

# Panel Cointegration Testing in the Presence of Linear Time Trends\*

Uwe Hassler<sup>†</sup> and Mehdi Hosseinkouchack  
Goethe University Frankfurt

July 22, 2016

## Abstract

We consider a class of panel tests covering tests for the null hypothesis of no cointegration as well as cointegration. All tests under investigation rely on single-equations estimated by least squares, and they may be residual-based or not. We focus on test statistics computed from regressions with intercept only (i.e. without detrending) while at least one of the regressors (integrated of order 1) is dominated by a linear time trend. In such a setting, often encountered in practice, the limiting distributions and critical values provided for and applied with the situation “with intercept only” are not correct. It is demonstrated that their usage results in size distortions growing with the panel size  $N$ . Moreover, we show which are the appropriate distributions, and how correct critical values can be obtained from the literature.

**Keywords** Single-equations; large  $N$  asymptotics; integrated series with drift

**JEL** C22; C23

## 1 Introduction

Most panel tests for the null hypothesis of (no) cointegration rely on single-equations, notable exceptions being Larsson, Lyhagen, and Löthgren (2001), Groen and Kleibergen

---

\*We thank Christoph Hanck and Joakim Westerlund for helpful comments. Moreover, comments by Katarina Juselius and three anonymous referees that improved the paper are gratefully acknowledged.

<sup>†</sup>Corresponding author: RuW Building, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany, tel: +49-69-798-34762. Email: hassler@wiwi.uni-frankfurt.de

(2003), Breitung (2005) and Karaman Örsal and Droge (2014) who proposed panel system approaches. These system approaches, however, rely on the assumption of independence across the panel. Recent single-equation tests by Chang and Nguyen (2012) or Demetrescu, Hanck, and Tarcolea (2014) rely on nonlinear instrumental variable estimation, while the vast majority of such panel tests builds on (fully modified or dynamic) least squares [LS]. Here, we study exactly this class of LS based single-equation panel tests for the null of either cointegration or no cointegration.

We focus on the situation where the test statistics are computed from regressions with an intercept only (i.e. without detrending) while at least one of the integrated regressors displays a linear time trend on top of the stochastic trend. Such a constellation is often met in practical applications, see for instance Coe and Helpman (1995) and Westerlund (2005b) on R&D spillovers (total factor productivity and capital stock), Larsson, Lyhagen, and Löthgren (2001) on log real consumption and income (per capita), or Hanck (2009) on prices and exchange rate series testing the weak purchasing power parity (PPP). The relevance of a linear trend in panel data has been addressed in Hansen and King (1998) when commenting on the link between health care expenditure and GDP, see McCoskey and Selden (1998); consequently, Blomqvist and Carter (1997), Gerdtham and Löthgren (2000) or Westerlund (2007) worked (partly) with detrended series, i.e. they included time as an explanatory variable in their panel regressions. Hansen (1992, p. 103), however, argues that “It seems reasonable that excess detrending will reduce the test’s power.” Therefore, we study the empirically relevant case where test statistics are computed from regressions with intercept only (i.e. without detrending) while at least one of the I(1) regressors displays a linear time trend, i.e. it is I(1) with drift.

Before becoming more technical, we want to explain our findings as a rule for empirical application. Let  $\bar{Z}^{(m)}$  denote a generic panel cointegration statistic computed from a regression with intercept only involving  $m = k + 1$  I(1) variables. The least squares regression may be static in levels,

$$y_{i,t} = \bar{\alpha}_i + \bar{\beta}_{i,1}x_{i,1,t} + \cdots + \bar{\beta}_{i,k}x_{i,k,t} + \bar{u}_{i,t}, \quad t = 1, \dots, T, \quad i = 1, \dots, N,$$

where  $\{u_{i,t}\}$  may be I(1) in the case of no cointegration, or  $\{u_{i,t}\}$  is I(0) under the null hypothesis of cointegration, see Remark 1 and 3 below, respectively. Alternatively,  $\bar{Z}^{(m)}$  may be from the error-correction regression,

$$\Delta y_{i,t} = \bar{\kappa}_i + \bar{\gamma}_i y_{i,t-1} + \bar{\theta}_{i,1}x_{i,1,t-1} + \cdots + \bar{\theta}_{i,k}x_{i,k,t-1} + \bar{\varepsilon}_{i,t}, \quad t = 1, \dots, T, \quad i = 1, \dots, N,$$

where contemporaneous differences  $\Delta x_{i,t}$  or additional lags of differences may be required as additional regressors to render  $\bar{\varepsilon}_{i,t}$  free of serial correlation, see Remark 2 below. The

test statistic may be constructed from pooling the data or from averaging individual statistics, see e.g. Pedroni (1999, 2004) or Westerlund (2007). Much of the nonstationary panel literature relies on sequential limit theory where  $T \rightarrow \infty$  is followed by  $N \rightarrow \infty$ , such that limiting normality can be established under the assumption that none of the I(1) regressors follows a deterministic time trend:

$$\sqrt{N} (\bar{Z}^{(m)} - \bar{\mu}_m) / \bar{\sigma}_m \sim \mathcal{N}(0, 1).$$

The constants  $\bar{\mu}_m$  and  $\bar{\sigma}_m$  required for appropriate normalization are typically tabulated for a selected number of values of  $m$ , see again Remark 1 through 3. A different set of such moments  $\tilde{\mu}_m$  and  $\tilde{\sigma}_m$  is typically given for detrended regressions, too, where the test statistic  $\tilde{Z}^{(m)}$  stems from regressions of the type ( $m = k + 1$ )

$$y_{i,t} = \tilde{\alpha}_i + \tilde{\delta}_i t + \tilde{\beta}_{i,1} x_{i,1,t} + \cdots + \tilde{\beta}_{i,k} x_{i,k,t} + \tilde{u}_{i,t},$$

or

$$\Delta y_{i,t} = \tilde{\kappa}_i + \tilde{\psi}_i t + \tilde{\gamma}_i y_{i,t-1} + \tilde{\theta}_{i,1} x_{i,1,t-1} + \cdots + \tilde{\theta}_{i,k} x_{i,k,t-1} + \tilde{\varepsilon}_{i,t}.$$

For generic  $\tilde{Z}^{(m)}$  from e.g. the tests mentioned in Remark 1 through 3 it holds irrespective of an eventual linear trend in the data that

$$\sqrt{N} (\tilde{Z}^{(m)} - \tilde{\mu}_m) / \tilde{\sigma}_m \sim \mathcal{N}(0, 1).$$

Our main contribution is twofold, where we assume that at least one of the I(1) regressors has a linear time trend, while the regressions are run with intercept only (without detrending). First, it is shown that the normalization with  $\bar{\mu}_m$  and  $\bar{\sigma}_m$  and the resulting critical values for  $\bar{Z}^{(m)}$  from the regression “with intercept only” are not correct in the presence of linear time trends in the data. It is analytically (Proposition 1) and numerically demonstrated that their usage results in size distortions growing with the panel size  $N$ . Second, we characterize the appropriate limiting distributions by showing that normalization of  $\bar{Z}^{(m)}$  with  $\tilde{\mu}_{m-1}$  and  $\tilde{\sigma}_{m-1}$  results in a standard normal limit, such that the size of the tests can be controlled (Theorem 1). Put differently, Theorem 1 means in non-technical terms: The limiting distribution arising from a regression on  $k$  I(1) variables with drift and on an intercept only amounts to the limiting distribution in the case of a regression on  $k - 1$  I(1) variables and an intercept plus a linear time trend. Such a rule is known in a pure time series context for the special case of the residual-based Phillips-Ouliaris test for no cointegration since Hansen (1992, p. 103): “[...] deterministic trends in the data affect the limiting distribution of the test statistics *whether or not we detrend the data*”; see also the expositions in Hamilton (1994, p. 596, 597) and Hassler (2016, Prop.

16.6). It is even more relevant in our panel framework since we illustrate numerically and analytically that the size distortions of an inappropriate normalization grow with the panel size  $N$  (to zero or one, depending on the specific test). Moreover, we compare our proposal to account for linear trends in the data with the more traditional method of detrending the regression. By simulation we show that power gains of our new strategy according to Theorem 1 over detrending may be considerable. We hence recommend this strategy as being superior to detrending.

The rest of the paper is organized as follows. The next section fixes some notation and assumptions. Section 3 establishes and discusses our asymptotic results and illustrates them with numerical evidence. It also compares our suggestion to account for linear trends with the conventional method of detrending. The last section discusses consequences for applied work. Mathematical proofs are relegated to the Appendix.

## 2 Notation and assumptions

Restricting our attention to the single-equation framework we partition the  $m$ -vector  $z_{i,t}$  of observables into a scalar  $y_{i,t}$  and a  $k$ -element vector  $x_{i,t}$ ,  $z'_{i,t} = (y_{i,t}, x'_{i,t})$ ,  $m = k + 1$ . As usual, the index  $i$  stands for the cross-section,  $i = 1, \dots, N$ , while  $t$  denotes time,  $t = 1, \dots, T$ . Each sequence  $\{z_{i,t}\}$ ,  $t = 1, \dots, T$ , is assumed to be integrated of order 1,  $I(1)$ , where we allow for a non-zero drift, and assume for simplicity a negligible starting value,  $z_{i,0} = 0$ . While  $\{z_{i,t}\}$  may be cointegrated or not, depending on the respective null hypothesis, we rule out cointegration among  $\{x_{i,t}\}$ . Technically, these assumptions translate as follows, where  $W_{i,m}(\cdot)$  denotes an  $m$ -dimensional standard Wiener process,  $[x]$  stands for the integer part of a number  $x$ , and  $\Rightarrow$  is the symbol for weak convergence.

**Assumption 1** *With obvious partitioning according to  $(y_{i,t}, x'_{i,t})'$  we assume ( $i = 1, \dots, N$ )*

$$z_{i,t} = \mu_i t + \sum_{j=1}^t e_{i,j} = \begin{pmatrix} \mu_{i,y} \\ \mu_{i,x} \end{pmatrix} t + \sum_{j=1}^t \begin{pmatrix} e_{i,y,j} \\ e_{i,x,j} \end{pmatrix}, \quad t = 1, \dots, T.$$

*The stochastic zero mean process  $\{e_{i,t}\}$  is integrated of order 0 in that it satisfies*

$$T^{-0.5} \sum_{t=1}^{\lfloor rT \rfloor} e_{i,t} \Rightarrow \Omega_i^{0.5} W_{i,m}(r) = \Omega_i^{0.5} \begin{pmatrix} W_{i,y}(r) \\ W_{i,x}(r) \end{pmatrix}, \quad r \in [0, 1],$$

*with*

$$\Omega_i = \begin{pmatrix} \omega_{i,yy}^2 & \omega'_{i,xy} \\ \omega_{i,xy} & \Omega_{i,xx} \end{pmatrix}$$

where  $\omega_{i,yy}^2 > 0$  and  $\Omega_{i,xx}$  is positive definite.

Now, we turn to assumptions with respect to the tests. Let  $\bar{S}_i^{(m)}$  and  $\tilde{S}_i^{(m)}$  stand again for generic test statistics computed from individual single-equation least squares regressions with “intercept only” and “intercept plus linear trend”, respectively. The superscript  $(m)$  stands for the dimension of the  $I(1)$  vector entering the equations. One route to panel testing relies on so-called group statistics averaging individual statistics. We denote them as follows:

$$\bar{G}^{(m)} = \frac{1}{N} \sum_{i=1}^N \bar{S}_i^{(m)} \quad \text{or} \quad \tilde{G}^{(m)} = \frac{1}{N} \sum_{i=1}^N \tilde{S}_i^{(m)}.$$

Similarly, panel statistics rely on pooling the data across the within dimension, i.e. summing over terms showing up in the numerator and denominator separately,

$$\bar{P}^{(m)} = g \left( \sum_{i=1}^N \bar{N}_{i,T}^{(m)}, \sum_{i=1}^N \bar{D}_{i,T}^{(m)} \right) \quad \text{or} \quad \tilde{P}^{(m)} = g \left( \sum_{i=1}^N \tilde{N}_{i,T}^{(m)}, \sum_{i=1}^N \tilde{D}_{i,T}^{(m)} \right).$$

A typical example for the mapping  $g$  is  $g(x, y) = x/\sqrt{y}$  in the case of t-type statistics. Here it is assumed that the generic  $\bar{N}_{i,T}^{(m)}$  and  $\bar{D}_{i,T}^{(m)}$  or  $\tilde{N}_{i,T}^{(m)}$  and  $\tilde{D}_{i,T}^{(m)}$  are computed from individually demeaned or detrended regressions, respectively. We allow for group and panel statistics by introducing the generic notation  $\bar{Z}^{(m)}$  and  $\tilde{Z}^{(m)}$ , and maintain for the panel the joint null hypothesis

$$H_0 = \bigcap_{i=1}^N H_{i,0}. \quad (1)$$

A distinction between the individual null hypotheses  $H_{i,0}$  of cointegration or absence of cointegration is not required, both cases are treated in the generic assumption as follows.

**Assumption 2** Consider linear single-equation least squares regressions ( $i = 1, \dots, N$ ,  $t = 1, \dots, T$ )

$$y_{i,t} = \bar{\alpha}_i + \bar{\beta}'_i x_{i,t} + \bar{u}_{i,t} \quad \text{and} \quad y_{i,t} = \tilde{\alpha}_i + \tilde{\delta}_i t + \tilde{\beta}'_i x_{i,t} + \tilde{u}_{i,t}, \quad (2)$$

or

$$\Delta y_{i,t} = \bar{\kappa}_i + \bar{\gamma}_i y_{i,t-1} + \bar{\theta}'_i x_{i,t-1} + \bar{\varepsilon}_{i,t} \quad \text{and} \quad \Delta y_{i,t} = \tilde{\kappa}_i + \tilde{\psi}_i t + \tilde{\gamma}_i y_{i,t-1} + \tilde{\theta}'_i x_{i,t-1} + \tilde{\varepsilon}_{i,t}, \quad (3)$$

where where contemporaneous differences  $\Delta x_{i,t}$  or lags of  $\Delta z_{i,t-j}$ ,  $j > 0$ , may be required as additional regressors in (3) to ensure residuals free of serial correlation. Let  $\bar{Z}^{(m)}$  and  $\tilde{Z}^{(m)}$  stand for group statistics  $\bar{G}^{(m)}$  and  $\tilde{G}^{(m)}$  or for panel statistics  $\bar{P}^{(m)}$  and  $\tilde{P}^{(m)}$  computed from regressions with “intercept only” and “intercept plus linear trend”, respectively. We

assume under the null hypothesis (1) that it holds that

$$\begin{aligned}\sqrt{N}(\bar{Z}^{(m)} - \bar{\mu}_m) &\Rightarrow \mathcal{N}(0, \bar{\sigma}_m^2) \quad \text{if } \mu_i = 0, \\ \sqrt{N}(\tilde{Z}^{(m)} - \tilde{\mu}_m) &\Rightarrow \mathcal{N}(0, \tilde{\sigma}_m^2) \quad \text{for all } \mu_i,\end{aligned}$$

as  $T \rightarrow \infty$  followed by  $N \rightarrow \infty$ .

Tests e.g. by Kao (1999), Pedroni (1999, 2004), Westerlund (2005a,b) or Westerlund (2007) meet Assumption 2 under different sets of restrictions, and they will be considered in the next section, see Remark 1 through 3. In particular, these authors tabulate values of  $(\bar{\mu}_m, \bar{\sigma}_m)$  and  $(\tilde{\mu}_m, \tilde{\sigma}_m)$  for the different tests,  $m \geq 2$ .

Much of the earlier panel (co)integration literature assumed independent units invoking a central limit theorem to establish Assumption 2, see e.g. Kao (1999) and Pedroni (1999, 2004). Cross-sectional independence, however, is not maintained in our Assumption 2. Westerlund (2005a, 2007) e.g. allows for cross-correlation driven by a common factor. To account for that he suggests to replace  $x_{i,t}$  and  $y_{i,t}$  by the sectorally demeaned time series,

$$x_{i,t} - \bar{x}_t \text{ and } y_{i,t} - \bar{y}_t, \quad \text{where } \bar{x}_t = \frac{1}{N} \sum_{i=1}^N x_{i,t}, \quad \bar{y}_t = \frac{1}{N} \sum_{i=1}^N y_{i,t}.$$

That way he establishes that the limiting results maintained under Assumption 2 are met under cross-sectional correlation.

## 3 Results

### 3.1 Asymptotic theory

The first paper allowing for linear time trends in a panel cointegration context was by Kao (1999). He considers a residual-based unit root test for the null hypothesis of no cointegration in the tradition of Phillips and Ouliaris (1990). His test builds on pooling the data while allowing for a individual-specific intercept. Kao (1999) does not consider regressions containing a linear time trend as additional regressor, but allows for a linear drift in the data when performing a regression with a fixed effect intercept. In the case of  $k = 1$  regressor (i.e.  $m = 2$ ), Kao (1999, eq. (15)) observed that the linear time trend dominates the I(1) component; hence, the limiting distribution amounts to that of the panel unit root test by Levin, Lin, and Chu (2002) upon detrending. To become precise: Let  $\tilde{\mu}_1$  and  $\tilde{\sigma}_1$  denote the normalizing constants provided by Levin et al. (2002) for detrended panel unit root tests; then one should use them for the pooled residual-based

panel cointegration statistic  $\bar{P}^{(2)}$  in a bivariate regression if the regressor is I(1) with drift, see Kao (1999, Thm. 4):

$$\sqrt{N} (\bar{P}^{(2)} - \tilde{\mu}_1) \Rightarrow \mathcal{N}(0, \tilde{\sigma}_1^2) \quad \text{under } \mu_{i,x} \neq 0. \quad (4)$$

In Theorem 1 we extend Kao's result for any panel or group statistics from static or dynamic regressions with  $m \geq 2$  computed from regressions with intercept only in the presence of linear time trends.

**Theorem 1** *Let the data satisfy Assumption 1, and the generic test statistic  $\bar{Z}^{(m)}$  meets Assumption 2 for  $m \geq 2$ . Under the null hypothesis (1) it holds true that*

$$\sqrt{N} (\bar{Z}^{(m)} - \tilde{\mu}_{m-1}) \Rightarrow \mathcal{N}(0, \tilde{\sigma}_{m-1}^2) \quad \text{if } \mu_{i,x} \neq 0$$

as  $N \rightarrow \infty$ , where  $(\tilde{\mu}_{m-1}, \tilde{\sigma}_{m-1})$  are from Assumption 2.

PROOF See Appendix.

Note that Assumption 2 does not impose any restriction on  $\mu_{i,y}$ . As is shown in the proof, Theorem 1 holds irrespective of whether  $\{y_{i,t}\}$  displays a linear trend or not ( $\mu_{i,y} \neq 0$  or  $\mu_{i,y} = 0$ ).

Two research strategies can be employed in the presence of linear time trends when dealing with statistics resulting from regressions with intercept only. The first one simply ignores the linear time trends in the data and standardizes  $\bar{Z}^{(m)}$  with  $\bar{\mu}_m$  and  $\bar{\sigma}_m$ . The second strategy accounts for the drift in the data according to Theorem 1; in other words it applies  $\bar{Z}^{(m)}$  upon standardizing with  $\tilde{\mu}_{m-1}$  and  $\tilde{\sigma}_{m-1}$ . We summarize as follows.

**Strategy  $S_I$ :** When  $\bar{Z}^{(m)}$  is computed from panel regressions without detrending, then compare  $\sqrt{N} (\bar{Z}^{(m)} - \bar{\mu}_m) / \bar{\sigma}_m$  with quantiles from the standard normal distribution, i.e. ignore the presence of linear trends in the data.

**Strategy  $S_A$ :** When  $\bar{Z}^{(m)}$  is computed from panel regressions without detrending, then compare  $\sqrt{N} (\bar{Z}^{(m)} - \tilde{\mu}_{m-1}) / \tilde{\sigma}_{m-1}$  with quantiles from the standard normal distribution, i.e. account for the presence of linear trends in the data.

For the rest of the paper we assume that an applied econometrician is able to distinguish between the two cases, whether a linear time trend underlies the variables (e.g. log income or log prices) or not (e.g. interest or inflation rates). Hence, we maintain the assumption behind Theorem 1: the researcher knows that at least one regressor is I(1) with drift ( $\mu_{i,x} \neq 0$ ). We assume that strategy  $S_A$  is only employed when linear trends are truly present and thus refrain from the discussion of misspecification: what happens if there are no linear time trends in the data, but one erroneously accounts for trends.

The situation analyzed in Theorem 1 has not been considered in the previous panel cointegration literature, with the notable exception of Kao (1999). Consequently, all applied papers we are aware of standardize  $\bar{Z}^{(m)}$  with  $\bar{\mu}_m$  and  $\bar{\sigma}_m$  ignoring the effect of deterministic trends in the series, which amounts to strategy  $S_I$ . The effect of strategies  $S_I$  under linear time trends is discussed for growing  $N$  in the following proposition. The resulting size distortions depend on whether the test is right-tailed or left-tailed (null hypothesis is rejected for too large or too small values, respectively).

**Proposition 1** *Let the assumptions from Thm. 1 hold true. Further assume*

$$\tilde{\mu}_{m-1} < \bar{\mu}_m. \quad (5)$$

*Under the null hypothesis one has the following results under strategy  $S_I$ .*

- a) For a left-tailed test, the probability to reject according to strategy  $S_I$  increases with growing  $N$  to 1 as long as  $\mu_{i,x} \neq 0$ ;*
- b) for a right-tailed test, the probability to reject according to strategy  $S_I$  decreases with growing  $N$  to 0 as long as  $\mu_{i,x} \neq 0$ .*

PROOF See Appendix.

We now discuss a couple of panel tests satisfying Assumption 2 and (5), such that Theorem 1 and Proposition 1 apply.

**Remark 1** The residual-based unit root tests for the null hypothesis of no cointegration proposed by Pedroni (1999, 2004) build on static regressions as in (2). The null hypothesis (1) is rejected for too negative values. The expected values and standard deviations  $(\bar{\mu}_m, \bar{\sigma}_m)$  and  $(\tilde{\mu}_m, \tilde{\sigma}_m)$  showing up in Assumption 2 are available from Pedroni (1999, Table 2) for  $m > 2$  and from Pedroni (2004, Corollary 1) for  $m = 2$ . In order to apply Theorem 1 (strategy  $S_A$ ) for  $m = 2$ , one requires  $\tilde{\mu}_1$  and  $\tilde{\sigma}_1$ . These values stem from the detrended Dickey-Fuller distribution in the case of group statistics and have been tabulated by Nabeya (1999, Table 4):  $\tilde{\mu}_1 = -2.18136$  and  $\tilde{\sigma}_1 = 0.74991$ . Throughout we observe  $\tilde{\mu}_{m-1} < \bar{\mu}_m < 0$ . Hence, Proposition 1 a) applies. If strategy  $S_I$  is employed under linear trends, then the probability to reject the true null hypothesis converges to one with growing panel size  $N$ . Alternatively, Westerlund (2005a) suggested group and panel variance ratio type tests along the lines of Breitung (2002). The null hypothesis of no cointegration is rejected again for too small values, and  $(\bar{\mu}_m, \bar{\sigma}_m)$  and  $(\tilde{\mu}_m, \tilde{\sigma}_m)$  showing up in Assumption 2 are given in Westerlund (2005a, Table 1) for  $m \geq 2$ . To apply Theorem 1 with  $m = 2$ , we need  $\tilde{\mu}_1$  and  $\tilde{\sigma}_1$ . For the detrended Breitung distribution we obtain



by simulation  $\tilde{\mu}_1 = 0.0110$  and  $\tilde{\sigma}_1 = 0.005197$ , which are value corresponding to the case of group statistics. Again, we observe  $0 < \tilde{\mu}_{m-1} < \bar{\mu}_m$ , so that (5) holds. Consequently, Proposition 1 a) applies, and the probability to reject the true null hypothesis under strategy  $S_I$  grows with  $N$  as long as there is a linear trend in the data. To sum up: In the case of residual-based tests for no cointegration, strategy  $S_I$  results in massive size distortions; numerical evidence for finite  $N$  is given in Table 1 below.

**Remark 2** The error-correction tests by Westerlund (2007) relies on regressions of type (3). It is again a left-tailed test: The null hypothesis of no cointegration is rejected for too negative t-values associated with  $\gamma$ . Values of  $(\bar{\mu}_m, \bar{\sigma}_m)$  and  $(\tilde{\mu}_m, \tilde{\sigma}_m)$  are tabulated in Westerlund (2007, Table 1) for  $m \geq 2$ . In case of  $m = 1$  (i.e. no  $x_{i,t}$  on the right-hand side), the limiting distributions are of the usual Dickey-Fuller type. Hence,  $\tilde{\mu}_1$  and  $\tilde{\sigma}_1$  for group statistics are again from detrended Dickey-Fuller-type distributions and given in Nabeya (1999, Table 4), see above. Comparing  $\bar{\mu}_m$  with  $\tilde{\mu}_m$ , we find  $\tilde{\mu}_{m-1} < \bar{\mu}_m < 0$  meeting (5) again. Consequently, strategy  $S_I$  is increasingly liberal in the presence of linear time trends, and the probability to reject the true null hypothesis approaches 1 in the limit as long as the series display a linear time trend. For numerical evidence see Table 2 below.

**Remark 3** We now flip the null and the alternative hypotheses. Westerlund (2005b) allowed to test the null hypothesis of cointegration. He proposed a CUSUM group test statistic for this null hypothesis to be applied with tabulated values  $(\bar{\mu}_m, \bar{\sigma}_m)$  and  $(\tilde{\mu}_m, \tilde{\sigma}_m)$ ,  $m \geq 2$ . To apply Theorem 1 for  $m = 2$ , we provide as moments of the univariate, detrended distribution by simulation:  $\tilde{\mu}_1 = 0.6367$  and  $\tilde{\sigma}_1 = 0.14595$ .<sup>1</sup> This test is right-tailed, and according to Westerlund (2005b, Table 1)  $0 < \tilde{\mu}_{m-1} < \bar{\mu}_m$ . So, this time Proposition 1 b) comes in. Under strategy  $S_I$  in the presence of linear trends, the test is increasingly undersized with growing  $N$ . Obviously, such a conservative behaviour implies low power under the alternative hypothesis.

### 3.2 Numerical evidence

The statements obtained from Proposition 1 may be quantified more precisely by means of equations (13) and (14) given in the Appendix. These rejection probabilities apply approximately (for large  $N$ ) under the null hypothesis at nominal significance level  $\alpha$ . We report results for the group t-tests by Pedroni (1999, 2004) and by Westerlund (2007) in Tables 1 and 2.

---

<sup>1</sup>The univariate distribution is the supremum over the absolute value of a so-called second-level Brownian bridge, which shows up with the detrended KPSS test, too; see Kwiatkowski, Phillips, Schmidt, and Shin (1992).

Table 1: Approximate effective size of the group t-test by Pedroni (1999, 2004) computed from (13) at nominal level  $\alpha$  under strategy  $S_I$  for  $\mu_{i,x} \neq 0$

$N =$		10	20	30	40	50
$k = 1$	$\alpha = 0.01$	0.030	0.053	0.079	0.107	0.137
	$\alpha = 0.05$	0.126	0.190	0.249	0.307	0.361
	$\alpha = 0.10$	0.227	0.314	0.389	0.455	0.515
$k = 2$	$\alpha = 0.01$	0.017	0.024	0.030	0.036	0.043
	$\alpha = 0.05$	0.080	0.102	0.122	0.141	0.159
	$\alpha = 0.10$	0.154	0.188	0.217	0.243	0.268
$k = 3$	$\alpha = 0.01$	0.014	0.017	0.020	0.022	0.025
	$\alpha = 0.05$	0.067	0.078	0.087	0.096	0.104
	$\alpha = 0.10$	0.130	0.148	0.162	0.175	0.187

Table 2: Approximate effective size of the group t-test by Westerlund (2007) computed from (13) at nominal level  $\alpha$  under strategy  $S_I$  for  $\mu_{i,x} \neq 0$

$N =$		10	20	30	40	50
$k = 1$	$\alpha = 0.01$	0.139	0.352	0.564	0.732	0.846
	$\alpha = 0.05$	0.394	0.669	0.836	0.924	0.967
	$\alpha = 0.10$	0.566	0.808	0.921	0.969	0.988
$k = 2$	$\alpha = 0.01$	0.089	0.208	0.344	0.478	0.598
	$\alpha = 0.05$	0.283	0.484	0.644	0.763	0.846
	$\alpha = 0.10$	0.436	0.645	0.783	0.870	0.924
$k = 3$	$\alpha = 0.01$	0.067	0.150	0.247	0.349	0.450
	$\alpha = 0.05$	0.232	0.392	0.531	0.647	0.738
	$\alpha = 0.10$	0.372	0.553	0.687	0.783	0.852

Generally, the size distortions in Table 1 and 2 grow with  $N$ , while decreasing with  $m = k + 1$  at the same time. The fact that  $S_I$  is too liberal is characteristic for these tests where we reject for too negative values. Overrejections are not the general case, however, as we see when reversing the null and alternative hypotheses. To quantify distortions for the CUSUM test from Remark 3 we use equation (14) from the Appendix. When evaluating  $S_I$  under  $\mu_{i,x} \neq 0$ , we observe rejection probabilities equal to zero up to three digits for  $N = 10, 20, \dots$ ; this strongly supports the limiting result from Proposition 1 b).

### 3.3 Detrending

For regressions with intercept only, strategy  $S_I$  is clearly predominant in the literature and applied with the tests mentioned in the remarks above. We have illustrated its failure to control size under the null hypothesis in the presence of a linear time trend. In practice, one may use two strategies to account for linear time trends. The first one is the new  $S_A$  according to Theorem 1 from regressions without detrending. The second one consists of detrending the series, or equivalently run detrended regressions, i.e. include  $\tilde{\delta}_i t$  and  $\tilde{\psi}_i t$  in (2) and (3), respectively. The empirical strategy then becomes the following.

**Strategy  $S_D$ :** Compute  $\tilde{Z}^{(m)}$  from detrended panel regressions and compare the normalization  $\sqrt{N} \left( \tilde{Z}^{(m)} - \tilde{\mu}_m \right) / \tilde{\sigma}_m$  with quantiles from the standard normal distribution.

By Assumption 2, this strategy will provide asymptotically correct size. However, tests from detrended regressions will be prone to power losses relative to strategy  $S_A$ , which is more parsimonious. For that reason we next investigate the price of strategy  $S_D$  relative to  $S_A$  in terms of power.

In Monte Carlo experiments we study in particular the error-correction test (group t-statistic) by Westerlund (2007). Before turning to a power analysis, we make sure that size is under control. As data-generating process (DGP) we consider hence the null hypothesis of no cointegration under linear time trends:

$$y_{i,t} = \delta_i t + x_{i,1,t} + x_{i,2,t} + \dots + x_{i,k,t} + r_{i,0,t}, \quad (6)$$

$$t = 1, 2, \dots, T, \quad i = 1, 2, \dots, N,$$

$$x_{i,j,t} = \mu_{i,j,x} + x_{i,j,t-1} + v_{i,j,t}, \quad j = 1, 2, \dots, k, \quad (7)$$

where  $\{v_{i,j,t}\}$  are normal iid sequences,  $\mathcal{N}(0, \sigma_{i,j}^2)$ , independent of each other. Finally,  $r_{i,0,t} = r_{i,0,t-1} + v_{i,0,t}$  is an independent random walk entering (6). The DGP under the alternative of cointegration becomes:

$$\Delta y_{i,t} = -0.02 (y_{i,t-1} - \delta_i (t-1) - x_{i,1,t-1} - x_{i,2,t-1} - \dots - x_{i,k,t-1}) + v_{i,0,t} \quad (8)$$

where  $x_{i,j,t}$  and  $v_{i,0,t}$  are generated as before. Using the regression

$$\Delta y_{i,t} = \bar{\kappa}_i + \bar{\gamma}_i y_{i,t-1} + \bar{\theta}'_i x_{i,t-1} + \bar{\phi}'_i \Delta x_{i,t} + \bar{\varepsilon}_{i,t} \quad (9)$$

we computed the group t-statistic proposed by Westerlund (2007). Strategy  $S_D$  is employed with

$$\Delta y_{i,t} = \tilde{\kappa}_i + \tilde{\psi}_i t + \tilde{\gamma}_i y_{i,t-1} + \tilde{\theta}'_i x_{i,t-1} + \tilde{\phi}'_i \Delta x_{i,t} + \tilde{\varepsilon}_{i,t}. \quad (10)$$

All reported rejection frequencies rely on 10,000 replications.

The leading case consists of the following parameterization where only the first component of the regressors  $\{x_{i,t}\}$  is driven by a linear time trend:

$$T = 250, \quad \mu'_{i,x} = (1, 0, \dots, 0), \quad \sigma^2_{i,0} = \sigma^2_{i,1} = \dots = \sigma^2_{i,k} = 1. \quad (11)$$

This mimicks with  $k = 2$  or  $k = 3$  a typical macro panel with monthly data and e.g. income, interest rates and inflation rates as regressors. Table 3 reports the frequencies of rejection for different values of  $\delta_i$  from (6), and rejection is based on strategy  $S_A$  according to Theorem 1. It illustrates how well the rule of Theorem 1 works: The experimental sizes are close to the nominal ones. This is particularly true for  $\delta_i = 1$ , while the test is mildly conservative for  $\delta_i = 0.1$ , and a bit more conservative for  $\delta_i = 0$ , in particular for  $N$  large relative to  $T = 250$ . Next we consider strategy  $S_D$  with the same data. The rejection frequencies are given in Table 4. We observe that the experimental size from detrended regressions is close to the nominal one under the null hypothesis of no cointegration, irrespective of  $\delta_i$ .

Since strategies  $S_A$  and  $S_D$  both hold the nominal size, the question which one is more powerful naturally arises. The results contained in Table 5 are very clear: first, the power increases with  $\delta_i$ ; second, strategy  $S_A$  always outperforms  $S_D$  considerably, and has e.g. rejection frequencies more than twice as large for  $N = 10$ .

We have varied the leading case with the parameterization from (11). First, we allowed for more and stronger trends in the regressors,

$$\mu'_{i,x} = (1, 1, \dots, 1), \quad \text{or} \quad \mu'_{i,x} = (1, 2, \dots, k),$$

with all other parameters fixed. This corrects the mild undersizedness of strategy  $S_A$  reported in Table 3 yielding empirical sizes very close to the nominal one. At the same time power relative to Table 5 is increased, with strategy  $S_A$  still dominating  $S_D$ . Second, we have increased the size of the random walks, namely  $\sigma^2_{i,0} = \sigma^2_{i,1} = \dots = \sigma^2_{i,k} = 4$ , while the other parameters are from (11) and  $\delta_i = 0$ , see Table 6. Here, the linear trends are less pronounced, such that  $S_A$  results in slightly more conservative tests (compared to the first

Table 3: Experimental size at nominal level  $\alpha$  under  $S_A$  according to Theorem 1 for (11); DGP: (6) and (7)

$N =$		10	20	30	40	50
		$\delta_i = 0$				
$k = 1$	$\alpha = 0.01$	0.010	0.009	0.008	0.008	0.008
	$\alpha = 0.05$	0.048	0.048	0.043	0.040	0.041
	$\alpha = 0.10$	0.093	0.095	0.087	0.080	0.083
$k = 2$	$\alpha = 0.01$	0.008	0.009	0.007	0.009	0.006
	$\alpha = 0.05$	0.046	0.040	0.039	0.040	0.034
	$\alpha = 0.10$	0.093	0.082	0.083	0.082	0.073
$k = 3$	$\alpha = 0.01$	0.010	0.010	0.012	0.010	0.011
	$\alpha = 0.05$	0.051	0.052	0.052	0.049	0.052
	$\alpha = 0.10$	0.101	0.101	0.102	0.098	0.098
		$\delta_i = 0.1$				
$k = 1$	$\alpha = 0.01$	0.009	0.010	0.009	0.008	0.008
	$\alpha = 0.05$	0.050	0.048	0.048	0.044	0.043
	$\alpha = 0.10$	0.096	0.095	0.096	0.089	0.088
$k = 2$	$\alpha = 0.01$	0.009	0.009	0.008	0.009	0.007
	$\alpha = 0.05$	0.044	0.047	0.045	0.046	0.040
	$\alpha = 0.10$	0.093	0.094	0.090	0.092	0.085
$k = 3$	$\alpha = 0.01$	0.013	0.013	0.013	0.013	0.011
	$\alpha = 0.05$	0.055	0.056	0.058	0.056	0.054
	$\alpha = 0.10$	0.107	0.112	0.110	0.107	0.108
		$\delta_i = 1$				
$k = 1$	$\alpha = 0.01$	0.009	0.010	0.008	0.009	0.009
	$\alpha = 0.05$	0.047	0.045	0.045	0.042	0.044
	$\alpha = 0.10$	0.095	0.092	0.095	0.090	0.091
$k = 2$	$\alpha = 0.01$	0.011	0.009	0.008	0.010	0.009
	$\alpha = 0.05$	0.051	0.046	0.045	0.044	0.042
	$\alpha = 0.10$	0.097	0.093	0.090	0.086	0.084
$k = 3$	$\alpha = 0.01$	0.011	0.013	0.011	0.013	0.013
	$\alpha = 0.05$	0.056	0.054	0.055	0.055	0.056
	$\alpha = 0.10$	0.105	0.108	0.108	0.107	0.107

Table 4: Experimental size at nominal level  $\alpha$  under  $S_D$  (detrrending) for (11); DGP: (6) and (7)

$N =$		10	20	30	40	50
		$\delta_i = 0$				
$k = 1$	$\alpha = 0.01$	0.010	0.010	0.008	0.010	0.010
	$\alpha = 0.05$	0.047	0.047	0.045	0.048	0.044
	$\alpha = 0.10$	0.098	0.095	0.091	0.091	0.087
$k = 2$	$\alpha = 0.01$	0.012	0.012	0.013	0.012	0.012
	$\alpha = 0.05$	0.057	0.057	0.056	0.059	0.057
	$\alpha = 0.10$	0.112	0.108	0.106	0.108	0.111
$k = 3$	$\alpha = 0.01$	0.013	0.008	0.012	0.013	0.010
	$\alpha = 0.05$	0.055	0.050	0.050	0.055	0.049
	$\alpha = 0.10$	0.104	0.990	0.101	0.098	0.099
		$\delta_i = 0.1$				
$k = 1$	$\alpha = 0.01$	0.009	0.009	0.010	0.009	0.010
	$\alpha = 0.05$	0.051	0.050	0.049	0.051	0.048
	$\alpha = 0.10$	0.097	0.103	0.100	0.100	0.096
$k = 2$	$\alpha = 0.01$	0.012	0.013	0.013	0.015	0.014
	$\alpha = 0.05$	0.059	0.062	0.061	0.058	0.063
	$\alpha = 0.10$	0.116	0.117	0.115	0.113	0.117
$k = 3$	$\alpha = 0.01$	0.011	0.012	0.012	0.012	0.013
	$\alpha = 0.05$	0.054	0.052	0.052	0.052	0.055
	$\alpha = 0.10$	0.107	0.104	0.104	0.104	0.107
		$\delta_i = 1$				
$k = 1$	$\alpha = 0.01$	0.011	0.010	0.013	0.011	0.010
	$\alpha = 0.05$	0.052	0.047	0.048	0.050	0.050
	$\alpha = 0.10$	0.104	0.095	0.096	0.098	0.095
$k = 2$	$\alpha = 0.01$	0.015	0.011	0.014	0.014	0.012
	$\alpha = 0.05$	0.062	0.061	0.060	0.062	0.058
	$\alpha = 0.10$	0.119	0.119	0.116	0.119	0.115
$k = 3$	$\alpha = 0.01$	0.013	0.015	0.012	0.014	0.013
	$\alpha = 0.05$	0.053	0.056	0.053	0.053	0.054
	$\alpha = 0.10$	0.110	0.106	0.101	0.112	0.105

Table 5: Experimental power at nominal level 5% for (11); DGP: (8) and (7)

$N =$	10	20	30	40	50
$S_A$ : Theorem 1 for $\delta_i = 0$					
$k = 1$	0.486	0.727	0.859	0.933	0.965
$k = 2$	0.284	0.433	0.559	0.672	0.751
$k = 3$	0.155	0.234	0.288	0.336	0.387
$S_D$ : Detrended regression for $\delta_i = 0$					
$k = 1$	0.235	0.379	0.497	0.615	0.707
$k = 2$	0.163	0.247	0.331	0.399	0.452
$k = 3$	0.087	0.118	0.133	0.156	0.173
$S_A$ : Theorem 1 for $\delta_i = 0.1$					
$k = 1$	0.536	0.772	0.904	0.953	0.981
$k = 2$	0.311	0.485	0.622	0.726	0.802
$k = 3$	0.174	0.252	0.319	0.384	0.439
$S_D$ : Detrended regression for $\delta_i = 0.1$					
$k = 1$	0.262	0.402	0.531	0.636	0.716
$k = 2$	0.170	0.261	0.337	0.417	0.476
$k = 3$	0.090	0.127	0.143	0.168	0.186
$S_A$ : Theorem 1 for $\delta_i = 1$					
$k = 1$	0.824	0.974	0.997	1.000	1.000
$k = 2$	0.526	0.759	0.890	0.950	0.979
$k = 3$	0.292	0.419	0.553	0.661	0.734
$S_D$ : Detrended regression for $\delta_i = 1$					
$k = 1$	0.395	0.605	0.757	0.852	0.911
$k = 2$	0.219	0.330	0.433	0.524	0.595
$k = 3$	0.101	0.129	0.145	0.170	0.189

Table 6: Experimental size and power at nominal level 5% for  $T = 250$ ,  $\mu'_{i,x} = (1, 0, \dots, 0)$ ,  $\delta_i = 0$ , and  $\sigma^2_{i,j} = 4$ ; DGP: (6) or (8) and (7)

$N =$	10	20	30	40	50
$S_A$ : size					
$k = 1$	0.043	0.039	0.037	0.036	0.034
$k = 2$	0.043	0.041	0.037	0.034	0.031
$k = 3$	0.050	0.049	0.045	0.045	0.042
$S_A$ : power					
$k = 1$	0.410	0.622	0.771	0.869	0.920
$k = 2$	0.252	0.383	0.501	0.604	0.683
$k = 3$	0.147	0.205	0.257	0.304	0.364
$S_D$ : size					
$k = 1$	0.048	0.047	0.049	0.048	0.049
$k = 2$	0.056	0.063	0.061	0.062	0.062
$k = 3$	0.058	0.052	0.053	0.051	0.057
$S_D$ : power					
$k = 1$	0.236	0.387	0.522	0.622	0.699
$k = 2$	0.178	0.273	0.350	0.437	0.493
$k = 3$	0.102	0.136	0.161	0.192	0.213

panel in Table 3), and similarly, power is reduced (compared to the first panel in Table 5). Still,  $S_A$  clearly dominates  $S_D$  in Table 6. Third, we simulated shorter panels,  $T = 100$ . This makes both strategies,  $S_A$  and  $S_D$ , conservative under  $H_0$ , which is accompanied by a loss of power.

## 4 Concluding remarks

In time series econometrics it has been known for a long time that “the deterministic trends in the data affect the limiting distributions of the test statistics *whether or not we detrend the data.*” (Hansen, 1992, p. 103). This has been shown for the residual-based Phillips-Ouliaris (or Engle-Granger) cointegration test by Hansen (1992), see also the exposition in Hamilton (1994, p. 596, 597). Analogous results have been given for other cointegration tests by Hassler (2000) and Hassler (2001), see also the summary by Hassler (2016, Prop. 16.6). In this paper these findings are carried to the panel framework, and they are shown to continue to hold for single-equation tests relying on least squares, no matter whether the null hypothesis is absence or presence of cointegration. In a regression involving  $m \geq 2$  variables, much of the panel cointegration theory relies on normalization with some  $\bar{\mu}_m$  and  $\bar{\sigma}_m$  and letting the panel dimension  $N$  to infinity to obtain a standard normal distribution. The numbers  $\bar{\mu}_m$  and  $\bar{\sigma}_m$  are tabulated for the



case of regressions with intercept only; different figures  $\tilde{\mu}_m$  and  $\tilde{\sigma}_m$  are tabulated for regressions with intercept and linear time trend. We show the following: When statistics are computed from regressions with  $m$  integrated variables with intercept only, but when one of the integrated regressors is dominated by a linear time trend, then normalization with  $\tilde{\mu}_{m-1}$  and  $\tilde{\sigma}_{m-1}$  is required to achieve asymptotically valid inference under the null hypothesis (Theorem 1). Normalization with  $\bar{\mu}_m$  and  $\bar{\sigma}_m$ , however, which has been the conventional strategy so far, results in a loss of size control under the null hypothesis. In fact, employing  $\bar{\mu}_m$  and  $\bar{\sigma}_m$  in the presence of linear time trends gives rejection probabilities converging with  $N$  to 1 or 0, depending on whether the null hypothesis is no cointegration or cointegration, respectively (see Proposition 1). To avoid such size distortions one may employ the strategy following Theorem 1, or one may work with detrended regressions. Detrending, however, comes at a price: a regression with intercept only will provide more powerful tests (see e.g. Hamilton, 1994, p. 598); according to our simulations power gains of our new strategy over detrending may be very considerable, and this also holds true if there is a linear trend superimposing the level relation. Our limited Monte Carlo evidence, however, is restricted to the case of testing for the null hypothesis of no cointegration.

Hence, we propose the following empirical strategy if at least one of the integrated regressors is driven by a linear time trend when testing for no cointegration. First, test the null hypothesis of no cointegration with our new strategy  $S_A$  from Theorem 1, since it is more powerful than tests relying on detrending. If the null hypothesis of no cointegration is rejected according to Theorem 1, then one may test in a second step, whether a linear time trend is present superimposing the level relation between  $y_{i,t}$  and  $x_{i,t}$ . If strategy  $S_A$  does not reject the null hypothesis of no cointegration, then one may of course try a test building on detrending although it will tend to be less powerful.

## Appendix

### Proof of Theorem 1

Since group statistics  $\bar{G}^{(m)}$  are computed from individual regressions and since panel statistics  $\bar{P}^{(m)}$  build on pooled regressions, we suppress the index  $i$  and consider the generic case with  $\{y_t\}$  and  $\{x_t\}$  satisfying Assumption 1. Further, we focus on the stochastic regressors and ignore the constant intercept without loss of generality. We will proceed in four steps.

Step 1: First, note that the regressors may be rotated in such a way that all linear trends are concentrated in one scalar component. To that end define the  $k$ -element vector

$\lambda_1 := \mu_x / \sqrt{\mu'_x \mu_x}$  of unit length with

$$\tau_t := \lambda'_1 x_t = \sqrt{\mu'_x \mu_x} t + \lambda'_1 \sum_{j=1}^t e_{x,j}.$$

At the same time there exist  $k-1$  linearly independent  $k$ -element columns collected in the  $k \times (k-1)$ -matrix  $\Lambda_2$ . Due to the Gram-Schmidt orthogonalization one may assume that the invertible matrix  $\Lambda := (\lambda_1, \Lambda_2)$  is orthogonal:  $\Lambda \Lambda' = I_k$ . All columns of  $\Lambda_2$  eliminate the linear trend in  $x_t$ :

$$\xi_t := \Lambda'_2 x_t = \Lambda'_2 \sum_{j=1}^t e_{x,j}.$$

Hence,  $\xi_t$  is a  $(k-1)$ -vector integrated of order 1 without drift. Now, we are able to write

$$\Lambda' x_t = \begin{pmatrix} \tau_t \\ \xi_t \end{pmatrix}.$$

Step 2: Second, we show that the deterministic term in  $\tau_t$  dominates the I(1) component, which is clear from

$$T^{-1} \tau_{\lfloor rT \rfloor} = \sqrt{\mu'_x \mu_x} \frac{\lfloor rT \rfloor}{T} + O_p(T^{-0.5}) \Rightarrow \sqrt{\mu'_x \mu_x} r, \quad 0 \leq r \leq 1.$$

More precisely, we can show that empirical moments involving  $\tau_t$  equal those with  $t\sqrt{\mu'_x \mu_x}$  up to  $O_p(T^{-0.5})$ . We have from Park and Phillips (1988, Lemma 2.1) that the row vector

$$\left( \frac{1}{T^2} \sum_{t=1}^T \tau_t, \frac{1}{T^3} \sum_{t=1}^T \tau_t^2, \frac{1}{T^{2.5}} \sum_{t=1}^T \tau_t \xi'_t, \frac{1}{T^2} \sum_{t=1}^T \tau_t \Delta x'_{t-j} \right)$$

equals

$$\sqrt{\mu'_x \mu_x} \left( \frac{1}{T^2} \sum_{t=1}^T t, \frac{\sqrt{\mu'_x \mu_x}}{T^3} \sum_{t=1}^T t^2, \frac{1}{T^{2.5}} \sum_{t=1}^T t \xi'_t, \frac{1}{T^2} \sum_{t=1}^T t \Delta x'_{t-j} \right) + O_p(T^{-0.5});$$

further, if  $\mu_y \neq 0$ ,

$$\left( \frac{1}{T^3} \sum_{t=1}^T \tau_t y_t, \frac{1}{T^2} \sum_{t=1}^T \tau_t \Delta y_t \right) = \sqrt{\mu'_x \mu_x} \left( \frac{1}{T^3} \sum_{t=1}^T t y_t, \frac{1}{T^2} \sum_{t=1}^T t \Delta y_t \right) + O_p(T^{-0.5}),$$

or if  $\mu_y = 0$

$$\left( \frac{1}{T^{2.5}} \sum_{t=1}^T \tau_t y_t, \frac{1}{T^{1.5}} \sum_{t=1}^T \tau_t \Delta y_t \right) = \sqrt{\mu'_x \mu_x} \left( \frac{1}{T^{2.5}} \sum_{t=1}^T t y_t, \frac{1}{T^{1.5}} \sum_{t=1}^T t \Delta y_t \right) + O_p(T^{-0.5}).$$

Now, we are equipped to deal with the two cases: Residual-based tests from (2) and t-type tests from (3).

Step 3: Consider the least squares regression (without intercept for brevity)

$$y_t = \widehat{\beta}' x_t + \widehat{u}_t$$

with

$$\begin{aligned} \widehat{\beta} &= \left( \sum_{t=1}^T x_t x_t' \right)^{-1} \Lambda \Lambda' \sum_{t=1}^T x_t y_t = \left( \sum_{t=1}^T \Lambda' x_t x_t' \right)^{-1} \sum_{t=1}^T \Lambda' x_t y_t \\ &= \left( \sum_{t=1}^T \begin{pmatrix} \tau_t \\ \xi_t \end{pmatrix} x_t' \right)^{-1} \sum_{t=1}^T \begin{pmatrix} \tau_t \\ \xi_t \end{pmatrix} y_t. \end{aligned}$$

Similarly,

$$\widehat{\beta}' \Lambda = \sum_{t=1}^T (\tau_t, \xi_t') y_t \left( \sum_{t=1}^T \begin{pmatrix} \tau_t \\ \xi_t \end{pmatrix} (\tau_t, \xi_t') \right)^{-1}.$$

Consequently, the empirical residuals are

$$\widehat{u}_t = y_t - \widehat{\beta}' \Lambda \Lambda' x_t = y_t - \sum_{t=1}^T (\tau_t, \xi_t') y_t \left( \sum_{t=1}^T \begin{pmatrix} \tau_t \\ \xi_t \end{pmatrix} (\tau_t, \xi_t') \right)^{-1} \begin{pmatrix} \tau_t \\ \xi_t \end{pmatrix}. \quad (12)$$

For  $y_t$  we have by assumption

$$\begin{aligned} y_t &= \delta t + \beta' x_t + u_t = \delta t + \beta' \Lambda \Lambda' x_t + u_t \\ &= \delta t + \beta' \lambda_1 \tau_t + \theta' \xi_t + u_t, \quad \theta := \Lambda_2' \beta. \end{aligned}$$

If  $\{u_t\}$  is I(1), then the series are not cointegrated. If  $\{u_t\}$  is I(0), then there is cointegration, where a linear time trend may superimpose the cointegrating relation ( $\delta \neq 0$ ) or not ( $\delta = 0$ ). In any case,  $\{y_t\}$  is composed of the  $(k-1)$ -vector  $\{\xi_t\}$ , which is I(1), and a linear time trend asymptotically (since  $\delta t + \beta' \lambda_1 \tau_t \approx (\delta + \beta' \lambda_1 \sqrt{\mu_x' \mu_x}) t$  in the sense of Step 2). Therefore, the residuals  $\{\widehat{u}_t\}$  behave asymptotically as if they were computed from a regression on  $(k-1)$  I(1) regressors and on a linear trend. This establishes Theorem 1 for the case of residual-based tests.

Step 4: Consider the dynamic least squares regression (without intercept and without (lagged) differences as further regressors for brevity):

$$\Delta y_t = \widehat{\gamma} y_{t-1} + \widehat{\theta}' x_{t-1} + \widehat{\varepsilon}_t.$$

In order to investigate error-correction tests relying on the t statistic  $t_\gamma$  we employ what is sometimes called the Frisch-Waugh-Lovell theorem. At first stage, regress both  $\Delta y_t$  and  $y_{t-1}$  on  $x_{t-1}$ , and denote the fitted values as  $f_{0,t}$  and  $f_{1,t}$ , respectively. At second stage, the regression of  $f_{0,t}$  on  $f_{1,t}$  produces a slope estimator that is numerically identical to  $\hat{\gamma}$ , and so are the residuals, while the t statistics differ negligibly due to differences in degrees of freedom. As in Step 3, eq. (12), one can argue that both  $f_{0,t}$  and  $f_{1,t}$  behave asymptotically as if they were computed from a regression on  $(k-1)$  I(1) regressors and on a linear trend. Hence, because of the Frisch-Waugh-Lovell theorem,  $t_\gamma$  behaves as if  $x_{t-1}$  in (3) had been replaced by a linear time trend as regressor plus  $(k-1)$  regressors that are I(1). This establishes Theorem 1 for the case of error-correction tests, and the proof is complete. ■

## Proof of Proposition 1

According to Theorem 1 the statistic  $\bar{Z}^{(m)}$  requires under  $\mu_{i,x} \neq 0$  normalization with  $\tilde{\mu}_{m-1}$  and  $\tilde{\sigma}_{m-1}$  in order to result in a standard normal distribution under  $H_0$ . Let  $z_{1-\alpha}$  denote a quantile from the standard normal distribution. In the case where the panel tests are left-tailed, the rejection probability of strategy  $S_I$  under the null hypothesis becomes approximately for large  $N$  (at nominal level  $\alpha$ ):

$$P\left(\sqrt{N} \frac{\bar{Z}^{(m)} - \bar{\mu}_m}{\bar{\sigma}_m} < -z_{1-\alpha}\right) = \Phi\left(\sqrt{N} \frac{\bar{\mu}_m - \tilde{\mu}_{m-1}}{\tilde{\sigma}_{m-1}} - \frac{\bar{\sigma}_m}{\tilde{\sigma}_{m-1}} z_{1-\alpha}\right). \quad (13)$$

Analogously for right-tailed tests, the rejection probability of strategy  $S_I$  becomes under  $\mu_{i,x} \neq 0$  according to Theorem 1 with growing  $N$ :

$$P\left(\sqrt{N} \frac{\bar{Z}^{(m)} - \bar{\mu}_m}{\bar{\sigma}_m} > z_{1-\alpha}\right) = 1 - \Phi\left(\sqrt{N} \frac{\bar{\mu}_m - \tilde{\mu}_{m-1}}{\tilde{\sigma}_{m-1}} + \frac{\bar{\sigma}_m}{\tilde{\sigma}_{m-1}} z_{1-\alpha}\right). \quad (14)$$

For  $N \rightarrow \infty$  ones gets the limits given in Proposition 1 from (13) and (14). ■

## References

- Blomqvist, A. G. and R. A. L. Carter (1997). Is Health Care Really a Luxury? *Journal of Health Economics* 16(2), 207–229.
- Breitung, J. (2002). Nonparametric Tests for Unit Roots and Cointegration. *Journal of Econometrics* 108(2), 343–363.
- Breitung, J. (2005). A parametric approach to the estimation of cointegration vectors in panel data. *Econometric Reviews* 24, 151–173.

- Chang, Y. and C. M. Nguyen (2012). Residual Based Tests for Cointegration in Dependent Panels. *Journal of Econometrics* 167(2), 504–520.
- Coe, D. T. and E. Helpman (1995). International R&D Spillovers. *European Economic Review* 39(5), 859–887.
- Demetrescu, M., C. Hanck, and A. I. Tarcolea (2014). IV-based cointegration testing in dependent panels with time-varying variance. *Journal of Time Series Analysis* 35(5), 393–406.
- Gerdtham, U.-G. and M. Löthgren (2000). On stationarity and cointegration of international health expenditure and GDP. *Journal Of Health Economics* 19, 461–475.
- Groen, J. J. J. and F. Kleibergen (2003). Likelihood-Based Cointegration Analysis in Panels of Vector Error-Correction Models. *Journal of Business & Economic Statistics* 21(2), 295–318.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- Hanck, C. (2009). Cross-sectional Correlation Robust Tests for Panel Cointegration. *Journal of Applied Statistics* 36, 817–833.
- Hansen, B. E. (1992). Efficient Estimation and Testing of Cointegrating Vectors in the Presence of Deterministic Trends. *Journal of Econometrics* 53(1-3), 87–121.
- Hansen, P. and A. King (1998). Health care expenditure and GDP: Panel data unit root test results - comment. *Journal of Health Economics* 17, 377–381.
- Hassler, U. (2000). Cointegration Testing in Single Error-Correction Equations in the Presence of Linear Time Trends. *Oxford Bulletin of Economics and Statistics* 62(5), 621–632.
- Hassler, U. (2001). The effect of linear time trends on the KPSS test for cointegration. *Journal of Time Series Analysis* 22(3), 283–292.
- Hassler, U. (2016). *Stochastic Processes and Calculus: An Elementary Introduction with Applications*. Springer.
- Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics* 90, 1–44.
- Karaman Örsal, D. D. and B. Droge (2014). Panel cointegration testing in the presence of a time trend. *Computational Statistics and Data Analysis* 76, 377–390.

- Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How sure are we that economic time series have a unit root? *Journal of Econometrics* 54(1-3), 159–178.
- Larsson, R., J. Lyhagen, and M. Löthgren (2001). Likelihood-based cointegration tests in heterogeneous panels. *Econometrics Journal* 4, 109–142.
- Levin, A., C.-F. Lin, and C.-S. J. Chu (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics* 108, 1–24.
- McCoskey, S. K. and T. M. Selden (1998). Health care expenditures and GDP: panel data unit root test results. *Journal of Health Economics* 17(3), 369–376.
- Nabeya, S. (1999). Asymptotic Moments Of Some Unit Root Test Statistics In The Null Case. *Econometric Theory* 15(01), 139–149.
- Park, J. Y. and P. C. Phillips (1988). Statistical Inference in Regressions with Integrated Processes: Part 1. *Econometric Theory* 4(03), 468–497.
- Pedroni, P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics Special Issue*, 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* 20, 597–625.
- Phillips, P. C. B. and S. Ouliaris (1990). Asymptotic Properties of Residual Based Tests for Cointegration. *Econometrica* 58(1), 165–93.
- Westerlund, J. (2005a). New Simple Tests for Panel Cointegration. *Econometric Reviews* 24(3), 297–316.
- Westerlund, J. (2005b). A panel CUSUM test of the null of cointegration. *Oxford Bulletin of Economics and Statistics* 67, 231–262.
- Westerlund, J. (2007). Testing for error correction in panel data. *Oxford Bulletin of Economics and Statistics* 69, 709–748.