Panel Cointegration Testing Using Nonlinear Instruments*

Matei Demetrescu, Adina I. Tarcolea
Goethe-University Frankfurt†

This version: December 5, 2006

Abstract
The thorny issue of panel cointegration under cross-sectional dependence is addressed. We estimate a single-equation, error-correction model using the nonlinear instruments proposed by Chang (2002, JoE) for the augmented Dickey-Fuller test. When testing the null of no cointegration for a single unit, this yields a test statistic having asymptotic standard normal distribution even in the presence of endogenous regressors, irrespective of the number of integrated covariates. In panels exhibiting correlation or cointegration across units, individual test statistics are shown to be asymptotically independent, which leads to a panel test statistic robust to dependence across units.

Key words
Cross-dependent panel, Endogeneity, Regularly integrable transformation.

JEL classification
C12; C22; C23

*Previous versions of this paper were presented at the Workshop "Unit Roots and Cointegration in Panel Data", October 22nd-23rd 2004, Goethe-University Frankfurt and at the 12th Conference on Panel Data, June 24th-26th 2005, Copenhagen. The authors would especially like to thank Yoosoon Chang, as well as Anindya Banerjee, Jörg Breitung, Uwe Hassler, Vladimir Kuzin, Jean-Pierre Urbain and Jürgen Wolters for very helpful comments and suggestions. The second author was supported in her research by the Deutsche Forschungsgemeinschaft (DFG) within the project "Integration and Cointegration Analysis with Panel Data".

†Statistics and Econometric Methods, Goethe-University Frankfurt, Gräfstr. 78 PF 76, D-60054 Frankfurt, Germany, tel: +49.69.798.23660, Fax: +49.69.798.23662, email: deme@wiwi.uni-frankfurt.de, tarcolea@wiwi.uni-frankfurt.de.
\section{Motivation}

Using panel data is generally viewed as a method of gaining power when testing economic hypotheses and estimating relevant parameters. Although panels obviously contain more information than single time series, tests in nonstationary panels are difficult to build because of cross-correlation among units, which may lead to serious biases, see e.g. O’Connell (1998). Moreover, cointegration across units was also found to distort panel test statistics (Banerjee, Marcellino and Osbat, 2004). Then, one must either account for such cross-unit dependence, or use methods that are robust to it.

Popular tests for panel cointegration, like those of Kao (1999), McCoskey and Kao (1998) or Pedroni (2004), do not cope well with cross-unit dependence. One obvious fix is to bootstrap the respective test statistics, e.g. in the manner of Maddala and Wu (1999). This, however, comes at a cost: bootstrapping time series is notoriously difficult. Factor models, as advocated by Bai and Ng (2004), are able to deal with cross-unit dependence as well.

In this paper, we take a different approach, by adopting the single equation framework advocated by Banerjee, Dolado and Mestre (1998). Instead of using OLS estimation and testing, as Banerjee, Dolado and Mestre (1998) did, we suggest to employ nonlinear instrumental variables [NIV] of the kind Chang (2002) used in estimating the augmented Dickey-Fuller test regression.

Our main contribution is to analyze the asymptotic behavior of the testing procedure proposed above. For a single unit, we show this NIV cointegration test to have asymptotic standard normal distribution, even when regressors are not weakly exogenous. In a panel context, individual statistics are shown to be asymptotically independent in the presence of cross-unit correlation or cross-unit cointegration. This leads to a panel test statistic not affected by cross-unit dependence; moreover, standard limiting distributions result without $N$ asymptotics. Work in progress by Chang (2005) looks into a NIV estimation procedure for residual-based cointegration tests.

The paper is structured as follows. First, we describe the model we work with. Section 3 analyzes the NIV cointegration test in the error correction framework and its extension to panels exhibiting cross-dependence. Monte Carlo evidence on small-sample behavior is given in Section 4. The main findings and comments are summarized in the final section.

\section{Model and assumptions}

To reduce notational burden, we first study a single unit. The assumptions for the panel case are discussed in Section 3.2.

We assume the data to be modelled by an integrated vector autoregressive process of order $p + 1$ with $K + 1$ components, $K \geq 1$. This process is possibly cointegrated:
Assumption 1 Let the observed data for one unit be generated as follows:

\[ \Delta w_t = \Pi w_{t-1} + \sum_{i=1}^{p} A_i \Delta w_{t-i} + \epsilon_t, \]

where \( w'_t = (w_{1,t}, \ldots, w_{K+1,t}) \) and \( w_0 = 0 \).

Denote with \( r \) the rank of \( \Pi \). Under no cointegration, one has \( \Pi = 0 \), or \( r = 0 \). Under cointegration, one has \( 0 < r < K + 1 \) and the known factorization of \( \Pi, \Pi = \alpha_r \beta_r' \), in two \((K + 1) \times r\) matrices of adjustment speed coefficients and of parameters of the long-run relations, respectively, holds for \( r > 0 \). Under the alternative hypothesis of cointegration, we assume \( r = 1 \). This is needed to motivate the test statistic, see Banerjee, Dolado and Mestre (1998). The following assumption guarantees that the process \( w_t \) either follows a stable vector autoregressive process in differences (no cointegration), or, when \( \Pi \neq 0 \), \( I(2) \) processes are avoided.

Assumption 2 Let the roots of the characteristic polynomial associated to \( w_t \) defined in Assumption 1 be either 1 or have absolute values larger than 1. Further, if \( \Pi = 0 \), let \( \det (\alpha_r' (I - \sum_{i=1}^{p} A_i) \beta_r') \neq 0 \), where \( \alpha_r \) and \( \beta_r \) w.r.t. \( R^{K+1} \).

Note that no additional restriction is imposed upon the elements of the matrices \( A_i \) or on \( \Pi \). Also, the innovations \( \epsilon_t \) are allowed to correlate:

Assumption 3 Let \( \epsilon_t \) obey following conditions: \( \epsilon_t \sim iid (0, \Sigma) \), with \( \Sigma \) any symmetric, positive definite \((K + 1) \times (K + 1)\) matrix. Furthermore, assume \( \exists l \geq 2 \) such that \( E \| \epsilon_t \|^l < \infty \) and let \( \epsilon_t \) have absolutely continuous distribution function (w.r.t. Lebesgue measure); let also \( \exists s > 0 \) such that \( \phi(\lambda) = o(\|\lambda\|^s) \), where \( \phi \) is the corresponding characteristic function.

Assumptions 2 and 3 together allow for lack of exogeneity. This is an important aspect, since the weak exogeneity assumption is the main source of criticism to the single equation approach. The conditions in Assumption 3 are stronger than the typical sets of assumptions under which an invariance principle for cumulated innovations holds, but are needed in order to establish asymptotic behavior of regularly integrable transformations of integrated processes, see Park and Phillips (1999, 2001), as well as Chang, Park and Phillips (2001) and de Jong and Wang (2005). It is this framework that leads to robustness against cross-unit dependence in panels (see Subsection 3.2).

Denote \( w_{i,t} = (w_{2,t}, \ldots, w_{K+1,t})' \). In the single equation framework, the error correction representation can be then written as follows:

\[ \Delta w_{1,t} = \alpha (w_{1,t-1} + \theta' w_{e,t-1}) + \delta (L) \Delta w_{1,t-1} + \gamma (L) \Delta w_{e,t-1} + \epsilon_t, \]  
\[ \Delta w_{e,t} = \alpha_e (w_{1,t-1} + \theta' w_{e,t-1}) + \delta_e (L) \Delta w_{1,t-1} + \Gamma (L) \Delta w_{e,t-1} + \nu_t, \]

for \( t = 1, 2, \ldots, T \), where \( \alpha \in \mathbb{R}, \alpha_e \in \mathbb{R}^K, \theta \in \mathbb{R}^K \), the respective lag polynomials and the innovations \( \epsilon_t \) and \( \nu_t \) are defined implicitly from Assumption 1.
Not including contemporaneous differences $\Delta w_{e,t}$ in Equation (1), as Banerjee, Dolado and Mestre (1998) do, is compensated by having allowed for correlated innovations. Equation (1) can be transformed to contain only levels of the integrated variables $w_{1,t}$ and $w_{e,t}$, if one is interested in the pure ADL representation of the model.

The null hypothesis in the single equation framework is $\alpha = 0$; in case of weak exogeneity, $\alpha_e = 0$ is an implicit additional assumption. However, we wish to extend the types of endogeneity our test copes with, and thus explicitly allow the error correction to affect Equations (2). Hence, we shall examine following null:

Null hypothesis: $\alpha = 0$.

Note that, when allowing for error correction to affect the other components of $w_t$, the null of the test is actually absence of error correction in the studied equation and not lack of cointegration between $w_{1,t}$ and $w_{e,t}$. This attribute is common to all approaches based on a single equation and not specific to our test. See Remark 6 further below for a simple solution to this problem.

Under the alternative, $\alpha$ needs to be negative if error-correction is present only in Equation (1). Otherwise, $\alpha$ may also be positive (see Johansen, 1995, p. 54, for an example). Thus, our alternative hypothesis is as follows:

Alternative hypothesis: $\alpha \neq 0$.

The basic idea of our test is to replace OLS estimation of the test Equation (1) with instrumental estimation using regularly integrable transformations. Specifically, $F(w_{1,t-1})$ is used as instrument for $w_{1,t-1}$, where $F(\cdot)$ is restricted as follows:

**Assumption 4** Let $F(\cdot)$ be continuous on $\mathbb{R}$ with $\int_{-\infty}^{\infty} x F(x) dx$ finite and non-zero. Assume further that $|F(\cdot)|$ is bounded by a function $R(\cdot)$, where $R(\cdot)$ is integrable, continuous on $\mathbb{R}$ and monotone on $(-\infty, 0)$ and $(0, \infty)$.

In what concerns the other integrated regressors, two possibilities arise. First, we may take them as instruments for themselves. Second, we may take regularly integrable transformations as instruments. We shall call the first case "partial instrumentalization", and the second will be denoted as "complete instrumentalization". However, Monte Carlo experiments (see Section 4) show the completely instrumentalized test to have very low power. Therefore, this paper focusses on partial instrumentalization.

The described data generating process does not exhibit deterministic components. These will be dealt with in Section 3.1.

## 3 Asymptotic results

As already mentioned, we first deal with the case of a single unit. The panel case is discussed in the second subsection.
3.1 Single unit test

The test regression (1) is reparameterized to match the usual notation of the single equation framework. Defining

\[ x_{t-1}' = (\Delta w_{1,t-1}, \ldots, \Delta w_{1,t-p}, \Delta w_{e,t-1}, \ldots, \Delta w_{e,t-p}, w_{e,t-1}) \]

the I(1) variables \((x_{t-1,1} = w_{e,t-1})\) are separated from the I(0) ones \((x_{t-1,0})\). The single-equation model becomes with \(y_t = w_{1,t}\)

\[ \Delta y_t = \alpha y_{t-1} + \beta' x_{t-1} + \varepsilon_t, \quad (3) \]

with a new parameter vector \(\beta', \beta_0', \gamma_1', \ldots, \gamma_p'\), where \(\delta_i, \gamma_i, i = 1, 2, \ldots, p\), are the respective coefficients of the lag polynomials from Equation (1). It is convenient to write \(\beta_0' = (\delta_1, \ldots, \delta_p, \gamma_1', \ldots, \gamma_p')\) and \(\beta_1' = \alpha \theta'\), in accordance to \(x_{t-1} = (x_{t-1,0}, x_{t-1,1})\).

The t statistic of the estimated parameter \(\hat{\alpha}\) remains the natural choice as a test statistic for the null \(\alpha = 0\), even with instrumental estimation. Note that, under the null hypothesis \(\alpha = 0\), it holds \(\beta_1 = 0\). The parameter vector \(\theta\) is not identified under the null hypothesis, but \(\alpha\) and \(\beta_1\) are identified under both null and alternative hypothesis. Hence, there is no impediment in testing this way.

For the case of partial instrumentalization, one obtains with the help of standard regression algebra,

\[ \hat{\alpha} - \alpha = Q^{-1} M, \]

where

\[ M = \sum_{t=1}^{T} F(y_{t-1}) \varepsilon_t - \sum_{t=1}^{T} F(y_{t-1}) x_{t-1}' \left( \sum_{t=1}^{T} x_{t-1} x_{t-1}' \right)^{-1} \sum_{t=1}^{T} x_{t-1} \varepsilon_t \]

and

\[ Q = \sum_{t=1}^{T} F(y_{t-1}) y_{t-1} - \sum_{t=1}^{T} F(y_{t-1}) x_{t-1}' \left( \sum_{t=1}^{T} x_{t-1} x_{t-1}' \right)^{-1} \sum_{t=1}^{T} x_{t-1} y_{t-1}. \]

For the t statistic, it holds under the null hypothesis \(\alpha = 0\)

\[ t_{\hat{\alpha}} = \frac{\hat{\alpha}}{\hat{\sigma}_\alpha}, \]

with \(\hat{\sigma}_\alpha\) the estimated standard deviation of \(\hat{\alpha}\):

\[ \hat{\sigma}^2_{\alpha} = \hat{\sigma}^2_{\varepsilon} Q^{-2} P, \]
where $\hat{\sigma}^2$ is a consistent estimator of the residual variance, $\sigma^2 = \text{Var}(\varepsilon_t)$, and
\[
P = \sum_{t=1}^{T} F(y_{t-1})^2 - \sum_{t=1}^{T} F(y_{t-1}) \left( \sum_{t=1}^{T} x_{t-1}x'_{t-1} \right)^{-1} \sum_{t=1}^{T} x_{t-1}F(y_{t-1}),
\]
which leads to
\[
t_\alpha = \frac{M}{\hat{\sigma}\sqrt{P}}.
\]
In the following, it will be more convenient to study the pivotal statistic
\[
t^* = \frac{\hat{\alpha} - \alpha}{\hat{\sigma}_{\hat{\alpha}}} = \frac{M}{\hat{\sigma}\sqrt{P}},
\]
irrespective of the value of $\alpha$.

The following proposition establishes convergence properties of the NIV estimators (see the Appendix for proof details).

**Proposition 1** For partially instrumentalized NIV estimation of test equation (3), it holds as $T \to \infty$ under Assumptions 1 through 4 and $\alpha = 0$ that

a) if $\alpha_e = 0$

\[
\hat{\alpha} - \alpha = O_p \left( T^{-0.75} \right), \tag{4}
\]

\[
\hat{\beta}_0 - \beta_0 = O_p \left( T^{-0.25} \right) \tag{5}
\]

and

\[
\hat{\beta}_1 - \beta_1 = O_p \left( T^{-0.75} \right) \tag{6}
\]

b) if $\alpha_e \neq 0$,

\[
\hat{\beta}_0 - \beta_0 = O_p \left( T^{-0.25} \right).
\]

However, the convergence rates of $\hat{\alpha}$ or $\hat{\beta}_1$ to the true values $\alpha$ and $\beta_1$ can be $T^{0.25}$, depending on the cointegrating vector $(1, \theta)'$.

**Proof:** See the Appendix.

This behavior of the estimators associated to the integrated regressors appears under both the null and the alternative hypothesis, since we allowed for lack of exogeneity in form of error-correction in the other equations of the error correction model besides the test Equation (1). The direct effect on the test is that one should not use NIV residuals when computing the residual variance, since convergence rates of the estimators may not be high enough to ensure consistent residuals, given that the regressors which $\hat{\alpha}$ and $\hat{\beta}_1$ are associated to are integrated. This seems to be the case even when $y_{t-1}$ and $x'_{t-1,1}$ do not cointegrate (and the convergence rates of the estimators are high enough); a Monte Carlo experiment for the simplest case of two independent random walks with no deterministic components or short-run dynamics suggests that the behavior of the estimators is not very reliable and the variance of the test
statistic under the null hypothesis is lower than unity by a factor of up to 1.4 for sample sizes between $T = 100$ and $T = 500$. One should resort to alternative residual variance estimators, such as using residuals from the OLS estimation of the model, which is employed throughout this paper.

Fortunately, the behavior of the pivotal statistic $t^*$ is only indirectly affected by this, by means of the residual variance estimator. The following lemma eases the discussion of the suggested test. Its proof can be found in the Appendix as well.

**Lemma 1** If using a consistent residual variance estimator, it holds under the assumptions of Proposition 1 that

$$ t^* = \frac{T^{-0.25} \sum_{t=1}^{T} F(y_{t-1}) \varepsilon_t}{\sigma_x \sqrt{T^{-0.5} \sum_{t=1}^{T} F(y_{t-1})^2}} + o_p(1), $$

regardless of whether $\alpha_e = 0$ or $\alpha_e \neq 0$.

**Proof:** See the Appendix.

**Remark 1** It is now clear why no restrictive weak exogeneity assumptions have to be made. This is because no regressor except the lagged dependent variable $y_{t-1}$ influences the test statistic for large $T$, as can be seen from Lemma 1. In contrast to that, OLS estimation of the test equation requires either weak exogeneity or inclusion of leads to account for second-order bias (see Banerjee, Dolado and Mestre, 1998).

The following proposition summarizes the asymptotic behavior of the proposed test statistic under both null and alternative hypothesis.

**Proposition 2** Under the assumptions of Lemma 1, it holds as $T \to \infty$:

a) if $\alpha = 0$, then

$$ t_{\alpha} \overset{d}{\to} \mathcal{N}(0,1); $$

b) if $\alpha \neq 0$, then

$$ |t_{\alpha}| \overset{p}{\to} \infty. $$

**Proof:** See the Appendix.

**Remark 2** Confidence intervals for the parameter $\alpha$ are straightforward to build, since the pivotal statistic $t^*$ has asymptotic standard normal distribution whatever the true value of $\alpha$ is, as long as a consistent residual variance estimator is used.

**Remark 3** Following Demetrescu (2006), Proposition 2 can also be established for the case where $\Delta w_t$ follows a general linear process with a weak summability condition, when an autoregressive approximation of order growing to infinity, but slower than $T$, is used. This rate should be $o(T^{0.25})$, but not slowly varying at infinity, see Demetrescu (2006) for further details.
Remark 4 In practice, the order of the autoregressive process capturing short-run dynamics is of course not known. Due to Lemma 1, data-driven lag order choice (such as sequential significance testing of the autoregressive parameters) will have no asymptotic effect on the test statistic; see Demetrescu, Kuzin and Hassler (2006) for a situation where this is not the case. However, information criteria should not be used, since these typically result in logarithmic rates.

When accounting for deterministic components, the lagged differences are either not affected by a non-zero mean in levels, or can be easily demeaned and deseasonalized, respectively. For the levels $y_{t-1}$, one must ensure they possess the martingale property, purpose to which we follow Chang (2002, p. 275) and resort to recursive (adaptive) schemes of demeaning, deseasonalizing, or detrending of $y_{t-1}$.

Recursive adjustment has the additional advantage that it does not matter where deterministics appear - e.g. a linear trend in the data may appear in levels, or due to a non-zero intercept in the cointegrating relation. This is not the case with the Johansen procedure, for instance, where different asymptotic distributions result under different sources of deterministics.

For a non-zero mean, this means that the NIV cointegration test has to be carried out in following test equation:

$$
\Delta y_t = \alpha y_{t-1} + \beta' x_{t-1} + \varepsilon_t,
$$

where the recursively demeaned lagged level $y_{t-1}^\mu$ is given for $t \geq 2$ by

$$
y_{t-1}^\mu = y_{t-1} - \frac{1}{t-1} \sum_{j=1}^{t-1} y_j,
$$

and the integrated regressors $x_{t-1,1}$ may also require demeaning, hence the notation $X_{t-1}^\mu$ in Equation (7). Usual projection on a constant is allowed for the integrated regressors, in contrast to the case of the lagged dependent variable. The stationary regressors, being differences, need no adjustment.

For a linear trend, one must use as test equation

$$
\Delta y_t = \alpha y_{t-1}^\tau + \beta' x_{t-1}^\tau + \varepsilon_t,
$$

where the recursively detrended lagged level $y_{t-1}^\tau$ is given for $t \geq 2$ by

$$
y_{t-1}^\tau = y_{t-1} + \frac{2}{t-1} \sum_{j=1}^{t-1} y_j - \frac{6}{t(t-1)} \sum_{j=1}^{t-1} y_j,
$$

and the integrated regressors may be detrended the usual way. The stationary regressors and the regressand $\Delta y_t$ only require usual demeaning.

For deseasonalizing, note that Kuzin (2005) shows that simply subtracting recursive seasonal means is not admissible, and gives a method that ensures asymptotic equivalence to the case of recursive demeaning. The extension to detrending follows from his work.
Then, one uses as instruments $F(y_{t-1})$ or $F(y_{t-1}^T)$. The employed asymptotic theory holds for these instruments as well, but in terms of recursively demeaned (detrended) Brownian motions, see Chang (2002). Thus, for the case of test Equations (7) and (8), the results analog to Proposition 1, Lemma 1 and Proposition 2, summarized in the following proposition, can be shown to hold true. Its proof is very similar to that of Proposition 2 and not given here.

**Proposition 3** Under the assumptions of Lemma 1 and recursive demeaning or detrending, it holds for the $t$ statistics from test equations (7) or (8) as $T \to \infty$:

a) if $\alpha = 0$, then $t_{\alpha} \overset{d}{\to} N(0, 1)$;

b) if $\alpha \neq 0$, then $|t_{\alpha}| \overset{p}{\to} \infty$.

**Proof:** Omitted.

**Remark 5** It can be shown that $x_{t-1, 1}$ themselves may be recursively demeaned (detrended) without affecting the asymptotics. This was found to perform better in small samples than projecting $x_{t-1, 1}$ on a constant (on a time trend).

### 3.2 Panel test

We now turn our attention to the panel case and deal with panels containing $N$ vector autoregressive processes representing $N$ cross-sectional units, where $N$ is finite. In what concerns the notation, each used symbol becomes an additional index $i$, $i = 1, 2, \ldots, N$, representing the respective unit. The innovations $(\varepsilon_{1, t}, \nu_{1, t}^T, \varepsilon_{2, t}, \nu_{2, t}^T, \ldots, \varepsilon_{N, t}, \nu_{N, t}^T)$ are allowed to correlate across units, as specified by following assumption.

**Assumption 5** Denote $\varepsilon_t^N = (\varepsilon_{1, t}, \nu_{1, t}^T, \varepsilon_{2, t}, \nu_{2, t}^T, \ldots, \varepsilon_{N, t}, \nu_{N, t}^T)$ and let $\varepsilon_t^N \sim iid(0, \Sigma_N)$, with $\Sigma_N$ any symmetric, positive definite $N(K+1) \times N(K+1)$ matrix. Furthermore, assume $\exists \ell > 2$ such that $E \|\varepsilon_t^N\|^\ell < \infty$ and let $\varepsilon_t^N$ have absolutely continuous distribution function (w.r.t. Lebesgue measure) such that $\exists s > 0$ with $\phi_N(\lambda) = o(\|\lambda\|^s)$, where $\phi_N$ is the corresponding characteristic function.

Assumptions 1, 2 and 4 are maintained for each single unit. Thus, we do not explicitly allow for cross-unit dynamics, but augmenting the test regressions with lagged differences from other cross-sectional units can obviously be allowed for. Moreover, this augmentation is desirable, since, when ignoring cross-unit dynamics, each unit $\Delta w_t$ follows (marginally) a general linear process, which requires approximation by means of an autoregressive process of order growing to infinity, see also Remark 3.

We assume here the number of cross-sections to be finite. Note that Im and Pesaran (2003) argue that $N$ may grow to infinity, but at a certain maximal
rate. However, their argumentation ignores the behavior of the $o_p(1)$ term from Lemma 1 as both $N$ and $T$ grow to infinity. To avoid this issue, we impose $N < \infty$.

The testing hypotheses are modified as follows:

**Null hypothesis**: $\alpha_i = 0, \ i = 1, 2, \ldots, N$.

The null is rejected if equilibrium adjustment is found for at least one unit:

**Alternative hypothesis**: $\exists i, 1 \leq i \leq N$ with $\alpha_i \neq 0$.

The panel test statistic is defined as

$$\tilde{t} = \sum_{i=1}^{N} t_{\alpha_i}^2,$$

(9)

where the single test statistics $t_{\alpha_i}$ may be computed with recursive demeaning or detrending. Due to Lemma 1, one is not constrained to use the same set of explanatory variables for each individual test. This panel test statistic follows a chi-square distribution with $N$ degrees of freedom asymptotically, as stated in the following proposition, and one rejects for too large values of the test statistic. This is because single test statistics are asymptotically independent even if the innovations $\varepsilon_{i,t}$ correlate across units, see the Appendix for proof details. The panel case is where asymptotics of nonlinear transformations of integrated series indeed comes into its own.

**Proposition 4** Under the assumptions of Proposition 2 with Assumption 3 replaced by Assumption 5, it holds for $\tilde{t}$ from (9) as $T \to \infty$

a) if $\alpha_i = 0 \forall i \in \{1, 2, \ldots, N\}$

$$\tilde{t} \xrightarrow{d} \chi^2_N,$$

b) if $\exists i$ such that $\alpha_i \neq 0$

$$\tilde{t} \xrightarrow{p} \infty.$$

**Proof**: See the Appendix.

**Remark 6** Proposition 4 also allows one to build a simple multi-equation test for single units. Together with Lemma 1, this proposition shows that test statistics from different equations for the same unit are asymptotically independent as well. Then, one can use as test statistic for a single unit $i$ the sum of the squared $t$ statistics for each equation of unit $i$. Obviously, a $\chi^2_K$ distribution results asymptotically. As panel test statistic, one uses the sum of single test statistics for each equation in all units, which follows a $\chi^2_{NK}$ distribution.

**Remark 7** The same reasoning as in the proof of Proposition 4 was used by Chang (2002) in building the panel NIV unit root test. Also, Chang’s work, together with Lemma 1, shows that panels may be unbalanced in the sense of Assumption 4.1 from Chang (2002).
Remark 8 The innovations \( \nu_{i,t} \) may also correlate across units, with \( \nu_{j,t} \) and \( \epsilon_{j,t} \). This is because Lemma 1 ensures that the terms containing integrated regressors and lagged differences are asymptotically negligible. Especially useful, the elements of \( w_{i,c,t-1} \) may cointegrate across units.

Remark 9 Should the dependent variables \( y_{i,t} \) cointegrate across units, asymptotic independence is no longer guaranteed. However, Chang and Song (2005) show that independence of single test statistics holds, if the instrument generating functions \( F_i \) satisfy certain orthogonality conditions. They suggest the use of Hermite polynomials, but argue that these need rescaling before using them as instrument generating functions (see Chang and Song, 2005, for a complete discussion).

If knowing that the non-zero \( \alpha_i \) are, for instance, negative under the alternative, one-sided testing results in more power. A simple way to build the one-sided panel test is to take the standardized sum of single test statistics,

\[
\tilde{t}^- = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} t_{\tilde{g}_i}.
\]

Asymptotic standard normality of \( \tilde{t}^- \) is easily proved in the same manner as Proposition 4, and one rejects for large negative values of this test statistic.

4 Small sample behavior

All simulations are carried out in GAUSS, with 10000 replications for each considered case. Compared to Chang’s (2002) simulations, we employ a slightly modified instrument generating function, \( F(x) = cx e^{-|cx|} \), where \( c \) is the inverse of the standard deviation of \( \Delta y_t \); due to its consistency, it does not affect the asymptotics, in contrast to Chang’s (2002) choice of \( c \), which is proportional to \( T^{-0.5} \), leading thus to different asymptotics (see also Im and Pesaran, 2003, for a discussion on this subject).

4.1 Behavior of the single unit test

First, we study our test for a single unit. We are particularly interested in the effects the lack of exogeneity has on the test statistic. Therefore, we forgo short-run dynamics, but include one lagged difference in the test equation. All integrated variables, \( y_{t-1} \) as well as \( x_{t-1,1} \), are recursively demeaned for both partial and complete instrumentalization. For the OLS residual variance estimation, a constant is included in the test equation instead of using recursive demeaning of the integrated variables. Preliminary experiments indicate that better behavior is to be expected if the residual variance estimator is adjusted for degrees of freedom. We take advantage of knowing what the true alternative is, and employ a one-sided test against \( \alpha < 0 \).
In a first series of experiments, we specify the data generating process with error-correction affecting the other equations, under both the null and the alternative hypothesis. We set $\alpha_e = (-0.1, \ldots, -0.1)$. The covariance matrix of the innovations is set equal to the unity matrix. Under $H_1$, we let $\alpha = -0.1$ and $\alpha = -0.2$, and, for simplicity, we set the cointegrating vector parameters to $\theta = -\alpha e$; it can be checked that, with these parameter values, Assumption 2 is satisfied for any $N$ and no I(2) process could emerge.

The size and power (at the nominal level of 5%) are given for $K = 1$ and $K = 2$ integrated covariates and sample sizes $T = 100$, $T = 200$ and $T = 500$ in Table 1. Both variants of the test are studied: the partially instrumentalized test statistic is denoted by $t^{p}_{\alpha}$, while the completely instrumentalized one is denoted by $t^{c}_{\alpha}$.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$T = 100$</th>
<th>$T = 200$</th>
<th>$T = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K = 1$</td>
<td>$K = 2$</td>
<td>$K = 1$</td>
</tr>
<tr>
<td>0</td>
<td>6.54</td>
<td>7.44</td>
<td>6.17</td>
</tr>
<tr>
<td>$t^{c}_{\alpha}$</td>
<td>5.34</td>
<td>6.01</td>
<td>5.15</td>
</tr>
<tr>
<td>-0.1</td>
<td>29.41</td>
<td>28.53</td>
<td>43.41</td>
</tr>
<tr>
<td>$t^{c}_{\alpha}$</td>
<td>24.01</td>
<td>19.14</td>
<td>31.00</td>
</tr>
<tr>
<td>-0.2</td>
<td>61.49</td>
<td>59.80</td>
<td>86.63</td>
</tr>
<tr>
<td>$t^{c}_{\alpha}$</td>
<td>48.59</td>
<td>37.29</td>
<td>61.10</td>
</tr>
</tbody>
</table>

Table 1: Size and power of single test statistic [%], $\Sigma = I_{K+1}$, $\alpha_e \neq 0$.

We observe the completely instrumentalized test to be mildly oversized for $T = 100$. For larger sample sizes, the size is close to the nominal level. For all studied sample sizes, the size distortions are of larger magnitude for the partially instrumentalized test and increase with growing $K$. On the other hand, the partially instrumentalized test is clearly superior in terms of power: $t^{p}_{\alpha}$ dominates $t^{c}_{\alpha}$ for all sample sizes and all values of $K$, being up to four times more powerful. The power of both $t^{p}_{\alpha}$ and $t^{c}_{\alpha}$ decreases with increasing $K$, mirroring the situation under the null. As expected, the power increases with growing sample size. The test by Banerjee, Dolado and Mestre (1998) is more powerful (even though exogeneity is not provided for, its rejection frequency, for instance for $T = 500$ and $\alpha = -0.1$, is 100%), but this should not come as a surprise, since using instrumental estimation instead of OLS usually leads to less power; in our particular situation, the NIV test statistic also diverges slower under the alternative than the one based on OLS.

For a second series of experiments, $x_{t-1,1}$ is not affected by error correction, i.e. we set $\alpha_e = 0$. In exchange, we let the lack of exogeneity to be caused by correlation between innovations $\varepsilon_t$ and $\nu_t$. More precisely, we use a covariance matrix of the innovations with constant correlation, i.e. we set $\Sigma$ from Assumption 3 to equal $\Sigma_{(K+1) \times (K+1)} = \{\rho_{ij}\}_{1 \leq i, j \leq K+1}$, with $\rho_{ij} = 1$ if $i = j$, and $\rho_{ij} = \rho$ if $i \neq j$. For $\Sigma$ to be positive definite, it must hold that $-1/K < \rho < 1$.

12
We choose for $\rho$ the value of 0.5, to account for positive correlation often observed between macroeconomic time series. The rejection frequencies for the 5% level are given in Table 2.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$T = 100$</th>
<th>$T = 200$</th>
<th>$T = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K = 1$</td>
<td>$K = 2$</td>
<td>$K = 1$</td>
</tr>
<tr>
<td>0</td>
<td>6.07</td>
<td>6.84</td>
<td>5.30</td>
</tr>
<tr>
<td>$t_{10}^p$</td>
<td>6.42</td>
<td>6.52</td>
<td>5.35</td>
</tr>
<tr>
<td>-0.1</td>
<td>22.50</td>
<td>19.96</td>
<td>38.26</td>
</tr>
<tr>
<td>$t_{10}^p$</td>
<td>19.72</td>
<td>16.06</td>
<td>28.96</td>
</tr>
<tr>
<td>-0.2</td>
<td>51.07</td>
<td>41.17</td>
<td>79.09</td>
</tr>
<tr>
<td>$t_{10}^p$</td>
<td>40.37</td>
<td>27.58</td>
<td>50.36</td>
</tr>
</tbody>
</table>

Table 2: Size and power of single test statistic [%], $\rho = 0.5$, $\alpha_e = 0$.

Here, the size distortions of the completely instrumentalized test are larger, especially for $T = 100$, but decrease with growing $T$. The overall image is basically the same, although the size distortions of the partially instrumentalized test decrease faster for increasing sample size than before. The power is marginally lower for each studied case.

Finally, both sources of lack of exogeneity are combined and the respective results are given in Table 3. The findings of the first two series of experiments are confirmed: the power is much higher for $t_{10}^p$ than for $t_{10}^c$, while $t_{10}^c$ has better size properties. The power is again marginally lower than for each source of lack of exogeneity studied separately.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$T = 100$</th>
<th>$T = 200$</th>
<th>$T = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K = 1$</td>
<td>$K = 2$</td>
<td>$K = 1$</td>
</tr>
<tr>
<td>$\alpha = 0$</td>
<td>5.99</td>
<td>6.98</td>
<td>6.02</td>
</tr>
<tr>
<td>$t_{10}^p$</td>
<td>5.82</td>
<td>6.94</td>
<td>5.25</td>
</tr>
<tr>
<td>-0.1</td>
<td>20.76</td>
<td>17.23</td>
<td>31.43</td>
</tr>
<tr>
<td>$t_{10}^p$</td>
<td>16.98</td>
<td>13.98</td>
<td>21.74</td>
</tr>
<tr>
<td>-0.2</td>
<td>46.57</td>
<td>35.51</td>
<td>70.38</td>
</tr>
<tr>
<td>$t_{10}^p$</td>
<td>35.04</td>
<td>22.78</td>
<td>39.26</td>
</tr>
</tbody>
</table>

Table 3: Size and power of single test statistic [%], $\rho = 0.5$, $\alpha_e \neq 0$.

**Summing up**, the partially instrumentalized test statistic performs well in terms of power, and is slightly oversized under the null hypothesis, although, compared to tests that do not account for cross-dependence, these size distortions are not worth mentioning. Opposed to that, the completely instrumentalized test statistic does not behave satisfactorily under the alternative hypothesis; in fact, the power is so low that the better size properties are overshadowed. Both variants of the test statistic are robust to absence of exogeneity.
4.2 Behavior of the panel test

For the panel situation, we examine how well the asymptotic independence of single test statistics is preserved in small samples. We first study a panel of \( N = 2 \) units in detail, and turn our attention to larger cross-sectional dimension afterwards. Having generated a one-sided alternative for the Monte Carlo analysis, we use the test statistic \( \tilde{t}^- \) from Equation (10) to obtain more power.

For this series of experiments, we use the same form of covariance matrix as above; it allows in the panel situation for lack of exogeneity in each unit as well as for dependence across units; \( \Sigma_N = \text{Cov}(\epsilon_{1,t}, \nu'_{1,t}, \epsilon_{2,t}, \nu'_{2,t}) \) is a \((2K + 2) \times (2K + 2)\) matrix,

\[
\text{Cov}(\epsilon_{1,t}, \nu'_{1,t}, \epsilon_{2,t}, \nu'_{2,t}) = \begin{pmatrix}
1 & \rho & \rho & \cdots & \rho & \rho \\
\rho & 1 & \rho & \cdots & \rho & \rho \\
\rho & \rho & \ddots & \cdots & \ddots & \rho \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
\rho & \rho & \cdots & \rho & 1 & \rho \\
\rho & \rho & \cdots & \rho & \rho & 1
\end{pmatrix}.
\]

Of course, any positive definite matrix \( \Sigma_N \) is allowed by our results. However, aside from an economic justification (see O’Connell, 1998), this is a particularly persistent correlation structure (within as well as across units): the largest eigenvalue of \( \Sigma \) is \( O(NK) \). In contrast, Chang’s (2002) Monte Carlo setup for the unit root test uses random eigenvalues between 0 and 1, which leads to weak cross-correlation, see also Im and Pesaran (2003).

The correlation \( \rho \) is again chosen to be 0.5; this implies a correlation of 0.5 between the innovations of the two units.

We only report the results in the situation more relevant for practical applications, i.e. with both sources of endogeneity. The size and power of the panel test are given in Table 4 for \( N = 2 \).

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>( T = 100 )</th>
<th>( T = 200 )</th>
<th>( T = 500 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K = 1 )</td>
<td>( K = 2 )</td>
<td>( K = 1 )</td>
<td>( K = 2 )</td>
</tr>
<tr>
<td>( \alpha = 0 )</td>
<td>( t^0_N )</td>
<td>5.83</td>
<td>6.99</td>
</tr>
<tr>
<td></td>
<td>( t^2_N )</td>
<td>5.10</td>
<td>6.43</td>
</tr>
<tr>
<td>( -0.1 )</td>
<td>( t^0_N )</td>
<td>32.52</td>
<td>24.86</td>
</tr>
<tr>
<td></td>
<td>( t^2_N )</td>
<td>25.48</td>
<td>18.30</td>
</tr>
<tr>
<td>( -0.2 )</td>
<td>( t^0_N )</td>
<td>69.68</td>
<td>54.79</td>
</tr>
<tr>
<td></td>
<td>( t^2_N )</td>
<td>51.09</td>
<td>32.86</td>
</tr>
</tbody>
</table>

Table 4: Size and power of panel test, \( \rho = 0.5, \alpha_e \neq 0, N = 2 \).

The power increases compared to the single unit case, even for the partially instrumentalized test statistic. The most interesting fact, however, is that the magnitude of the size distortions under the null hypothesis is approximatively
the same as for the single test, in spite of a correlation of 0.5 between the innovations of the two units! This suggests that the property of asymptotic independence of single test statistics is well maintained in small samples. This intuition is confirmed by Table 5, where we report mean and standard deviation of a single test statistic ($\mu_i$ and $\sigma_i$), as well as mean and standard deviation of the panel test statistic ($\mu$ and $\sigma$) for $N = 2$, together with the correlation ($\rho$) between the two single test statistics based on which the panel test statistic is computed (Monte Carlo estimates based on 10000 samples).

$$T_i = 100$$

$$T_i = 200$$

$$T_i = 500$$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$K_i = 1$</th>
<th>$K_i = 2$</th>
<th>$K_i = 1$</th>
<th>$K_i = 2$</th>
<th>$K_i = 1$</th>
<th>$K_i = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$t_{\bar{\alpha}}$</td>
<td>$\mu_i$</td>
<td>0.055</td>
<td>0.016</td>
<td>0.029</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>$\sigma_i$</td>
<td>1.098</td>
<td>1.131</td>
<td>1.072</td>
<td>1.082</td>
<td>1.053</td>
</tr>
<tr>
<td>2</td>
<td>$t_{\bar{\alpha}}$</td>
<td>$\mu_i$</td>
<td>0.065</td>
<td>0.001</td>
<td>0.041</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>$\sigma_i$</td>
<td>1.093</td>
<td>1.101</td>
<td>1.048</td>
<td>1.058</td>
<td>1.018</td>
</tr>
<tr>
<td></td>
<td>$\mu$</td>
<td>0.078</td>
<td>0.022</td>
<td>0.041</td>
<td>0.002</td>
<td>-0.023</td>
</tr>
<tr>
<td></td>
<td>$\sigma$</td>
<td>1.199</td>
<td>1.144</td>
<td>1.073</td>
<td>1.094</td>
<td>1.049</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.020</td>
<td>0.022</td>
<td>0.014</td>
<td>0.021</td>
<td>-0.007</td>
</tr>
</tbody>
</table>

Table 5: Characteristics of single and panel ($N = 2$) test statistics under the null hypothesis.

Indeed, the correlation between single test statistics is practically negligible. Even for the worst case ($T = 100$ and $K = 1$), it is rather small: 0.023 compared to its cause, $\rho = 0.5$, and decreases with growing $T$. The moments of the asymptotic distribution are better approximated with complete instrumentalization than with partial instrumentalization, thus explaining the better size properties of $t_{\bar{\alpha}}$.

For the case $N > 2$, we note that the main problem resides not with cross-dependence, but with the distortions single test statistics exhibit. Non-zero mean of single test statistics has the worst effect, since, due to the definition of the panel test statistic, it is multiplied by $\sqrt{N}$. Inflated standard deviations, however, do not affect that much the size of the panel test statistic as the number of units grows. A small-sample correction for single test statistics seems appropriate, especially for centering. Finally, for very large $N$, even small correlations between single test statistics will have negative effect on the size of the panel test, see also Im and Pesaran (2003).

Table 6 studies the size properties of the panel test for $N = 5$ and $N = 10$, together with the power under the alternatives $\alpha_i = -0.1$ and $\alpha_i = -0.2$. The results suggest that complete instrumentalization should be used, if at all, in panels with larger cross-sectional dimension, where a large $N$ may compensate
Table 6: Size and power of panel test, $N = 5$ and $N = 10$. 

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\alpha_i$</th>
<th>$T_i = 100$</th>
<th>$T_i = 200$</th>
<th>$T_i = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$t_0^p$</td>
<td>6.36</td>
<td>7.76</td>
<td>6.27</td>
</tr>
<tr>
<td></td>
<td>$t_0^c$</td>
<td>5.33</td>
<td>6.95</td>
<td>4.85</td>
</tr>
<tr>
<td>5</td>
<td>$-0.1$</td>
<td>$t_5^p$</td>
<td>55.31</td>
<td>42.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_5^c$</td>
<td>42.48</td>
<td>28.82</td>
</tr>
<tr>
<td></td>
<td>$-0.2$</td>
<td>$t_5^p$</td>
<td>92.84</td>
<td>80.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_5^c$</td>
<td>75.94</td>
<td>51.64</td>
</tr>
<tr>
<td>10</td>
<td>$-0.1$</td>
<td>$t_{10}^p$</td>
<td>74.15</td>
<td>60.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_{10}^c$</td>
<td>60.08</td>
<td>40.98</td>
</tr>
<tr>
<td></td>
<td>$-0.2$</td>
<td>$t_{10}^p$</td>
<td>98.06</td>
<td>91.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_{10}^c$</td>
<td>89.43</td>
<td>69.62</td>
</tr>
</tbody>
</table>

for the lack of power of single tests, and the better size properties of single tests result in a better behavior of the panel test under the null hypothesis. For $N = 10$, the power, although increasing compared to $N = 2$ and $N = 5$, is still relatively low even for $T = 500$. For partial instrumentalization, the size distortions are caused by the mean of the test statistic. For instance, with $N = 10$, $T = 100$ and $K = 1$, this mean equals 0.162 for the partially instrumentalized test (estimation from the same Monte Carlo experiment), larger than the mean of the single test statistic by a proportionality factor of about 2.945, close to $\sqrt{10}$. The size properties become less acceptable with growing $K$ and $N$. On the other hand, the power increases with $N$, as expected. **To sum up**, partial instrumentalization is to be preferred as long as $N \ll T$.

## 5 Conclusions

Until recently, panel cointegration studies have been carried out under the assumption of independent units. This not being a plausible assumption, especially in macroeconomic panels, it is expected that such studies lead to biased conclusions. The present paper proposes a test that does not exhibit such shortcoming.

Our test is obtained from the error correction representation of Banerjee, Dolado and Mestre (1998), to which the nonlinear instrumental variable method was applied in a manner similar to Chang (2002).

The proposed test statistic is shown to be asymptotically standard normal distributed and requires no exogeneity assumptions. We find, however, that the residual variance should not be estimated using NIV estimators of the parameters, since these may not converge fast enough to ensure consistent residuals.
Instead, we use the OLS residual variance estimator.

In cross-correlated as well as in cross-cointegrated panels, individual test statistics are shown to be asymptotically independent; thus, a panel cointegration test robust to cross-dependence can be built. Up to a certain degree, panels may also be unbalanced, and no $N$ asymptotics is required.

The included Monte Carlo evidence shows that, in small samples, the property of asymptotic independence of single test statistics is well maintained. This is true only to a lesser degree for the approximation of the small sample distribution of single test statistics by their asymptotic distribution. Our test should be used in panels with small cross-sectional dimension (relative to the time dimension): a typical application would consist in multi-country studies, with several countries observed over a couple of decades.

Appendix

The proofs of the results stated in the paper require following lemma.

**Lemma A** Under the assumptions of Proposition 1, it holds as $T \to \infty$:

A.1
\[
\frac{1}{T^{0.5}} \sum_{t=1}^{T} F^2 (y_{t-1}) \xrightarrow{d} \mathcal{L} (1, 0) \int_{-\infty}^{\infty} F^2 (s) \, ds;
\]

A.2
\[
\frac{1}{T^{0.5}} \sum_{t=1}^{T} y_{t-1} F (y_{t-1}) \xrightarrow{d} \mathcal{L} (1, 0) \int_{-\infty}^{\infty} sF (s) \, ds;
\]

A.3
\[
\frac{1}{T^{0.25}} \sum_{t=1}^{T} F (y_{t-1}) \varepsilon_t \xrightarrow{d} \sigma_t \sqrt{\mathcal{L} (1, 0) \int_{-\infty}^{\infty} F^2 (s) \, ds} \cdot W (1);
\]

A.4
\[
\frac{1}{T^{0.5}} \sum_{t=1}^{T} F (y_{t-1}) \chi_{t-1,0} = O_p (1),
\]

where $W (r)$ is a standard Brownian motion and $\mathcal{L} (t, s)$ is the local Brownian time associated with another Brownian motion $U$, independent of $W$ and having as variance the long-run variance of $\Delta y_{t-1}$:

\[
\mathcal{L}(t, s) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_{0}^{t} \mathbb{1} (|U (r) - s| < \epsilon) \, dr,
\]
with $1(\cdot)$ the usual indicator function, $1(A) = 1$, if proposition $A$ is true, and 0, otherwise.

If, additionally, $\alpha_e = 0$, it holds

$$A.5 \quad \frac{1}{T} \sum_{t=1}^{T} F(y_{t-1}) x'_{t-1,1} = O_p(1);$$

if $\alpha_e \neq 0$, a different behavior of this sample cross-moment can emerge, depending on $\theta$:

$$A.6 \quad \frac{1}{T} \sum_{t=1}^{T} F(y_{t-1}) x'_{t-1,1} = O_p(T^{-0.5}).$$

**Proof of Lemma A**

A.1 follows from de Jong and Wang (2005), Theorem 2, since $F^2$ satisfies their conditions when $F$ obeys our Assumption 4.

A.2 also follows from de Jong and Wang (2005).

A.3 is proven by Demetrescu (2006), see his Lemma 1, item d).

A.4 is a direct consequence of Demetrescu’s Lemma 1, item e).

A.5 is shown to hold true under no cointegration (i.e. $\alpha = 0$ and $\alpha_e = 0$) by Chang, Park and Phillips (2001), Lemma 5.

A.6 Under cointegration of $y_{t-1}$ and $x_{t-1,1}$ (as implied by $\alpha_e \neq 0$), each element of $x_{t-1,1}$ can be expressed as a linear combination of $I(1)$ variables that are either cointegrated with $y_{t-1}$ or not. When at least one of the $I(1)$ variables that are not cointegrated with $y_{t-1}$ is present in the linear combinations, the cross-moment has order $O_p(1)$ due to A.5; otherwise, when elements of $x_{t-1,1}$ equal $y_{t-1}$ plus $I(0)$ noise, the $O_p(T^{-0.5})$ order emerges due to A.2 and A.4.

**Proof of Proposition 1**

a) Recall, $\hat{\alpha} - \alpha = Q^{-1}M$, with $M$ and $Q$ defined in the text. Let us now examine the behavior of $M$. The second term on the right-hand side of the equation defining $M$ can be written as

$$\left( \sum_{t=1}^{T} F(y_{t-1}) x'_{t-1} \right) D_T^{-1} D_T \left( \sum_{t=1}^{T} x_{t-1} x'_{t-1} \right)^{-1} D_T D_T^{-1} \left( \sum_{t=1}^{T} x_{t-1} \varepsilon_t \right)$$
with $D_T$ a \(((K + 1)p + K) \times ((K + 1)p + K)\) diagonal matrix partitioned according to the stationary and integrated components of $x_{t-1}$:

$$D_T = \begin{pmatrix} T^{0.5} & 0 \\ 0 & T \end{pmatrix}, \quad D_T^{-1} = \begin{pmatrix} T^{-0.5} & 0 \\ 0 & T^{-1} \end{pmatrix}. $$

It follows that

$$\left( \sum_{t=1}^{T} F(y_{t-1}) x_{t-1}^{\prime} \right) \left( \begin{array}{cc} T^{-0.5} & 0 \\ 0 & T^{-1} \end{array} \right) = \left( \begin{array}{cc} T^{-0.5} \sum_{t=1}^{T} F(y_{t-1}) x_{t-1,0}^{\prime} & T^{-1} \sum_{t=1}^{T} F(y_{t-1}) x_{t-1,1}^{\prime} \\ \end{array} \right) = (O_p(1), O_p(1))$$

and

$$\left( \begin{array}{cc} T^{-0.5} & 0 \\ 0 & T^{-1} \end{array} \right) \left( \sum_{t=1}^{T} x_{t-1} x_{t-1}^{\prime} \right)^{-1} \left( \begin{array}{cc} T^{-0.5} \sum_{t=1}^{T} x_{t-1,0} x_{t-1,0}^{\prime} & T^{-1} \sum_{t=1}^{T} x_{t-1,1} x_{t-1,1}^{\prime} \\ \end{array} \right)^{-1} = \left( \begin{array}{cc} O_p(1) & O_p(T^{-0.5}) \\ O_p(T^{-0.5}) & O_p(1) \end{array} \right),$$

see Lemma A.4 and A.5. Then,

$$D_T \left( \sum_{t=1}^{T} x_{t-1} x_{t-1}^{\prime} \right)^{-1} D_T^{-1} = \left( \begin{array}{cc} T^{-0.5} \sum_{t=1}^{T} x_{t-1,0} x_{t-1,0}^{\prime} & T^{-1} \sum_{t=1}^{T} x_{t-1,1} x_{t-1,1}^{\prime} \\ \end{array} \right) = \left( \begin{array}{cc} O_p(1) & O_p(T^{-0.5}) \\ O_p(T^{-0.5}) & O_p(1) \end{array} \right),$$

due to continuity of matrix inversion and nonsingularity. Hence,

$$M = \sum_{t=1}^{T} F(y_{t-1}) \varepsilon_t - (O_p(1), O_p(1)) \left( \begin{array}{cc} O_p(1) & O_p(T^{-0.5}) \\ O_p(T^{-0.5}) & O_p(1) \end{array} \right) = O_p(T^{0.25}).$$

For $Q$, we only need to examine

$$\left( \begin{array}{cc} T^{-0.5} & 0 \\ 0 & T^{-1} \end{array} \right) \left( \sum_{t=1}^{T} x_{t-1} y_{t-1} \right) = \left( \begin{array}{cc} T^{0.5} \left( \sum_{t=1}^{T} x_{t-1,0} y_{t-1} \right) & T \left( \sum_{t=1}^{T} \sum_{t=1}^{T} x_{t-1,1} y_{t-1} \right) \\ \end{array} \right) = \left( \begin{array}{cc} O_p(T^{0.5}) & O_p(T) \end{array} \right),$$

19
which, under no cointegration, leads to

\[
Q = \sum_{t=1}^{T} F(y_{t-1})y_{t-1} - (O_{p}(1), O_{p}(1)) \left( \begin{array}{cc} O_{p}(1) & O_{p}(T^{-0.5}) \\ O_{p}(T^{-0.5}) & O_{p}(1) \end{array} \right) \left( \begin{array}{c} O_{p}(T^{0.5}) \\ O_{p}(T) \end{array} \right)
\]

\[
= O_{p}(T^{0.5}) - (O_{p}(1), O_{p}(1)) \left( \begin{array}{c} O_{p}(T^{0.5}) \\ O_{p}(T) \end{array} \right)
\]

\[
= O_{p}(T^{0.5}) - (O_{p}(1) + O_{p}(T)) = O_{p}(T),
\]

from which the convergence rate for \( \hat{\alpha} \) follows directly,

\[
\hat{\alpha} - \alpha = O_{p}(T^{-1}) O_{p}(T^{0.25}) = O_{p}(T^{-0.75}).
\]

For \( \beta \), we have

\[
\hat{\beta} - \beta = J^{-1} R
\]

with \( R \) a column vector:

\[
R = \sum_{t=1}^{T} x_{t-1} \xi_{t} - \sum_{t=1}^{T} x_{t-1}y_{t-1} \left( \sum_{t=1}^{T} F(y_{t-1})y_{t-1} \right)^{-1} \sum_{t=1}^{T} F(y_{t-1}) \xi_{t}
\]

and \( J \) a matrix

\[
J = \sum_{t=1}^{T} x_{t-1}x'_{t-1} - \sum_{t=1}^{T} x_{t-1}y_{t-1} \left( \sum_{t=1}^{T} F(y_{t-1})y_{t-1} \right)^{-1} \sum_{t=1}^{T} F(y_{t-1})x'_{t-1}.
\]

Split \( R \) and \( J \) corresponding to the stationary and integrated regressors. Then, it is straightforward to check that for \( R = (R_{0}, R_{1})' \) it holds

\[
R_{0} = O_{p}(T^{0.75}) \quad \text{and} \quad R_{1} = O_{p}(T^{1.75}).
\]

Then, for

\[
J = \left( \begin{array}{cc} A & B \\ C & D \end{array} \right),
\]

it holds due to Lemma A that \( A = O_{p}(T) \), \( B = O_{p}(T^{1.5}) \), \( C = O_{p}(T^{2}) \) and \( D = O_{p}(T^{2.5}) \). Using formulae for inverting partitioned matrices (e.g. Lütkepohl, 1996, p. 147), one obtains, after some algebra,

\[
J^{-1} = \left( \begin{array}{cc} O_{p}(T^{-1}) & O_{p}(T^{-2}) \\ O_{p}(T^{-1.5}) & O_{p}(T^{-2.5}) \end{array} \right),
\]

from which the desired convergence rates follow.

b) The result follows along the same lines, but now A.6 could hold instead of A.5 (see the proof of Lemma A for details). While \( M \) remains in any case of order \( O_{p}(T^{0.25}) \), \( Q \) can be of order \( O_{p}(T^{0.5}) \) instead of \( O_{p}(T) \), resulting in a convergence order for \( \hat{\alpha} \) of \( O_{p}(T^{0.25}) \). In what concerns \( \hat{\beta} \), \( B \) may be \( O_{p}(T) \) and

20
D may be $O_p(T^2)$, if A.6 holds. Thus, a behavior similar to that of $\hat{\alpha}$ emerges for $\hat{\beta}_1$, while the behavior of $\hat{\beta}_0$ is unaffected.

**Proof of Lemma 1** From the proof of Proposition 1, it follows

$$M = \frac{1}{T^{0.25}} \sum_{i=1}^{T} F(y_{i-1}) \varepsilon_i - o_p(1).$$

We also have by arguments similar to those in the proof of Proposition 1 that

$$P = \sum_{i=1}^{T} F(y_{i-1})^2 - (O_p(1), O_p(1)) \left( \begin{array}{cc} O_p(1) & O_p(T^{-0.5}) \\ O_p(T^{-0.5}) & O_p(1) \end{array} \right) \left( \begin{array}{c} O_p(1) \\ O_p(1) \end{array} \right),$$

and thus

$$\frac{1}{T^{0.5}} P = \frac{1}{T^{0.5}} \sum_{i=1}^{T} F(y_{i-1})^2 - o_p(1),$$

which leads to the desired result, the numerator of the $t$ statistic being different from zero with probability 1. It is obvious that this relation holds irrespective of which behavior, A.5 or A.6, holds.

**Proof of Proposition 2** The result follows directly from Lemma 1, if joint convergence of A.3 and A.1 holds. This is indeed the case, since the proof of A.3 given by Demetrescu (2006) establishes A.3 as an implication of A.1.

**Proof of Proposition 4** Chang (2002) argues that, as $T \to \infty$,

$$\frac{1}{\sqrt{T}} \sum_{i=1}^{T} F_i (y_{i,t-1}) \varepsilon_{i,t} \approx_d \sqrt{T} \int_{0}^{1} F_i \left( \sqrt{T} U_i \right) dW_i,$$

where $\approx_d$ stands for equivalence in distribution. This also applies for recursively demeaned (detrended) data, $y_{i,t-1}$ or $y_{j,t-1}$, but in relation with the corresponding recursively demeaned (detrended) Brownian motions. The left-hand side of equation above is the numerator of the individual test statistic. For two units $i \neq j$, the equivalent right-hand side representations are independent if and only if their quadratic covariation, given by

$$\sigma_{ij} \sqrt{T} \int_{0}^{1} F_i \left( \sqrt{T} U_i \right) F_j \left( \sqrt{T} U_j \right) ds,$$

where $\sigma_{ij}$ is the covariance of $\varepsilon_{i,t}$ and $\varepsilon_{j,t}$, disappears almost surely as $T \to \infty$ (see Chang, Park and Phillips, 2001). For independent units, this condition
holds trivially. But even if $\sigma_{ij} \neq 0$, it is known (see Kasahara and Kotani, 1979) that

$$\int_0^1 F_i \left( \sqrt{T}U_i \right) F_j \left( \sqrt{T}U_j \right) ds = O_p \left( \frac{\ln T}{T} \right).$$

The denominators are asymptotically uncorrelated, since, as Chang, Park and Phillips (2001, Lemma 5, item j) show, $T^{-0.5} \sum_{t=1}^T F_i^2 (y_{i,t-1}) F_j^2 (y_{j,t-1}) = o_p(1)$, the function $F^2(\cdot)$ being itself regularly integrable, if $F(\cdot)$ is. This establishes asymptotic independence, from which the first result follows. Consistency is obvious given Proposition 1, item b).

References


