promise in the panel data context, much more so than in the time series context. In fact, the use of the covariates not only leads to increased power, but can actually have an order effect on the shrinking neighbourhoods around unity for which power is non-negligible.

Keywords: panel data, unit root test, Gaussian power envelope, covariates.

JEL codes: C32.

1. INTRODUCTION

Owing to their effects on both the econometric method used and the economic interpretation of the model examined, it is very common to pre-test the data for unit roots. This is typically done by using univariate unit root tests. Yet, in applications it is extremely rare that variables are observed in isolation. In fact, most of the time, the model of interest is a multivariate one, and the purpose of the pre-testing is to determine whether the estimation should be carried out with variables in levels or in differences. Moreover, the variables are correlated, for otherwise there would be little or no need for a multivariate model. Against this background, an obvious question that arises is: What gains can be made by accounting for the multiplicity of variables also in the unit root pre-testing stage? In the time series literature, researchers have been asking themselves this question for many years (see, e.g., Hansen, 1995; Elliott and Jansson, 2003; Jansson, 2004; and Westerlund, 2013). The answer that emerges is: More power! Indeed, as Hansen (1995, p. 1168) writes, ‘power is dramatically improved’ when one variable is tested while accounting for its covariation with other variables, as opposed to when the testing is carried out using only one variable at a time. Intuitively, if there are covariates present, by accounting for their variation we get more precise estimates, and hence more powerful tests. This is made precise in Hansen (1995), and Elliott and Jansson (2003), who study the local asymptotic power envelope for the unit root testing problem with covariates. According to their results, the key parameter driving power is
the correlation between the covariates and the variable being tested; the larger this correlation, the higher is the power envelope. What is more, the gain in power tends to increase as more deterministic trend terms are added, which are known to have a detrimental effect on power. The potential of the covariates is therefore largest in the cases that matter most in practice.

As the discussion in the previous paragraph suggests, the effect of covariates for unit root testing in time series is by now well understood, and there is a rich optimality theory devoted to this issue. Interestingly, while in recent years the problem of testing for unit roots in panels has attracted much attention, little is known regarding the potential gains of accounting for covariates in this context. In fact, the only works known to us are those of Costantini and Lupi (2013), and Westerlund (2015), in which the authors propose panel extensions of already existing time series tests with covariates. However, there is as of yet no optimality theory. In fact, it is only very recently that researchers have started to consider the local power of panel unit root tests. An important contribution in this regard is that of Moon et al. (2007), who derive the local power envelope for tests of a unit root in an univariate panel data model with Gaussian errors. Their main finding is that the neighbourhoods around unity for which the power envelope is defined depend critically on the presence of a linear trend; if the trend is absent, the power envelope is defined within $N^{-1/2}T^{-1}$-neighbourhoods of unity, whereas if the trend is present, the power envelope is defined within $N^{-1/4}T^{-1}$-neighbourhoods. This is the so-called ‘incidental trend problem’, which has been shown to have a substantial effect on the power of univariate panel unit root tests (see Moon et al., 2015 for a recent survey). In fact, the effect can be so detrimental that some authors have even recommended against the use of their tests in the linear trend case (see, e.g., Moon and Perron, 2004, and Moon and Perron, 2008). This begs the question: What power gains can researchers possibly hope to achieve by accounting for covariates in the panel data context?

The purpose of the present study is to provide the answer to the above question. This is done by deriving the local asymptotic power envelope of the panel unit root testing problem with covariates. Section 2 sets the stage by introducing the model and the assumptions maintained in Sections 2 and 4. Four specifications of the deterministic component of the data are considered: (D0) no deterministic component; (D1) a constant in the variable being tested; (D2) a constant in both the variable being tested and the covariates; and (D3) a constant and trend in the variable being tested and a constant in the covariates. But while the deterministic component is quite general, in order to make the analysis of the local asymptotic power envelope manageable, and in keeping with the bulk of the existing literature on local power (see, e.g., Moon et al., 2007; Moon et al., 2014; and Becheri et al. 2015a), some of the other assumptions are quite restrictive. As we demonstrate in Section 5, however, the restrictive conditions are not needed in empirical applications.

Section 3 lays out the hypothesis to be tested and derives the resulting power envelope for each deterministic specification. The envelopes under D0 and D1 are shown to be identical, implying that the presence of a constant in the variable being tested does not affect maximal achievable power. Consistent with the results of Moon et al. (2007) for the case without covariates, the envelope under D0 and D1 is defined within $N^{-1/2}T^{-1}$-neighbourhoods of unity. However, as in the time series case, the use of the covariates creates a dependence on the correlation between the covariates and the variable being tested, with the no-covariate envelope arising naturally in the special case of no correlation. Hence, by increasing the correlation, the envelope with covariates can be made arbitrarily high when compared to the envelope without covariates. Unlike in D1 when the constant is added to the tested variable, the envelope in D2 when said constant is added to the covariates lies strictly below the envelope in D0. However, the envelope with covariates is still at least as high as the envelope without covariates. The results for D3 are materially different.
from those reported by Ploberger and Phillips (2002) and Moon et al. (2007) for the linear trend case without covariates. Specifically, while the envelope without covariates is defined within $N^{-1/4}T^{-1}$-neighbourhoods of unity, provided that there is at least some correlation, the envelope with covariates is defined within $N^{-1/2}T^{-1}$-neighbourhoods, which is the same as in D0–D2. The use of the covariates therefore resolves the incidental trend problem. In fact, similarly to before, by taking the correlation to one, the envelope with covariates can be made arbitrarily high relative to the no-covariate envelope without any deterministic constant or trend terms. The gain in power brought about by the covariates is therefore able to outweigh the loss incurred by the inclusion of the linear trend.

The main contribution of this paper lies with the analysis of the local power envelope in the presence of covariates. However, we also propose a feasible test that is obtained as a by-product of the local power analysis. Section 4 is devoted to this test, whose small-sample properties are investigated my means of Monte Carlo simulations in the supplementary appendix. The results indicate that sizeable power gains are available even when the covariates are only weakly correlated, and that these gains are especially pronounced for D3 with incidental trends included, where tests without covariates do not have any power beyond nominal size.

In our empirical application, presented in Section 5, we consider as an example purchasing power parity (PPP). The main purpose is to illustrate how the feasible test derived under rather restrictive conditions can be readily modified to accommodate more general data-generating processes.

Section 6 concludes. All proofs are provided in the supplementary appendix.

2. MODEL AND ASSUMPTIONS

Consider the $[(1 + m) \times 1]$ vector $z_{i,t}$, observable for $t = 1, \ldots, T$ time periods and $i = 1, \ldots, N$ cross-section units. It is convenient to decompose this variable as $z_{i,t} = (y_{i,t}, x_{i,t}')'$, where $y_{i,t}$ is a scalar variable of main interest and $x_{i,t}$ is an $[m \times 1]$ vector of covariates. The data-generating process of $z_{i,t}$ is assumed to be given by

$$z_{i,t} = D_t \beta_i + u_{i,t},$$

where $D_t \beta_i$ and $u_{i,t} = (u_{yi,t}, u_{xi,t}')'$ represent the deterministic and stochastic components of $z_{i,t}$, respectively. The matrix $D_t$ has dimension $[(m + 2) \times (m + 1)]$ and is given by

$$D_t = \begin{pmatrix} d_{yt} & 0_{2 \times m} \\ 0_{m \times 1} & d_{xt} I_m \end{pmatrix},$$

where $d_{yt} = (1, t)'$ and $d_{xt} = 1$ contain the deterministic terms that enter the equation for $y_{i,t}$ and $x_{i,t}$, respectively. Also, $\beta_i = (\beta_{0yi}, \beta_{1yi}, \beta_{0xi})'$, where $\beta_{0yi}$ is $[m \times 1]$, while $\beta_{0xi}$ and $\beta_{1yi}$ are scalars. Similarly to in Elliott and Jansson (2003), the following deterministic specifications will be considered (where parameters are free unless otherwise stated):

D0. $\beta_i = 0_{(m + 2) \times 1}$ (no constant or trend terms);
D1. $\beta_{1yi} = 0$ and $\beta_{0xi} = 0_{m \times 1}$ (a constant in $y_{i,t}$);
D2. $\beta_{1yi} = 0$ (a constant in both $y_{i,t}$ and $x_{i,t}$);
D3. $\beta_i$ unrestricted (a constant and trend in $y_{i,t}$ and constant in $x_{i,t}$).

For easy reference to each of these specifications, it is useful to write $Dd$, where $d \in \{0, 1, 2, 3\}$. The maintained assumptions that we will be working under are given below, which are

© 2018 Royal Economic Society.
stated in terms of $e_{i,t} = ((1 - \alpha_i L)u_{y_{i,t}, t}, u'_{x_{i,t}, t})'$, where $L$ is the lag operator and $\alpha_i$ is the largest autoregressive root of $u_{y_{i,t}, t}$.

**Assumption 2.1.** $e_{i,t}$ is independently distributed across both $i$ and $t$ with $E(e_{i,t}) = 0 (I + m) \times 1$.

$$E(e_{i,t}e'_{i,t}) = \Sigma_i = \begin{pmatrix} \Sigma_{11i} & \Sigma_{12i} \\ \Sigma_{12i}' & \Sigma_{22i} \end{pmatrix}$$

is a known positive definite matrix, and $E(\|e_{i,t}\|^k) < \infty$ for $k \geq 4$.

**Assumption 2.2.** $u_{y_{i,0}} = \ldots = u_{y_{N,0}} = 0$.

**Assumption 2.3.** $\alpha_i = 1 + c_i N^{-\kappa}T^{-1}$, where $\kappa > 0$, $c_i$ is independently distributed with $E(c_i^k) < \infty$ for all $k \geq 1$ and $E(c_i^0) = 1$. Also, $c_i$ and $e_{i,t}$ are mutually independent.

**Remark 2.1.** The above assumptions are rather restrictive but are standard in this type of optimality work (see, e.g., Elliott et al., 1996; Moon et al., 2007; Moon et al., 2014; and Becheri et al., 2015a). In fact, when compared to studies such as Moon et al. (2007), and Moon et al. (2014), Assumptions 2.1–2.3 are actually quite weak in that $e_{i,t}$ is not required to be cross-section homoskedastic, the support of $c_i$ need not be bounded, and $\kappa$ is not pre-specified a priori. Moreover, some of the assumptions are not needed when considering the unit root null hypothesis, and are there just to facilitate a manageable asymptotic analysis under the local alternative. The assumptions needed in practice are therefore much less restrictive, as we explain in detail in Section 5.

### 3. POWER ENVELOPES

Define $\tilde{\alpha}_i = 1 + \tilde{c}_i N^{-\kappa}T^{-1}$, where $\tilde{c}_i$ and $\tilde{\kappa}$ are tuning parameters to be set by the researcher, the selection of which will be discussed in Section 4. According to the Neyman–Pearson lemma, if $e_{i,t}$ is Gaussian, rejecting for ‘large’ values of the likelihood ratio (LR) test statistic leads to the uniformly most powerful test of the hypotheses of

$$H_0 : c_1 = \ldots = c_N = 0 \quad \text{versus} \quad H_1 : c_i = \tilde{c}_i < 0 \text{ for all } i.$$ 

In order to describe this test statistic, we follow, e.g., Elliott and Jansson (2003), and concentrate the Gaussian log-likelihood function with respect to $\beta_1, \ldots, \beta_N$. Let us therefore define the ‘quasi-differenced’ version of $z_{i,t}$ as $z_{i,t}(\tilde{\alpha}_i) = ((1 - \tilde{\alpha}_i L)y_{i,t}, x'_{i,t})'$ for $t \geq 2$ and $z_{i,1}(\tilde{\alpha}_i) = (y_{i,1}, x'_{i,1})'$. The matrix of deterministic regressors is denoted $D_t(\tilde{\alpha}_i, d)$. As reflected by the dependence on $d$, the exact form of this matrix depends on the deterministic specification. In D3,

$$D_t(\tilde{\alpha}_i, 3) = \begin{pmatrix} d_{yt}(\tilde{\alpha}_i) & 0_{2 \times m} \\ 0_{m \times 1} & d_{x_{1,t}}I_m \end{pmatrix} \quad \text{for } t > 1 \quad \text{and} \quad D_1(\tilde{\alpha}_i, 3) = \begin{pmatrix} d_{y1} & 0_{2 \times m} \\ 0_{m \times 1} & d_{x1}I_m \end{pmatrix},$$

where $d_{yt}(\tilde{\alpha}_i) = (1 - \tilde{\alpha}_i L)d_{yt}$.

The appropriate specifications under D1 and D2 are obtained by simply removing the rows corresponding to the restrictions imposed. In this notation, the

\footnote{Note how $d_{yt}(\tilde{\alpha}_i)$ satisfies the difference equation $d_{yt}(\tilde{\alpha}_i) = \Delta d_{yt} + \tilde{\alpha}_i d_{yt-1}(\tilde{\alpha}_i)$ with initial condition $d_{yt}(\tilde{\alpha}_i) = d_{yt}$. $(1 - \tilde{\alpha}_i L)y_{i,t}$ may be written in the same form.}

© 2018 Royal Economic Society.
generalized least squares (GLS) estimator of $\beta_i$ under D1–D3 is given by
\[
\hat{\beta}_i(\tilde{\alpha}, d) = \left( \sum_{t=1}^{T} D_t(\tilde{\alpha}, d) \Sigma_i^{-1} D_t(\tilde{\alpha}, d)' \right)^{-1} \sum_{t=1}^{T} D_t(\tilde{\alpha}, d) \Sigma_i^{-1} z_{i,t}(\tilde{\alpha}, d).
\]
The associated GLS residual is $\hat{u}_{i,t}(\tilde{\alpha}, d) = z_{i,t}(\tilde{\alpha}) - D_t(\tilde{\alpha}, d)' \hat{\beta}_i(\tilde{\alpha}, d)$. In D0, $D_t(\tilde{\alpha}, 0) = \{ \emptyset \}$ for all $t$, and so $\hat{u}_{i,t}(\tilde{\alpha}, 0) = z_{i,t}(\tilde{\alpha})$. The LR test statistic for testing $H_0$ versus $H_1$ is given by
\[
LR(\bar{\alpha}, d) = \sum_{i=1}^{N} \sum_{t=1}^{T} [\hat{u}_{i,t}(\bar{\alpha}, d)' \Sigma_i^{-1} \hat{u}_{i,t}(\bar{\alpha}, d) - \hat{u}_{i,t}(1, d)' \Sigma_i^{-1} \hat{u}_{i,t}(1, d)],
\]
where $\bar{\alpha} = (\bar{\alpha}_1, \ldots, \bar{\alpha}_N)'$ is $[N \times 1]$. The test statistic that we are going to consider is the following standardized version of $LR(\bar{\alpha}, d)$:
\[
\psi(\bar{\alpha}, d) = \frac{LR(\bar{\alpha}, d) - a(\bar{\alpha}, d)}{b(\bar{\alpha}, d)},
\]
where $a(\bar{\alpha}, d)$ and $b(\bar{\alpha}, d)$ are the asymptotic mean and variance of $LR(\bar{\alpha}, d)$ under the unit root null hypothesis.

The results reported in this paper are based on the following asymptotic expansion of $\psi(\bar{\alpha}, d)$:
\[
\psi(\bar{\alpha}, d) = -\pi(\alpha, \bar{\alpha}, d) + \xi_{NT} + o_P(1),
\]
where $\xi_{NT} \to_d \xi \sim N(0, 1)$ as $N, T \to \infty$, with $\to_d$ signifying convergence in distribution, and
\[
\pi(\alpha, \bar{\alpha}, d) = \frac{\mu(\alpha, \bar{\alpha}, d)}{\sqrt{b(\bar{\alpha}, d)}},
\]
with $\alpha = (\alpha_1, \ldots, \alpha_N)'$ and where $\mu(\alpha, \bar{\alpha}, d)$ is a certain function that depends on $\alpha$, and that is zero under $H_0$. Among the terms that appear on the right-hand side of (3.1), $\pi(\alpha, \bar{\alpha}, d)$ is the only one that depends on $\alpha$. The extent of the local power is therefore determined by this term. The local power envelope is obtained by choosing $\tilde{\alpha}$ so as to maximize $\pi(\alpha, \bar{\alpha}, d)$. Let us therefore define
\[
\bar{\pi}(\alpha, d) = \sup_{\tilde{\alpha} \in \mathbb{R}^N} \pi(\alpha, \tilde{\alpha}, d).
\]
In what follows, this will be referred to as the ‘drift’ of the envelope. Let us further introduce $z_\phi$, which is such that $P(\xi \leq -z_\phi) = \phi$. The expansion in (3.1) implies that the asymptotic local power envelope when $H_0$ is tested against $H_1$ is given by
\[
\Phi\left( \text{plim} \bar{\pi}(\alpha, d) - z_\phi \right),
\]
where $\Phi(\cdot)$ is the cumulative distribution function of $\xi$.

The derivation of $\bar{\pi}(\alpha, d)$ requires us to evaluate $a(\bar{\alpha}, d)$, $b(\bar{\alpha}, d)$, and $\mu(\alpha, \bar{\alpha}, d)$ for each deterministic specification, and in the remainder of this section we do just that. We begin by considering D0 and D1, the results for which are reported in Theorem 3.1. Here and throughout, the multiple correlation coefficient between $(1 - \alpha_i L)u_{i,t}^{(1)}$ and $u_{i,t}^{(1)}$, as defined by
\[
R_i^2 = \Sigma_{1i}^{-1} \Sigma_{2i} \Sigma_{2i}^{-1} \Sigma_{1i},
\]
2 If $\Sigma_{1i} = 0_{(1+m) \times 1}$ (covariates are uninformative), $\Sigma_i$ is block-diagonal and $LR(\bar{\alpha})$ coincides with the test statistic of Moon et al. (2007).

© 2018 Royal Economic Society.
will be used as a measure for the ‘importance’ of the covariates.

**Theorem 3.1** Consider D0 and D1, and suppose that \( \tilde{\kappa} > 0 \), \( \min(\tilde{\kappa} + \kappa, 2\kappa) > 1/2 \) and \( \sqrt{N}/T = o(1) \). Under Assumptions 2.1–2.3,

\[
a(\tilde{\alpha}, d) = N^{1-2\kappa} \frac{1}{2N} \sum_{i=1}^{N} (1 - R_i^2)^{-1} \tilde{c}_i^2,
\]

\[
b(\tilde{\alpha}, d) = 4 \times a(\tilde{\alpha}, d),
\]

\[
\mu(\alpha, \tilde{\alpha}, d) = N^{1-(\kappa+\kappa)} \frac{1}{N} \sum_{i=1}^{N} (1 - R_i^2)^{-1} \tilde{c}_i c_i,
\]

with \( d \in \{0, 1\} \).

The restrictions that \( \tilde{\kappa} > 0 \), \( \min(\tilde{\kappa} + \kappa, 2\kappa) > 1/2 \) and \( \sqrt{N}/T = o(1) \) ensure that the remainder in (3.1) is indeed \( o_P(1) \) under D0 and D1. Interestingly, these are substantially weaker than the restrictions considered by Moon et al. (2007) in their no fixed effects and incidental intercepts cases, which require \( \tilde{\kappa} = \kappa = 1/2 \) and \( N/T = o(1) \). The appropriate formula for \( \pi(\alpha, \tilde{\alpha}, d) \) under the former set of assumptions is given by

\[
\pi(\alpha, \tilde{\alpha}, d) = \frac{u(\alpha, \tilde{\alpha}, d)}{\sqrt{b(\alpha, d)}} = N^{1/2-\kappa} \frac{\frac{1}{N} \sum_{i=1}^{N} (1 - R_i^2)^{-1} \tilde{c}_i c_i}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (1 - R_i^2)^{-1} \tilde{c}_i^2}}
\]

\[
\leq N^{1/2-\kappa} \frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (1 - R_i^2)^{-1} \tilde{c}_i^2}}{\sqrt{2}},
\]

where the last step is due to the Cauchy–Schwarz inequality. This means that the power envelope is obtained by setting \( \tilde{c}_1 = c_i \) for all \( i \) (or even \( \tilde{c}_i = \delta c_i \) for some \( \delta > 0 \)). Note in particular how \( \tilde{c}_1 = \ldots = \tilde{c}_N = \tilde{c} \) is optimal if \( c_1 = \ldots = c_N = c \). Moreover, provided that \( \tilde{\kappa} > 0 \), \( \pi(\alpha, \tilde{\alpha}, d) \) does not depend on \( \tilde{\kappa} \), which is of course very convenient in practice. For power to be truly local in the sense that it should not depend on the sample size, we require \( \kappa = 1/2 \). Direct substitution into the above specification of \( \pi(\alpha, \tilde{\alpha}, d) \) gives

\[
\pi(\alpha, \tilde{\alpha}, d) \leq \sqrt{\frac{1}{N} \sum_{i=1}^{N} (1 - R_i^2)^{-1} \tilde{c}_i^2}.
\]

The power envelope under D0 and D1 is therefore given by (3.2) with

\[
\tilde{\pi}(\alpha, d) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (1 - R_i^2)^{-1} \tilde{c}_i^2}.
\]

The fact that \( \tilde{\pi}(\alpha, d) \) is strictly increasing in \( R_i^2 \) illustrates the power-increasing potential of the covariates. The minimal value is obtained by setting \( R_1^2 = \ldots = R_N^2 = 0 \), in which case \( \tilde{\pi}(\alpha, d) \) reduces to \( \sqrt{\frac{1}{N} \sum_{i=1}^{N} \tilde{c}_i^2}/2 \), which is identically the drift of the no-covariate envelope of Moon et al. (2007).

\[\text{The fact that } \pi(\alpha, \tilde{\alpha}, d) \text{ does not depend on } \tilde{\kappa} \text{ explains why we only have to bound it from below; the larger } \tilde{\kappa} \text{ is, the smaller the effect of } \tilde{c}_i N^{-\kappa} T^{-1} \text{ on } \tilde{\alpha}_i, \text{ which does not affect the results.}\]
Remark 3.1 The fact that the results for D0 and D1 are the same is in agreement with the findings of Moon et al. (2007). The inclusion of a unit-specific intercept in the equation for \( y_{i,t} \) therefore has no effect on the local power envelope.

Remark 3.2 Becheri et al. (2015b) consider a model that is very similar to the one considered here under D1. But while similar, their model differs from ours in at least one important respect, namely, that the regressors are assumed to be cross-section common, i.e., \( x_{i,t} = x_t \) for all \( i \). According to their results, this change makes \( \psi(\bar{\alpha}, 1) \) asymptotically mixed normal, as opposed to the asymptotically Gaussian theory derived in the present paper.

Remark 3.3 It is should be noted that \( \mu(\alpha, \bar{\alpha}, d) \) is not really the true drift in \( LR(\bar{\alpha}, d) \), but just a first-order approximation. As an illustration of the importance of the distinction, consider the case when \( \bar{c}_1 = \ldots = \bar{c}_N = \bar{c} \) and \( c_i \neq 0 \) is heterogeneous such that \( (1 - R_i^2)^{-1} c_i = \mu(\alpha, \bar{\alpha}, d) \) ‘average out’ in the sense that \( \sum_{i=1}^{N} (1 - R_i^2)^{-1} c_i = o_p(N) \). This should not be taken to imply that \( \bar{\pi}(\alpha, d) = 0 \) for all \( \kappa \) and \( \bar{\kappa} \). The reason is that the expression given in Theorem 3.1 for \( \mu(\alpha, \bar{\alpha}, d) \) is only accurate up to a first-order approximation in \( N^{-(\bar{\kappa} + \kappa)} \bar{c}_i c_i \). The appropriate second-order term is in this case given by \( N^{1-(\bar{\kappa} + 2\kappa)} \frac{2}{3} \bar{\kappa} \sum_{i=1}^{N} (1 - R_i^2)^{-1} \bar{c}_i^2 c_i^2 \), which means that while power is negligible for \( \kappa = 1/2 \), this is not the case if \( \kappa = 1/4 \).

Theorem 3.2 Consider D2, and suppose that \( \bar{\kappa} > 0 \), \( \min(\bar{\kappa} + \kappa, 2\kappa) > 1/2 \) and \( N/T = o(1) \). Under Assumptions 2.1–2.3,

\[
a(\bar{\alpha}, 2) = N^{1-2\bar{\kappa}} \frac{1}{2N} \sum_{i=1}^{N} (1 - R_i^2)^{-1} \left( 1 - \frac{2}{3} R_i^2 \right) \bar{c}_i^2,
\]

\[
b(\bar{\alpha}, 2) = 4 \times a(\bar{\alpha}, 2),
\]

\[
\mu(\alpha, \bar{\alpha}, 2) = N^{1-(\bar{\kappa} + \kappa)} \frac{1}{N} \sum_{i=1}^{N} (1 - R_i^2)^{-1} \left( 1 - \frac{2}{3} R_i^2 \right) \bar{c}_i c_i.
\]

The condition placed on \( N/T \) is stronger in Theorem 3.2 than in Theorem 3.1, which is a reflection of the increased approximation error brought about by the estimation of the incidental intercepts in \( x_{i,t} \). Provided that the more restrictive condition is satisfied, however, according to Theorem 3.2,

\[
\pi(\alpha, \bar{\alpha}, 2) = N^{1/2-\kappa} \frac{1}{\sqrt{N}} \frac{1}{\sqrt{2}} \frac{\sum_{i=1}^{N} (1 - R_i^2)^{-1} (1 - \frac{2}{3} R_i^2) \bar{c}_i c_i}{\sqrt{\sum_{i=1}^{N} (1 - R_i^2)^{-1} (1 - \frac{2}{3} R_i^2) \bar{c}_i^2}}.
\]

implying that

\[
\bar{\pi}(\alpha, 2) = \sqrt{\frac{1}{N}} \frac{1}{\sqrt{2}} \frac{\sum_{i=1}^{N} (1 - R_i^2)^{-1} (1 - \frac{2}{3} R_i^2) \bar{c}_i^2}{\sqrt{\sum_{i=1}^{N} (1 - R_i^2)^{-1} (1 - \frac{2}{3} R_i^2) \bar{c}_i^2}},
\]

under \( \kappa = 1/2 \). Since \( (1 - 2R_i^2/3) \leq 1 \), we have that \( \bar{\pi}(\alpha, 2) \leq \bar{\pi}(\alpha, d) \) for \( d \in \{0, 1\} \). The power envelope under D2 is therefore not larger than under D0 and D1. In other words, while adding a unit-specific intercept to the equation for \( y_{i,t} \) has no effect, unless \( R_i^2 = \ldots = R_N^2 = 0 \),

---

4 The presence of cancelling out effects is also the reason why the common point optimal test discussed in the previous remark may fail to attain the power envelope (see Becheri et al., 2016 for a discussion in the absence of covariates).
power is reduced if the same intercept is added to the equation for \( x_{i,t} \). Hence, not only is the requirement on \( N/T \) relatively stronger under D2, which is a restriction under both the null and the alternative hypotheses, but there is also lower power. However, since \((1 - R_t^2)^{-1} (1 - 2R_t^2)/3 = 1 + R_t^2 (1 - R_t^2)^{-1} / 3 \geq 1\), we have that the power envelope is again increasing in \( R_t^2 \) with the minimal value of \( \sqrt{\sum_{i=1}^{N} c_i^2 / 2} \) obtained by setting \( R_1^2 = \ldots = R_N^2 = 0\). Therefore, while lower than in D0 and D1, the power envelope under D2 with covariates is still higher than if the covariates are absent.

The results reported for D0–D2 suggest that the use of the covariates can have a substantial effect on power. Interestingly, the effect under D3 is potentially much larger.

**Theorem 3.3** Consider D3, and suppose that \( N/T = o(1) \), \( \kappa > 1/4 \) and \( \bar{\kappa} > \max(1/6, 1/2 - \kappa) \). Under Assumptions 2.1–2.3,

\[
a(\bar{\alpha}) = N^{1-\kappa} \frac{1}{N} \sum_{i=1}^{N} \bar{c}_i + N^{1-2\kappa} \frac{1}{6N} \sum_{i=1}^{N} \bar{c}_i^2 - N^{1-3\kappa} \frac{1}{12N} \sum_{i=1}^{N} \bar{c}_i^3 R_i^2 (1 - R_t^2)^{-1},
\]

\[
b(\bar{\alpha}) = N^{1-2\kappa} \frac{1}{3N} \sum_{i=1}^{N} R_i^2 (1 - R_t^2)^{-1} \bar{c}_i^2,
\]

\[
\mu(\alpha, \bar{\alpha}) = N^{1-(\bar{\kappa} + \kappa)} \frac{1}{6N} \sum_{i=1}^{N} R_i^2 (1 - R_t^2)^{-1} \bar{c}_i \bar{c}_i.
\]

According to Theorem 3.3,

\[
\pi(\alpha, \bar{\alpha}, 3) = N^{1/2 - \kappa} \frac{1}{6N} \sum_{i=1}^{N} R_i^2 (1 - R_t^2)^{-1} \bar{c}_i \bar{c}_i \leq N^{1/2 - \kappa} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \bar{R}_i^2 (1 - R_t^2)^{-1} \bar{c}_i^2} \leq \sqrt{12}.
\]

implying that the power envelope is again obtained by setting \( \bar{c}_i = c_i \) for all \( i \). The most interesting implication of this, though, is that the envelope is again obtained by setting \( \kappa = 1/2 \), in which case

\[
\bar{\pi}(\alpha, 3) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \bar{R}_i^2 (1 - R_t^2)^{-1} \bar{c}_i^2}.
\]

The fact that the envelope is defined for \( \kappa = 1/2 \), and not for \( \kappa = 1/4 \) as in Ploberger and Phillips (2002), and Moon et al. (2007), is very interesting because it means that the covariates actually resolve the incidental trends problem. Note in particular how \( \bar{\pi}(\alpha, 3) > 0 \) for any \( R_t^2 > 0 \), which means that \( R_t^2 \) does not have to be ‘large’ for the problem to go away. What is more, the envelope is strictly increasing in \( R_t^2 \). Hence, not only do the covariates entail negligible power within a relatively small \( N^{-1/2}T^{-1} \)-neighbourhood around the null, by increasing the value of \( R_t^2 \) they can also push the envelope upwards. This is in stark contrast to the time series case, where the covariates affect the shape of the envelope but not the neighbourhood. The intuition behind this result goes as follows. In time series, the need to detrend the data leads to increased variance and hence reduced power for all known unit root tests. The increase in variance is, however, not unbounded, which means that power is non-negligible within the same shrinking neighbourhood around the null hypothesis had the data not been detrended. Conversely, if there are covariates that are correlated with the innovations driving the variable of interest, as Hansen (1995) points out,
this will lead to decreased variance and increased power. Now, in the type of panels considered here, the trend slopes are incidental parameters whose number grows with $N$. Thus, while the variance contribution of each cross-section unit is finite, the accumulated effect is not. This is why the shrinking neighbourhood of the local power envelope without covariates is $N^{-1/4}T^{-1}$ and not $N^{-1/2}T^{-1}$. But the effect of the covariates accumulates too, and so power within $N^{-1/2}T^{-1}$-neighbourhoods is again possible. To take an example, suppose that $R_i^2 = R_t^2$ for all $i$. As already pointed out, the drift in the constant-only envelope without covariates is given by $\sqrt{\frac{1}{N} \sum_{i=1}^N c_i^2}/2$. Hence, since $\hat{\pi}(\alpha, 3) = \sqrt{R^2(1 - R^2) - \frac{1}{N} \sum_{i=1}^N c_i^2}/12 \geq \sqrt{\frac{1}{N} \sum_{i=1}^N c_i^2}/2$ whenever $R^2 \geq 6/7 \approx 0.86$, the covariates can more than outweigh the loss of power incurred by the incidental trends.

**Remark 3.4** The fact that $\hat{\pi}(\alpha, 3) = 0$ whenever $R_1^2 = \ldots = R_{N}^2 = 0$ is due to the fact that Theorem 3.3 only reports the leading drift term, which is of order $O(N^{1/2 - \epsilon})$. The next term is of order $O(N^{1/2 - 2\kappa})$, which is negligible under our assumption that $\kappa > 1/4$. In the supplementary appendix, we show that if $R_1^2 = \ldots = R_{N}^2 = 0$ and $\kappa = 1/4$, then $\hat{\pi}(\alpha, 3)$ reduces to $\sqrt{\frac{1}{N} \sum_{i=1}^N c_i^2}/(6\sqrt{5})$, the no-covariate drift of Moon et al. (2007). This is due to the fact that when $R_1^2 = \ldots = R_{N}^2 = 0$ the detrending wipes out not only the trend but also the leading drift term.

### 4. A FEASIBLE TEST

It is possible to construct a feasible test that asymptotically attains the power envelope at pre-specified points. In this paper, we follow Elliott and Jansson (2003), and propose a four-step implementation approach.

1. **Compute the first-step estimator** $\hat{\Sigma}_i^{SL}$ of $\Sigma_i$ as $\hat{\Sigma}_i^{SL} = T^{-1} \sum_{t=1}^T \hat{u}_{i,t}^{SL}(1, d)\hat{u}_{i,t}^{SL}(1, d)'$. Here $\hat{u}_{i,t}^{SL}(1, d) = z_{i,t}(1) - D_t(1, d)\hat{\beta}_{i,t}^{SL}(1, d)$ for $d \in \{1, 2, 3\}$ and $\hat{u}_{i,t}^{SL}(1, 0) = z_{i,t}(1)$, where $\hat{\beta}_{i,t}^{SL}(1, d)$ is the ordinary least squares (OLS) estimator in a regression of $z_{i,t}(1)$ onto $D_t(1, d)$.

2. **Compute** $\hat{\Sigma}_i^{S2} = T^{-1} \sum_{t=1}^T \hat{u}_{i,t}^{S2}(1, d)\hat{u}_{i,t}^{S2}(1, d)'$. Here $\hat{u}_{i,t}^{S2}(1, d) = z_{i,t}(1) - D_t(1, d)\hat{\beta}_{i,t}^{S2}(1, d)$ for $d \in \{1, 2, 3\}$ and $\hat{u}_{i,t}^{S2}(1, 0) = z_{i,t}(1)$, where $\hat{\beta}_{i,t}^{S2}(1, d)$ is $\hat{\beta}_{i,t}^{SL}(1, d)$ based on using $\hat{\Sigma}_i^{SL}$ in place of $\Sigma_i$.

3. **The feasible LR statistic is given by**

\[
\hat{L}R(\hat{\alpha}, d) = \sum_{i=1}^N \sum_{t=1}^T \left[ \hat{u}_{i,t}^{S2}(\hat{\alpha}, d)'(\hat{\Sigma}_i^{S2})^{-1}\hat{u}_{i,t}^{S2}(\hat{\alpha}, d) - \hat{u}_{i,t}^{S2}(1, d)'(\hat{\Sigma}_i^{S2})^{-1}\hat{u}_{i,t}^{S2}(1, d) \right],
\]

where $\hat{u}_{i,t}^{S2}(\alpha, d)$ is defined analogous to $\hat{u}_{i,t}^{SL}(1, d)$.

4. **Denote by** $\hat{a}(\hat{\alpha}, d)$ and $\hat{b}(\hat{\alpha}, d)$ the values of $a(\hat{\alpha}, d)$ and $b(\hat{\alpha}, d)$, respectively, with $R_i^2$ replaced by $\hat{R}_i^2 = (\hat{\Sigma}_{i11}^{S2})^{-1}\hat{\Sigma}_{i12}^{S2}(\hat{\Sigma}_{i22}^{S2})^{-1}(\hat{\Sigma}_{i21}^{S2})$, where $\hat{\Sigma}_{i11}^{S2}, \hat{\Sigma}_{i12}^{S2}$, and $\hat{\Sigma}_{i22}^{S2}$ are the submatrices of $\hat{\Sigma}_i^{S2}$, which is partitioned conformably with $\Sigma_i$. The feasible point-
optimal LR statistic is given by

\[ \hat{\psi}(\bar{\alpha}, d) = \frac{\hat{L}R(\bar{\alpha}, d) - \hat{a}(\bar{\alpha}, d)}{\sqrt{\hat{b}(\bar{\alpha}, d)}}. \]

The computation of \( \hat{u}_{S_2i,t}(\bar{\alpha}_i, \bar{\kappa}_i) \) in Step 3 requires a choice of \( \bar{\alpha}_i \) and \( \bar{\kappa}_i \). Because the local power of \( \psi(\bar{\alpha}, d) \) only depends on \( R_{i}^2, \bar{c}_i, c_i, \) and \( \kappa \), the value of \( \bar{\kappa} \) is not expected to make much difference, and we have unreported Monte Carlo results that confirm this. With \( c_i \) unknown, there is little or no reason for choosing one value of \( \bar{c}_i \) over another. Moon et al. (2007) use Monte Carlo simulations to investigate the finite-sample performance of their feasible test under various specifications of \( \bar{c}_i \). According to their results, \( \bar{c}_i = -1 \) leads to the best overall performance. In Section 5, we therefore set \( \bar{\kappa} = 1/4 \) and \( \bar{c}_i = -1 \) (see also the Monte Carlo study in the supplementary appendix).

The only difference between \( \hat{\psi}(\bar{\alpha}, d) \) and \( \psi(\bar{\alpha}, d) \) is that \( \hat{\Sigma}^{S_1}_{i} \) is estimated in the former test statistic. But \( \hat{\Sigma}^{S_1}_{i} \) and \( \hat{\Sigma}^{S_2}_{i} \) are consistent, which means that \( \hat{\psi}(\bar{\alpha}, d) \) is asymptotically equivalent to \( \psi(\bar{\alpha}, d) \). Theorem 4.1 summarizes this.

**Theorem 4.1** Suppose that \( N/T = o(1) \), and that \( \kappa \) and \( \bar{\kappa} \) satisfy the conditions of Theorems 3.1–3.3. Then, under Assumptions 2.1–2.3,

\[ \hat{\psi}(\bar{\alpha}, d) = \psi(\bar{\alpha}, d) + o_P(1). \]

**Remark 4.1** Because \( \hat{\Sigma}^{S_1}_{i} \) and \( \hat{\Sigma}^{S_2}_{i} \) are both consistent, asymptotically it does not matter which one is being used. Step 2 is therefore not necessary for Theorem 4.1 to hold. This explains why in Moon et al. (2007, 2014) there is no need to compute \( \hat{\Sigma}^{S_2}_{i} \). The reason for using \( \hat{\Sigma}^{S_2}_{i} \) rather than \( \hat{\Sigma}^{S_1}_{i} \) is therefore only a matter of small-sample performance.

**Remark 4.2** Theorem 3.1 only requires that \( \sqrt{N/T} = o(1) \). The more restrictive condition in Theorem 4.1 that \( N/T = o(1) \) in D0 and D1 is due to the estimation of \( \Sigma_i \), leading to increased approximation error.

The results reported in Section 4 and Theorem 4.1 imply that under Assumptions 2.1–2.3, and the relevant conditions on \( N/T \) and \( \bar{\kappa} \),

\[ \hat{\psi}(\bar{\alpha}, d) = \xi_{NT} + o_P(1). \]

The asymptotic null distribution of \( \hat{\psi}(\bar{\alpha}, d) \) is therefore given by \( N(0, 1) \). Corollary 1 formalizes this.

**Corollary 4.1** Suppose that \( H_0 \) is true, and that the conditions of Theorems 3.1–3.3 and 4.1 are met. Then,

\[ \hat{\psi}(\bar{\alpha}, d) \xrightarrow{d} N(0, 1). \]

The choice of critical region is dictated by the sign of the leading term in \( \mu(\alpha, \bar{\alpha}, d) \), which determines the extent of local power under the alternative. The sign of this term is determined by the sign of \( c_i \bar{c}_i \). In Section 5, we set \( \bar{c}_i = \bar{c} = -1 \), which means that under our formulation \( H_1 \) with \( c_i < 0 \) for all \( i \), we have \( c_i \bar{c}_i \geq 0 \). In this case, the test should be set up as one-sided, using only the right tail of the \( N(0, 1) \) distribution to reject the null. Of course, in practice we never really know if in fact \( c_i < 0 \) for all \( i \), and it is not unlikely that some of \( c_1, ..., c_N \) are positive. Hence, unless one is willing to make (potentially false) assumptions regarding these parameters,
we recommend using a two-tailed test, and to reject whenever \(|\hat{\psi}(\alpha, d)| > z_{2\phi}\), where \(z_{2\phi}\) is as before.

5. EMPIRICAL APPLICATION

In this section, we revisit the exchange rate data set of Amara and Papell (2006), Chang and Song (2009), and Costantini and Lupi (2013), which comprises U.S. dollar exchange rates for twenty industrialized countries for the period 1973Q1–1998Q4. The data are taken from the International Monetary Fund’s International Financial Statistics database, and include both nominal and real rates, where the latter has been deflated using the consumer prices index. Both rates are measured in logs. The purpose here is the same as in, e.g., Costantini and Lupi (2013), and Elliott and Pesavento (2006), which is to test for a unit root in the real exchange rate \((y_{i,t})\) using the first difference of the nominal exchange rate as covariate \((x_{i,t})\).

Panel unit root tests have become very popular in the purchasing power parity (PPP) literature. The main reason for this is the increased number of observations that can be brought to bear on the PPP hypothesis, leading to tests with relatively high power. It is therefore not very surprising to find that many earlier studies were in fact able to reject the unit root, or no PPP, null hypothesis. However, it has since become clear that the cross-section independence assumption underlying many of the tests employed is likely to be violated, and that some of the rejections might therefore be due to size distortions (O’Connell, 1998). Consistent with this story, the evidence provided by more recent studies is not very supportive of PPP (see, e.g., Harris et al., 2005, and the references provided therein). One such study is that of Chang and Song (2009), who apply a battery of cross-section correlation robust panel unit root tests. The conclusion is that once the cross-correlation is taken into account, the evidence of PPP disappears.

Of course, robustness is not for free, but comes at a cost in terms of precision. Thus, while correctly sized, robust tests are expected to suffer from lower power when compared with non-robust tests. This means that a non-rejection of the unit root null need not be interpreted as providing evidence against PPP, but could just as well be due to a lack of power. This leaves us with an intricate dilemma: we would like to be able to exploit the power that becomes available when using tests that impose cross-section independence; however, when we do this we run the risk of obtaining spurious results due to omitted cross-correlation.

The \(\hat{\psi}(\alpha, d)\) test considered in Section 4 is very powerful, but is based on a number of restrictive assumptions that are questionable in empirical work. The main purpose of this section is therefore to illustrate how to modify the test so as to improve upon its applicability. The main issues are: (i) nonzero initialization, (ii) serial correlation, and (iii) cross-section dependence, which are all ruled out by Assumptions 2.1 and 2.2. All other implementation issues (including the appropriate values of \(\bar{\kappa}\) and \(\bar{c}_i\) to consider) are dealt with as in the Monte Carlo study presented in the supplementary appendix. The only exception is that since in applications the data-generating process is unknown, following the recommendation of Section 4, in this section the test is set up as two-sided.

The assumption of zero initial value (Assumption 2.2) is critical under D0, but is irrelevant under D1–D3 when incidental intercepts (and trends) are fitted to \(y_{i,t}\), provided that \(u_{yi,0} = O_P(1)\).\(^6\) In this section, we follow Costantini and Lupi (2013) and focus on the results for D1.

\(^5\) We refer to Chang and Song (2009, Section 5) for a more detailed description of the data.

\(^6\) If \(u_{yi,0}\) is ‘large’ in the sense that it is growing with the size of the sample, then Corollary 1 might not be valid anymore. This will be the case if \(u_{yi,0}\) is made a draw from \(N(0, (1 - \alpha_i^2)^{-1})\), which under Gaussianity is identically the

© 2018 Royal Economic Society.
The results for D2 (see, e.g., Amara and Papell, 2006, and Elliott and Pesavento, 2006) were very similar and are available upon request.

The presence of serially correlated errors is treated as in, e.g., Elliott and Jansson (2003), and Moon et al. (2007), which means that all contemporaneous covariance matrix estimators are replaced by their heteroskedasticity and autocorrelation consistent (HAC) counterparts. For example, instead of $\hat{\Sigma} \cdot \hat{\Sigma}^{1}$ in Step 1 of the step-wise implementation approach of Section 4, we use $\hat{\Sigma} \cdot \hat{\Sigma}^{1} = \hat{\Sigma} \cdot \hat{\Sigma}^{1} + (\hat{\Sigma} \cdot \hat{\Sigma}^{1}')$, where $(\hat{\Sigma} \cdot \hat{\Sigma}^{1}) = \sum_{t=1}^{T} H^{(1)}(\cdot)$ is a kernel function, and $M > 0$ is the associated kernel bandwidth parameter. In this section, we use the Bartlett kernel with $M$ set equal to the integer part of $4(7100)^{2/9}$, as recommended by Newey and West (1994). As Moon et al. (2014) show, provided that their linear process assumption is met, the local power envelope is unaffected by the HAC modification to account for serial correlation.

A common approach to accommodate cross-sectional dependence is to assume that the data admits to a common factor representation (see, e.g., Bai and Ng, 2004). In our case, this means that (2.1) should be replaced with

$$z_{i,t} = D_{i} \beta_{i} + F_{i} \lambda_{i} + u_{i,t}, \tag{5.1}$$

where $D_{i}$ and $u_{i,t}$ are as before, $F_{i}$ is a vector of common factors, and $\lambda_{i}$ is the associated vector of factor loadings. As Pedroni et al. (2015) explain, the common factor representation in (5.1) imposes only mild restrictions on the types of cross-sectional dependencies that can be permitted, and we will therefore use it here in our application. Another equally important reason for considering (5.1) is the common observation that the evidence on PPP is to a large extent driven by a bubble-like episode in the 1980s, when there was a large unexplained appreciation of the U.S. dollar, followed by an equally large offsetting depreciation (see Papell, 2002 for a detailed discussion). This rise and fall of the dollar had a substantial effect on all U.S. dollar-based exchange rates. Papell (2002), and Harris et al. (2005) try to account for this effect by treating it as a structural break in the deterministic component of $y_{i,t}$. Specifically, a model with a break in constant and linear trend is assumed, which is shown to capture some of the observed exchange rate behaviour. The approximation is, however, far from perfect, and the deterministic modelling approach is ad hoc.

In the present application, we treat the rise and fall of the dollar as a part of $F_{i}$, which can be estimated from the data. We then ‘defactor’ $z_{i,t}$, and apply $\tilde{\psi}(\alpha, 1)$ to the resulting estimated idiosyncratic errors. Two approaches will be employed, which differ only in how the factors are estimated. Specifically, while the PANIC approach of Bai and Ng (2004) is based on estimated principal component factors, the PANICCA approach of Reese and Westerlund (2016) instead takes the cross-sectional average of the observables as the estimated factors. In order to avoid spurious results when $z_{i,t}$ in non-stationary, the estimation is carried out using the first-differenced data, and the estimated first-differenced idiosyncratic errors are then accumulated up to levels. This gives $\hat{u}_{i,t}$, which is consistent for $u_{i,t}$ under very general conditions (see Bai and Ng, 2004, and Reese and Westerlund, 2016). Testing $\hat{u}_{i,t}$ is therefore asymptotically equivalent to testing $u_{i,t}$, which motivates our approach. The estimated level factors (obtained by accumulating the estimated first-differences) are similarly consistent for (the space spanned by) $F_{i}$.

unconditional distribution when $c_{i} < 0$ for all $i$ (see Moon et al., 2007 for an illustration). The reason is that under this assumption, $u_{i,t} = O(N^{\kappa/2}T^{1/2})$, which under $\kappa > 0$ means that the initial value is of larger order in magnitude than the observed sample. However, as Moon et al. (2007) argue, this assumption is not very realistic in the sense that $u_{i,t}$ is growing with $N$. © 2018 Royal Economic Society.
Figure 1. Plot of the estimated real exchange rate factors over time.

Unreported evidence suggests that the first principal components factor explains most of the common variation of $z_{i,t}$ (50–90%), and that the contribution of additional factors is only marginal. We therefore estimate one factor from both $y_{i,t}$ and $x_{i,t}$. The resulting estimated real exchange rate factor is plotted in Figure 1 for both PANIC and PANICCA. We see that the lines representing the two factor estimates almost coincide, and that there is a marked bump midway through the sample, which coincides with the rise and fall of the dollar in the 1980s. We also see that, except for the bump, both factor estimates seem to be hovering around a constant mean. We therefore interpret our estimated factors as representing the rise and fall of the dollar.

In what remains, we test for a unit root in the defactored exchange rate data, as explained earlier. Because the HAC modification to account for serial correlation does not affect the envelope and because of the consistency of the estimated idiosyncratic errors, the modified $\hat{\psi}(\bar{\alpha}, 1)$ statistic considered here is expected to retain the optimality properties of the original test statistic. We begin by considering some results obtained when applying the serial correlation robust version of the $\hat{V}_{fe1,nT}(C)$ test of Moon et al. (2007) (see Moon et al., 2014 for details) to the first-differenced nominal exchange rate. This is relevant, because Assumption 2.1 requires that the stochastic component of $x_{i,t}$ is stationary. The results reported in Table 1 suggest that the unit root null hypothesis can be rejected, which is just as expected given the results of Costantini and Lupi (2013), and Elliott and Pesavento (2006). We take this as evidence in favour of the stationary covariate requirement. Looking next at the results for the real exchange rate, we see that while strongly rejected by $\hat{\psi}(\bar{\alpha}, 1)$, the evidence based on the $\hat{V}_{fe1,nT}(C)$ test is only significant at the
Table 1. Empirical LR unit root test values.

<table>
<thead>
<tr>
<th>Test</th>
<th>Approach</th>
<th>Statistic</th>
<th>p-value</th>
<th>Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{V}_{fe1,nT}(C)$</td>
<td>PANICCA</td>
<td>−2.236</td>
<td>0.025</td>
<td>−3.767</td>
<td>0.000</td>
</tr>
<tr>
<td>$\hat{V}_{fe1,nT}(C)$</td>
<td>PANIC</td>
<td>−2.059</td>
<td>0.040</td>
<td>−3.772</td>
<td>0.000</td>
</tr>
<tr>
<td>$\hat{\psi}(\bar{\alpha}, 1)$</td>
<td>PANICCA</td>
<td>−5.308</td>
<td>0.000</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>$\hat{\psi}(\bar{\alpha}, 1)$</td>
<td>PANIC</td>
<td>−4.896</td>
<td>0.000</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

Notes. “FD nominal rate” refers to the first-differenced nominal exchange rate, which is used as covariate. $\hat{V}_{fe1,nT}(C)$ refers to the normalized feasible LR test statistic with (without) covariates and $t_i = −1$. See the text for an explanation of the PANIC and PANICCA approaches.

5% level (but not at the 1% level). The evidence against the unit root null therefore increases quite substantially when the covariate is included in the test.

In order to shed some light on the importance of the covariate and the extent to which the relatively strong evidence against the null can be attributed to higher power, we computed $R^2_i$ for each exchange rate in our sample. The values for PANICCA (PANIC) ranges between 0.03 and 0.83 (0.01 and 0.82) with an average of 0.37 (0.35). The average $R^2_i$ is therefore higher than the largest value considered for $R^2_i$ in the Monte Carlo simulations of the supplementary appendix. In view of this and the simulation results reported for D1, $\hat{\psi}(\bar{\alpha}, 1)$ is expected to be substantially more powerful than $\hat{V}_{fe1,nT}(C)$, which is consistent with our empirical results.

The finding that the real exchange rate is stationary after controlling for the rise and fall of the dollar stands in sharp contrast to the results reported by Papell (2002), and Harris et al. (2005), who find little or no such evidence. One possible explanation for this difference in the results is that the deterministic approach of Papell (2002), and Harris et al. (2005) is too rough, and that the factor approach considered here is better at capturing the rise and fall of the dollar, leading to more powerful unit root tests. But if this is indeed the case, $\hat{V}_{fe1,nT}(C)$ should be able to reject, which it is not. The treatment of the rise and fall of the dollar is therefore not the whole story. Indeed, since this is the only difference between $\hat{\psi}(\bar{\alpha}, 1)$ and $\hat{V}_{fe1,nT}(C)$, the inclusion of the covariate is clearly an important contributing factor. The factor treatment of the rise and fall of the dollar alone is therefore not enough to unlock the evidence in favor of a stationary real exchange rate, but there is a need to account for the information contained in the first difference of the nominal exchange rate.

6. CONCLUSION

Economic variables are correlated. This information has been shown to be extremely valuable when testing for a unit root in the time series context (see, e.g., Elliott and Jansson, 2003). In this paper, we show that in the panel data context even greater gains are possible. Our main findings can be summarized as follows. First, consistent with the results of Moon et al. (2007) for the case without covariates, the envelope under D0 and D1 is defined within $N^{-1/2}T^{-1}$-neighbourhoods of unity. However, unlike the envelope without covariates, which is flat in $R^2_i$, the envelope with covariates is increasing in $R^2_i$. Hence, by making $R^2_i$ larger, the envelope with covariates can be made arbitrarily high when compared with the envelope without covariates. Second, unlike in D1 when the constant is added to the variable being tested, the envelope is reduced in D2 when said constant is added to the covariates. However, the envelope with covariates is still at least as high as the envelope without covariates. Third, in D3 the use of correlated covariates not only leads to increased power, but also affects the shrinking neighbourhood around unity in which power is non-negligible.

© 2018 Royal Economic Society.
ACKNOWLEDGEMENTS

Previous versions of this paper were presented at the 2016 Amsterdam Panel Data Workshop, NY Camp Econometrics XI, University of Groningen, IHS in Vienna, the 2016 NESG in Leuven, the 22nd International Panel Data Conference in Perth, and the 9th York Econometrics Symposium in 2016. The authors would like to thank conference and seminar participants, and in particular Peter Boswijk, Kees Jan van Garderen, Michael Jansson (Editor), Benoit Perron, Hashem Pesaran, Werner Ploberger, Eric Renault, and two anonymous referees for helpful comments and suggestions. Part of this paper was written while the first author enjoyed the hospitality of the Department of Economics at Lund University with the financial support of the C. Willems Stichting and the NWO MaGW and Veni grants. The second author would like to thank the Knut and Alice Wallenberg Foundation for financial support through a Wallenberg Academy Fellowship, and the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number P2014–0112:1.

REFERENCES


**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online version of this article at the publisher’s website:

Online Appendix
Replication Package

*Co-editor Michael Jansson handled this manuscript.*