

Panel Data Cointegration Testing with Structural Instabilities

Anindya Banerjee^a and Josep Lluís Carrion-i-Silvestre^b

^aDepartment of Economics, University of Birmingham Edgbaston, Birmingham, UK; ^bUniversity of Barcelona, Barcelona, Spain

ABSTRACT

Spurious regression analysis in panel data when the time series are cross-section dependent is analyzed in the article. The set-up includes (possibly unknown) multiple structural breaks that can affect both the deterministic and the common factor components. We show that consistent estimation of the long-run average parameter is possible once cross-section dependence is controlled using cross-section averages in the spirit of Pesaran's common correlated effects approach. This result is used to design individual and panel cointegration test statistics that accommodate the presence of structural breaks that can induce parameter instabilities in the deterministic component, the cointegration vector and the common factor loadings.

ARTICLE HISTORY

Received November 2021
Accepted February 2024

KEYWORDS

Common factors;
Cross-section dependence;
Spurious regression;
Structural breaks

1. Introduction

The literature on nonstationary panel data has experienced important developments in recent times, analyzing the properties of parameter estimation under both spurious regression and cointegration. One interesting feature of working in a nonstationary panel data framework is that it is possible to obtain consistent estimates of statistical relationships regardless of whether there is cointegration or spurious regression—see Phillips and Moon (1999). The idea behind the consistent estimation of (the long-run average) parameters in a spurious panel relies on the fact that panel units add information that lead to detect a stronger overall signal than that of the pure time series case—see Phillips and Moon (2000). This article adds to this body of the panel data literature by defining a model specification under spurious regression that allows for the presence of multiple structural breaks affecting the level, trend and/or the slope parameters of the model, as well as in the factor structure generating the cross-section dependence. Panel data units are assumed to be cross-section dependent which is captured by an approximate common factor model.


There are a number of contributions in the literature that are relevant for our proposal. Gengenbach, Urbain, and Westerlund (2016) and Banerjee and Carrion-i-Silvestre (2017) design panel data cointegration tests with cross-section dependence, but without structural breaks. Although the two papers differ in the degree of heterogeneity that is allowed for in the potential cointegration vector—the former paper considers a heterogeneous cointegration vector across the units of the panel, whereas the latter assumes homogeneity—both follow the spirit of the common correlated effects (CCE) estimator advocated in Pesaran (2006) and show that the common factors that drive the cross-section dependence can be proxied through the use of cross-section averages of the variables that appear in the model.

CCE-methodology offers the advantage of avoiding the need to estimate the number of common factors (m^0), although m^0 is required to be bounded by the number of observable stochastic processes $(1 + k)$ that appear in the model specification—the so-called *rank condition*, $m^0 \leq 1 + k$.

Westerlund and Edgerton (2008) and Banerjee and Carrion-i-Silvestre (2015) define panel data cointegration statistics accounting for multiple structural breaks—that can affect the deterministic component and/or the slope parameters of the model—and cross-section dependence—modeled with a common factor model. The framework of these papers allows for complete heterogeneity in the sense that all parameters are panel unit specific—that is, the break dates, the deterministic component and the slope parameters are allowed to be heterogeneous across units. The common factors are estimated using principal components as in Bai and Ng (2004) so that it is possible to estimate the number of common factors using information criteria. However, it is worth noting that Westerlund and Edgerton (2008) assume that all common factors are $I(0)$ stationary, whereas Banerjee and Carrion-i-Silvestre (2015) allow for a combination of $I(0)/I(1)$ common factors.

The contribution of the present article can be summarized as follows. First, the article shows that it is possible to obtain a consistent pooled estimate of the long-run average coefficient, which captures a statistical relationship among non-individually cointegrated variables—see Phillips and Moon (1999). The model specification allows for the presence of multiple structural breaks that can affect the long-run average coefficient. This result is of special relevance from an empirical point of view, since it implies that practitioners can obtain consistent estimates of the slope parameters using a panel data pooled estimator regardless of whether the variables define a cointegration relationship. This is in direct contrast to the findings

CONTACT Anindya Banerjee  banerjeeanindya250@gmail.com  Department of Economics, University of Birmingham Edgbaston, Birmingham, B15 2TT, UK.

 Supplementary materials for this article are available online. Please go to www.tandfonline.com/UBES.

© 2024 The Authors. Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

in the papers mentioned in the previous paragraph that assume heterogeneous potential cointegration vectors and estimate the parameters unit by unit, which delivers inconsistent estimators of the parameters under spurious regression. Second, the fact that the panel cointegration statistic is based on the pooled estimation of the potential cointegration vector, removes the presence of the stochastic regressors in the limiting distribution of the cointegration statistic—that is, we move from a panel data cointegration analysis framework to a panel data unit root one. This produces a spin-off contribution, since the statistic that is proposed here extends the Pesaran (2007) panel data unit root test by allowing for structural breaks in the same way as Perron (1990) and Kim and Perron (2009) extended the Dickey-Fuller framework in time series analysis. Third, our setup also deals with the case where the factor loadings can be affected by structural breaks, a situation that has not been previously investigated in panel data cointegration testing analysis. This adds more flexibility and increases the interest of the model from an empirical point of view. Fourth, the common factors are estimated using the CCE approach which avoids the need to estimate the number of common factors as long as the rank condition is satisfied. While the CCE approach does not pose any problem for the computation of the pooled estimator of the slope coefficients, we show there are potentially important issues at the testing stage. This leads us to suggest a conservative testing strategy that mitigates the need to know both the number and order of integration of the common factors. Finally, the article proves that in some cases it is possible to obtain a consistent estimate of the break fraction parameters under spurious regression when the break dates are unknown. This is an interesting result per se, adding to the results obtained in Phillips and Moon (1999) concerning the consistent estimation of the long-run average coefficient under spurious regression.

Taken together, our article provides a panel cointegration testing framework that explicitly models the effects of structural breaks. Westerlund (2018) had argued that an approximate common factor model—estimated using the CCE approach—can result in a simple device that might capture unattended structural breaks. In principle this is an interesting feature that could act as a model-safeguarding device, avoiding the need to define a particular deterministic specification, at the risk of violating the rank condition. We argue however that if the presence of breaks is a reasonable feature of the data generation process, it would be better to undertake explicit modeling of these breaks and free the common factors to capture other potential misspecification errors and dependences that might affect the model.

The article is organized as follows. Section 2 describes the model, Section 3 defines the pooled CCE estimator of the slope parameters and deals with the estimation of the structural break dates. The consistency of the pooled CCE estimator provides the ground upon which individual and panel data cointegration statistics are proposed in Section 4. Section 5 conducts an extensive Monte Carlo simulation experiment to analyse the finite sample performance of the statistics that have been proposed in the article. An empirical illustration that focuses on house prices and per capita disposable income for the U.S. states is conducted in Section 6. Finally, Section 7 concludes. Supplementary material is collected in the appendices with the proofs, tables of critical values, and tables and figures that summarize the finite

sample performance of the statistics. A GAUSS code program is available to implement the proposal that is designed in the article.

2. The model

Let $Y_{i,t} = (y_{i,t}, x'_{i,t})'$ be a $(1+k)$ -vector of $I(1)$ stochastic processes with data generating process (DGP):

$$Y_{i,t} = \Theta_i D_t + \pi_{i,t} F_t + U_{i,t} \quad (1)$$

$$(I - L) F_t = v_t \quad (2)$$

$$(I - L) U_{i,t} = e_{i,t}, \quad (3)$$

where Θ_i is a $((1+k) \times (J^0 + 1))$ matrix of coefficients with rows defined as $\theta_{i,l} = (\theta_{i,l,1,1}, \dots, \theta_{i,l,J^0+1,1}, \theta_{i,l,1,2}, \dots, \theta_{i,l,J^0+1,2})$, $l = 1, \dots, (1+k)$, $D_t = (1, DU'_t, t, DT'_t)'$, $DU_t = (DU_{1,t}, \dots, DU_{J^0,t})'$, $DT_t = (DT_{1,t}, \dots, DT_{J^0,t})'$ with $DU_{j,t} = 1(t > T_j^0)$, $DT_{j,t} = (t - T_j^0)1(t > T_j^0)$, $1(\cdot)$ the indicator function, $T_j^0 = \lfloor \lambda_j^0 T \rfloor$ the j th break date, $\lambda_j^0 = T_j^0/T$ the break fraction and $\lfloor \cdot \rfloor$ the integer part, $j = 1, \dots, J^0$, with the convention that $T_0^0 = 0$ and $T_{J^0+1}^0 = T$ —that is, $\lambda_0^0 = 0$ and $\lambda_{J^0+1}^0 = 1$. Throughout the article, the superscript “0” indicates the true value of the corresponding figure. The deterministic component covers two variants: (a) a constant with level shifts—henceforth, constant case—and (b) a time trend with both level and slope shifts—in what follows, time trend case. The set of generic break fraction parameters vector is defined as $\Lambda = \{\lambda = (\lambda_1, \dots, \lambda_{J^0}) | \epsilon < \lambda_1 < \dots < \lambda_{J^0} < 1 - \epsilon \text{ and } |\lambda_j - \lambda_{j+1}| > \epsilon\}$, with ϵ denoting the amount of trimming that is specified—popular choices are $\epsilon \in \{0.1, 0.15, 0.2, 0.25\}$. At this stage, the model set-up assumes that the number and position of the common structural breaks are known, but the procedure that can be implemented to estimate them is discussed below.

To begin with, cross-section dependence in the panel is defined by F_t which is a m^0 -vector of observable common factors and $\pi_{i,t} = \pi_{i,j}$, $T_{j-1}^0 < t \leq T_j^0$, the $((1+k) \times m^0)$ matrix of factor loadings. Denote by $K = \sigma(F_0, \dots, F_t, \dots)$ the sigma field generated by the sequence $\{F_t\}_{t=0}^\infty$ so that, conditionally on K , $U_{i,t} = (U_{y,i,t}, U'_{x,i,t})'$ are independent across i —see Urbain and Westerlund (2011). Let $M < \infty$ be a generic positive number, not depending on T and N , and define the Euclidean norm of a generic matrix A as $\|A\| = \text{trace}(A'A)^{1/2}$. The vector of stochastic processes $V_{i,t} = (v'_t, e'_{i,t})'$, $i = 1, \dots, N$, $t = 1, \dots, T$, is assumed to satisfy the following assumptions—see Bai and Ng (2004) and Banerjee and Carrion-i-Silvestre (2015).

Assumption 1. (i) $v_t = C_j^F(L) w_t$, $w_t \sim \text{iid}(0, \Sigma_{w_j})$, $E\|w_t\|^4 \leq M$, $T_{j-1} < t \leq T_j$, and (ii) $\text{var}(\Delta F_t) = \sum_{l=0}^\infty C_{j,l}^F \Sigma_{w_j} C_{j,l}^{F'} > 0$, $T_{j-1} < t \leq T_j$, (iii) $\sum_{l=0}^\infty l \|C_{j,l}^F\| < M$; and (iv) $C_j^F(1)$ has rank $m_1^0 \forall j$, $0 \leq m_1^0 \leq m^0$, $j = 1, \dots, J^0 + 1$.

Assumption 2. (i) For each i , $e_{i,t} = C_{ij}^U(L) \varepsilon_{i,t}$, $\varepsilon_{i,t} \sim \text{iid}(0, \sigma_{\varepsilon_{ij}}^2)$, $E\|\varepsilon_{i,t}\|^8 \leq M$, $\sum_{l=0}^\infty l \|C_{ij,l}^U\| < M$, $\omega_{ij}^2 = C_{ij}^U(1)^2 \sigma_{\varepsilon_{ij}}^2 > 0$; (ii) $E(\varepsilon_{i,t} \varepsilon'_{i,t}) = \tau_{i,l}$ with $\sum_{i=1}^N \|\tau_{i,l}\| \leq M$ for all l , $T_{j-1} < t \leq T_j$; (iii) $E\left\|\frac{1}{\sqrt{N}} \sum_{i=1}^N [\varepsilon_{i,t} \varepsilon'_{i,t} - E(\varepsilon_{i,t} \varepsilon'_{i,t})]\right\|^4 \leq M$, for every (t, s) ;

and (iv) $C_{ij}^U(1)$ is positive definite almost surely for all i and j , $j = 1, \dots, J^0 + 1$.

Assumption 3. The errors $\varepsilon_{i,t}$, w_t and loadings π_i are three mutually independent groups across i , t and $(1 + k)$ dimensions.

Assumption 4. $E\|F_0\| \leq M$, and for every $i = 1, \dots, N$, $E\|U_{i,0}\| \leq M$.

Assumption 5. As $T_j^0 \rightarrow \infty$ and $T \rightarrow \infty$, $T_j^0/T \rightarrow \lambda_j^0$, $j = 1, \dots, J^0$, with $0 < \lambda_1^0 < \dots < \lambda_{J^0}^0 < 1$.

Assumptions 1–5 ensure that the central limit theorem (CLT) holds for $V_{i,t}$, so that, for the most general specification that is considered in this article, we have:

$$T^{-1/2} \sum_{t=1}^{[rT]} V_{i,t} \Rightarrow \sum_{k=1}^{j-1} [C_{i,k}(1) - C_{i,k+1}(1)] W_i(\lambda_k^0) + C_{i,j}(1) W_i(r); r \in [\lambda_{j-1}^0, \lambda_j^0],$$

as $T \rightarrow \infty$ for all i , where $W_i(r)$ is a $(m^0 + 1 + k)$ -vector of standard Brownian motion on $r \in [\lambda_{j-1}^0, \lambda_j^0]$, and

$$\Omega_{ij} = C_{ij}(1) C_{ij}(1)' = \begin{bmatrix} C_j^F(1) C_j^F(1)' & C_j^F(1) C_{ij}^U(1)' \\ C_{ij}^U(1) C_j^F(1)' & C_{ij}^U(1) C_{ij}^U(1)' \end{bmatrix},$$

is the long-run conditional covariance matrix with expected value $E(\Omega_{ij}) = \Omega_j$, $j = 1, \dots, J^0 + 1$, the so-called long-run conditional average covariance matrix—see Lemma 1 in the appendix. The matrix Ω_{ij} can be partitioned to define $\Omega_{U_{ij}U_{ij}} = C_{ij}^U(1) C_{ij}^U(1)'$, $\Omega_{U_{x_{ij}}U_{x_{ij}}} = C_{ij}^{U_x}(1) C_{ij}^{U_x}(1)'$ and $\Omega_{U_{x_{ij}}U_{y_{ij}}} = C_{ij}^{U_x}(1) C_{ij}^{U_y}(1)'$, $j = 1, \dots, J^0 + 1$.

The model considers the case where $x_{i,t}$ are assumed to be either cross-section independent—imposing all, but the first, rows of $\pi_{i,t}$ to be equal to zero—or cross-section dependent with dependence driven by F_t . Furthermore, it is possible to assume that the set of observable common factors affecting the endogenous variable $y_{i,t}$ is different from those affecting $x_{i,t}$, a situation that is covered if we define $\pi_{i,t}$ to be a block-diagonal matrix.

Despite the presence of the operator $(I - L)$ in (2), F_t can be $I(0)$, $I(1)$, or a combination of both, depending on the rank of $C_j^F(1)$, $j = 1, \dots, J^0 + 1$. Let m_0^0 and m_1^0 be the number of $I(0)$ and $I(1)$ common factors, respectively, so that $m^0 = m_0^0 + m_1^0$. If $C_j^F(1) = 0$, then F_t is $I(0)$ and $m_0^0 = m^0$. If $C_j^F(1)$ is of full rank, then each component of F_t is $I(1)$ and $m_1^0 = m^0$. If $C_j^F(1) \neq 0$, but not full rank, then some components of F_t are $I(1)$ and some are $I(0)$. Following Bai, Kao and Ng (2009) and Banerjee and Carrion-i-Silvestre (2015), the inclusion of common factors in the potential long-run relationship implies a change in the standard concept of cointegration. The usual definition of cointegration among $Y_{i,t}$ requires F_t to be $I(0)$, so that $Y_{i,t}$ captures all the common stochastic trends. However, allowing F_t to be $I(1)$ is also relevant from an empirical point of view since F_t might be accounting for effects that are not captured by $Y_{i,t}$ alone. Another interesting feature is that (1)

Table 1. Model specifications.

Model	Deterministic component Constant	Trend	Slope parameters	Factor loadings
A1	$\mu_{ij,1} \neq \mu_{ij+1,1} \forall j$	$\mu_{ij,2} = 0 \forall j$	$\beta_j = \beta_0 \forall j$	$\eta_{ij} = \eta_{i,0} \forall j$
A2	$\mu_{ij,1} \neq \mu_{ij+1,1} \forall j$	$\mu_{ij,2} \neq \mu_{ij+1,2} \forall j$	$\beta_j = \beta_0 \forall j$	$\eta_{ij} = \eta_{i,0} \forall j$
B1	$\mu_{ij,1} \neq \mu_{ij+1,1} \forall j$	$\mu_{ij,2} = 0 \forall j$	$\beta_j \neq \beta_{j+1} \forall j$	$\eta_{ij} = \eta_{i,0} \forall j$
B2	$\mu_{ij,1} \neq \mu_{ij+1,1} \forall j$	$\mu_{ij,2} \neq \mu_{ij+1,2} \forall j$	$\beta_j \neq \beta_{j+1} \forall j$	$\eta_{ij} = \eta_{i,0} \forall j$
C1	$\mu_{ij,1} \neq \mu_{ij+1,1} \forall j$	$\mu_{ij,2} = 0 \forall j$	$\beta_j \neq \beta_{j+1} \forall j$	$\eta_{ij} \neq \eta_{ij+1} \forall j$
C2	$\mu_{ij,1} \neq \mu_{ij+1,1} \forall j$	$\mu_{ij,2} \neq \mu_{ij+1,2} \forall j$	$\beta_j \neq \beta_{j+1} \forall j$	$\eta_{ij} \neq \eta_{ij+1} \forall j$

specifies time dependent factor loadings so that the structural breaks can also affect the way in which the common factors affect panel units. Finally, the DGP also allows for the possibility that the structural breaks do not affect all elements in $Y_{i,t}$, and also covers intermediate situations where, for instance, the deterministic component and/or the loadings do not change across regimes for some of the elements in $Y_{i,t}$.

The most general model specification that is admitted in our framework is

$$y_{i,t} = \mu_{ij,1}^y + \mu_{ij,2}^y(t - T_{j-1}^0) + x_{i,t}'\beta_j + \eta_{ij}^y F_t + \xi_{i,t}^y \quad (4)$$

$$x_{i,t} = \mu_{ij,1}^x + \mu_{ij,2}^x(t - T_{j-1}^0) + \eta_{ij}^x F_t + \xi_{i,t}^x, \quad (5)$$

$\mu_{ij,k} = (\mu_{ij,k}^y, \mu_{ij,k}^x)'$, $k \in \{1, 2\}$, $\eta_{ij} = (\eta_{ij}^y, \eta_{ij}^x)'$, $T_{j-1}^0 < t \leq T_j^0$, $j = 1, \dots, J^0 + 1$. Combining (4) and (5) we obtain:

$$\begin{pmatrix} y_{i,t} \\ x_{i,t} \end{pmatrix} = \begin{pmatrix} \beta_j' \mu_{ij,1}^x + \mu_{ij,1}^y & \beta_j' \mu_{ij,2}^x + \mu_{ij,2}^y \\ \mu_{ij,1}^x & \mu_{ij,2}^x \end{pmatrix} d_t + \begin{pmatrix} \beta_j' \eta_{ij}^x + \eta_{ij}^y \\ \eta_{ij}^x \end{pmatrix} F_t + \begin{pmatrix} \beta_j' \xi_{i,t}^x + \xi_{i,t}^y \\ \xi_{i,t}^x \end{pmatrix},$$

that is,

$$\begin{pmatrix} y_{i,t} \\ x_{i,t} \end{pmatrix} = \begin{pmatrix} \mu_{ij,1}^{y*} & \mu_{ij,2}^{y*} \\ \mu_{ij,1}^{x*} & \mu_{ij,2}^{x*} \end{pmatrix} d_t + \begin{pmatrix} \pi_{i,t}^{y*} \\ \pi_{i,t}^{x*} \end{pmatrix} F_t + \begin{pmatrix} \xi_{i,t}^{y*} \\ \xi_{i,t}^{x*} \end{pmatrix}, \quad (6)$$

$d_t = (1, (t - T_{j-1}^0))'$, which corresponds to (1) given that $\theta_{i,1,j,k} = \theta_{ij,k}^y = \sum_{l=1}^j \mu_{i,l,k}^{y*}$, $\mu_{i,l,k}^{y*} = \beta_l' \mu_{i,l,k}^x + \mu_{i,l,k}^y$, $\theta_{i,j,k}^x = \sum_{l=1}^j \mu_{i,l,k}^{x*}$, $k \in \{1, 2\}$, $\pi_{i,t} = (\pi_{i,t}^{y*}, \pi_{i,t}^{x*})'$, $\pi_{i,t}^{y*} = \beta_j' \eta_{ij}^x + \eta_{ij}^y$, $\pi_{i,t}^{x*} = \eta_{ij}^x$, and $U_{i,t} = (\xi_{i,t}^{y*}, \xi_{i,t}^{x*})'$, $\xi_{i,t}^{y*} = \sum_{l=1}^j \beta_l' \xi_{i,t}^x + \xi_{i,t}^y$, $T_{j-1}^0 < t \leq T_j^0$, $j = 1, \dots, J^0 + 1$. Although the model has been written as a pure structural change model in which all parameters change, we can impose parameter constraints in (4) to obtain partial structural change specifications. Our article distinguishes six different model specifications depending on whether the structural breaks affect the deterministic component, the slope parameters and/or the factor loadings. Table 1 summarizes all the cases covered in the article. Model A assumes that the structural breaks affect only the deterministic component, Model B considers that the structural breaks affect both the deterministic component and the slope parameters and, finally, Model C permits the structural breaks to affect the deterministic component, the slope parameters and the factor loadings. For each model specification we consider the two variants of the deterministic component—designated by a 1 for the constant case and by a 2 for the time trend case. This defines Models A1, A2, B1, B2, C1, and C2.

The block-orthogonal model specification given by (4) for each panel unit is:

$$y_i = D_l(\lambda^0)\mu_{i,1}^y + D_\tau(\lambda^0)\mu_{i,2}^y + x_i^l(\lambda^0)\beta + F^l(\lambda^0)\eta_i^{y'} + \xi_i^y, \quad (7)$$

with $l \in \{A, B, C\}$ denoting the model and y_i and ξ_i being T -vectors of the dependent variable and disturbance term, respectively, for the i th panel unit. $D_l(\lambda^0) = [D_{l,1} \cdots D_{l,J^0+1}]_{T \times (J^0+1)} = \text{diag}(\iota_1, \dots, \iota_{J^0+1})$ and $D_\tau(\lambda^0) = [D_{\tau,1} \cdots D_{\tau,J^0+1}]_{T \times (J^0+1)} = \text{diag}(\tau_1, \dots, \tau_{J^0+1})$ are diagonal matrices defined with the elements between parentheses, ι_j is a $(T_j^0 - T_{j-1}^0)$ -vector of ones, $\tau_j = (1, 2, \dots, T_j^0 - T_{j-1}^0)'$ is the time trend of the j th regime, and $\mu_{i,k}^y = (\mu_{i,1,k}^y, \dots, \mu_{i,J^0+1,k}^y)'$, $k \in \{1, 2\}$. For Model A, $x_i^A(\lambda^0) = x_i = [x_{i,1} \cdots x_{i,k}]_{(T \times k)}$ and $\beta = \beta_0$ is a k -vector of parameters—subscript “0” denotes that the parameters are not affected by structural breaks. For Models B and C:

$$x_i^l(\lambda^0) = [x_i(1) \ x_i(2) \ \dots \ x_i(J^0 + 1)]_{(T \times k(J^0+1))}; \quad l \in \{B, C\}, \quad (8)$$

with $x_i(j) = [(D_{i,j} \odot x_{i,1}) \cdots (D_{i,j} \odot x_{i,k})]$, $j = 1, \dots, J^0 + 1$, where \odot is the element-wise (Hadamard) product and $\beta = (\beta_1', \dots, \beta_{J^0+1}')'$. Similarly, $F^l(\lambda^0) = F = [F_1 \cdots F_{m^0}]_{(T \times m^0)}$, $l \in \{A, B\}$, and $\eta_i^y = \eta_{i,0}^y$ is a m^0 -row vector of loadings, whereas for Model C:

$$F^C(\lambda^0) = [F(1) \ F(2) \ \dots \ F(J^0 + 1)]_{(T \times m^0(J^0+1))}, \quad (9)$$

with $F(j) = [(D_{i,j} \odot F_1) \cdots (D_{i,j} \odot F_{m^0})]$, $j = 1, \dots, J^0 + 1$, and $\eta_i^y = (\eta_{i,1}^y, \dots, \eta_{i,J^0+1}^y)$ is the $m^0(J^0 + 1)$ -row vector of factor loadings. The estimation of β in (7) can be performed using the panel data pooled estimator, although it would require both the common factors and the structural break dates to be known. Certainly, there are some cases where the common factors can be thought to be observable as discussed in Pesaran (2015), but this situation is rarely found in practice. The same applies to the structural break dates, so that the next section designs an estimation procedure that will allow the empirical implementation of our proposal when both the common factors and the structural breaks are unknown—the known common factors and/or structural breaks situation can be seen as particular cases.

3. Unknown Common Factors and Structural Breaks

The discussion that is conducted in this section focuses on Model C2, since the other specifications are obtained as particular cases. To ease the derivations, at this stage it is assumed that the number of structural breaks J^0 (but not their position) is known, an assumption that will be relaxed below. We follow Pesaran (2006, 2007) and use cross-section averages to capture the unobserved common factors, which can be written in matrix notation from the block-orthogonal model specification as $\bar{Y} = D(\lambda)\bar{\mu}' + F(\lambda)\bar{\pi}' + \bar{U}$, where $D(\lambda) = \text{diag}([\iota_1, \tau_1], \dots, [\iota_{J^0+1}, \tau_{J^0+1}])$, $F(\lambda) = F^C(\lambda)$, $\bar{\pi} = (\bar{\pi}_1, \dots, \bar{\pi}_{J^0+1})'$, $\bar{Y}_t = N^{-1} \sum_{i=1}^N Y_{i,t}$, $\bar{\mu} = N^{-1} \sum_{i=1}^N \mu_i$, $\bar{\pi}_j = N^{-1} \sum_{i=1}^N \pi_{i,j}$ and $\bar{U}_t = N^{-1} \sum_{i=1}^N U_{i,t}$.

Assumption 6. $(\bar{\pi}_j) = m^0 \leq (1+k)$ for all N as $N \rightarrow \infty \forall j$, $j = 1, \dots, J^0 + 1$.

If the rank condition established in Assumption 6 is met, we have:

$$F = (\bar{Y} - D(\lambda)\bar{\pi}' - \bar{U}) \bar{\pi} (\bar{\pi}'\bar{\pi})^{-1}, \quad (10)$$

with $\bar{U}_t = O_p(T^{1/2}N^{-1/2})$ so that $\bar{U}_t \xrightarrow{p} 0$ for (i) fixed T and $N \rightarrow \infty$, and (ii) $T \rightarrow \infty$ and $N \rightarrow \infty$ with $T/N \rightarrow 0$. Then, the observable cross-section averages $\bar{h}_t = (D_t, D_{t,1}, \bar{Y}_t', \dots, D_{t,J^0+1,t}, \bar{Y}_t')'$ can be used to proxy the unobserved factors. Adding and subtracting (10) in (7) we have:

$$y_i = D_l(\lambda)\mu_{i,1}^y + D_\tau(\lambda)\mu_{i,2}^y + x_i(\lambda)\beta + F(\lambda)\pi_i^{y'} \pm (\bar{Y} - D(\lambda)\bar{\pi}' - \bar{U}) \bar{\pi} (\bar{\pi}'\bar{\pi})^{-1} \pi_i^{y'} + \xi_i^y = D_l(\lambda)\mu_{i,1}^y + D_\tau(\lambda)\mu_{i,2}^y + x_i(\lambda)\beta + \bar{Y}\delta_i^y + v_i, \quad (11)$$

with $\delta_i^y = \bar{\pi} (\bar{\pi}'\bar{\pi})^{-1} \pi_i^{y'}$ and $v_i = \xi_i^y + (F(\lambda) - \bar{Y}\bar{\pi} (\bar{\pi}'\bar{\pi})^{-1}) \pi_i^{y'}$. This defines an extension of the so-called cross-section augmented regression model in Holly, Pesaran, and Yamagata (2010) and Banerjee and Carrion-i-Silvestre (2017):

$$y_{i,t} = \mu_{i,j,1}^y + \mu_{i,j,2}^y (t - T_{j-1}) + x_{i,t}'\beta_j + \bar{Y}_t'\delta_{i,j}^y + v_{i,t}, \quad (12)$$

$T_{j-1} < t \leq T_j$, $j = 1, \dots, J^0 + 1$, where $T_B = (T_1, \dots, T_{J^0})'$ is a vector of generic break dates. Note that it is implicitly assumed that $m^0 = k + 1$ since all $k + 1$ cross-section averages are included in (12). Let us define the projection matrix $\bar{M}(\lambda) = I - \bar{H}(\lambda) (\bar{H}(\lambda)' \bar{H}(\lambda))^{-1} \bar{H}(\lambda)'$, with $\bar{H}(\lambda) = [D(\lambda) \bar{z}(\lambda)]$, $\bar{z}(\lambda) = [\bar{z}(1) \cdots \bar{z}(J^0 + 1)]_{(T \times (k+1)(J^0+1))}$, and $\bar{z}(j) = [(D_{i,j} \odot \bar{x}_1) \cdots (D_{i,j} \odot \bar{x}_k) (D_{i,j} \odot \bar{y})]$, $j = 1, \dots, J^0 + 1$. Then, the estimation of β in (11) can be done using the pooled CCE estimator (PCCE):

$$\hat{\beta}_{PCCE}(\lambda) = \left(\sum_{i=1}^N x_i(\lambda)' \bar{M}(\lambda) x_i(\lambda) \right)^{-1} \left(\sum_{i=1}^N x_i(\lambda)' \bar{M}(\lambda) y_i \right). \quad (13)$$

Following Bai (2010), Kim (2011, 2014), and Baltagi, Feng, and Kao (2016, 2019), the estimation of the break dates is carried out minimizing the global sum of squared residuals (SSR), so that:

$$\hat{T}_B = \arg \min_{\lambda \in \Lambda} SSR(\lambda), \quad (14)$$

where $\hat{T}_B = (\hat{T}_1, \dots, \hat{T}_{J^0})'$, $\hat{\lambda} = \hat{T}_B/T$, $SSR(\lambda) = \sum_{i=1}^N SSR_i(\lambda)$, $SSR_i(\lambda)$ being the SSR for the i th panel unit associated with the estimation of (12). Thus, the estimation of the break dates can be obtained carrying out a grid search over all potential combinations of J^0 structural breaks.¹ The following theorem analyzes the consistency of $\hat{\beta}_{PCCE}(\hat{\lambda})$ and $\hat{\lambda}$. For completeness, we also include the known breaks case.

¹A GAUSS computer code is available upon request to estimate the break dates in panel data with up to two structural breaks, although the Bai and Perron (1998) efficient estimation algorithm which allows global minimizers of SSR that is of order $O(T^2)$ for any $J^0 \geq 2$ can also be adapted to our panel data framework. In this regard, Ditzén, Karavias and Westerlund (2023) extend the Bai and Perron (1998) methodology to panel data with the development of a toolbox for Stata software using the Bai and Perron (1998) efficient estimation algorithm adapted to panel data.

Theorem 1. Let $Y_{i,t}$ be a vector of $(1 + k)$ stochastic processes with DGP given by (1)–(3) and satisfying Assumptions 1–6. Then, as $(T, N) \rightarrow \infty$:

1. Known break dates case. For Models A, B, and C, $\hat{\beta}_{PCCE}(\lambda^0) \xrightarrow{P} \beta = \Omega_{U_x U_x}^{-1} \Omega_{U_x U_y}$.
2. Unknown break dates case:
 - 2.a. Model A1: $(\hat{\lambda} - \lambda^0)$ does not converge to zero in probability.
 - 2.b. Models A2, B1, B2, C1, and C2: (i) $(\hat{\lambda} - \lambda^0) \xrightarrow{P} 0$, with $(\hat{\lambda} - \lambda^0) = o_p(T^{-1/2})$ for Models A2, B2 and C2, $(\hat{\lambda} - \lambda^0) = o_p(N^{-1/2})$ for Models B1 and C1, and (ii) $\hat{\beta}_{PCCE}(\hat{\lambda}) \xrightarrow{P} \beta = \Omega_{U_x U_x}^{-1} \Omega_{U_x U_y}$.

See the companion appendix for the proof. Theorem 1 establishes that $\hat{\lambda}$ is not consistent for Model A1 since the effects of the structural changes are dominated, in the limit, by the stochastic trend components,² whereas it is consistent for the other models. In addition, it is shown that the PCCE estimator converges toward the long-run average coefficient β in the known breaks case for all models, and in the unknown breaks case for the models for which a consistent break fraction vector estimate can be obtained. This establishes the basis for the derivation of our panel data cointegration test. Finally, it is worth noting that Theorem 1 establishes the consistency of the break fraction estimates for Models A2, B1, B2, C1, and C2. As for the break dates estimates, the results in Theorem 1 show that it is not possible to obtain consistent estimates of the break dates for Models A2, B2, and C2, consistency only applies to break fraction estimates. For Models B1 and C1, consistent break dates can be obtained if $TN^{-1/2} \rightarrow 0$ is assumed. Although there are some contributions in the panel literature that derive consistent break dates estimation procedures—see Bai (2010), Kim (2011, 2014) and Baltagi, Feng and Kao (2016, 2019), among others—these approaches assume that the disturbance term of the model is $I(0)$, whereas here we deal with the novel panel data spurious regression with structural breaks case. Consequently, the framework under which those results are obtained is not comparable to our panel data spurious regression set-up.

It is possible to define an alternative estimation of the break dates that relies on the minimization of the SSR of the model:

$$y_{i,t} = \mu_{i,j,1}^y + \mu_{i,j,2}^y (t - T_{j-1}) + x'_{i,t} \hat{\beta}_j + v_{i,t}^*, \quad (15)$$

instead of using (12), where $\hat{\beta}_j$ refers to the j th regime parameters in $\hat{\beta}_{PCCE}(\lambda)$ for a generic T_B . The alternative estimation of the break dates is then defined as:

$$\tilde{T}_B = \arg \min_{\lambda \in \Lambda} SSR^*(\lambda), \quad (16)$$

where $\tilde{T}_B = (\tilde{T}_1, \dots, \tilde{T}_{j_0})'$, $\tilde{\lambda} = \tilde{T}_B/T$, $SSR^*(\lambda) = \sum_{i=1}^N SSR_i^*(\lambda)$, $SSR_i^*(\lambda)$ being the SSR for the i th panel unit associated with the estimation of (15). It is worth noting that this procedure ignores the common factors in (15), although they are considered in the estimation of $\hat{\beta}_{PCCE}(\lambda)$. Therefore, this estimator can be seen as an hybrid method in which the common factors are used in the estimation of $\hat{\beta}_{PCCE}(\lambda)$, but not in the minimization of the SSR of (15). The following theorem summarizes the main features of this alternative estimator.

²This situation is equivalent to the one found in time series analysis, see Perron (1990).

Theorem 2. Let $Y_{i,t}$ be a vector of $(1 + k)$ stochastic processes with DGP given by (1)–(3) and satisfying Assumptions 1–6. Then, as $(T, N) \rightarrow \infty$, $(\tilde{\lambda} - \lambda^0) = o_p(N^{-1/4}T^{-1/2})$ for Models B1 and C1, and $(\tilde{\lambda} - \lambda^0) = o_p(N^{-1/3}T^{-1})$ for Models A2, B2, and C2.

The proof is given in the companion appendix. As can be seen, the rate of convergence of $\tilde{\lambda}$ is faster than the $\hat{\lambda}$ one for a given model, which suggests that better performance of the former in finite samples are to be expected. This result is due to the fact that the alternative estimator avoids projecting the observable variables in the model against the common factors, which leads us to obtain a stronger signal around the combination of break dates that minimizes the SSR of (15)—see the appendix. Finally, consistent break dates can be obtained for Models A2, B2, and C2—although at a slow rate of convergence—and for Models B1 and C1 if it is assumed that $TN^{-1/2} \rightarrow 0$. As mentioned above, the existing panel data structural break estimation techniques that have been recently proposed in the literature assume that the disturbance term of the model is $I(0)$, which leads to consistent estimates of the break dates with higher rates of convergence.

The estimation of the number of structural breaks can be carried out using panel information criteria similar to the ones proposed in Bai and Ng (2002, 2004). To be specific, let us define an information criteria of the form:

$$IC(\hat{\lambda}, J) = \ln \hat{\sigma}^2(\hat{\lambda}, J) + ((1 + J)k)g(N, T),$$

with $\hat{\sigma}^2(\hat{\lambda}, J) = N^{-1}T^{-1} \sum_{i=1}^N \sum_{t=1}^T \Delta \hat{v}_{i,t}^2$, $J = 0, 1, \dots, J_{\max}$, where $\hat{v}_{i,t}$ denotes the estimated residuals from (12) and $g(N, T)$ is the penalty function—one possibility suggested in Bai and Ng (2002) is the panel BIC that is established with $g(N, T) = \ln(NT / (N + T))(N + T)/NT$. Then, $\hat{J} = \arg \min_{J=0, \dots, J_{\max}} IC(\hat{\lambda}, J)$. Similarly, we can define \tilde{J} if $\tilde{\lambda}$ is used instead of $\hat{\lambda}$ in the estimation of (12). The following theorem shows that the suggested information criterion provides consistent estimation of the number of structural breaks.

Theorem 3. Let $Y_{i,t}$ be a vector of $(1 + k)$ stochastic processes with DGP given by (1)–(3) and satisfying Assumptions 1–6. Then, $\lim_{N, T \rightarrow \infty} \Pr(\hat{J} = J^0) = 1$, $\hat{J} \in \{\hat{J}, \tilde{J}\}$, if (i) $g(N, T) \rightarrow 0$ and (ii) $C_{NT}^2 g(N, T) \rightarrow \infty$ as $(T, N) \rightarrow \infty$, where $C_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$.

The proof is outlined in the appendix. This result defines a framework general enough to be able to treat all the elements of the model in either an endogenous or exogenous way.

4. PCCE-based Cointegration Test Statistic

Following Pesaran (2007), Holly, Pesaran and Yamagata (2010), and Banerjee and Carrion-i-Silvestre (2017), we design a cointegration statistic that approximates the unobserved common factors using cross-section averages of the observable variables. The simplicity of this approach comes at the cost of restricting the framework defined above since Pesaran (2007) and Pesaran, Smith and Yamagata (2013) assume that both the common factor and the idiosyncratic components are $I(1)$ under the null

hypothesis of unit root—that is, $m^0 = m_1^0$. In what follows, this section focuses on the most general specification given by Model C2 using $\hat{\lambda}$, although the procedure is valid for the other model specifications and the $\tilde{\lambda}$ estimator.

Since $\hat{\beta}_{PCCCE}(\hat{\lambda})$ is a consistent estimate of β —for all models except A1—we define $\hat{y}_{i,t} = y_{i,t} - x'_{i,t}\hat{\beta}_{PCCCE}(\hat{\lambda})$ and base the testing of the null hypothesis of no cointegration for each unit on the cross-section augmented ADF (CADF) type regression equation:³

$$\begin{aligned} \Delta \hat{y}_{i,t} = & \sum_{j=0}^{\hat{j}} \theta_{i,j} DU_{j,t} + \sum_{j=1}^{\hat{j}} \gamma_{i,j} D(\hat{T}_j)_t + \sum_{j=0}^{\hat{j}} \vartheta_{i,j} DT_{j,t} + \alpha_{i,0} \hat{y}_{i,t-1} \\ & + \sum_{l=1}^{p_i} \alpha_{i,l} \Delta \hat{y}_{i,t-l} + \sum_{j=0}^{\hat{j}} \varphi'_{i,j} (DU_j \odot \bar{A})_{t-1} \\ & + \sum_{j=0}^{\hat{j}} \sum_{l=0}^{p_i} \kappa'_{i,j,l} (DU_j \odot \Delta \bar{A})_{t-l} + v_{i,t}, \end{aligned} \quad (17)$$

with $D(\hat{T}_j)_t = 1$ for $t = \hat{T}_j + 1$, 0 otherwise, $j = 1, \dots, \hat{j}$. Note that the order of augmentation p_i in (17) is heterogeneous, which can be selected using the modified information criteria in Ng and Perron (2001)—homogeneous p_i as in Pesaran (2007) can be imposed, if desired. The pseudo t-ratio statistic of $\hat{\alpha}_{i,0}$ in (17), $t_{\hat{\alpha}_{i,0}}(\hat{\lambda})$, is used to test the null hypothesis of no cointegration.⁴ The number of common factors that is assumed defines \bar{A}_t , so that $\bar{A}_t = \bar{y}_t$ for $m^0 = 1$, whereas $\bar{A}_t = (\bar{y}_t, \bar{x}_{1,t}, \dots, \bar{x}_{k,t})'$ for $m^0 = 1 + k$. For the intermediate cases where $m^0 < 1 + k$, \bar{A}_t will be defined with \bar{y}_t and $m^0 - 1$ elements of \bar{x}_t . The following theorem provides the limiting distribution of $t_{\hat{\alpha}_{i,0}}(\hat{\lambda})$ for both the known and unknown breaks cases.

Theorem 4. Let $Y_{i,t}$ be a vector of $(1 + k)$ stochastic processes with DGP given by (1)–(3) and satisfying Assumptions 1–6. The $t_{\hat{\alpha}_{i,0}}(\hat{\lambda})$ statistic, $\hat{\lambda} \in \{\hat{\lambda}, \tilde{\lambda}\}$, converges sequentially (if $N \rightarrow \infty$ first, then $T \rightarrow \infty$, that is, $\Rightarrow_{(N,T)_{\text{seq}}}$) and jointly (if $(N, T) \xrightarrow{j} \infty$ with $\sqrt{T}/N \rightarrow 0$, that is, $\Rightarrow_{(N,T)_j}$) to:

$$\begin{aligned} t_{\hat{\alpha}_{i,0}}(\hat{\lambda}) & \Rightarrow \frac{\int_0^1 W_i(r) dW_i(r) - \omega_{iF}^l(\lambda^0)' G_F^l(\lambda^0)^{-1} \pi_{iF}^l(\lambda^0)}{\left(\int_0^1 W_i^2(r) dr - \pi_{iF}^l(\lambda^0)' G_F^l(\lambda^0)^{-1} \pi_{iF}^l(\lambda^0) \right)^{1/2}} \\ & \equiv R(\lambda^0, l), \end{aligned}$$

where $W_i(r)$ denotes a scalar standard Brownian motion, $\omega_{iF}^l(\lambda^0)$, $G_F^l(\lambda^0)$ and $\pi_{iF}^l(\lambda^0)$ are functions of a m^0 -vector of

standard Brownian motions associated with the common factors that are defined in the appendix, with $l \in \{A1, A2, B1, B2, C1, C2\}$ for the known breaks case ($\hat{\lambda} = \tilde{\lambda} = \lambda^0$) and, for the unknown breaks case, $l \in \{B1, C1\}$ when $\hat{\lambda}$ is used and $l \in \{A2, B1, B2, C1, C2\}$ when $\tilde{\lambda}$ is applied.

The proof is given in the appendix and some remarks are in order. First, the limiting distribution of $t_{\hat{\alpha}_{i,0}}(\hat{\lambda})$ and $t_{\hat{\alpha}_{i,0}}(\tilde{\lambda})$ depends on the number and position of structural breaks (λ^0) and on the number of I(1) nonstationary common factors—note that it is assumed that $m^0 = m_1^0$. If analysts have knowledge—for example, based on economic theory—about the number of common factors, one can consider such information when \bar{A}_t is defined. In this regard, we could follow the strategy in Pesaran et al. (2013) and compute the statistic using all possible combinations of m^0 cross-section averages available in the system as a way of obtaining robust conclusions. When m^0 is unknown, we can follow a conservative strategy and assume that the rank condition is satisfied with equality. The price that we would pay if $m^0 < 1 + k$, but we impose $m^0 = 1 + k$, is to have a test statistic with empirical size smaller than the nominal size accompanied by loss of power. The advantage is to allow us to remain agnostic about the number of integrated stochastic trends driving the data.⁵ Second, the appendix shows that the limiting distribution of $t_{\hat{\alpha}_{i,0}}(\lambda^0)$ for Models A and B is the same, for each variant of the deterministic component, whereas it is different for Model C.

Third, for the unknown breaks case and Models B1 and C1, $t_{\hat{\alpha}_{i,0}}(\hat{\lambda})$ converges to the same limiting distribution as for the known structural breaks case. Unfortunately, this does not occur for Models A2, B2, and C2. In these cases, we suggest the implementation of the modified procedure in Kim and Perron (2009) that relies on the use of trimmed data. In brief, the idea is based on establishing a window of observations around a consistent estimate of λ —that is, $(\hat{\lambda} - \lambda^0) = O_p(T^{-a})$ for some $0 < a < 1$ —of length $2\omega(T)$ with $\omega(T) \equiv \kappa T^\delta$, $\kappa > 0$ and $-1 < -a < \delta < 0$. The window defines the set of observations compressed between $\hat{T}_l + 1 \equiv T(\hat{\lambda} - \omega(T)) + 1$ and $\hat{T}_h \equiv T(\hat{\lambda} + \omega(T))$ and has the characteristics of: (i) increasing slowly enough to be asymptotically negligible relative to T and, (ii) increasing fast enough to include the true structural break dates since $\hat{T}_l - \hat{T}_B = T(\hat{\lambda} - \omega(T)) - T\lambda^0 = (T^{-a-\delta}T^a(\hat{\lambda} - \lambda^0) - \kappa)T^{\delta+1} \xrightarrow{P} -\infty$ and $\hat{T}_B - \hat{T}_h = T\lambda^0 - T(\hat{\lambda} + \omega(T)) = (T^{-a-\delta}T^a(\lambda^0 - \hat{\lambda}) - \kappa)T^{\delta+1} \xrightarrow{P} -\infty$. Once \hat{T}_l and \hat{T}_h have been specified, Kim and Perron (2009) define a new dataset as:⁶

$$\hat{y}_{i,t}^n = \hat{y}_{i,t+\hat{T}_{j-1,h}-\hat{T}_{j-1,l}} - S(\hat{y}_{i,t}, j-1) \text{ for } \hat{T}_{j-1} < t \leq \hat{T}_j,$$

with $S(\hat{y}_{i,t}, j-1) = \hat{y}_{i,\hat{T}_{j-1,h}} - \hat{y}_{i,\hat{T}_{j-1,l}}$, $\hat{T}_{0,h} = \hat{T}_{0,l} = 0$ and $S(\hat{y}_{i,t}, 0) = 0$ —the same transformation applies to obtain \bar{A}_t^n with $S(\bar{A}_t, j-1) = \bar{A}_{\hat{T}_{j-1,h}} - \bar{A}_{\hat{T}_{j-1,l}}$. The new dataset removes

³For Models A and B the factor loadings are assumed to be constant across regimes so that the sixth and seventh element of the right hand side of (17) are not affected by the structural breaks—that is, $\sum_{j=0}^{\hat{j}} \varphi'_{i,j} (DU_j \odot \bar{A})_{t-1}$

has to be replaced by $\varphi'_i \bar{A}_{t-1}$ and $\sum_{j=0}^{\hat{j}} \sum_{l=0}^{p_i} \kappa'_{i,j,l} (DU_j \odot \Delta \bar{A})_{t-l}$ by $\sum_{l=0}^{p_i} \kappa'_{i,l} \Delta \bar{A}_{t-l}$.

⁴We abuse notation when using the Hadamard product in the expressions $DU_j \odot \bar{A}$ and $DU_j \odot \Delta \bar{A}$ that appear in (17) since, in general, the involved matrices might have different column dimensions. In this case, it should be understood that for any given $A_{m \times n}$ and $B_{m \times o}$ matrices $A \odot B = [(A_1 \odot B_1) \dots (A_1 \odot B_o) \dots (A_n \odot B_1) \dots (A_n \odot B_o)]_{m \times no}$. This convention applies throughout the article.

⁵An alternative strategy in the case of unknown number of factors, as followed by Pesaran et al. (2013), is to undertake the testing for all permissible values of m^0 (using all combinations of m^0 cross-section averages for each choice of m^0). The size properties of such a procedure are not clear nor are the likely conclusions if one accepts the null hypothesis for some values of m^0 and rejects for others. This is a topic for future research.

⁶For the empirical implementation of this procedure we can follow Kim and Perron (2009) and use a trimming window of 6 observations.

the data points between $\hat{T}_l + 1$ and \hat{T}_h and reconnects the remaining data shifting down the data after the window with the $S(\cdot, j - 1)$ function. Then, the pseudo t -ratio statistic of $\hat{\alpha}_{i,0}$ in (17) is computed using the trimmed data $\hat{y}_{i,t}^n$ and \bar{A}_t^n with the break dates given by \hat{T}_l . The resulting statistic is denoted as $t_{\hat{\alpha}_{i,0}}(\hat{\lambda}_{tr})$, which limiting distribution is given in the following corollary.

Corollary 1. Let $Y_{i,t}$ be a vector of $(1 + k)$ stochastic processes with DGP given by (1)–(3) and satisfying Assumptions 1–6. For the unknown structural breaks case and as either $(N, T)_{\text{seq}}$ or $(N, T)_j$, $t_{\hat{\alpha}_{i,0}}(\hat{\lambda}_{tr}) \Rightarrow R(\lambda^0, l)$, $l \in \{A2, B2, C2\}$.

The proof derives from the developments in Theorem 4. As can be seen, the limiting distribution of $t_{\hat{\alpha}_{i,0}}(\hat{\lambda}_{tr})$ equals the limiting distribution that is obtained for the known structural breaks case. Finally, it should be stressed that the limiting distributions in Theorem 4 and Corollary 1 also correspond to the distributions obtained for a panel unit root test with multiple structural breaks that affect the level and/or the slope of the time trend. Therefore, as side contributions, this article (i) extends the panel data unit root tests in Pesaran (2007) and Pesaran, Smith and Yamagata (2013) to the multiple structural breaks case—affecting the deterministic component and/or the factor loadings—and (ii) provides an extension of Perron (1989, 1990) and Kim and Perron (2009) unit root tests to panel data framework.

The limiting distributions in Theorem 4 are approximated by Monte Carlo simulation. A limited set of critical values for $t_{\hat{\alpha}_{i,0}}(\lambda^0)$ are reported in Tables B.1 to B.6 for Models A1, B1, and C1, and in Tables B.7 to B.12 for Models A2, B2, and C2, with $J = 1$ and $m^0 \leq k + 1 \in \{2, 3, 4\}$ common factors. Pesaran (2007) and Pesaran, Smith and Yamagata (2013) also propose a truncated version of the $t_{\hat{\alpha}_{i,0}}(\lambda^0)$ statistic in order to ensure that the statistic has finite moments, although these papers provide evidence that for $T > 15$ the empirical distributions of the truncated and untruncated statistics are equivalent. Therefore, in this article we do not consider the truncated version of $t_{\hat{\alpha}_{i,0}}(\lambda^0)$.

The combination of the individual test statistics defines the cross-section augmented ADF panel cointegration statistics $CIPS(\hat{\lambda}) = N^{-1} \sum_{i=1}^N t_{\hat{\alpha}_{i,0}}(\hat{\lambda})$, $\hat{\lambda} \in \{\hat{\lambda}, \tilde{\lambda}, \hat{\lambda}_{tr}\}$, as proposed in Pesaran (2007). This statistic allows us to test the null hypothesis of no panel data cointegration against the alternative that there is a fraction of panel units for which cointegration holds. The critical values for the $CIPS(\lambda^0)$ statistic are presented in Tables B.13 to B.18 for Models A1, B1, and C1, and in Tables B.19–B.24 for Models A2, B2, and C2, with $J = 1$ and $m^0 \leq k + 1 \in \{2, 3, 4\}$. Interestingly, the critical values seem to be symmetrical around $\lambda^0 = 0.5$ for large T , a characteristic that is more evident for the panel data cointegration statistic critical values. This feature has also been found for the ADF unit root test with one structural break proposed in Perron (1989, 1990).⁷

Westerlund, Hosseinkouchack and Solberger (2016) derive the local-to-unity asymptotic power functions of Pesaran (2007)

type-test statistics, and show that these statistics have non-negligible power in a neighborhood of T^{-1} . This is a distinctive feature of these statistics when compared to other proposals in the literature, that have been shown to have non-negligible power in the neighborhood given by $N^{-\kappa} T^{-1}$, $\kappa > 0$ —see, for instance, Moon and Perron (2008). This implies that as N increases Pesaran (2007) type-test statistics will tend to be dominated by other panel data tests having non-negligible local power for $\kappa > 0$.

5. Finite Sample Performance

Let us consider the DGP defined by:

$$\begin{aligned} y_{i,t} &= \theta_{i,1} DU_t + \theta_{i,2} DT_t + x'_{i,t} \varsigma_0 + DU_t x'_{i,t} \varsigma_1 + F'_t \varphi_{i,0}^y \\ &\quad + DU_t F'_t \varphi_{i,1}^y + u_{i,t} \\ \Delta x_{i,t} &= \Delta F'_t \varphi_{i,0}^x + \Delta (DU_t F'_t) \varphi_{i,1}^x + v_{i,t}; \\ F_{j,t} &= \rho F_{j,t-1} + w_{j,t}; \\ u_{i,t} &= \phi_i u_{i,t-1} + \varepsilon_{i,t}, \end{aligned} \quad (18)$$

where $\theta_{i,1} \sim N(10, 1)$, $\theta_{i,2} \sim N(0.3, 1)$, $\varsigma_0 = 1$, $\varsigma_1 \in \{1, 5\}$, $\varphi_{i,0}^y \sim U[0, 1]$, $\varphi_{i,1}^y \sim U[0, 2]$, $\varphi_{i,0}^x \sim U[0, 1]$, $\varphi_{i,1}^x \sim U[0, 3]$, $v_{i,t} \sim N(0, 1)$, $w_{j,t} \sim N(0, 1)$, $j \in \{1, 2\}$, and $\varepsilon_{i,t} \sim N(0, 1)$ are mutually independent groups. Under the null hypothesis of no cointegration $\phi_i = 1 \forall i$, whereas under the alternative hypothesis of cointegration we set $\phi_i = 0.9 \forall i$. This section investigates the performance of the proposed statistics considering one unknown structural break—additional simulation results for known structural breaks are available upon request. We distinguish three different cases of interest. Case 1 specifies a DGP in which both the deterministic component and the cointegration vector change, but the loadings remain constant across regimes—that is, $\varsigma_1 \in \{1, 5\}$ and $\varphi_{i,1}^y = \varphi_{i,1}^x = 0 \forall i$ in (18). Case 2 uses the Case 1 DGP but the computations are carried out allowing for a nonexistent change in factor loadings. This is done to proxy situations where the investigator may be uncertain about whether there are changes in the factor loadings and wishes to protect herself (in terms of size of the test) against this possibility. Finally, Case 3 deals with the pure structural break case where all parameters (deterministic component, cointegration vector and factor loadings) change across regimes. Note that for each of these cases we consider the two variants of the deterministic component—that is, the constant and the time trend cases. The break date is specified at $\lambda^0 = 0.5$ and we consider up to two common factors, $m^0 \in \{1, 2\}$, with $\rho \in \{1, 0.99, 0.95, 0.9\}$. The time dimension is set at $T \in \{50, 100, 200\}$ and the cross-section dimension is $N \in \{20, 50\}$. The nominal size is set at 5% and the critical values tabulated in the previous section are used.

5.1. Constant Case

5.1.1. Performance of the Break Fraction and Pooled Estimators

This section investigates the performance of $\hat{\lambda}$ and $\tilde{\lambda}$ for Models B and C with the deterministic component given by the constant case—that is, $\theta_{i,2} = 0 \forall i$ in (18). Let us first focus on the $\hat{\lambda}$ estimator. Figures C.1–C.4 present histograms of $\hat{\lambda}$ for Model B1

⁷ A GAUSS program, available upon request, allows the computation of critical values for the multiple structural breaks case for both the individual and panel data cointegration statistics.

(Case 1) when $\varsigma_1 = 1$ for all possible combinations that consider that the true number of common factors is $m^0 \in \{1, 2\}$ and the imposed number of common factors is $m \in \{1, 2\}$. It is worth stressing that consistency of the $\hat{\lambda}$ and $\tilde{\lambda}$ estimators has been established under the null hypothesis of spurious regression situation, but for completeness we also include the histograms for these estimators under cointegration.

In general and regardless of the order of integration of the common factors, histograms are concentrated around λ^0 with higher probability as either T or N increases. It is interesting to note that allowing for $m > m^0$ does not affect the performance of $\hat{\lambda}$, whereas the probability mass of $\hat{\lambda}$ around λ^0 increases with m^0 . As expected, as the magnitude of the structural change that affect the slope parameters increases, the performance of $\hat{\lambda}$ shows a substantial improvement, with almost all probability mass located on λ^0 , regardless of m^0 , m and the order of integration of the idiosyncratic and the common factor components—see Figures C.5–C.8 with the histograms of $\hat{\lambda}$ for Model B1 when $\varsigma_1 = 5$. These features are also found when the results for Cases 2 and 3 are analyzed—see Figures C.9–C.24. The use of the $\tilde{\lambda}$ estimator produces better results since now, and even for $\varsigma_1 = 1$, the probability mass of $\tilde{\lambda}$ around λ^0 is higher than the $\hat{\lambda}$ one—see Figures C.25–C.48 that depict the histograms of $\tilde{\lambda}$ for Cases 1 to 3, $m \in \{1, 2\}$, $m^0 \in \{1, 2\}$ and $\varsigma_1 \in \{1, 5\}$. This simulation evidence is demonstration of the statement made in [Theorems 1 and 2](#).

Finally, we have also analyzed the performance of $\hat{\beta}_{PCCE}(\hat{\lambda})$ and $\hat{\beta}_{PCCE}(\tilde{\lambda})$ and found evidence that supports that its application leads to a consistent estimation of β . To be specific, the mean squared error (MSE) decreases as N and/or T increase, and as ρ moves away from one—detailed results are available upon request.

5.1.2. Empirical Size and Power of the Panel Cointegration Statistic

Table C.1 summarizes the performance of the $CIPS(\hat{\lambda})$ statistic. When the magnitude of the structural change affecting the slope is $\varsigma_1 = 1$, we can observe size distortion problems for the panel cointegration statistic for Case 1 when the order of integration of both the idiosyncratic and the common factor components is the same, regardless of the (true and assumed) number of common factors. These distortions reduce as the model specification becomes more flexible—that is, Cases 2 and 3—and even disappear in some situations—see the results for $m^0 = m = 1$. As the magnitude of the structural break increases to $\varsigma_1 = 5$, the empirical size tends to the nominal one in all cases. It is worth highlighting that the misspecification error of either under-specifying ($0 < m < m^0$) or over-specifying ($m^0 < m$) the number of common factors does not cause high size distortion problems. The panel cointegration $CIPS(\hat{\lambda})$ statistic becomes conservative under the null hypothesis of spurious regression if the common factor is $I(0)$. This is a consequence of the violation of the key assumption underlying the design of the statistic—see [Section 4](#). The empirical power of $CIPS(\hat{\lambda})$ increases with N and T , and as ρ moves away from one. As expected, the more flexible model specifications—either because the type of model is more general (Case 2) or because the number of common factors is over-specified ($m^0 < m$)—show lower empirical power values.

The performance of the statistical inference improves when $\tilde{\lambda}$ is used. Table C.2 evidences a clear enhancement of the empirical size figures when $\varsigma_1 = 1$ —the exception is found for Case 3 with $m^0 = m = 2$ and $T = 200$, although the size distortion almost disappears for $\varsigma_1 = 5$. However and contrary to what has been found for $CIPS(\hat{\lambda})$, the under-specification of the number of common factors causes mild over-size distortions for Model C—see the results for Case 3. This results is somehow to be expected since we are missing the presence of common factors in the model specification. As above, $CIPS(\tilde{\lambda})$ becomes conservative if the assumption of common order of integration of the idiosyncratic and common components under the null hypothesis of spurious regression is not satisfied. The $CIPS(\tilde{\lambda})$ statistic encompasses $CIPS(\hat{\lambda})$ in term of empirical power, since the former statistic shows similar, if not higher, empirical power values—note that the higher power shown by $CIPS(\hat{\lambda})$ when $\varsigma_1 = 1$ are due to the mild size distortions that have been documented.

To the best of our knowledge, there are no other proposals in the literature that can be used to establish a direct comparison with the test statistics that are proposed in this article. The closest proposal might be the PANIC-based panel cointegration test statistic that is developed by Banerjee and Carrion-i-Silvestre (2015). That paper designed a panel data statistic to test the joint null hypothesis of spurious regression without structural breaks against the alternative hypothesis of panel cointegration with structural breaks for large T compared to N . As mentioned in the introduction, the cross-section dependence was introduced in the specification using an approximate common factor model. Tables C.3 and C.4 present the empirical size and power for the panel ADF statistic that is computed for the estimated idiosyncratic component (Z_c^*) and the MQ_c^c statistic to estimate the number of $I(0)$ and $I(1)$ common stochastic trends. The simulation results are reported for Cases 1 and 3, considering $m^0 \in \{1, 2\}$ and allowing for up to $m_{\max} = 6$ common factors—the number of common factors is estimated using the panel BIC statistic in Bai and Ng (2004). The Z_c^* statistic shows important size distortions, the higher the value of ς_1 , which in fact can be interpreted in terms of empirical power since this situation is a mere consequence of the violation of the DGP that is assumed under null hypothesis that is assumed in this proposal. Thus, note that the DGP defined in (18) allows for the presence of structural breaks under the null hypothesis, so that the joint hypothesis of spurious regression without structural breaks that is assumed by Banerjee and Carrion-i-Silvestre (2015) is not fulfilled. The behavior of the MQ_c^c statistic is also affected by this feature, although it is fair to note that the MQ_c^c statistic tries to capture that misspecification error detecting more common factors than really exist—in general, it can be shown that the estimated number of common factors is double the true number.

Finally and following the suggestion in Westerlund (2018), we have also studied the performance of the $CIPS$ panel cointegration test statistic when the structural breaks are ignored. Table C.5 summarizes the empirical size and power of the $CIPS$ proposed in Banerjee and Carrion-i-Silvestre (2017) for Cases 1 and 3, and $\varsigma_1 \in \{1, 5\}$. As can be seen, the $CIPS$ presents size distortions, the higher the value of ς_1 . For $\varsigma_1 = 1$ and $m^0 = 1$, the size distortions can be reduced if we over-specify the number of common factors, although this desirable feature disappears as ς_1

increases. The violation of the assumption of common order of integration of the idiosyncratic and common components under the null hypothesis of spurious regression leads to a conservative test statistic, but only when $\varsigma_1 = 1$ and $m > m^0$. For the other cases, the statistic shows over-rejection distortions. These results are in line with the literature that shows that the null hypothesis of panel unit root is rejected when cross-section dependence is ignored—see Banerjee, Marcellino and Osbat (2005). Here the cross-section dependence is considered, but to some extent the omission of common structural breaks might leave one source of cross-section dependence that links the panel units uncaptured. This shows that modelling the presence of structural breaks is required if misleading statistical inference is to be avoided.

5.1.3. Estimation of the Number of Structural Breaks

Table C.10 provides the frequency of the estimated number of structural breaks that derives from the use of the panel BIC statistic, considering up to two structural breaks. The ability of the panel BIC to select the correct number of structural breaks improves as N , T and ς_1 increase, with the statistic that is based on $\tilde{\lambda}$ the one that produces better results.

Let us first focus on the results for $\varsigma_1 = 1$. Under the spurious regression case with $\rho = 1$, $J(\hat{\lambda})$ tends to under-estimate J for small N and T , with correct number of structural breaks detection frequencies that range between 0.26 and 0.48 for Model B1, and between 0.48 and 0.6 for Model C1. As can be seen, $J(\tilde{\lambda})$ outperforms $J(\hat{\lambda})$ with correct frequencies in the range 0.52–0.86 for Model B1, and 0.57–0.7 for Model C1. Further, note that part of the under-estimation frequency of $J(\hat{\lambda})$ moves to over-estimation as N increases for a given T —this feature is hardly observed for $J(\tilde{\lambda})$. When $\rho = 0.9$, $J(\hat{\lambda})$ still shows a tendency to under-estimate J for Model B1, although this behavior is attenuated for Model C1. As for the $J(\tilde{\lambda})$ estimator, the probability of correct detection remains unchanged as N increases, regardless of T , with a tendency of reallocate part of the under-estimation frequency on the over-estimation one as T increases. Under the cointegration scenario, the behavior of both $J(\tilde{\lambda})$ and $J(\hat{\lambda})$ statistics improve with correct J estimation frequencies that tend to one as T increases, and with $J(\tilde{\lambda})$ outperforming $J(\hat{\lambda})$ in all cases. Finally, when the magnitude of the structural break increases to $\varsigma_1 = 5$ both estimators provide good results.

Table C.12 summarizes the performance of the panel BIC when there are no structural breaks ($J^0 = 0$) affecting the parameters of the model. The frequency of correct selection of the number of breaks for both $J(\tilde{\lambda})$ and $J(\hat{\lambda})$ statistics is similar, and higher than the one obtained for the one structural break case. In general, higher correct selection frequencies are obtained for Model C1. This results indicates that the proposed panel BIC is useful to detect the presence of structural breaks in our model set-up.

5.2. Time Trend Case

5.2.1. Performance of the Break Fraction and Pooled Estimators

This section investigates the performance of $\hat{\lambda}$ and $\tilde{\lambda}$ for Models B and C with the deterministic component given by the time

trend case—that is, $\theta_{i,1} \neq \theta_{i,2} \neq 0 \forall i$ in (18). Figures C.49–C.52 depict the histograms of $\hat{\lambda}$ for Model B2 (Case 1) when $\varsigma_1 = 1$ for all possible combinations that consider that the true number of common factors is $m^0 \in \{1, 2\}$ and the imposed number of common factors is $m \in \{1, 2\}$. In general, we can observe that the probability mass concentrates around λ^0 as T and/or N increase for a given set of ϕ_i and ρ parameters. The probability around λ^0 is slightly reduced for Cases 2 and 3, especially for small T , although it tends toward one as T and N get large—see Figures C.53–C.60. Note that when this estimation bias appears, the estimates are located around λ^0 . As for the constant case, the use of $\tilde{\lambda}$ produces better results, since the probability mass of $\tilde{\lambda}$ around λ^0 is, in general, similar or larger than the $\hat{\lambda}$ one—see Figures C.61–C.72.

The simulation experiment has also studied the performance of $\hat{\beta}_{PCCE}(\hat{\lambda})$ and has found that the MSE decreases as N and/or T increase, and as ρ moves away from one—detailed results are available upon request. All these results reinforce the theoretical analysis that has been detailed in Theorem 1.

5.2.2. Empirical Size and Power of the Panel Cointegration Statistic

Table C.6 collects the empirical size and power of $CIPS(\hat{\lambda}_{tr})$. The statistic shows good performance, with an empirical size that is close to 0.05 when $\phi_i = \rho = 1$. Under-specification of the number of common factors does not seem to cause major size distortions for Case 1, although this is not the case for Case 3 and, to a lesser extent, for Case 2, for large N and T . As for the constant case, the statistic becomes conservative if the assumption of common order of integration of the idiosyncratic and common factor components is not met. The empirical power increases as T and N increase, and decreases as the model specification gets more complicated—this is something to be expected due to the presence of a higher number of parameters to be estimated. For completeness, we have also included the results that are based on the badly-behaved $CIPS(\hat{\lambda})$ statistic, which behaves similar to $CIPS(\hat{\lambda}_{tr})$. This result might be surprising, but we need to take into account that the term that invalidates the use of $CIPS(\hat{\lambda})$ in the limit also depends on the magnitude of the change in the slope of the time trend, so that small values of $\|\theta_{i,2}\|$ might have little effect on the limiting distribution.

We have also analyzed the performance of the statistics under local-to-zero breaks through the definition of shrinking break magnitudes for all parameters related to the structural instability. To be specific, $\theta_{i,1}$, $\theta_{i,2}$, ς_1 , $\phi_{i,1}^y$, and $\phi_{i,1}^x$ defined above have been rescaled by $T^{-1/2-\delta}$, $\delta > 0$ —we have used an arbitrarily small positive value of $d = 0.01$ —so that in the limit the effects of the structural break tend to zero.⁸ Table C.8 indicates that the performance of the $CIPS(\hat{\lambda}_{tr})$ statistic under shrinking breaks is similar to the fixed breaks one—the empirical size is close to the nominal one, whereas in some cases a mild drop in the empirical power is observed, although the empirical power equals one as T gets large.

The finite sample performance of $CIPS(\tilde{\lambda})$ is investigated in Table C.7, which reveals a test statistic with an empirical size that

⁸The specification of the shrinking factor $T^{-1/2-\delta}$, $\delta > 0$, is motivated by the fact that $(\hat{\lambda} - \lambda^0) = O_p(T^{-1/2})$ for Models A2, B2, and C2.

is close to 0.05 in most cases—with the exception of a mild over-rejection distortion found for Case 3 when $m^0 = m = 2$ with large T . The values for the empirical power of $CIPS(\tilde{\lambda})$ are higher than the $CIPS(\hat{\lambda}_{tr})$ ones—and similar to the $CIPS(\hat{\lambda})$ ones—for a given set of ρ , T , and N values. As for the shrinking breaks configuration, we observe that the empirical size of $CIPS(\tilde{\lambda})$ is close to 0.05, although the empirical power is reduced when compared to the results that are based on fixed break magnitudes. To some extent, this is something to be expected since in this case, although $\tilde{\lambda}$ is consistent, the rate at which $\tilde{\lambda}$ tends to λ^0 is reduced—it has been shown that $(\tilde{\lambda} - \lambda^0) = o_p(N^{-1/3}T^{-1})$ for the fixed breaks case. From this point of view, it seems that $CIPS(\hat{\lambda}_{tr})$ outperforms $CIPS(\tilde{\lambda})$, although the empirical power equals one as T gets large.

Contrary to what has been done for the constant case, here we cannot compare the performance of the statistic with other existing proposal in the literature, since the case of panel cointegration testing with unknown structural changes that affect the slope of the time trend have not been previously addressed in the literature—Banerjee and Carrion-i-Silvestre (2015) only deal with the unknown breaks date situation for the constant case. However, we have investigated the properties of the statistical inference when the structural breaks are ignored with the computation of the panel cointegration statistic in Banerjee and Carrion-i-Silvestre (2017), which results are summarized in Table C.9. Unfortunately, in this case the empirical size is not controlled since, for instance, for Case 1 with $m^0 = m = 1$, the empirical size tends toward one as T and N increase. The size distortion is less noticeable for Case 3 with $m^0 = m = 1$, although it is still important in some cases—see results for $m^0 = m = 2$ and $T = 200$. In addition, the violation of the assumption of common order of integration of the idiosyncratic and common components under the null hypothesis of spurious regression does not seem to lead to a conservative test statistic. Finally, even in those cases where the size distortions are not very large—that is, Case 1 with $m^0 = m = 1$, $N = 20$ and $T = 50$ —the empirical power of the statistic is only slightly above the empirical size—that is, 0.07 versus 0.09. Again, this suggests that empirical analyses should include parameter instabilities in the model specification if there is evidence that structural changes might have affected the model.

5.2.3. Estimation of the Number of Structural Breaks

Table C.11 provides the frequency of the estimated number of structural breaks that derives from the use of the panel BIC statistic, considering up to two structural breaks. The ability of the panel BIC to select the correct number of structural breaks is very good, with a correct detection frequency that range between 0.92 and 1 regardless of N , T , ϕ_i , and ρ , for both $\hat{\lambda}$ and $\tilde{\lambda}$ based estimators, and model specifications. Table C.12 presents the frequency of the number of structural breaks estimation for model specifications that do not include structural breaks ($J^0 = 0$). For Model B2 the frequency of correct classification lies in the ranges [0.73, 0.92] for $J(\hat{\lambda})$ and [0.69, 0.92] for $J(\tilde{\lambda})$ under the null hypothesis of spurious regression. For Model C2 the classification ranges narrow to [0.93, 0.96] for $J(\hat{\lambda})$ and [0.90, 0.97] for $J(\tilde{\lambda})$ under the null hypothesis of spurious regression. Under the alternative hypothesis of cointegration,

the frequency of correct estimation of the number of structural breaks equals one for both $J(\hat{\lambda})$ and $J(\tilde{\lambda})$ statistics. This leads us to suggest the use of the panel BIC statistic in empirical applications.

6. Empirical Illustration

A well-functioning housing market has been shown to be very relevant for the proper evolution of credit and financial markets, which in turn has effects on macroeconomic variables such as output, fiscal deficit and unemployment. The empirical evidence that analyses the potential relationship between housing prices and real disposable income per capita is mixed, and depends both on the scope (national or regional) and on the period of analysis. Following Holly, Pesaran and Yamagata (2010), we focus on the US economy considering the 48 contiguous U.S. States and the District of Columbia ($N = 49$) using annual data between 1975 to 2019 ($T = 45$)—see Holly, Pesaran and Yamagata (2010) for the sources of the statistical information—with the model:

$$hp_{i,t} = \alpha_i + \beta y_{i,t} + u_{i,t}, \quad (19)$$

where $hp_{i,t}$ denotes the logarithm of the real housing prices index and $y_{i,t}$ is the logarithm of the real disposable income per capita. They argue that the use of panel data cointegration analysis can provide better statistical inference given the short time period of the available information. The preliminary analysis conducted in their paper reveals that cross-section dependence is present in the dataset.

As discussed in Holly, Pesaran and Yamagata (2010), the boom house prices started in early 2000 in the United States and accelerated during 2003–2006, something that has been interpreted in the literature as a housing price bubble. Figures C.73 and C.74 depict the variables, which show hump-shaped behavior in some of the house price time series during mid-eighties and in the middle of the first decade of 2000. If this is the case, it is safe to assume that the potential long-run relationship between housing prices and disposable income per capita might have been affected. This in turn implies that parameter instabilities should be accounted for in the model. The empirical specification that is used here generalizes the one in (19) by considering the effects of structural breaks as follows:

$$hp_{i,t} = \mu_{i,j,1} + \beta_j y_{i,t} + \bar{Y}_t' \delta_{ij}^y + v_{i,t} \quad T_{j-1} < t \leq T_j, \quad (20)$$

with $\bar{Y}_t = (\bar{h}_{p,t}, \bar{y}_t)'$, $\bar{h}_{p,t} = N^{-1} \sum_{i=1}^N hp_{i,t}$, $\bar{y}_t = N^{-1} \sum_{i=1}^N y_{i,t}$ and $j = 1, \dots, J$.

Table 2 summarizes the results that are based on both break dates estimation procedures discussed in Section 3, although it has to be born in mind that \tilde{T}_B clearly outperforms \hat{T}_B , so that more weight should be imposed on the conclusions drawn from the former estimator. The order of augmentation p_i in (17) is selected using the modified Akaike's information criterion (MAIC) in Ng and Perron (2001) and Perron and Qu (2007) for each panel unit with up to five lags, and the trimming is set at $\epsilon = 0.2$. Let us first focus on the specification that considers one structural break. As can be seen, there is strong evidence of panel

Table 2. Parameter estimates and *CIPS* panel cointegration test statistic.

<i>J</i>	Model	\hat{T}_B	\tilde{T}_B	$\hat{\beta}$	Panel BIC	<i>CIPS</i>	
						<i>m</i> = 1	<i>m</i> = 2
0				1.23	−6.28	−2.76**	−2.82**
1	B1	1984		(0.78, 0.87)′	−6.00	−2.87**	−3.39**
1	C1	1991		(0.62, 0.69)′	−6.14	−1.79	−2.99**
1	B1		1985	(0.91, 0.89)′	−5.98	−2.50**	−3.09**
1	C1		1986	(0.27, 0.73)′	−6.29	−2.26	−2.87**
2	B1	(1984, 1993)′		(0.77, 0.94, 0.90)′	−5.81	−2.58	−2.71
2	C1	(1988, 2004)′		(0.43, 0.82, 0.45)′	−6.07	−3.15**	−1.99
2	B1		(1984, 2009)′	(0.86, 0.91, 0.77)′	−5.76	−3.58**	−3.20**
2	C1		(1986, 2003)′	(0.27, 0.70, 0.54)′	−6.04	−3.12**	−3.02

** and * denote rejection of the null hypothesis at the 5% and 10% significance levels, respectively

cointegration for Model B1, since the null hypothesis of spurious regression is rejected at the 5% significance level regardless of *m*. Evidence of cointegration is also found for Model C1, although only when *m* = 2. This situation illustrates the potential effect on the empirical power of *CIPS* when *m* < *m*⁰, which can be reduced if a nonstationary common factor is not accounted for. Although the estimated break date— $\hat{T}_B = 1984$ and $\tilde{T}_B = 1985$ for Model B1, and $\hat{T}_B = 1991$ and $\tilde{T}_B = 1986$ for Model C1—is not located around the 2003–2006 housing price bubble period mentioned above, it does reflect important changes that experience the U.S. real state market—that is, the deregulations of the mortgage market, the development of a secondary mortgage market and the increasing role of the government-sponsored enterprises that affected the U.S. economy during the first half of the eighties; see Gerardi, Harvey and Willen (2010) and Ahamada and Diaz-Sanchez (2013). It is interesting to note that all estimated elasticities are below one, with the major change observed for Model C1 with \tilde{T}_B .

The conclusions that are obtained for the two structural breaks depend on the break estimation procedure. Evidence of panel cointegration is quite weak when using \hat{T}_B —that is, the null hypothesis of spurious regression is only rejected for Model C1 with *m* = 1—and is stronger when using \tilde{T}_B —panel cointegration is found except for Model C1 with *m* = 2. It is worth noting that for the latter, the first estimated break date detects the policy changes that experienced the U.S. housing market during the eighties, whereas the second break date is close to the global financial crisis (Model B1) or to the housing prices bubble discussed above (Model C1). As for the parameter estimates, income elasticity shows a mild (large) increase from the first to the second regimes Model B1 (Model C1), to experience a decrease from the second to the third regimes for both models.

Finally, we have also reported the value of the panel BIC statistic that has been proposed in the article as a way to estimate the number of structural breaks. The model specification that minimizes the panel BIC is the one given by Model C1 with *J* = 1 and \tilde{T}_B . This result reinforces the analysis the overall discussion about the existence of a long-run relationship between the logarithm of the real housing prices index and the logarithm of the real disposable income per capita for the U.S. States, a conclusion that is robust to the accommodation of one unknown structural break.

7. Conclusions

The article has shown that a consistent estimate of the long-run average coefficient can be obtained when cross-section dependence is present among the panel data units. The type of cross-section dependence that is considered in the article is strong, which is accounted for using an approximate common factor model. The model specification is quite flexible and allows for multiple structural breaks that can affect the deterministic component, the cointegrating vector and/or the loadings of the common factors. The estimation procedure that is applied is based on the CCE approach in Pesaran (2006), which approximates the unobserved common factors using cross-section averages of the observable variables of the model. Our result contributes to the literature of nonstationary panel data analysis, where consistent estimation of the parameters of the model is feasible in a spurious regression framework. The article conducts an extensive simulation exercise to study the finite sample performance of the estimator and test statistic that has been proposed in the article.

The application of the procedures that are designed in the article is illustrated with a model that defines a potential relationship between housing prices and real disposable income per capita. The analysis builds upon the use of U.S. regional data that covers a long time period. The main conclusion that is drawn indicates that robust evidence on the existence of a long-run relationship between housing prices and real disposable income per capita can be obtained once the presence of structural breaks that capture relevant events for the U.S. housing market are allowed for in the model specification.

Supplementary Materials

Appendix A. Mathematical appendix: Collects the proofs of the theorems.

Appendix B. Tables of critical values: Presents the tables of critical values for the individual and panel data cointegration test statistics, considering one structural break.

Appendix C. Monte Carlo experiment and empirical illustration: Pooled CCE estimator, break fraction histograms and empirical size and power of the panel cointegration statistics. Plots of the variables used in the empirical application.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

Carrion-i-Silvestre acknowledges the financial support from the grant PID2020-114646RB-C41 funded by MCIN/AEI/10.13039/501100011033.

References

- Ahamada, I., and Diaz-Sanchez, J. L. (2013), "A Retrospective Analysis of the House Price Macro-Relationship in the United States," *International Journal of Central Banking*, December, 153–174. [132]
- Bai, J. (2010), "Common Breaks in Means and Variances for Panel Data," *Journal of Econometrics*, 157, 78–92. [125,126]
- Bai, J., and Perron, P. (1998), "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47–78. [125]
- Bai, J., Kao, C., and Ng, S. (2009), "Panel Cointegration with Global Stochastic Trends," *Journal of Econometrics*, 149, 82–99. [124]
- Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221. [126]
- (2004), "A PANIC Attack on Unit Roots and Cointegration," *Econometrica*, 72, 1127–1177. [122,123,126,129]
- Baltagi, B. H., Feng, Q., and Kao, C. (2016), "Estimation of Heterogeneous Panels with Structural Breaks," *Journal of Econometrics*, 191, 176–195. [125,126]
- (2019), "Structural Changes in Heterogeneous Panels with Endogenous Regressors," *Journal of Applied Econometrics*, 34, 883–892. [125,126]
- Banerjee, A., and Carrion-i-Silvestre, J. L. (2015), "Cointegration in Panel Data with Breaks and Cross-Section Dependence," *Journal of Applied Econometrics*, 30, 1–23. [122,123,124,129,131]
- (2017), "Testing for Panel Cointegration Using Common Correlated Effects Estimators," *Journal of Time Series Analysis*, 38, 610–636. [122,125,126,129,131]
- Banerjee, A., Marcellino, M., and Osbat, C. (2005), "Testing for PPP: Should We Use Panel Methods?" *Empirical Economics*, 30, 77–91. [130]
- Ditzen, J., Karavias, Y., and Westerlund, J. (2023), "Multiple Structural Breaks in Interactive Effects Panel Data and the Impact of Quantitative Easing on Bank Lending," working paper arXiv:2211.06707. [125]
- Gengenbach, C., Urbain, J. P., and Westerlund, J. (2016), "Error Correction Testing in Panels with Common Stochastic Trends," *Journal of Applied Econometrics*, 31, 982–1004. [122]
- Gerardi, K. S., Harvey, S. R., and Willen, P. S. (2010), "The Impact of Deregulation and Financial Innovation on Consumers: The Case of the Mortgage Market," *The Journal of Finance*, LXV, 1, 333–360. [132]
- Holly, S., Pesaran, M. H., and Yamagata, T. (2010), "A Spatio-Temporal Model of House Prices in the USA," *Journal of Econometrics*, 158, 160–173. [125,126,131]
- Kim, D. (2011), "Estimating a Common Deterministic Time Trend Break in Large Panels with Cross-Sectional Dependence," *Journal of Econometrics*, 164, 310–330. [125,126]
- (2014), "Common Breaks in Time Trends for Large Panel Data with a Factor Structure," *Econometrics Journal*, 17, 301–337. [125,126]
- Kim, D., and Perron, P. (2009), "Unit Root Tests Allowing for a Break in the Trend Function under both the Null and Alternative Hypotheses," *Journal of Econometrics*, 148, 1–13. [123,127,128]
- Moon, H. R., and Perron, B. (2008), "Asymptotic Local Power of Pooled t-ratio Tests for Unit Roots in Panels with Fixed Effects," *Econometrica Journal*, 11, 80–104. [128]
- Ng, S., and Perron, P. (2001), "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," *Econometrica*, 69, 1519–1554. [127,131]
- Perron, P. (1989), "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, 57, 1361–1401. [128]
- (1990), "Testing for a Unit Root in a Time Series with a Changing Mean," *Journal of Business & Economic Statistics*, 8, 153–162. [123,126,128]
- Perron, P., and Qu, Z. (2007), "A Simple Modification to Improve the Finite Sample Properties of Ng and Perron's Unit Root Tests," *Economics Letters*, 94, 12–19. [131]
- Perron, P., and Zhu, X. (2005), "Structural Breaks with Deterministic and Stochastic Trends," *Journal of Econometrics*, 129, 65–119.
- Pesaran, M. H. (2006), "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74, 967–1012. [122,125,132]
- (2007), "A Simple Panel Unit Root Test in the Presence of Cross Section Dependence," *Journal of Applied Econometrics*, 22, 265–312. [123,125,126,127,128]
- (2015), *Time Series and Panel Data Econometrics*, Oxford: Oxford University Press. [125]
- Pesaran, M. H., Smith, L. V., and Yamagata, T. (2013), "Panel Unit Root Tests in the Presence of a Multifactor Error Structure," *Journal of Econometrics*, 175, 94–115. [126,127,128]
- Phillips, P. C. B., and Moon, H. R. (1999), "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057–1111. [122,123]
- (2000), "Nonstationary Panel Data Analysis: An Overview of Some Recent Developments," *Econometric Reviews*, 19, 263–286. [122]
- Urbain, J. P., and Westerlund, J. (2011), "Least Squares Asymptotics in Spurious and Cointegrated Panel Regressions with Common and Idiosyncratic Stochastic Trends," *Oxford Bulletin of Economics and Statistics*, 73, 119–139. [123]
- Westerlund, J. (2018), "CCE in Panels with General Unknown Factors," *Econometrica Journal*, 21, 264–276. [123,129]
- Westerlund, J., and Edgerton, D. (2008), "A Simple Test for Cointegration in Dependent Panels with Structural Breaks," *Oxford Bulletin of Economics and Statistics*, 70, 665–704. [122]
- Westerlund, J., Hossainkouchack, M., and Solberger, M. (2016), "The Local Power of the CADF and CIPS Panel Unit Root Tests," *Econometric Reviews*, 35, 845–870. [128]