Cointegration in panel data with breaks and cross-section dependence

Anindya Banerjee*
Department of Economics
European University Institute

Josep Lluís Carrion-i-Silvestre
Department of Econometrics, Statistics and Spanish Economy
University of Barcelona

December 11, 2006

Abstract

The power of standard panel cointegration statistics may be affected by misspecification errors if structural breaks are not accounted for. We allow for one structural break when testing the null hypothesis of no cointegration that retain good properties in terms of empirical size and power. Response surfaces to approximate the finite sample moments that are required to implement the statistics are provided. Since panel cointegration statistics rely on the assumption of cross-section independence, a generalisation of the tests to the common factor framework is carried out in order to allow for dependence among the units of the panel.

Keywords: Panel cointegration, structural break, common factors, cross-section dependence
JEL Codes: C12, C22

*Corresponding author: Professor Anindya Banerjee. Department of Economics. European University Institute. Villa San Paolo. Via della Piazzuola 43. 50133 Florence, Italy. Tel: +39-055-4685-956; Fax: +39-055-4685-902; E-mail: anindya.banerjee@eui.eu. Banerjee thanks the Research Council of the European University Institute for financial assistance for this research, and Carrion-i-Silvestre acknowledges the financial support from the Generalitat de Catalunya under BE-2004 grant. We also thank participants at seminars in Copenhagen, Ente Einaudi (Rome), City University (London), Maastricht University, and University of Algarve (Portugal) for their comments. For useful discussions we also thank Jushan Bai, Peter Boswijk, Jesús Gonzalo, Soren Johansen, Pierre Perron and Jean-Pierre Urbain.

1 Introduction

The theory of cointegration establishes that there exist linear combinations of integrated variables that cancel out common stochastic trends. This phenomenon gives rise to equilibrium relationships among integrated variables, which means that in the long run these variables show co-movement or are cointegrated with each other. Although a large part of the traditional theory has been based upon the assumption of structural stability, the concept of cointegration *per se* does not rule out the possibility that both the cointegrating vector(s) and the deterministic component(s) of the long run relationship might change along the analyzed time period. In fact, Hansen (1992), and Quintos and Phillips (1993) propose test statistics to assess the stability of the cointegration relationship. More interestingly, it is well known that if no account is taken of changes in the parameters of the model, inference concerning the presence of cointegration can be affected by misspecification errors. This in turn can bias conclusions towards accepting the null hypothesis of no cointegration – *e.g.* see Campos, Ericsson and Hendry (1996), and Gregory and Hansen (1996).

All these considerations have motivated the search to design procedures to test for cointegration allowing for structural breaks. Thus, Gregory and Hansen (1996) generalized the standard cointegration approach in Engle and Granger (1987) to allow for the presence of structural breaks that might affect either the deterministic component or the cointegration vector of the long run relationship. Hao (1996), Bartley, Lee and Strazicich (2001), and Carrion-i-Silvestre and Sansó (2006) use the multivariate version of the KPSS statistic in Harris and Inder (1994), and Shin (1994) to test for the null of cointegration with one structural break. Finally, Hansen and Johansen (1999), and Busetti (2002) propose methods to estimate the cointegration rank in a multivariate framework.

These proposals are extremely relevant for the imperatives that arise in empirical modelling where structural breaks are very common. Gregory and Hansen (1996) and Gabriel, Da Silva and Nunes (2002) investigate the long run money demand for the U.S. and Portugal, respectively. Busetti (2002) conducts two illustrations using road casualties in Great Britain, and macroeconomic data for the UK. Finally, Clemente, Marcuello, Montañés and Pueyo (2004) focus on health care expenditure demand functions. The main conclusion that arises from these applications is that inference on cointegration analysis can be affected by the presence of structural breaks. Other applications that may be envisaged for this methodology include looking at models of convergence, real exchange rates, exchange rate pass through and the issue of the solvency of the current account and its relation to the budget deficit, the so-called Feldstein-Horioka puzzle.

The literature on panel data econometrics with integrated data has experienced rapid development since 1990s. The driving force behind the popularity of the use of the panel data techniques is the idea that the power of tests for unit roots and cointegration might be increased by combining the information that comes from the cross-section ($i = 1, ..., N$) and the time ($t = 1, 2, \ldots, T$) dimensions, especially when the time dimension is restricted by the lack of availability of long series of reliable time-series data. As a result, new statistics to assess the stochastic properties of panel data sets have appeared in the literature – see Banerjee (1999), Baltagi (2005) and Breitung and Pesaran (2005) for an overview of the field.

Surprisingly, the issue of instability has not received a great deal of attention in the panel data cointegration framework. In this regard, Kao and Chiang (2000) analyze instability in cointegration relationships assuming that cointegration is present, with a homogeneous cointegrating vector in all the units of the panel – although it is possible to split the panel into two sub-panels using a bootstrap scheme – and a common break date. Breitung (2005) proposes a VAR-based panel data cointegration procedure that permits the introduction of dummy variables outside the long run relationship. Finally, Westerlund (2006) extends the LM statistic in McCoskey and Kao (1998) by allowing for structural breaks.

As may be seen, the scope of the literature that addresses the panel data cointegration hypothesis testing allowing for structural breaks is fairly limited. The first contribution of our paper is therefore to generalize the approach in Pedroni (1999, 2004) to account for one structural break that may affect the long run relationship in a number of different ways. Our proposal applies more generally to the class of static-equation-based panel tests for cointegration but does not extend to cointegrated vector error correction models (VECM) for panels with integrated data for which more work is needed in order to develop feasible procedures.

Pedroni proposes seven statistics depending on the way that the individual information is combined to define the panel tests. The statistics can be grouped into either parametric or non-parametric statistics, depending on the way that autocorrelation and endogeneity bias are treated. In this paper we focus only on the parametric statistics, since these are at least asymptotically equivalent to their non-parametric counterparts. A Monte Carlo study, which could be constructed straightforwardly, would reveal the behaviour of non-parametric tests in finite samples when compared to the parametric tests, but is not included here solely for the sake of concision.

One important feature to consider in these tests is cross-section dependence. Most panel data statistics – including those due to Pedroni – assume cross-section independence, except for common time effects. This is in many contexts a highly restrictive assumption to make. As our second contribution, we address this concern by using a factor model approach due to Bai and Ng (2004) to generalize the degree of permissible cross-section dependency to allow for idiosyncratic responses to multiple common factors.

Taken together we thereby generalize the class of panel cointegration tests to allow for both structural breaks and cross-section dependence. The limiting distributions of the statistics are derived and new sets of critical values are computed wherever required.

Our paper takes the following shape. In Section 2 the interest of our proposal is motivated through Monte Carlo simulations. Section 3 presents the models and statistics for the null hypothesis of no cointegration with power against the alternative of broken cointegration. The moments that are required for the computation of the panel data statistics are computed in this section. In this regard, we estimate response surfaces to approximate these moments for whichever sample size. Section 4 extends the approach to the common factor framework. Section 5 focuses on the finite sample properties of the statistics. Section 6 provides an empirical illustration of the use of our tests using data on exchange rate pass through. The issue of the degree of exchange rate pass through is an important focus of investigation in the macroeconomics literature, although much of the testing has been undertaken under severely restrictive assumptions. Finally, Section 7 concludes with some remarks.

Some of the proofs of the results are collected in the Appendix. For the remaining proofs, the reader is referred either to existing sources in the literature or to a companion appendix available from the authors on request. The same remark applies to the tables not reported in this paper.

2 Motivation

Pedroni (1999, 2004) proposes seven statistics to test the null hypothesis of no cointegration using single-equation methods based on the estimation of static regressions. Since the statistics are based on single-equation methods the cointegrating rank for each unit is either 0 or 1, with a heterogeneous cointegrating vector for each unit. After estimating individual static regressions for each unit, the cointegrating residuals are used to compute each of the statistics. The seven statistics are classified into two different groups depending on whether they are within-dimension-based statistics – homogeneity is assumed when computing the cointegration test statistics – or between-dimension-based statistics where heterogeneous behaviour (across the units of the panel) is allowed. As mentioned in the introduction, we are concerned only with the parametric version of the statistics, i.e. the normalized bias and the pseudo t-ratio statistics.

To motivate our proposal we analyze the effects of structural breaks on the parametric group of Pedroni statistics through Monte Carlo simulations. Only the results for the pseudo t-ratio statistics are reported in the Appendix. Similar results are available upon request for the normalized bias. First, we focus on the case where there is cointegration but the deterministic component changes at a point in time. Subsequently we also consider the case of an unstable cointegrating vector.

The Data Generating Process (DGP) is given by:

$$
\begin{aligned}
y_{i,t} &= f_i(t) + x_{i,t}'\delta_{i,t} + z_{i,t} \\
\Delta x_{i,t} &= \varepsilon_{i,t} \\
z_{i,t} &= \rho_i z_{i,t-1} + v_{i,t} \\
\zeta_{i,t} &= (\varepsilon_{i,t}, v_{i,t})' \sim iid\ N(0, I_2),
\end{aligned}
$$

where $f_i(t)$ denotes the deterministic component.

Four different cases are considered. Firstly, we have $f_i(t) = \mu_i + \theta_i DU_{i,t}$ with $DU_{i,t} = 1$ for $t > T_{bi}$ and 0 otherwise, where $T_{bi} = \lambda_i T$, $\lambda_i \in (0,1)$, denotes the date of the break. The parameter set is given by $\mu_i = 1$, $\theta_i = \{0, 1, 3, 5, 10\}$, $\delta_{i,t} = \delta_i = 1$, and $\lambda_i = \{0.25, 0.5, 0.75\}$. The autoregressive parameter comes from the set $\rho_i = \{0, 0.5\}$. The sample size is $T = \{100, 200\}$, the number of units is $N = \{20, 40\}$ and the results are based on $1,000$ replications. For simplicity but without loss of generality, we have specified a common break date for all units in all the simulations. The model that has been estimated to compute the pseudo t-ratio Pedroni panel data cointegration test statistics includes a constant term (individual effects) as the deterministic component.

Secondly, we have also analyzed the case where the structural break changes both the level and the slope of the time trend. The deterministic function is given by $f_i(t) = \mu_i + \theta_i DU_{i,t} + \beta_i t + \gamma_i DT^*_{i,t}$, where $\mu_i = 1$, $\theta_i = 3$, $\beta_i = 0.3$ and $DT^*_{i,t}$ is the dummy variable defined above. Note that in this case the pseudo t-ratio statistic has been computed using a time trend as the deterministic component.

The third case studies the effects of a change both in the level and in the cointegrating vector. As before, the deterministic component is $f_i(t) = \mu_i + \theta_i DU_{i,t}$, with $\mu_i = 1$ and $\theta_i = \{0, 3\}$. Now we focus on the change in the cointegrating vector specifying $\delta_{i,t} = \delta_{i,1} = 1$ for $t \leq T_{bi}$ and $\delta_{i,t} = \delta_{i,2} = \{0, 2, 3, 4, 5, 10\}$ for $t > T_{bi}$. The model estimated to compute the (pseudo t-ratio) Pedroni panel data cointegration statistic includes a constant term as the deterministic component.

Finally, the fourth case considers a change in the level and time trend, that defines the deterministic component, together with a change in the cointegrating vector. In this case $f_i(t) = \mu_i + \theta_i DU_{i,t} + \beta_i t + \gamma_i DT^*_{i,t}$, with $\mu_i = 1$, $\theta_i = 3$, $\beta_i = 0.3$, $\gamma_i = 0.5$, and $\delta_{i,t} = \delta_{i,1} = 1$ for $t \leq T_{bi}$ and $\delta_{i,t} = \delta_{i,2} = \{0,\ 2, 3, 4, 5,\ 10\}$ for $t > T_{bi}$. The model estimated to compute the pseudo t-ratio Pedroni panel data cointegration statistic includes individual and time effects.

Detailed results of the simulations for all four cases are available from us in Tables A.1 to A.3 of the companion appendix to this paper. In the first case, our results show that the effect of a change in level only matters in those situations where the magnitude of the change is large and the break date is located at the end of the time period. Therefore, we can conclude that for small and moderate changes in level the misspecification error of the deterministic component does not damage the power of Pedroni statistic. However, in the second case the consequences of the misspecification error are more serious, since the empirical power approaches zero as the magnitude of the change in trend (γ_i) increases when the break date is placed either in the middle ($\lambda_i = 0.5$) or at the end ($\lambda_i = 0.75$) of the period. In the third case, for the empirical power to diminish the change in the cointegrating vector has to be either moderate or large, and be located in the middle ($\lambda_i = 0.5$) or at the end ($\lambda_i = 0.75$) of the period. Notice that this conclusion is reached irrespective of the change in level that affects the constant term.

Finally, when the level, time trend and the cointegrating vector change, and a model estimated to compute the pseudo t-ratio Pedroni panel data cointegration statistic includes individual and time effects, the change in the trend implies further reductions on the empirical power of the statistic when the break date is located in the middle and at the end of the period.

In summary, we may conclude that misspecification errors due to the lack of accounting for a structural break can reduce the power of the panel data cointegration test in Pedroni (2004) in those cases where the break date is placed in the middle or at the end of the time period. Therefore, we observe a bias towards the spurious non-rejection of the null hypothesis of no cointegration. A relevant feature is that the power distortions seem to appear only when the break changes either the slope of the time trend or the cointegrating vector, but no effects are seen when the break only affects the constant term.

3 Models and test statistics

In order to consider the issues described above more formally, let $\{Y_{i,t}\}$ be a $(m \times 1)$-vector of non-stationary stochastic process whose elements are individually I(1). Moreover, let us assume that the DGP that describes $Y_{i,t}$ is given by the following triangular representation

$$\Delta x_{i,t} = \varepsilon_{i,t}$$
$$y_{i,t} = f_i(t) + x'_{i,t}\delta_{i,t} + e_{i,t}$$

where $Y_{i,t} = \left(y_{i,t}, x'_{i,t}\right)'$ is conveniently partitioned into a scalar $y_{i,t}$ and $x_{i,t}$ $(r \times 1)$-vector, $i = 1, \ldots, N$, $t = 1, \ldots, T$. The disturbance terms $\xi_{i,t} = \left(\varepsilon'_{i,t}, e_{i,t}\right)'$ are assumed to satisfy the multivariate invariance principle in Phillips and Durlauf (1986), and to be independent across units, i.e. $E\left(\xi_{i,t}, \xi'_{j,s}\right) = 0\ \forall i, j, t, s$.

The general functional form for the deterministic term $f(t)$ is given by

$$f_i(t) = \mu_i + \beta_i t + \theta_i DU_{i,t} + \gamma_i DT^*_{i,t}, \tag{1}$$

where

$$DU_{i,t} = \begin{cases} 0 & t \leq T_{bi} \\ 1 & t > T_{bi} \end{cases} ; DT^*_{i,t} = \begin{cases} 0 & t \leq T_{bi} \\ (t - T_{bi}) & t > T_{bi} \end{cases},$$

with $T_{bi} = \lambda_i T$ denoting the time of the break for the i-th unit, $i = 1, \ldots, N$, $\lambda_i \in \Lambda$, where Λ is a closed subset of $(0, 1)$.[1] Note also that the cointegrating vector is specified as a function of time so that

$$\delta_{i,t} = \begin{cases} \delta_{i,1} & t \leq T_{bi} \\ \delta_{i,2} & t > T_{bi} \end{cases}.$$

Using these elements, we propose up to six different model specifications:

- Model 1. Constant term with a change in level but stable cointegrating vector:

$$y_{i,t} = \mu_i + \theta_i DU_{i,t} + x'_{i,t}\delta_i + e_{i,t} \tag{2}$$

- Model 2. Time trend with a change in level but stable cointegrating vector:

$$y_{i,t} = \mu_i + \beta_i t + \theta_i DU_{i,t} + x'_{i,t}\delta_i + e_{i,t} \tag{3}$$

- Model 3. Time trend with change in both level and trend but stable cointegrating vector:

$$y_{i,t} = \mu_i + \beta_i t + \theta_i DU_{i,t} + \gamma_i DT^*_{i,t} + x'_{i,t}\delta_i + e_{i,t} \tag{4}$$

- Model 4. Constant term with change in both level and cointegrating vector:

$$y_{i,t} = \mu_i + \theta_i DU_{i,t} + x'_{i,t}\delta_{i,t} + e_{i,t} \tag{5}$$

[1] For instance, Gregory and Hansen (1996) follow the previous literature and define $\Lambda = [0.15, 0.85]$.

- Model 5. Time trend with change in both level and cointegrating vector (the slope of trend does not change):

$$y_{i,t} = \mu_i + \beta_i t + \theta_i DU_{i,t} + x'_{i,t}\delta_{i,t} + e_{i,t} \tag{6}$$

- Model 6. The time trend and the cointegrating vector change:

$$y_{i,t} = \mu_i + \beta_i t + \theta_i DU_{i,t} + \gamma_i DT^*_{i,t} + x'_{i,t}\delta_{i,t} + e_{i,t} \tag{7}$$

Using any one of these specifications we propose testing the null hypothesis of no cointegration against the alternative hypothesis of cointegration (with break) using the ADF test statistic applied to the residuals of the cointegration regression as in Engle and Granger (1987) and Gregory and Hansen (1996) but in the panel data framework developed in Pedroni (1999, 2004). In fact, Gregory and Hansen (1996) propose the specifications given by Models 1, 2 and 4 above, so that the specifications in Models 3, 5 and 6 allow us to extend their approach.

Our proposal can be described in the following steps. First and following Gregory and Hansen (1996), we proceed to the OLS estimation of one of the models given in (2) to (7) and run the following ADF type-regression equation on the estimated residuals ($\hat{e}_{i,t}(\lambda_i)$):

$$\Delta\hat{e}_{i,t}(\lambda_i) = \rho_i \hat{e}_{i,t-1}(\lambda_i) + \sum_{j=1}^{k} \phi_{i,j}\Delta\hat{e}_{i,t-j}(\lambda_i) + \varepsilon_{i,t}. \tag{8}$$

Note that the notation that is used refers to the break fraction (λ_i) parameter, which (if it exists) is in most cases unknown. In order to get rid of the dependence of the statistics on the break fraction parameter, Gregory and Hansen (1996) suggest estimating the models given in (2) to (7) for all possible break dates, subject to trimming, obtaining the estimated OLS residuals and computing the corresponding ADF statistic. With the sequence of ADF statistics in hand, we can also estimate the break point for each unit as the date that minimizes the sequence of individual ADF test statistics – either the t-ratio, $t_{\hat{\rho}_i}(\lambda_i)$, or the normalized bias, computed as $T\hat{\rho}_i(\lambda_i) = T\hat{\rho}_i\left(1 - \hat{\phi}_{i,1} - \cdots - \hat{\phi}_{i,k}\right)^{-1}$ – see Hamilton (1994), pp. 523. Gregory and Hansen (1996) derive the limiting distribution of $t_{\hat{\rho}_i}\left(\hat{\lambda}_i\right) = \inf_{\lambda_i\in\Lambda} t_{\rho_i}(\lambda_i)$ and $T\hat{\rho}_i\left(\hat{\lambda}_i\right) = \inf_{\lambda_i\in\Lambda} T\hat{\rho}_i(\lambda_i)$, which are shown not to depend on the break fraction parameter. Specifically, Gregory and Hansen (1996) show that $T\hat{\rho}_i\left(\hat{\lambda}_i\right) \Rightarrow \inf_{\lambda_i\in\Lambda} \int_0^1 Q(\lambda_i, s)\, dQ(\lambda_i, s) \Big/ \int_0^1 Q(\lambda_i, s)^2 ds$, and $t_{\hat{\rho}_i}\left(\hat{\lambda}_i\right) \Rightarrow \inf_{\lambda_i\in\Lambda} \int_0^1 Q(\lambda_i, s)\, dQ(\lambda_i, s) \Big/ \left[\int_0^1 Q(\lambda_i, s)^2 dr \left(1 + \varrho(\lambda_i)' D(\lambda_i)\, \varrho(\lambda_i)\right)\right]^{1/2}$, where $\Rightarrow$ denotes weak convergence, $Q(\lambda_i, s)$ and $\varrho(\lambda_i)$ are functions of Brownian motions and the deterministic component, and $D(\lambda_i)$ depends on the model – see the Theorem in Gregory and Hansen (1996) for further details. As mentioned above Gregory and Hansen (1996) deal only with some of the specifications in this paper, although their developments can be easily extended and similar limiting distributions obtained for the statistics. Note that the estimation of the break date $\hat{T}_{bi}$ is conducted as

$$\hat{T}_{bi} = \arg\min_{\lambda_i\in\Lambda} t_{\hat{\rho}_i}(\lambda_i)\,; \quad \hat{T}_{bi} = \arg\min_{\lambda_i\in\Lambda} T\hat{\rho}_i(\lambda_i)\,,$$

$\forall i = 1, \ldots, N$. At this point we could either follow Gregory and Hansen (1996) and test the null hypothesis for each unit or decide to combine the unit-specific information in a panel data statistic.

The panel statistics on which we focus in order to test the null hypothesis are given by the $Z_{\hat{\rho}_{NT}}$ and $Z_{\hat{t}_{NT}}$ tests in Pedroni (1999, 2004), which can be thought as analogous to the residual-based tests in Engle and Granger (1987). These test statistics are defined by pooling the individual ADF tests, so that they belong to the class of between-dimension test statistics. Specifically, they are computed as:

$$N^{-1/2} Z_{\hat{\rho}_{NT}} \left(\hat{\lambda}\right) = N^{-1/2} \sum_{i=1}^{N} T \hat{\rho}_i \left(\hat{\lambda}_i\right) \tag{9}$$

$$N^{-1/2} Z_{\hat{t}_{NT}} \left(\hat{\lambda}\right) = N^{-1/2} \sum_{i=1}^{N} t_{\hat{\rho}_i} \left(\hat{\lambda}_i\right), \tag{10}$$

where $\hat{\rho}_i \left(\hat{\lambda}_i\right)$ and $t_{\hat{\rho}_i} \left(\hat{\lambda}_i\right)$ are the estimated coefficient and associated t-ratio from (8) and

$$\hat{\lambda} = \left(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_i, \ldots, \hat{\lambda}_N\right)'$$

is the vector of estimated break fractions.

Note that in this framework we allow for a high degree of heterogeneity since the cointegrating vector, the short run dynamics and the break date estimate might differ among units. The use of the panel data cointegration test aims to increase the power of the statistical inference when testing the null hypothesis of no cointegration, but some heterogeneity is preserved when conducting the estimation of the parameters individually.

Following Pedroni (1999, 2004), the panel test statistics are shown to converge to standard Normal distributions once they have been properly standardized.

Theorem 1 *Let Θ and Ψ denote the mean and variance for the vector Brownian motion functional $\Upsilon' \equiv (\inf_{\lambda_i \in \Lambda} \int_0^1 Q(\lambda_i, s)\, dQ(\lambda_i, s) \left[\int_0^1 Q(\lambda_i, s)^2\, ds\right]^{-1}, \inf_{\lambda_i \in \Lambda} \int_0^1 Q(\lambda_i, s)\, dQ(\lambda_i, s) \times [\int_0^1 Q(\lambda_i, s)^2\, ds \ \ (1 + \varrho(\lambda_i)' D(\lambda_i) \varrho(\lambda_i))]^{-1/2})$. Then, as $(T, N \to \infty)_{\text{seq}}$ we have that under the null hypothesis of no cointegration the asymptotic distribution of the statistics $Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$ and $Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$ defined in (9) and (10), respectively, are given by*

$$N^{-1/2} Z_{\hat{\rho}_{NT}} \left(\hat{\lambda}\right) - \Theta_1 \sqrt{N} \Rightarrow N(0, \Psi_1)$$
$$N^{-1/2} Z_{\hat{t}_{NT}} \left(\hat{\lambda}\right) - \Theta_2 \sqrt{N} \Rightarrow N(0, \Psi_2).$$

Following Pedroni (2004), in order to prove Theorem 1 we require only the assumption of finite second moments of the random variables characterized as Brownian motion functionals, which will allow to apply the Lindberg-Levy Central Limit Theorem as $N \to \infty$. The moments of the limiting distributions, $\Theta_1, \Psi_1, \Theta_2$ and Ψ_2, are approximated by Monte Carlo simulation for the different specifications and allowing up to seven stochastic regressors in the cointegrating relationship – i.e. the dimension of the $Y_{i,t}$ $(m \times 1)$-vector goes from (2×1) to (8×1). Table

1 presents the moments of the limit distributions based on $T = 1,000$. As can be seen, the moments of the distribution depends both on the specification and the number of stochastic regressors.

Since the limit distribution of the tests can provide a poor approximation in finite samples, we have approximated the moments of the test statistics for different values of the sample size, specifically $T = \{30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 400, 500, 1,000\}$. In addition, the finite sample distributions depend on the procedure that is applied when selecting the order (k) of the parametric correction in (8). The results reported in Table 2 are for lag length selection using the t-sig criterion in Ng and Perron (1995) with a $k_{\max} = 5$ as the maximum order of lags. Since reporting the moments of the finite sample distribution for the different values of T and the number of stochastic regressors $p = (m-1)$. The general functional form that has been estimated is

$$g(T,p) = \sum_{l=0}^{3} \left(\beta_{0,l} + \beta_{1,l}\frac{1}{T} + \beta_{2,l}\frac{1}{T^2} + \beta_{3,l}\frac{1}{T^3} \right) p^l,$$

where $g(T,p)$ refer to $\Theta_1, \Psi_1, \Theta_2$ and Ψ_2, for the different model specifications. These functions have been estimated by OLS using the Newey-West robust covariance disturbance matrix to assess the individual significance of the regressors – the level of significance is 10%. Table 2 reports the estimated coefficients of the response surfaces. In all simulations 10,000 replications were used to simulate the moments. We also have the results for the finite sample distributions for fixed $k = 0$, 2 and 5. These are reported in the companion appendix in Tables A.4 to A.6.

4 Common factors in panel cointegration

In the sections above, we have generalized static-regression-based tests for cointegration to include structural breaks in the deterministic components of the processes. These derivations are valid only under the assumption that the units are cross-sectionally independent. However, this requirement is rarely likely to be satisfied in empirical economic applications where countries or regions depend each other. Therefore, in order to generalize the framework and applicability of the paper further, we have extended our approach to allow for cross-section dependence. We model such dependence by using common factors as in Bai and Ng (2004). In addition to dependence, our tests also can accommodate the presence of structural breaks.[2] We deal first with the case where the break date is known and then proceed to the more realistic scenario of an unknown break date.

[2] An alternative approach to dealing with cross-sectional dependence is proposed by Chang (2005) using a non-linear IV technique.

4.1 Break date(s) known

In this framework the model is given in structural form as:

$$y_{i,t} = f_i(t) + x'_{i,t}\delta_{i,t} + u_{i,t} \quad (11)$$
$$u_{i,t} = F'_t\pi_i + e_{i,t} \quad (12)$$
$$(I-L)F_t = C(L)w_t \quad (13)$$
$$(1-\rho_i L)e_{i,t} = H_i(L)\varepsilon_{i,t} \quad (14)$$
$$(I-L)x_{i,t} = G_i(L)v_{i,t}, \quad (15)$$

$t = 1,\dots,T$, $i = 1,\dots,N$, where $C(L) = \sum_{j=0}^{\infty} C_j L^j$, and $f_i(t)$ denotes the deterministic component (which may be broken as in (1) above), F_t denotes a $(r\times 1)$-vector containing the common factors, with π_i the vector of loadings. Despite the operator $(1-L)$ in equation (13), F_t does not have to be I(1). In fact, F_t can be I(0), I(1), or a combination of both, depending on the rank of $C(1)$. If $C(1) = 0$, then F_t is I(0). If $C(1)$ is of full rank, then each component of F_t is I(1). If $C(1) \neq 0$, but not full rank, then some components of F_t are I(1) and some are I(0). Our analysis is based on the same set of assumptions in Bai and Ng (2004), and Bai and Carrion-i-Silvestre (2005). Let $M < \infty$ be a generic positive number, not depending on T and N:

Assumption A: (i) for non-random π_i, $\|\pi_i\| \le M$; for random π_i, $E\|\pi_i\|^4 \le M$, (ii) $\frac{1}{N}\sum_{i=1}^{N}\pi_i\pi_i' \xrightarrow{p} \Sigma_\Pi$, a $(r\times r)$ positive definite matrix.

Assumption B: (i) $w_t \sim iid(0,\Sigma_w)$, $E\|w_t\|^4 \le M$, and (ii) $Var(\Delta F_t') = \sum_{j=0}^{\infty} C_j\Sigma_w C_j' > 0$, (iii) $\sum_{j=0}^{\infty} j\|C_j\| < M$; and (iv) $C(1)$ has rank r_1, $0 \le r_1 \le r$.

Assumption C: (i) for each i, $\varepsilon_{i,t} \sim iid\left(0,\sigma^2_{\varepsilon,i}\right)$, $E|\varepsilon_{i,t}|^8 \le M$, $\sum_{j=0}^{\infty} j|H_{i,j}| < M$, $\omega_i^2 = H_i(1)^2\sigma^2_{\varepsilon,i} > 0$; (ii) $E(\varepsilon_{i,t}\varepsilon_{j,t}) = \tau_{i,j}$ with $\sum_{i=1}^{N}|\tau_{i,j}| \le M$ for all j;
(iii) $E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\varepsilon_{i,s}\varepsilon_{i,t} - E(\varepsilon_{i,s}\varepsilon_{i,t})]\right|^4 \le M$, for every (t,s).

Assumption D: The errors $\varepsilon_{i,t}$, w_t, and the loadings π_i are three mutually independent groups.

Assumption E: $E\|F_0\| \le M$, and for every $i = 1,\dots,N$, $E|e_{i,0}| \le M$.

Assumption F: (i) $v_{i,t} \sim iid(0,\Sigma_v)$, $E\|v_{i,t}\|^4 \le M$, and (ii) $Var\left(\Delta x'_{i,t}\right) = \sum_{j=0}^{\infty} G_{i,j}\Sigma_v G'_{i,j} > 0$, (iii) $\sum_{j=0}^{\infty} j\|G_{i,j}\| < M$; and (iv) $G(1)$ has full rank.

Assumption G: (i) $E(e_{i,t}|v_{i,t}) = 0$ when stochastic regressors are assumed to be strictly exogenous or (ii) $E(e_{i,t}|v_{i,t}) = \Delta x'_{i,t}A_i(L) + \xi_{i,t}$, with $A_i(L)$ being a $(k\times 1)$-vector of lags and leads polynomials of finite orders and $\xi_{i,t} \sim iid(0,\Sigma_\xi)$, when stochastic regressors are non-strictly exogenous.

Assumption A ensures that the factor loadings are identifiable. Assumption B establishes the conditions on the short and long run variance of ΔF_t – i.e. the short-run variance matrix is positive definite and the long run variance matrix may have reduced rank in order to accommodate stationary linear combinations of $I(1)$ factors. Assumption C(i) allows for some weak serial correlation in $(1-\rho_i L)e_{i,t}$, whereas C(ii) and C(iii) allow for weak cross-section correlation. Assumption E defines the initial conditions. Assumption F establishes conditions on the first differences of the stochastic regressors. Finally, Assumption G defines two situations depending on whether the stochastic regressors are strictly exogenous regressors or non-strictly exogenous. This distinction is important here, because in the common factor framework the limiting distributions of the statistics do not depend on the number of stochastic regressors if strict exogeneity holds. However, this is no longer true when correlation between $e_{i,t}$ and $v_{i,s}$ is allowed and modifications need to be introduced to account for endogenous regressors. Here we suggest using the DOLS estimation method in Stock and Watson (1993) to account for endogeneity, where we assume that the number of leads and lags is fixed as in Stock and Watson (1993), although they can be chosen using a BIC information criterion.[3]

For ease of exposition, we assume strictly exogenous stochastic regressors, although the Appendix contains a discussion of the more general case. The estimation of the common factors is done as in Bai and Ng (2004). We compute the first differences:

$$\Delta y_{i,t} = \Delta f_i\left(t\right) + \Delta x_{i,t}'\delta_{i,t} + \Delta F_t'\pi_i + \Delta e_{i,t},$$

and take the orthogonal projections:

$$\begin{aligned} M_i\Delta y_i &= M_i\Delta F\pi_i + M_i\Delta e_i \\ &= f\pi_i + z_i, \end{aligned} \tag{16}$$

with $M_i = I - \Delta x_i^d\left(\Delta x_i^{d\prime}\Delta x_i^d\right)^{-1}\Delta x_i^{d\prime}$ being the idempotent matrix, and $f = M_i\Delta F$ and $z_i = M_i\Delta e_i$. The superscript d in Δx_i^d indicates that there are deterministic elements. The estimation of the common factors and factor loadings can be done as in Bai and Ng (2004) using principal components. Specifically, the estimated principal component of $f = (f_2, f_3, \ldots, f_T)$, denoted as $\tilde{f}$, is $\sqrt{T-1}$ times the r eigenvectors corresponding to the first r largest eigenvalues of the $(T-1)\times(T-1)$ matrix $y^*y^{*\prime}$, where $y_i^* = M_i\Delta y_i$. Under the normalization $\tilde{f}\tilde{f}'/\left(T-1\right) = I_r$, the estimated loading matrix is $\tilde{\Pi} = \tilde{f}'y^*/\left(T-1\right)$. Therefore, the estimated residuals are defined as

$$\tilde{z}_{i,t} = y_{i,t}^* - \tilde{f}_t'\tilde{\pi}_i. \tag{17}$$

We can recover the idiosyncratic disturbance terms through cumulation, i.e. $\tilde{e}_{i,t} = \sum_{j=2}^{t}\tilde{z}_{i,j}$, and test the unit root hypothesis ($\alpha_{i,0} = 0$) using the ADF regression equation

$$\Delta\tilde{e}_{i,t}\left(\hat{\lambda}_i\right) = \alpha_{i,0}\tilde{e}_{i,t-1}\left(\hat{\lambda}_i\right) + \sum_{j=1}^{k}\alpha_{i,j}\Delta\tilde{e}_{i,t-j}\left(\hat{\lambda}_i\right) + \varepsilon_{i,t}. \tag{18}$$

The null hypothesis of unit root can be tested using the pseudo t-ratio $t_{\tilde{e}_i}^j\left(\lambda_i\right)$, $j = c, \tau, \gamma$, for testing $\alpha_{i,0} = 0$ in (18), where c denotes the models that do not include a time trend and the structural change can affect either the level and/or the cointegrating vector, τ stands for those models that include a linear time trend with a structural break that can affect either the level

[3] In the more standard panel cointegration framework without common factors, as discussed above, the distributions of the test statistics have been computed without making any assumption about strict exogeneity. The distributions depend on the number of regressors in the model and this is reflected in Table 1 for the asymptotic moments and the response surfaces in Table 2, both computed for varying m.

and/or the cointegrating vector, and γ denotes those specifications that include change in the trend – note that in this case, λ_i is assumed to be homogeneous across the units.

When $r = 1$ we can use an ADF-type equation to analyze the order of integration of F_t as well. However, in this case we need to proceed in two steps. In the first step we regress $\tilde{F}_t$ on the deterministic specification and the stochastic regressors. In the second step we estimate the ADF regression equation using the detrended common factor $\left(\tilde{F}_t^d\right)$, i.e. the residuals of the first step:

$$\Delta \tilde{F}_t^d = \delta_0 \tilde{F}_{t-1}^d + \sum_{j=1}^{k} \delta_j \Delta \tilde{F}_{t-j}^d + u_t,$$

and test if $\delta_0 = 0$.

Finally, if $r > 1$ we should use one of the two statistics proposed in Bai and Ng (2004) to fix the number of common stochastic trends (q). As before, let $\tilde{F}_t^d$ denote the detrended common factors. Start with $q = r$ and proceed in three stages – we reproduce these steps here for completeness:

1. Let $\tilde{\beta}_\perp$ be the q eigenvectors associated with the q largest eigenvalues of $T^{-2}\sum_{t=2}^{T} \tilde{F}_t^d \tilde{F}_t^{d\prime}$.
2. Let $\tilde{Y}_t^d = \tilde{\beta}_\perp \tilde{F}_t^d$, from which we can define two statistics:
   (a) Let $K(j) = 1 - j/(J+1)$, $j = 0, 1, 2, \ldots, J$:
      i. Let $\tilde{\xi}_t^d$ be the residuals from estimating a first-order VAR in $\tilde{Y}_t^d$, and let
      $$\tilde{\Sigma}_1^d = \sum_{j=1}^{J} K(j) \left(T^{-1} \sum_{t=2}^{T} \tilde{\xi}_t^d \tilde{\xi}_t^{d\prime} \right).$$
      ii. Let $\tilde{v}_c^d(q) = \frac{1}{2}\left[\sum_{t=2}^{T}\left(\tilde{Y}_t^d \tilde{Y}_{t-1}^{d\prime} + \tilde{Y}_{t-1}^d \tilde{Y}_t^{d\prime}\right) - T\left(\tilde{\Sigma}_1^d + \tilde{\Sigma}_1^{d\prime}\right)\right]\left(\sum_{t=2}^{T} \tilde{Y}_{t-1}^d \tilde{Y}_{t-1}^{d\prime}\right)^{-1}$.
      iii. Define $MQ_c^d(q) = T\left[\tilde{v}_c^d(q) - 1\right]$ for the case of no change in the trend and $MQ_c^d(q, \lambda) = T\left[\tilde{v}_c^d(q, \lambda) - 1\right]$ for the case of a change in the trend.
   (b) For p fixed that does not depend on N and T:
      i. Estimate a VAR of order p in $\Delta \tilde{Y}_t^d$ to obtain $\tilde{\Pi}(L) = I_q - \tilde{\Pi}_1 L - \ldots - \tilde{\Pi}_p L^p$. Filter $\tilde{Y}_t^d$ by $\tilde{\Pi}(L)$ to get $\tilde{y}_t^d = \tilde{\Pi}(L)\tilde{Y}_t^d$.
      ii. Let $\tilde{v}_f^d(q)$ be the smallest eigenvalue of
      $$\Phi_f^d = \frac{1}{2}\left[\sum_{t=2}^{T}\left(\tilde{Y}_t^d \tilde{Y}_{t-1}^{d\prime} + \tilde{Y}_{t-1}^d \tilde{Y}_t^{d\prime}\right)\right]\left(\sum_{t=2}^{T} \tilde{Y}_{t-1}^d \tilde{Y}_{t-1}^{d\prime}\right)^{-1}.$$
      iii. Define the statistic $MQ_f^d(q) = T\left[\tilde{v}_f^d(q) - 1\right]$ for the case of no change in the trend and $MQ_f^d(q, \lambda) = T\left[\tilde{v}_f^d(q, \lambda) - 1\right]$ for the case of a change in the trend.
3. If $H_0 : r_1 = q$ is rejected, set $q = q - 1$ and return to the first step. Otherwise, $\tilde{r}_1 = q$ and stop.

The following Theorem offers the main results concerning these statistics.

Theorem 2 *Let $\{y_{i,t}\}$ the stochastic process with DGP given by (11) to (15). The following results hold as $(N,T)_{\text{seq}} \to \infty$. Let k be the order of autoregression chosen such that $k \to \infty$ and $k^3/\min[N,T] \to 0$.*

(1) Under the null hypothesis that $\rho_i = 1$ in (14),

(1.a) for the specification that does not include a time trend, with or without change in level:

$$t_{\tilde{e}_i}^{c}(\lambda_i) \Rightarrow \frac{\frac{1}{2}\left(W_i(1)^2 - 1\right)}{\left(\int_0^1 W_i(s)^2 ds\right)^{1/2}},$$

(1.b) for those specifications including a time trend with or without change in level:

$$t_{\tilde{e}_i}^{\tau}(\lambda_i) \Rightarrow -\frac{1}{2}\left(\int_0^1 V_i(s)^2 ds\right)^{-1/2},$$

where $V_i(s) = W_i(s) - sW_i(1)$.

(1.c) for those specifications including a time trend with change in trend:

$$t_{\tilde{e}_i}^{\gamma}(\lambda) \Rightarrow -\frac{1}{2}\left(\lambda^2 \int_0^1 V_i(b_1)^2 dr + (1-\lambda)^2 \int_0^1 V_i(b_2)^2 dr\right)^{-1/2},$$

where $V_i(b_j) = W_i(b_j) - b_j W_i(1)$, $j = 1, 2$, are two independent detrended Brownian processes.

(2) When $r = 1$, under the null hypothesis that F_t has a unit root and no change in trend:

$$t_{\tilde{F}}^{d} \Rightarrow \frac{\int_0^1 W_w^d(s)\, dW_w^d(s)}{\left(\int_0^1 W_w^d(s)^2 ds\right)^{1/2}},$$

where $W_w^d(s)$ denotes the detrended Brownian motion, while when we allow for change in trend:

$$t_{\tilde{F}}^{d}(\lambda) \Rightarrow \frac{\int_0^1 W_w^d(s,\lambda)\, dW_w^d(s,\lambda)}{\left(\int_0^1 W_w^d(s,\lambda)^2 dr\right)^{1/2}},$$

where $W_w^d(s,\lambda)$ is the detrended Brownian motion and λ denotes the break fraction parameter.

(3) When $r > 1$, let W_q be a q-vector of standard Brownian motion and W_q^d the detrended counterpart. Let $v_*^d(q)$ be the smallest eigenvalues of the statistic computed for a model that does not include change in trend. Then:

$$\Phi_*^d = \frac{1}{2}\left[W_q^d(1)\, W_q^d(1)' - I_p\right]\left[\int_0^1 W_q^d(s)\, W_q^d(s)'\, ds\right]^{-1},$$

and letting $v_*^d(q,\lambda)$ be the smallest eigenvalues of the statistic computed for the model that includes change in trend:

$$\Phi_*^d(\lambda) = \frac{1}{2}\left[W_q^d(1,\lambda)\, W_q^d(1,\lambda)' - I_p\right]\left[\int_0^1 W_q^d(s,\lambda)\, W_q^d(s,\lambda)'\, ds\right]^{-1},$$

(3.1) Let J be the truncation lag of the Bartlett kernel, chosen such that $J \to \infty$ and $J/\min\left[\sqrt{N}, \sqrt{T}\right] \to 0$. Then, under the null hypothesis that F_t has q stochastic trends, $MQ_c^d(q) \xrightarrow{d} v_*^d(q)$ and $MQ_c^d(q,\lambda) \xrightarrow{d} v_*^d(q,\lambda)$.
(3.2) Under the null hypothesis that F_t has q stochastic trends with a finite VAR($\bar{p}$) representation and a VAR(p) is estimated with $p \geq \bar{p}$, $MQ_f^d(q) \xrightarrow{d} v_*^d(q)$ and $MQ_f^d(q,\lambda) \xrightarrow{d} v_*^d(q,\lambda)$.

The proof is outlined in the Appendix. Some remarks are in order in order to explain the theorem.

Remark 1: The limiting distributions of the statistics do not depend on the stochastic regressors because these are assumed to be orthogonal to the factors and strictly exogenous to the idiosyncratic errors.[4] This implies in particular that Theorem 2 covers the case where a break in the cointegrating vector is also allowed in the model.

Remark 2: Breaks do not affect the limiting distributions of either the ADF and Φ_*^d statistics, as long as they do not involve a change in the trend. Note that this also implies (a) that multiple changes in the constant and/or the slopes of the cointegrating vector are allowed in the specification of the model and (b) the imposed break dates in each unit can be heterogeneous. Note that this feature implies that in the particular case of no structural breaks, valid panel cointegration tests can be computed that have the same limiting distributions presented in Theorem 2 for the cases where the structural break does not affect the trend.

Remark 3: When the models involve a change in trend, the distributions of the statistics depend on the number and location of the breaks. In particular, within the common factor framework, for the tests to be correctly sized, this implies that the matrix of projections M_i that is used above cannot depend on i, which means that all elements that are defined in Δx_i^d should be the same across i. Thus, to warrant that M_i does not (asymptotically) depend on i we have to assume common break dates, i.e. we assume that the break dates are the same for all units.

This restriction can be seen as a limitation of our analysis, but in fact it is due to the definition of the common factors framework. Thus, (16) specifies a common factor structure for all units, so that f_t cannot depend on i. If we look at the definition of $f = M_i \Delta F$ we can see that the specification of heterogeneous structural breaks implies that the idempotent matrix M_i depends on i. The only way to overcome this situation is to impose $M_i = M \ \forall i$ so that the structural breaks are the same for all units. This is the reason why in Theorem 2 we have not included any subscript on λ for the units.

This is not a problem when changes in trend are not present because the regressors that are specific to each individual do not play a role in the asymptotic distributions. This is shown in the relevant part of the proof of the theorem given in the Appendix.

Remark 4: The limiting distributions for $t_{\tilde{e}_i}^c(\lambda_i)$ and $t_{\tilde{F}}^d$ derived in *(1.a)* and *(2)* are the standard Dickey-Fuller distributions for constant and constant and trend respectively. The

[4] As noted above, in the case where strict exogeneity does not hold, we recommend the use of DOLS estimation. Bai and Carrion-i-Silvestre (2005) follow the approach in Bai (2005) and consider the case where the regressors are not orthogonal to the factors.

moments for $t^{c}_{\tilde{e}_i}(\lambda_i)$, $t^{\tau}_{\tilde{e}_i}(\lambda_i)$ and $t^{\gamma}_{\tilde{e}_i}(\lambda)$ for different sample sizes are reported in Table 3. Note that the asymptotic moments for $t^{c}_{\tilde{e}_i}(\lambda_i)$ and $t^{\tau}_{\tilde{e}_i}(\lambda_i)$ do not depend on the location of the break while those for $t^{\gamma}_{\tilde{e}_i}(\lambda)$ are a function of the (common) break date.

Remark 5: The ADF statistic when there is one structural break given by $t^{d}_{\tilde{F}}(\lambda)$ derived in *(2)* can be found in Perron (1989) for the specification denoted as Model C. The limiting distributions of the MQ test in *(3)* with no change in trend may be found in Bai and Ng (2004), while the corresponding distributions for a single known break date in trend, $\Phi^{d}_{*}(\lambda)$, have been simulated by us and are reported in Table 4. The asymptotic critical values reported in this table depend both on the number of stochastic common trends and on the break fraction. Note however that these critical values correspond to the case of only one structural break, though our approach can be easily extended to multiple changes in trend.

The individual ADF statistics for the idiosyncratic disturbance terms can be pooled to define a panel data cointegration tests using the moments for $t^{c}_{\tilde{e}_i}(\lambda_i)$, $t^{\tau}_{\tilde{e}_i}(\lambda_i)$ and $t^{\gamma}_{\tilde{e}_i}(\lambda)$. We can thus define, for the case of no change in trend, the statistics

$$N^{-1/2}Z^{\tilde{e}j}_{\hat{t}_{NT}}(\lambda) - \Theta^{\tilde{e}j}_{2}\sqrt{N} \Rightarrow N\left(0, \Psi^{\tilde{e}j}_{2}\right) \qquad (19)$$

$(T, N \to \infty)_{\text{seq}}$, where $Z^{\tilde{e}j}_{\hat{t}_{NT}}(\lambda) = \sum_{i=1}^{N} t^{j}_{\tilde{e}_i}(\lambda_i)$, $j = c, \tau$, and $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_i, \ldots, \lambda_N)'$. The moments $\Theta^{\tilde{e}j}_{2}$ and $\Psi^{\tilde{e}j}_{2}$ are the same as the ones for the statistics in Bai and Ng (2004).[5] For the case of change in trend, the statistic is computed as

$$N^{-1/2}Z^{\tilde{e}\gamma}_{\hat{t}_{NT}}(\lambda) - \Theta^{\tilde{e}\gamma}_{2}(\lambda)\sqrt{N} \Rightarrow N(0, \Psi^{\tilde{e}\gamma}_{2}(\lambda)),$$

$(T, N \to \infty)_{\text{seq}}$, where $Z^{\tilde{e}\gamma}_{\hat{t}_{NT}}(\lambda) = \sum_{i=1}^{N} t^{\gamma}_{\tilde{e}_i}(\lambda)$ and the break date is common to all units, i.e. $\lambda_1 = \lambda_2 = ... = \lambda_N$.

4.2 Break date(s) unknown

So far the developments in this section have been based on the implicit assumption of known break date. When the break date is unknown we can proceed to estimate it using the infimum functional as described above. As above, we need to distinguish between the cases with or without a change in trend.

4.2.1 No change in trend

Within this case there are two sub-cases to consider: first, where the break dates (across the units) are heterogeneous, and, second, where the break date is homogeneous.

Heterogeneous break dates The break dates for each unit can be estimated by minimizing the sum of squares over all possible break dates using (11) in first-differences. Using the estimated break dates, the factors are then estimated as in section 4.1. The standardized test

[5] Note that Bai and Ng (2004) prefer to combine individual p-values instead of using these moments.

statistic is then constructed as in (19) using the estimated break dates. Since the moments do not depend on the location of the break, the same corrections can be used for the cases where the (heterogeneous) breaks are known or unknown.

Homogeneous break dates Here we compute the $Z^{e}_{\hat{t}_{NT}}(\lambda)$ statistic for each break date, where the break date is the same for each unit, using the idiosyncratic disturbance terms. The break date is estimated as the argument that minimizes the sequence of standardized $Z^{e}_{\hat{t}_{NT}}(\lambda)$ statistics. Thus, the test statistic that is used to test the null hypothesis of non-cointegration for the idiosyncratic disturbance term is given by

$$Z^{\tilde{e}^j}_{\hat{t}_{NT}}\left(\hat{\lambda}\right) = \inf_{\lambda\in\Lambda}\left(\frac{N^{-1/2}Z^{\tilde{e}^j}_{\hat{t}_{NT}}(\lambda) - \Theta^{\tilde{e}^j}_2\sqrt{N}}{\sqrt{\Psi^{\tilde{e}^j}_2}}\right),\ j = c,\tau. \tag{20}$$

The estimated break date denoted $\hat{T}_b$ is given by

$$\hat{T}_b = \arg\min_{\lambda\in\Lambda}\left(\frac{N^{-1/2}Z^{\tilde{e}^j}_{\hat{t}_{NT}}(\lambda) - \Theta^{\tilde{e}^j}_2\sqrt{N}}{\sqrt{\Psi^{\tilde{e}^j}_2}}\right).$$

The limiting distribution of $Z^{\tilde{e}^j}_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$ is given in the following Theorem.

Theorem 3 *Let $\{y_{i,t}\}$ the stochastic process with DGP given by (11) to (15). Then, as $(T,N\to\infty)_{\text{seq}}$ the $Z^{\tilde{e}^j}_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$ test in (20) converges to*

$$Z^{\tilde{e}^j}_{\hat{t}_{NT}}\left(\hat{\lambda}\right) \Rightarrow N(0,1),\ j = c,\tau.$$

The proof follows from noting that under the null $Z^{\tilde{e}^j}_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$, $j = c,\tau$, is the infimum of a sequence of perfectly correlated random variables which are asymptotically standard normal. The perfect correlation arises from the fact that the distributions of the statistics under the null do not depend on λ. It is easy to establish – see Embrechts, Klüppelberg and Mikosch (1997), page 210 – that the infimum is also standard normal. The first two blocks of Panel A of Table 5 provide the critical values for (20) obtained by simulation for different values of T and for $N = 100$, and confirm the validity of the limiting result.

4.2.2 Change in trend

For reasons discussed earlier (see Remark 3 above), when there is a change in trend, we only consider the case of homogeneous break dates. The statistic is given by

$$Z^{\tilde{e}^\gamma}_{\hat{t}_{NT}}\left(\hat{\lambda}\right) = \inf_{\lambda\in\Lambda}\left(\frac{N^{-1/2}Z^{\tilde{e}^\gamma}_{\hat{t}_{NT}}(\lambda) - \Theta^{\tilde{e}^\gamma}_2(\lambda)\sqrt{N}}{\sqrt{\Psi^{\tilde{e}^\gamma}_2(\lambda)}}\right) \tag{21}$$

where $Z_{\hat{t}_{NT}}^{\tilde{e}\gamma}\left(\hat{\lambda}\right)$ is now the infimum of a correlated sequence of statistics, each of which is standard normally distributed. Note that the correlation of the sequence of statistics comes from the fact that the statistics in the sequence depend on the same time series information.

Theorem 4 *Let $\{y_{i,t}\}$ the stochastic process with DGP given by (11) to (15). Then, as $(T, N \to \infty)_{\text{seq}}$ the $Z_{\hat{t}_{NT}}^{\tilde{e}\gamma}\left(\hat{\lambda}\right)$ test in (21) converges to*

$$Z_{\hat{t}_{NT}}^{e}\left(\hat{\lambda}\right) \Rightarrow \inf_{\lambda\in\Lambda} \kappa\left(\lambda\right),$$

where $\kappa\left(\lambda\right)$ denotes a standard Normal distribution for a given λ.

The proof is outlined in the Appendix. The critical values for (21) are obtained by simulation for different values of T and for $N = 100$ – see third block of panel A of Table 5.[6] Here the deviation from normality is clear and does not disappear even in large samples.

We need to consider finally the case of testing for unit roots in the common factors when the break is not known. As shown above, this matters only when there is a change in trend. Our procedure would then involve estimating the break date by using the statistic given in (21). This break date is then used to compute the ADF and the MQ tests for the common factors. Critical values are reported in panel B of Table 5.

5 Monte Carlo simulation

In this section we analyze by conducting simulation experiments the finite sample performance of the statistics that have been proposed in the paper. We begin by considering a DGP where the units are not cross-section dependent, so that our results in section 3 can be used. We then consider a DGP with cross-section dependence which uses our results in section 4.

5.1 Cross-section independent

The empirical size of the tests is studied regressing two independent random walks, which have been generated as the cumulated sum of *iid* $N(0,1)$ processes. The sample size has been set equal to $T = \{50, 100, 250\}$ and the number of units at $N = \{20, 40\}$. The results reported in Table 6 are obtained from $1,000$ replications, assuming that the break date is unknown and using the estimated response surfaces of the previous section. As can be seen, the empirical size of both the normalized bias and the pseudo t-ratio statistics is close to the nominal size irrespective of T and N.

The empirical power of the statistics is assessed using the DGP given by:

$$\begin{aligned} y_{i,t} &= \mu_i + \theta_i DU_{i,t} + \beta_i t + \gamma_i DT^*_{i,t} + x'_{i,t}\delta_{i,t} + z_{i,t} \\ z_{i,t} &= \rho_i z_{i,t-1} + v_{i,t}, \end{aligned}$$

[6] Following the previous literature, we introduce trimming at the end points of the sample so that $\lambda \in \Lambda$, with $\Lambda = [0.15, 0.85]$.

where $v_{i,t} \sim iid\ N(0,1)\ \forall i\,,\ i=1,\ldots,N$. The specification of the values of the parameters depends on the model under consideration. In general, the constant and, when required, the slope of the trend are set equal to $\mu_i = 1$ and $\beta_i = 0.3$, respectively. When there is a change in the level the magnitude is set equal to $\theta_i = 3$, while for the change in trend we consider $\gamma_i = 0.5$. The change in the cointegrating vector is given by $\delta_{i,t} = \delta_{i,1} = 1$ for $t \leq T_{bi}$ and $\delta_{i,t} = \delta_{i,1} = 3$ for $t > T_{bi}$, for a break date randomly located at $\lambda_i \sim U(0.15, 0.85)$, $\forall i\,,\ i=1,\ldots,N$, where U denotes the uniform distribution – the same results are obtained when break fraction is fixed either at $\lambda_i = 0.25$, $\lambda_i = 0.5$ or $\lambda_i = 0.75\ \forall i$. Simulations were performed for two autoregressive coefficients $\rho_i = \{0.5, 0.8\}$, although we only report results for $\rho_i = 0.8$ to save space. The computation of the statistics controls the autocorrelation in the disturbance term including up to $k_{\max} = 5$ lags using the t-sig criterion to select the order of the autoregressive correction. Results in Tables 7 and 8 show the empirical power of both statistics, respectively, for different combinations of DGP's and estimated models when $\rho_i = 0.8$. Thus, we can assess the empirical power of the statistics when DGP does not coincide with the model that is estimated. When the DGP and estimated model coincide both statistics show good power, which increases with T and N – see bold-typed columns in Tables 7 and 8. However, $Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$ outperforms $Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$ since for the former statistic the power equals one in all cases. In general, when the estimated model is misspecified and misspecification involves the cointegrating vector, the empirical power of $Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$ decreases. For instance, when DGP is given by Model 1 and we estimate Model 4, the power of the $Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$ statistic is reduced – note that the converse is also true. The same is found when either the DGP is given by Model 2 and we estimate Model 5, or when the DGP is given by Model 3 and we estimate Model 6. However, this feature is not found for the $Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$ statistic, which does not lose any power when this sort of misspecification occurs. Finally, misspecification due to lack of accounting for time trend – i.e. DGP given by Models 2, 3, 5 and 6, and estimation of specifications given by Models 1 and 4 – reduces the power of both statistics as T increases, although for the $Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$ statistic the specifications that allow for a change in the cointegrating vector always show higher power.

In all, simulations lead us to conclude that $Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$ statistic outperforms $Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$ in all situations that have been considered. Furthermore, overparameterisation of the estimated model does not cause loss of power for the $Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$ statistic. These features indicate that $Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$ should be preferred to $Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$ in empirical applications.

5.2 Cross-section dependent

In order to deal with the situation with common factors, to mimic the impact of cross-sectional dependence, consider the DGP given by a bivariate system:

$$
\begin{aligned}
y_{i,t} &= f_i(t) + x'_{i,t}\delta_{i,t} + u_{i,t} \\
u_{i,t} &= F_t\pi_i + e_{i,t} \\
F_t &= \phi F_{t-1} + \sigma_F w_t \\
e_{i,t} &= \rho_i e_{i,t-1} + \varepsilon_{i,t} \\
\Delta x_{i,t} &= v_{i,t},
\end{aligned}
$$

where $(w_t, \varepsilon_{i,t}, v_{i,t})'$ follow a mutually *iid* standard multivariate Normal distribution for $\forall i, j$ $i \neq j$ and $\forall t, s$ $t \neq s$. In this paper we consider two different situations depending on the number of common factors, i.e. $r = \{1, 3\}$, and specify three values for the autoregressive parameters $\phi = \{0.8, 0.9, 1\}$ and $\rho_i = \{0.95, 0.99, 1\}$ $\forall i$. Note that these values allows to analyze both the empirical size and power of the statistics. The importance of the common factors is controlled through the specification of $\sigma_F^2 = \{0.5, 1, 10\}$. The number of common factors is estimated using the panel BIC information criterion in Bai and Ng (2002) with $r_{\max} = 6$ as the maximum number of factors. We consider $N = 40$ units and $T = \{50, 100, 250\}$ time observations.

The simulation results for size and power for the case with no breaks (with one or more factors) are close to those reported by Bai and Ng (2004) and are therefore not included in the paper but are available in the companion appendix in Tables A.7 to A.9. From these results it may be seen that the empirical size of the ADF pooled idiosyncratic t-ratio statistic $\left(Z^{e}_{\hat{t}_{NT}}\right)$ and the ADF statistic of the common factor – when there is only one factor in the DGP – is close to the nominal size, which is set at the 5% level of significance. As expected the power of the tests increases as the autoregressive parameter moves away from unity. These results do not change when specifying three common factors. Furthermore, simulations available upon request in the companion appendix indicate that these conclusions are obtained irrespective of the deterministic specification.

Let us turn now to the results for the case where there is one unknown structural break under the alternative hypothesis.[7] In this case we only report simulations for Models 3 and 6 with $\lambda_i = 0.5$ $\forall i$, where the common break point is estimated as described in Section 4.2.2 – results for the other specifications that have been considered in the paper are available in the companion appendix.[8]

Column 5 in Tables 9 and 10 reports results for the empirical size for the $Z^{\tilde{e}^{\gamma}}_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$ test when there is only one common factor. Looking at these results we can conclude that, in general, mild underrejection is found, so the test tends to be conservative. It is worth mentioning that the $Z^{\tilde{e}^{\gamma}}_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$ test overrejects for Model 6 with $\sigma_F^2 = 10$, although the empirical size approaches the nominal one as T increases. The empirical power is close to the nominal size for $\rho_i = 0.99$

[7] Results for the case where the break point is known are available from the authors. The tests on the idiosyncratic disturbance terms show good properties in terms of empirical size and power. The ADF statistic on the common trend, when there is one single common factor, shows the right size although, as expected, it has low power when the autoregressive parameter is close to unity and the sample size is small. Our results for three factors are comparable to the one-factor case.

[8] The results are presented in Tables A.10 to A.13 for one common factor and Tables A.14 to A.25 for three common factors, with persistence parameters for the factors given by 1, 0.99 and 0.95 for each of the model specifications.

The results may be summarized as follows. For Model 1, the tests for both the idiosyncratic component and for the common factors show the correct size and have good power even when the value of the stationary roots under the alternative hypothesis is 0.99. When a change in trend is also allowed, i.e. Model 2, the tests show a certain loss in power for values of the stationary root equal to 0.99, but performance improves considerably when the root falls to 0.95.

For Model 4, with one common factor, the test on the idiosyncratic component tends to be over-sized if $\sigma_F^2 = 10$, while the behaviour of the test of the common trend depends on the stochastic properties of the idiosyncratic component. Similar conclusions hold in the case of three common factors, with the MQ statistics showing a tendency to over-estimate the number of common trends.

Finally, for Model 5, the behaviour of the tests for the idiosyncratic and common trends are similar to their behaviour under Model 4. There is undersizing when the stationary root is equal to 0.99, with reasonable power being obtained when T is relatively large and the root is equal to 0.95.

– see Column 8 in Tables 9 and 10 – which is to be expected provided that the alternative is close to the null hypothesis. When the autoregressive coefficient moves away form the null hypothesis the $Z_{\hat{t}_{NT}}^{\tilde{e}^{\gamma}}\left(\hat{\lambda}\right)$ test attains reasonable power values – see Column 11 in Tables 9 and 10. Concerning the $t_{\tilde{F}}^{d}$ statistic on the estimated common factor, we can see that we require T to be large for the statistic to show good properties in terms of empirical size and power. This is not surprising, since this statistic relies on only one source of information, that is, the one coming from the time dimension whereas the $Z_{\hat{t}_{NT}}^{\tilde{e}^{\gamma}}\left(\hat{\lambda}\right)$ test pools across the N dimension also.

Let us now focus on the results for the three common factors. Tables 11 to 13, and Tables 14 to 16 present the results of the empirical size for the $Z_{\hat{t}_{NT}}^{\tilde{e}^{\gamma}}\left(\hat{\lambda}\right)$ test for Models 3 and 6, respectively. On the one hand, mild underrejection is found in general for the $Z_{\hat{t}_{NT}}^{\tilde{e}^{\gamma}}\left(\hat{\lambda}\right)$ of both models. On the other hand, overrejection for Model 6 when $\sigma_F^2 = 10$ is observed, although the nominal size is attained as T increases. Note that this feature has been also found for the one common factor case. Regarding the $MQ_c^d(q)$ test, we can see that the statistic selects the correct number of stochastic trends when T is large and σ_F^2 is small, but it tends to detect more non-stationary common stochastic trends for T small and $\sigma_F^2 = 10$ – in Tables 11 to 13, and Tables 14 to 16, MQ(3) denotes the frequency that the $MQ_c^d(q)$ statistic has detected three common stochastic trends, MQ(2) for the frequency corresponding to two stochastic trends, MQ(1) for one stochastic trend and, finally, MQ(0) denotes the times that the statistic has not detected any stochastic trend. When computing the $MQ_c^d(q)$ test, the bandwidth for the Bartlett spectral window is set as $J = 4ceil\left[\min\left[N,T\right]/100\right]^{1/4}$. The conclusions about the empirical power of the $Z_{\hat{t}_{NT}}^{\tilde{e}^{\gamma}}\left(\hat{\lambda}\right)$ test are similar to those found for the one common factor case. The $Z_{\hat{t}_{NT}}^{\tilde{e}^{\gamma}}\left(\hat{\lambda}\right)$ test shows low power when the alternative hypothesis is close to the null hypothesis ($\rho_i = 0.99$), although the test shows good power for $\rho_i = 0.95$.

As before, the $MQ_c^d(q)$ test overestimates the number of common stochastic trends unless T is large for Model 3. Regarding the performance of the test statistic for Model 6, we can see that the statistic selects the correct number of stochastic trends when T is large, σ_F^2 is small and $\rho_i = 1$, but it tends to detect more non-stationary common stochastic trends for T small, $\sigma_F^2 = 10$ and $\rho_i = 1$. In general, we can observe that the $MQ_c^d(q)$ test detects more non-stationary common stochastic trends than exist when the idiosyncratic component is stationary in variance ($\rho_i < 1$) regardless of the value of α.

To sum up, the simulations that have been conducted in this section reveal that the $Z_{\hat{t}_{NT}}^{\tilde{e}^{\gamma}}\left(\hat{\lambda}\right)$ test statistic has good properties in terms of empirical size and power, but the $MQ_c^d(q)$ test tends to overestimate the number of common stochastic trends. Thus, if we consider the $Z_{\hat{t}_{NT}}^{\tilde{e}^{\gamma}}\left(\hat{\lambda}\right)$ test to be a test for cointegration in panels – having controlled for the presence of either integrated or stationary common trends – the good size and power properties of this test are encouraging for its use in empirical applications such as the one reported in the section below.

6 Empirical illustration - exchange rate pass through in the euro area

Campa and González-Mínguez (2006) (henceforth CM), Campa, Goldberg and González-Mínguez (2006) (henceforth CGM), and Frankel et al. (2005) *inter alia* have investigated the issue of exchange rate pass through (ERPT) of foreign to domestic prices. Studies of ERPT have been conducted both for the United States and for countries of the euro area to study the importance institutional arrangements (such as the inauguration of the euro area) in generating responses to exchange rate institutions and changes.

An important feature missing from the discussion is a connection between the theoretical arguments surrounding the key determinants of pass through, and the inappropriate techniques used to estimate equations measuring import or export exchange rate pass. For example, while almost all the theories contain a long run or steady-state relationship in the levels of a measure of import unit values (in domestic currency), the exchange rate (relating the domestic to the numeraire currency) and a measure of foreign prices (unit values in the numeraire currency, typically US dollars), this long run is routinely disregarded in most of the empirical implementations.

There is a substantial consensus in the literature that the time series being studied are integrated variables. Therefore, one way of defining the long run is in the sense of Engle and Granger (1987), henceforth EG, where the long run is given by the so-called cointegrating relationship. A reason often given for ignoring this long run, and substituting it by an ad hoc measure, is a failure to find evidence in the data for cointegration. Two problems arise from such an ad hoc reinterpretation – first the contradiction between a theoretical prediction of a steady state that cannot be found in the data, and, second, the ad hoc measure proposed being no more than an extended version of the estimate of the short-run (where the pass through equation is estimated in differences) and strongly dominated by this estimated short-run.

An alternative explanation is that the estimation method used – typically single-equation autoregressive distributed lag (ARDL) models – may not be powerful enough to verify the theory for the span of data available. Therefore, instead of looking for a new definition of the long run, a more satisfactory approach is to look for the long run relationship using more appropriate and powerful methods, such as those which allow for changes in the long run or use more powerful panel data methods. In doing so, it is also important to allow for structural breaks in the long run theoretical relationship to take due account of pass-through rates in response to changes in financial regime such as the introduction of the euro in January 1999.

6.1 Exchange Rate Pass Through into Import Prices

By definition,[9] import prices for any type of goods j, MP_t^j are a transformation of export prices of a country's trading partners XP_t^j using the bilateral exchange rate ER_t and dropping superscript j for clarity:

$$MP_t = ER_t \cdot XP_t. \tag{22}$$

[9] This section is based on Campa, Goldberg and González-Mínguez (2005).

In logarithms (depicted in lower case):

$$mp_t = er_t + xp_t, \tag{23}$$

where the export price consists of the exporters marginal cost and a mark-up:

$$XP_t = FMC_t \cdot FMKUP_t. \tag{24}$$

So that in logarithms we have:

$$xp_t = fmc_t + fmkup_t. \tag{25}$$

Substituting for xp_t^j into equation (23) yields:

$$mp_t = er_t + fmkup_t + fmc_t. \tag{26}$$

The literature on industrial organization offers explanations for why the effect of the change in er_t on mp_t may differ from one, using determinants of the mark-up such as competitive conditions among exporters in the destination markets. Mark-up responsiveness depends on the market share of domestic producers relative to foreign producers, the form of competition that takes place in the market for the industry, and the extent of price discrimination. Other factors affecting pass-through are the currency denomination of exports and structure and importance of intermediate goods markets.

For example, the empirical setup of CGM is based on (26) assumes unity translation of exchange rate movements. If the pass-through is complete (producer currency pricing), and their mark-ups do not fluctuate in response to fluctuations of the exchange rates, this leads to a pure currency translation. At the other extreme, they can decide not to vary the prices in the destination country currency (local currency pricing or pricing to market) and absorb the fluctuations within the mark-up. Thus, mark-ups in an industry are assumed to consist of a component specific to the type of good, independent of the exchange rate and a reaction to exchange rate movements:

$$fmkup_t = a + \Phi er_t. \tag{27}$$

It is also important to consider the effects working through the marginal cost. These are a function of demand conditions in the importing country, marginal costs of production (labor wages) in the exporting country, and the commodity prices denominated in foreign currency:

$$fmc_t = c_0 \cdot y_t + c_1 \cdot fw_t + c_2 \cdot er_t + c_3 \cdot fcp_t. \tag{28}$$

Substituting (28) and (27) into (26), we have:

$$mp_t = a + \underbrace{(1 + \Phi + c_2)}_{b} er_t + c_0 \cdot y_t + c_1 \cdot fw_t + c_3 \cdot fcp_t + \varepsilon_t,$$

where the coefficient b on the exchange rate er_t is the pass-through elasticity. In the CGM 'integrated world market' specification, $c_0 \cdot y_t + c_1 \cdot fw_t + c_3 \cdot fcp_t$, independent of the exchange rate, is called the opportunity cost of allocating those same goods to other customers and is reflected in the world price of the product fp_t in the world currency (here taken to be the US dollar).[10] Thus the final equation can be re-written as follows:

$$mp_t = \widehat{a} + \widehat{b} \cdot er_t + \widehat{c} \cdot fp_t + \varepsilon_t, \tag{29}$$

which gives the long run relation between the import price, exchange rate and a measure of foreign price.

6.2 ERPT - estimation

Both economic theory and relevant tests lead us to think of each of the series (import price, exchange rate and world price) as being characterized by a unit root. However, despite the underlying levels equation (29), CM are not able to reject the null hypothesis of the non-existence of a cointegrating relationship among the three series for countries of the euro area. Hence, they proceed by estimating equation (29) in first differences:

$$\Delta mp_t = a + \sum_{k=0}^{4} b_k \cdot \Delta er_{t-k} + \sum_{k=0}^{4} c_k \cdot \Delta fp_{t-k} + \varepsilon_t, \tag{30}$$

for industrial sector i in country j – the superscripts have been omitted for clarity. Next, they define the coefficient b_0 and the sum of coefficients $\sum_{k=0}^{4} b_k$ as the short-run and long run ERPT respectively.

Since CM do not find evidence of the long run in the Engle-Granger (1987) sense, they propose their own working definition of the long run. The CM definition of the long run pass through, which is constructed by summing the estimated coefficients for the first five lags (i.e. lag 0 to lag 4), is somewhat arbitrary, and thus rather inadequate for the purpose of enquiring about the actual long run effect of the exchange rate on import prices.

An alternative route is to use the panel cointegration technology developed in this paper, where for each (i, j) pair. Given that we have 10 countries and 9 industrial sectors, a panel-based test could use up to approximately $9 \times 10 \times 110$ observations)[11] and allowing for heterogeneity, we should in principle obtain a far clearer idea of the common trends underlying the series and hence of the long run. In the spirit of the discussion above, any such estimation procedure in panels would of course need to allow for structural change. In addition it would also need to allow for dependence among the units of the panel. We look at these issues in turn after a brief consideration of the data.

[10] The integrated market hypothesis in CM is based on the assumption that there exists a single world market for each good. Therefore, regardless of the origin of the product, on the world market, it has one world price. This price constitutes the opportunity cost of selling to a local market. Thus, in the CM setup for the integrated market and, consequently, in ours, it proxies for the foreign price. The currency denomination does not in fact matter, as long as the exchange rate for the local currency is taken vis-a-vis this 'world' currency. In the CM case the extra-euro area imports denominated in US dollars are taken as a proxy for the world price.

[11] See the data appendix for full details. The number of observations in the panel is dependent on our need to use a balanced panel. In order to obtain the longest time dimension three countries, Austria, Finland and Portugal, need to be deleted from the whole sample since we do not have observations before 1996 for these countries. In order to maximise the cross-section dimension, so that no country is dropped, our sample needs to start in 1996:1 and end in 2004:12. The estimation results reported below are for this choice of the sample since fewer observations are lost under this configuration.

6.3 Data

A sample of 1995-2005 for the euro area from Eurostat is used. The construction of the variables follows CM, and is described in the data appendix. The indicator we use for import prices, the index of import unit values (IUV), has a series of caveats associated with their use but we are constrained in our investigations by the quality of the publicly available data.

It is also important to emphasize that there are a number of reasons why we expect there may be a change in the long run ERPT within our sample. Firstly, on the 1^{st} of January 1999 eleven European countries fixed their exchange rates by adopting the euro.[12] This constituted a change in monetary policy, especially for countries that previously had less credible policy regimes. This is more notorious in countries with previously rather less successful monetary policy, the perceived stabilization of monetary policy may well have induced the producers to change their pricing strategies and would thus be expected to have an influence on the long run ERPT. Moreover the adoption of a common currency has changed the competitive conditions, by increasing the share of goods denominated in the (new) domestic currency. Finally, virtually all the currencies were depreciating against the US dollar in the period 1995-2000, and especially since 1996. Moreover, after a short period of a stable euro dollar exchange rate, the euro currency(ies) started appreciating, till the end of our sample. This asymmetry of exchange rate developments may have different implications for ERPT.

6.4 Panel cointegration tests

There are essentially three ways of proceeding in order to construct panels from the data sets - (i) creating country panels of industry cross-sections, (ii) industry panels with country cross-sections, and (iii) a pooled panel in which every country and industry combination constitutes a separate unit. In search of the existence of a cointegrating relationship in the series we try to maximize the dimensions of our panel, and thus will focus on (iii). Results for (i) and (ii) are available from us upon request.

The results from both the modified Pedroni test and the test developed in this paper are given in Table 17 below. As noted earlier, results are presented for the longest available panel which includes all countries. The panel headed ‘Cross-section independent’ reports the results from the panel cointegration tests that assume cross-section independence and accommodates for one structural breaks. The results for all six model specifications – for all of which heterogeneous break dates are permitted – are reported in this panel. As can be seen, all statistics when compared to the standard normal distribution lead to strong rejection of the null hypothesis of non-cointegration. However, the assumption of independence is not likely to be valid in this case, since the countries involved in the panel have been tied to the same process of monetary convergence. Furthermore, the strong rejection that is found might be due to the unattended cross-section dependence – see Banerjee, Marcellino and Osbat (2004).

The panel headed ‘Cross-section dependent’ provides the results for the panel statistics where cross-sectional dependence is allowed for through the consideration of common factors. The test results are reported in Table 17 for the idiosyncratic components along with the number

[12] Greece failed to fulfil the Maastricht Treaty criteria, and therefore joined 2 years later, effective 1^{st} of January 2001.

of non-stationary in variance common factors ($\hat{r}_1$) that is obtained using the non-parametric $MQ_c^d(q)$ and $MQ_c^d(q,\lambda)$ statistics – the maximum number of factors allowed is $r_{\max} = 6$ and we have used the panel BIC in Bai and Ng (2002) to estimate the number of common factors ($\hat{r}$). For Models 3 and 6, we are restricted to imposing a homogeneous unknown break date. The remaining models allow for both heterogeneous and homogeneous breaks. The break dates detected by the cross-sectional dependent test, when homogeneous breaks are imposed, are also given.

It is clear than in each case the tests overwhelmingly reject the null of no cointegration. The reported breaks all occur in the neighborhood of the introduction of the euro in 1999 or the beginning of its strong appreciation in 2002 relative to the dollar. A refinement of our tests to allow for multiple breaks would perhaps allow us to detect both these 'regime' changes although it may be argued that the time dimension of the panel in our case will not permit such detailed discrimination. Finally, if one were to think of model 4 as the most plausible choice of model, there is very clear evidence, under every configuration, for a long run relationship with a structural break in 2000.

6.5 Concluding remarks to empirical illustration

The results show ample evidence for an EG cointegrating relationship between the variables in levels - as in the underlying theoretical equation (29). We can therefore redefine the long run effect of exchange rate fluctuations on prices to be consistent with the theoretical literature. Instead of a rather arbitrary sum of (mainly insignificant, of opposite signs) coefficients on lags of the exchange rate[13], which we discussed in Section 6.2, we propose using the EG cointegrating equation coefficient. The use of the standard measure of pass through should be viewed with caution, or re-interpreted substantially.

There are a number of further issues to consider in detail, including the magnitude of the pass through coefficient and the direction of the change (if any) in response to new monetary arrangements, These need to be investigated within the context of countries and individual sectors. ERPT in the long run is found to be equal to one or close to one in the commodity sectors, throughout the entire sample, while it tends to be rather lower than one in the manufacturing, food, beverages and tobacco and chemical sectors.

Allowing for a structural break in the relationship we find that ERPT has generally increased in the vicinity of the euro introduction and the change is especially evident in Southern European countries. This may be the effect of perceived stabilization in the monetary regime, which led to less noise in exchange rate developments. Moreover the increase in ERPT in the second part of our sample may be due to specific exchange rate developments (euro/dollar depreciation till 2000, and subsequent appreciation) which may suggest asymmetrical response of the import prices. Further detail on these results are available upon request.

[13] This would, as we have argued, make the estimates problematic to interpret in any case - regardless of whether the notion of a long run, defined as the sum of four or five short run effects, is coherent.

7 Conclusions

This paper has shown that inference based on parametric Pedroni panel cointegration test statistics can be affected by the presence of structural breaks. Monte Carlo evidence indicates that in some situations the power of the tests drops as the magnitude of the structural break increases. Specifically, when the structural break affects either the slope of the time trend or the cointegrating vector the power approaches zero as T, N and the magnitude of the break increases. In contrast, the power of the standard parametric Pedroni panel cointegration statistics is affected to a much lesser extent when the structural break only changes the level – we require a large magnitude of structural breaks located at the end of the time period to reduce the power of the statistics.

These features have motivated our proposal, and have led us to design statistical procedures to account for the presence of structural breaks when testing for cointegration. Six different specifications have been introduced depending on the effect of structural breaks on the long run relationship. Finite sample and asymptotic moments have been computed that allow defining panel cointegration statistics for the specifications considered.

The issue of cross-section dependence is addressed in the paper by assuming an approximate common factor structure. We derive the limiting distributions of statistics in two situations of interest, i.e. (i) for the case of no structural break, and (ii) when there are changes in level and trend. The performance of the approach is investigated through Monte Carlo simulations, from which we conclude that the statistics show good performance once the procedures have accounted for structural breaks. An empirical example based on looking at exchange rate pass through in the euro area helps to illustrate the value of our theoretical proposals.

References

[1] Bai, J. (2005): "Panel Data Models with Interactive Effects". Mimeo, Department of Economics, New York University.

[2] Bai, J. and J. L. Carrion-i-Silvestre (2005): "Panel cointegration and common factors", *Mimeo*, Departament d'Econometria, Estadística i Economia Espanyola. Universitat de Barcelona.

[3] Bai, J. and Ng, S. (2002): "Determining the Number of Factors in Approximate Factor Models", *Econometrica*, 70, 191-221.

[4] Bai, J. and S. Ng (2004): "A PANIC Attack on Unit Roots and Cointegration", *Econometrica*, 72, 1127-1177.

[5] Baltagi, B. H. (2005): *Econometric Analysis of Panel Data*, 3rd edition, John Wiley & Sons, New York.

[6] Banerjee, A. (1999): "Panel Data Unit Roots and Cointegration: An Overview", *Oxford Bulletin of Economics and Statistics*, Special issue, 607–629.

[7] Banerjee, A., Marcellino, M. and C. Osbat (2004): "Some Cautions on the Use of Panel Methods for Integrated Series of Macro-Economic Data", *Econometrics Journal*, 7, 322-340.

[8] Banerjee, A., Marcellino, M. and C. Osbat (2005): "Testing for PPP: Should We Use Panel Methods?". *Empirical Economics*, 30, 77-91.

[9] Bartley, W. A., J. Lee and M. C. Strazicich (2001): "Testing the Null of Cointegration in the Presence of a Structural Break", *Economics Letters*, 73, 315–323.

[10] Breitung, J. (2005): "A Parametric Approach to the Estimation of Cointegration Vectors in Panel Data", *Econometric Reviews*, 24, 151-173.

[11] Breitung, J. and M. H. Pesaran, 2005, Unit roots and cointegration in panels". Forthcoming in Matyas, L., Sevestre, P., Eds., The Econometrics of Panel Data, Klüver Academic Press.

[12] Busetti, F. (2002): "Testing for (Common) Stochastic Trends in the Presence of Structural Breaks", *Journal of Forecasting*, 21, 81–105.

[13] Campa, J. M. and J. M. González-Mínguez (2006): "Difference in Exchange Rate Pass-through in the Euro Area", *European Economic Review*, 50, 121-145.

[14] Campa, J. M., L. S. Goldberg and J. M. González-Mínguez (2005): "Exchange-rate Pass-through to Import Prices in the Euro Area", NBER working paper 11632.

[15] Campos, J., N. Ericsson and D. F. Hendry (1996): "Cointegration Tests in the Presence of Structural Breaks", *Journal of Econometrics*, 70, 187-220.

[16] Carrion-i-Silvestre, J. Ll. and A. Sansó (2006): "Testing the Null of Cointegration with Structural Breaks", *Oxford Bulletin of Economics and Statistics*, 68, 623-646.

[17] Chang, Y. (2005): "Residual Based Tests for Cointegration in Dependent Panels", *Mimeo.* Department of Economics, Rice University.

[18] Embrechts, P., C. Klüppelberg and T. Mikosch (1997): *Modelling Extremal Events.* Springer-Verlag. Berlin.

[19] Engle, R. F. and C. W. J. Granger (1987): "Cointegration and Error Correction: Representation, Estimation, and Testing", *Econometrica*, 55, 251-276.

[20] Frankel, J., D. Parsely, and S. Jin-Wei (2005): "Slow Passthrough Around the World: A New Import for Developing Countries", CID Working Paper No. 116.

[21] Gregory, A. W. and B. E. Hansen (1996): "Residual-based Tests for Cointegration in Models with Regime Shifts", *Journal of Econometrics*, 70, 99-126.

[22] Hamilton, J. D. (1994): *Time Series Analysis.* Princeton University Press.

[23] Hansen, B. E. (1992): "Tests for Parameter Instability in Regressions with I(1) Processes", *Journal of Business and Economic Statistics*, 10, 321-335.

[24] Hansen, H. and S. Johansen (1999): "Some Tests for Parameter Constancy in Cointegrated-VAR Models", *Econometrics Journal*, 2, 306-333.

[25] Hao, K. (1996): "Testing for Structural Change in Cointegrated Regression Models: some Comparisons and Generalisations", *Econometric Reviews*, 15, 401–429.

[26] Harris, D. and B. Inder (1994): "A Test of the Null Hypothesis of Cointegration". In Colin P. Hargreaves (Ed.) *Nonstationary Time Series Analysis and Cointegration.* Oxford University Press, Oxford.

[27] Jansen, W. J. (2000): "International Capital Mobility", *Journal of International Money and Finance*, 19, 507-511.

[28] Kao, C. and M. H. Chiang (2000): "Testing for Structural Change of a Cointegrated Regression" *Mimeo.* Department of Economics, Syracuse University.

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

[30] Ng, S. and P. Perron (1995): "Unit Root Test in ARMA models with Data-Dependent Methods for the Selection of the Truncation Lag", *Journal of the American Statistical Association*, 90, 268-281.

[31] Pedroni, P. (1999): "Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors", *Oxford Bulletin of Economics and Statistics*, Special Issue, 653-670.

[32] Pedroni, P. (2004): "Panel Cointegration. Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis", *Econometric Theory*, 20, 3, 597-625.

[33] Perron, P. (1989): "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis", *Econometrica*, 57, 1361-1401.

[34] Perron, P. (1997): "Further Evidence on Breaking Trend Functions in Macroeconomic Variables", *Journal of Econometrics*, 80, 355-385.

[35] Phillips, P. C. B. and S. N. Durlauf (1986): "Multiple Time Series Regression with Integrated Processes", *Review of Economic Studies*, 53, 473-495.

[36] Quintos, C. and P. C. B. Phillips (1993): "Parameter Constancy in Cointegrating Regressions", In *New Developments in Time Series Econometrics*, J. M. Dufour and B. Raj (Eds.). Studies in Empirical Economics. Physica-Verlag. Springer Verlag.

[37] Shin, Y. (1994): "A Residual-based Test of the Null of Cointegration against the Alternative of no Cointegration", *Econometric Theory*, 10, 91–115.

[38] Stock, J. H. and M. W. Watson (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems", *Econometrica*, 64, 783-820.

[39] Westerlund, J. (2006): "Testing for Panel Cointegration with Multiple Structural Breaks", *Oxford Bulletin of Economics and Statistics*, 68, 101-132.

[40] Zivot, E., and D. W. K. Andrews (1992): "Further Evidence on the Great Crash, the Oil Price Shock, and the Unit-Root Hypothesis", *Journal of Business & Economic Statistics 10*, 251–270.

A Mathematical Appendix

For the sake of simplicity let us first assume that the stochastic regressors are strictly exogenous. Once the main result is derived, we show how these derivations can be extended to account for non-strictly exogenous regressors.

A.1 Proof of statement *(1.a)* of Theorem 2

Let us assume the model given by (11) and (12). Alternatively, the model can be expressed as:

$$y_{i,t} = x_{i,t}'\delta_i + F_t\pi_i + e_{i,t}.$$

As can be seen, the model assumes that residuals from the static regression follows a factor structure as defined in Bai and Ng (2004). Note that if we introduce (16) in (17) we obtain

$$\begin{aligned} \tilde{z}_{i,t} &= z_{i,t} + f_t\pi_i - \tilde{f}_t\tilde{\pi}_i \\ &= z_{i,t} - v_t H^{-1}\pi_i - \tilde{f}_t d_i, \end{aligned} \tag{31}$$

where $v_t = \tilde{f}_t - f_t H$ and $d_i = \tilde{\pi}_i - H^{-1\prime}\pi_i$, where H is an $(r \times r)$ matrix defined as follows $H = V_{NT}^{-1}\left(\hat{f}'f/T\right)(\Pi'\Pi/N)$ with V_{NT} the $(r \times r)$ diagonal matrix of the first r largest eigenvalues of $(NT)^{-1}\, y^* y^{*\prime}$ in decreasing order. The computation of the partial sum processes of (31) gives:

$$T^{-1/2}\sum_{j=2}^{t}\tilde{z}_{i,j} = T^{-1/2}\sum_{j=2}^{t} z_{i,j} - T^{-1/2}\sum_{j=2}^{t} v_j H^{-1}\pi_i - T^{-1/2}\sum_{j=2}^{t}\tilde{f}_j d_i. \tag{32}$$

Let us analyse each element of (32) separately. The left-hand side of (32) is equal to

$$\begin{aligned} T^{-1/2}\sum_{j=2}^{t}\tilde{z}_{i,j} &= T^{-1/2}\sum_{j=2}^{t}\left[M_i\Delta\tilde{e}_i\right]_j \\ &= T^{-1/2}\sum_{j=2}^{t}\Delta\tilde{e}_{i,j} - T^{-1/2}\sum_{j=2}^{t}\left[P_i\Delta\tilde{e}_i\right]_j, \end{aligned} \tag{33}$$

where $\left[P_i\Delta\tilde{e}_i\right]_j$ denotes the j-th element of the vector $P_i\Delta\tilde{e}_i$, and $P_i = I_{T-1} - M_i$. The first element on the right-hand side of (33) is equal to

$$T^{-1/2}\sum_{j=2}^{t}\Delta\tilde{e}_{i,j} = T^{-1/2}\tilde{e}_{i,t} - T^{-1/2}\tilde{e}_{i,1} = T^{-1/2}\tilde{e}_{i,t} + O_p(1),$$

so that by the invariance principle

$$T^{-1/2}\sum_{j=2}^{[sT]}\Delta\tilde{e}_{i,j} \Rightarrow \sigma_i W_i(s),$$

with $t = [sT]$. The second element on the right-hand side of (33) is

$$T^{-1/2} \sum_{j=2}^{t} [P_i \Delta \tilde{e}_i]_j = T^{-1/2} (x_{i,t} - x_{i,1})' \left(\Delta x_i' \Delta x_i \right)^{-1} \Delta x_i' \Delta \tilde{e}_i.$$

Note that $(\Delta x_i' \Delta x_i)^{-1} \Delta x_i' \Delta \tilde{e}_i = \left(T^{-1} \Delta x_i' \Delta x_i \right)^{-1} \left(T^{-1} \Delta x_i' \Delta \tilde{e}_i \right) = o_p(1)$, since $\left(T^{-1} \Delta x_i' \Delta x_i \right) \to^p Q_{\Delta x_i \Delta x_i}$, the variance-covariance matrix of $\Delta x_i' \Delta x_i$, and $T^{-1} \Delta x_i' \Delta \tilde{e}_i \xrightarrow{p} 0$ since these elements are orthogonal by definition. On the other hand, $T^{-1/2} x_{i,t} \Rightarrow \Omega_{22,i}^{1/2} W_{i,k}(s)$ and $T^{-1/2} x_{i,1} \xrightarrow{p} 0$ by assumption. These derivations lead us to

$$T^{-1/2} \sum_{j=2}^{t} \tilde{z}_{i,j} = T^{-1/2} \tilde{e}_{i,t} + o_p(1),$$

since $T^{-1/2} x_{i,t} (\Delta x_i' \Delta x_i)^{-1} \Delta x_i' \Delta \tilde{e}_i = o_p(1)$. The same result can be achieved for $T^{-1/2} \sum_{j=2}^{t} z_{i,j}$, i.e.

$$T^{-1/2} \sum_{j=2}^{t} z_{i,j} = T^{-1/2} e_{i,t} + o_p(1).$$

This indicates that the presence of stochastic regressors does not have any effect on the partial sum processes. Regarding the term involving $\{v_t\}$ we see from Eq. (A.3) in Bai and Ng (2004) that

$$T^{-1/2} \sum_{j=2}^{t} v_j = O_p\left(C_{NT}^{-1} \right),$$

where $C_{NT} = \min \left\{ N^{-1/2}, T^{-1/2} \right\}$. Moreover and as shown in Bai and Ng (2004), the term $d_i = O_p\left(C_{NT}^{-1} \right)$ and $T^{-1/2} \sum_{j=2}^{t} \tilde{f}_j = O_p(1)$, so that

$$T^{-1/2} \sum_{j=2}^{t} \tilde{z}_{i,j} = T^{-1/2} \sum_{j=2}^{t} z_{i,j} + O_p\left(C_{NT}^{-1} \right).$$

The DF statistic computed on the idiosyncratic disturbance term computed from the usual expression

$$t_{\tilde{e}_i}^{c}(\lambda_i) = \frac{\sum_{t=2}^{T} \tilde{e}_{i,t-1} \Delta \tilde{e}_{i,t}}{\sqrt{\tilde{\sigma}_i^2 \sum_{t=2}^{T} \tilde{e}_{i,t-1}^2}},$$

From all these results it follows that

$$t_{\tilde{e}_i}^{c}(\lambda_i) \Rightarrow \frac{\frac{1}{2}\left(W_i(1)^2 - 1 \right)}{\left(\int_0^1 W_i(s)^2 \, ds \right)^{1/2}},$$

that is, the limiting distribution is the same derived in Bai and Ng (2004) for the constant case – see Bai and Ng (2004) for the proof. The same result is found for the ADF test. This implies that the presence of stochastic regressors does not affect the limiting distribution of the statistic.

A.2 Proof of statement *(1.b)* of Theorem 2

The generalization that includes a time trend can be carried out as well. In this case the model (11) is replaced by

$$y_{i,t} = \mu_i + \beta_i t + x'_{i,t}\beta_i + u_{i,t}.$$

Note that as before we are not dealing with the structural break case since we are defining the benchmark limiting distributions. Contrary to the previous specification, taking first differences does not remove the deterministic elements, since now the trend becomes a constant. This is a relevant feature since the limiting distribution of the ADF-type statistic varies. However, the asymptotic distribution of the statistic is the same as the one derived in Bai and Ng (2004) for the trend case. The proof follows similar steps above. Now the first difference of regressors defines the following idempotent matrix

$$M_i = I_{T-1} - \Delta x_i^d \left(\Delta x_i^{d\prime} \Delta x_i^d\right)^{-1} \Delta x_i^{d\prime},$$

where the Δx_i^d matrix is defined by the row vectors $\left(1, \Delta x'_{i,t}\right)'$. Note that as before the first element of (33) converges to

$$T^{-1/2} \sum_{j=2}^{[sT]} \Delta \tilde{e}_{i,j} \Rightarrow \sigma_i W_i (s).$$

The limiting expression of the second element in (33) has to be derived in several steps. First, note that $T^{-1}\Delta x_i^{d\prime}\Delta x_i^d$ converges to variance-covariance matrix of Δx_i^d, so that all these elements are $O_p(1)$. The first element of the vector $T^{-1}\Delta x_i^{d\prime}\Delta\tilde{e}_i$ is given by $T^{-1/2}\left(T^{-1/2}\sum_{t=1}^{T}\Delta\tilde{e}_{i,t}\right) = T^{-1/2}\left(T^{-1/2}\left(\tilde{e}_{i,T} - \tilde{e}_{i,1}\right)\right)$, where $T^{-1/2}\left(\tilde{e}_{i,T} - \tilde{e}_{i,1}\right) \Rightarrow \sigma_i W_i(1)$ since $T^{-1/2}\tilde{e}_{i,1} \to^p 0$. Note that the extra rescaling term $T^{-1/2}$ would be used below. The rest of the elements in $T^{-1}\Delta x_i^{d\prime}\Delta\tilde{e}_i$ involve cross-products among the first difference of the stochastic regressors and $\Delta\tilde{e}_i$ that converges to zero since we have assumed independence. Therefore,

$$\left(\Delta x_i^{d\prime}\Delta x_i^d\right)^{-1}\Delta x_i^{d\prime}\Delta\tilde{e}_i = \left[\begin{array}{c} E\, T^{-1/2}\left(T^{-1/2}\left(\tilde{e}_{i,T} - \tilde{e}_{i,1}\right)\right) + o_p(1) \\ \left(-D^{-1}CE\right)\, T^{-1/2}\left(T^{-1/2}\left(\tilde{e}_{i,T} - \tilde{e}_{i,1}\right)\right) + o_p(1) \end{array}\right]$$

where $E = \left(A - BD^{-1}C\right)^{-1}$ and $A = 1, B = T^{-1}\iota'\Delta x_i, C = B'$ and $D = T^{-1}\Delta x_i'\Delta x_i$ denote the elements of the partitioned matrix $T^{-1}\Delta x_i^{d\prime}\Delta x_i^d$, with $\iota = (1, \ldots, 1)'$. The partial sum process of $\Delta x_{i,t}^d$ is

$$T^{-1/2}\sum_{j=2}^{t}\Delta x_{i,j}^d = \left[T^{-1/2}t \quad T^{-1/2}\left(x_{i,t} - x_{i,1}\right)'\right],$$

so that

$$T^{-1/2}\sum_{j=2}^{t}\Delta x_{i,j}^d\left(\Delta x_i^{d\prime}\Delta x_i^d\right)^{-1}\Delta x_i^{d\prime}\Delta\tilde{e}_i = \frac{t}{T}E\,\left(T^{-1/2}\left(\tilde{e}_{i,T} - \tilde{e}_{i,1}\right)\right) + o_p(1),$$

since $T^{-1}(x_{i,t}-x_{i,1})' = o_p(1)$. Moreover, the matrix E can be expressed as

$$
\begin{aligned}
\left(A - BD^{-1}C\right)^{-1} &= A^{-1} + A^{-1}B\left(D - CA^{-1}B\right)^{-1}CA^{-1} \\
&= 1 + B\left(D - B'B\right)^{-1}B'.
\end{aligned}
$$

Note that $B = T^{-1}\iota'\Delta x_i \to^p 0$ so that $\left(A - BD^{-1}C\right)^{-1} = 1 + o_p(1)$. Therefore,

$$
\begin{aligned}
T^{-1/2}\sum_{j=2}^{t}\Delta x_{i,j}^{d}\left(\Delta x_i^{d\prime}\Delta x_i^{d}\right)^{-1}\Delta x_i^{d\prime}\Delta\tilde{e}_i &= \frac{t}{T}\left(T^{-1/2}\left(\tilde{e}_{i,T}-\tilde{e}_{i,1}\right)\right) + o_p(1) \\
&\Rightarrow r\ \sigma_i W_i(1).
\end{aligned}
$$

From Bai and Ng (2004), the terms $T^{-1/2}\left\|\sum_{j=2}^{t} v_j\right\| = O_p\left(C_{NT}^{-1}\right)$, $\|d_i\| = O_p\left(C_{NT}^{-1}\right)$ and $T^{-1/2}\left\|\sum_{j=2}^{t}\tilde{f}_j\right\| = O_p(1)$. These derivations lead us to

$$
\begin{aligned}
T^{-1/2}\sum_{j=2}^{[sT]}\tilde{z}_{i,j} &= T^{-1/2}\tilde{e}_{i,t} - \frac{s}{T}T^{-1/2}\tilde{e}_{i,T} + O_p\left(C_{NT}^{-1}\right) \\
&\Rightarrow \sigma_i\left(W_i(s) - s\ W_i(1)\right) \equiv \sigma_i V_i(s).
\end{aligned}
$$

The DF statistic is

$$
t_{\tilde{e}}^{\tau}(i) = \frac{T^{-1}\sum_{t=2}^{T}\tilde{e}_{i,t-1}\Delta\tilde{e}_{i,t}}{\left(\tilde{\sigma}_i^2 T^{-2}\sum_{t=2}^{T}\tilde{e}_{i,t-1}^2\right)^{1/2}}.
$$

Note that the following identity holds

$$
T^{-1}\sum_{t=2}^{T}\tilde{e}_{i,t-1}\Delta\tilde{e}_{i,t} = \frac{\tilde{e}_{i,T}^2}{2T} - \frac{\tilde{e}_{i,1}^2}{2T} - \frac{1}{2T}\sum_{t=2}^{T}\left(\Delta\tilde{e}_{i,t}\right)^2,
$$

which shows that $T^{-1}\tilde{e}_{i,T}^2 \Rightarrow \sigma_i^2 V_i(1)^2 = 0$, $T^{-1}\tilde{e}_{i,1}^2 = 0$ and $T^{-1}\sum_{t=2}^{T}\left(\Delta\tilde{e}_{i,t}\right)^2 \to^p \sigma_i^2$, from which it follows that $T^{-1}\sum_{t=2}^{T}\tilde{e}_{i,t-1}\Delta\tilde{e}_{i,t} \to^p -\sigma_i^2/2$ and $T^{-2}\sum_{t=2}^{[sT]}\tilde{e}_{i,t-1}^2 \Rightarrow \sigma_i^2\int_0^1 V_i(s)^2\,ds$ – see Bai and Ng (2004), Lemma G.4. Using these elements it is straightforward to see that

$$
t_{\tilde{e}_i}^{\tau}(\lambda_i) \Rightarrow -\frac{1}{2}\left(\int_0^1 V_i(s)^2\,ds\right)^{-1/2},
$$

where $V_i(s) = W_i(s) - s\ W_i(1)$, *i.e.* the limiting distribution is the same derived in Bai and Ng (2004) for the trend case. Although the proof is more involved, the same result is achieved for the ADF test. As before, this implies that the presence of stochastic regressors does not affect the limiting distribution of the statistic. Note that this result is also achieved when there are level shifts in the model, since the impulse dummies do not affect the limiting distribution of the $t_{\tilde{e}_i}^{\tau}(\lambda_i)$ statistic.

A.3 Proof of statement *(1.c)* of Theorem 2

The model is given by the following deterministic specification

$$f_i(t) = \mu_i + \beta_i t + \theta_i DU_{i,t} + \gamma_i DT^*_{i,t},$$

which implies that $\Delta f_i(t) = \beta_i + \theta_i D\left(T^i_b\right)_t + \gamma_i DU_{i,t}$ and $\Delta x^d_{i,t} = \left(1, D\left(T^i_b\right)_t, DU_{i,t}, \Delta x'_{i,t}\right)$. In order to simplify the steps of the proof, we deal with the equivalent specification that does not include the impulse dummy, *i.e.* $\Delta x^d_{i,t} = \left(1, DU_{i,t}, \Delta x'_{i,t}\right)$. This simplifies derivations, although it does not imply loss of generality. Moreover, note that the subspace spanned by $\left(1, DU_{i,t}, \Delta x'_{i,t}\right)$ is equivalent to the one spanned by $\left(DU^1_{i,t}, DU^2_{i,t}, \Delta x'_{i,t}\right)$ where $DU^1_{i,t} = 1$ for $t \leq T_b$ and 0 otherwise, and $DU^2_{i,t} = 1$ for $t > T_b$ and 0 otherwise. This redefinition makes $DU^1_{i,t}$ and $DU^2_{i,t}$ to be orthogonal. Note that as before the first element of (33) converges to

$$T^{-1/2}\sum_{j=2}^{[sT]} \Delta\tilde{e}_{i,j} \Rightarrow \sigma_i W_i(s).$$

The limiting expression of the second element in (33) has to be derived in several steps. First, note that $T^{-1}\Delta x^{d\prime}_i \Delta x^d_i$ converges to variance and covariance matrix of Δx^d_i, so that all these elements are $O_p(1)$. The first element of the vector $T^{-1}\Delta x^{d\prime}_i \Delta\tilde{e}_i$ is given by $T^{-1/2}\left(T^{-1/2}\sum_{t=1}^{T_b}\Delta\tilde{e}_{i,t}\right) = T^{-1/2}\left(T^{-1/2}\left(\tilde{e}_{i,T_b} - \tilde{e}_{i,1}\right)\right)$, where $T^{-1/2}\left(\tilde{e}_{i,T_b} - \tilde{e}_{i,1}\right) \Rightarrow \sigma_i W_i(\lambda)$ since $T^{-1/2}\tilde{e}_{i,1} \to^p 0$. The second element is $T^{-1/2}\left(T^{-1/2}\sum_{t=T_b+1}^{T}\Delta\tilde{e}_{i,t}\right) = T^{-1/2}\left(T^{-1/2}\left(\tilde{e}_{i,T} - \tilde{e}_{i,T_b}\right)\right)$, where $T^{-1/2}\left(\tilde{e}_{i,T} - \tilde{e}_{i,T_b}\right) \Rightarrow \sigma_i W_i(1) - \sigma_i W_i(\lambda)$. Note that as before the extra rescaling term $T^{-1/2}$ would be used below. Finally, the third set of elements in the product is $T^{-1}\Delta x'_i \Delta\tilde{e}_i$ that converges to zero since we have assumed independence. Therefore,

$$\left(\Delta x^{d\prime}_i \Delta x^d_i\right)^{-1} \Delta x^{d\prime}_i \Delta\tilde{e}_i = \begin{bmatrix} E\; T^{-1/2}\left(T^{-1/2}\left(\tilde{e}_{i,T_b} - \tilde{e}_{i,1}\right), T^{-1/2}\left(\tilde{e}_{i,T} - \tilde{e}_{i,T_b}\right)\right)' + o_p(1) \\ \left(-D^{-1}CE\right)\; T^{-1/2}\left(T^{-1/2}\left(\tilde{e}_{i,T_b} - \tilde{e}_{i,1}\right), T^{-1/2}\left(\tilde{e}_{i,T} - \tilde{e}_{i,T_b}\right)\right)' + o_p(1) \end{bmatrix}$$

where $E = \left(A - BD^{-1}C\right)^{-1}$ and $A = diag(\lambda, 1-\lambda)$, $B = T^{-1}\left[DU^1_i, DU^2_i\right]'\Delta x_i$, $C = B'$ and $D = T^{-1}\Delta x'_i \Delta x_i$ denote the elements of the partitioned matrix $T^{-1}\Delta x^{d\prime}_i \Delta x^d_i$. Moreover, following the steps given above $\left(A - BD^{-1}C\right)^{-1} = A^{-1} + o_p(1)$, since $B \to^p 0$. The partial sum process of $\Delta x^d_{i,t}$ for $t \leq T_b$ is

$$T^{-1/2}\sum_{j=2}^{t}\Delta x^d_{i,j} = \left[T^{-1/2}t \quad 0 \quad T^{-1/2}\left(x_{i,t} - x_{i,1}\right)'\right],$$

while for $t > T_b$ is

$$T^{-1/2}\sum_{j=2}^{[sT]}\Delta x^d_{i,j} = \left[T^{-1/2}T_b \quad T^{-1/2}(s - T_b) \quad T^{-1/2}\left(x_{i,t} - x_{i,1}\right)'\right],$$

so that for $t \leq T_b$

$$
\begin{aligned}
T^{-1/2}\sum_{j=2}^{[sT]}\Delta x_{i,j}^{d}\left(\Delta x_i^{d\prime}\Delta x_i^{d}\right)^{-1}\Delta x_i^{d\prime}\Delta\tilde{e}_i &= \frac{s}{T}\frac{1}{\lambda}\left(T^{-1/2}\left(\tilde{e}_{i,T_b}-\tilde{e}_{i,1}\right)\right)+o_p\left(1\right) \\
&\Rightarrow \frac{s}{\lambda}\sigma_i W_i\left(\lambda\right),
\end{aligned}
$$

since $T^{-1}\left(x_{i,t}-x_{i,1}\right)' = o_p\left(1\right)$. Therefore, for $t \leq T_b$

$$
\begin{aligned}
T^{-1/2}\sum_{j=2}^{[sT]}\tilde{z}_{i,j} &= T^{-1/2}\tilde{e}_{i,t}-\frac{s}{T}T^{-1/2}\tilde{e}_{i,T}+O_p\left(C_{NT}^{-1}\right) \\
&\Rightarrow \sigma_i\left(W_i\left(s\right)-\frac{s}{\lambda}W_i\left(\lambda\right)\right),
\end{aligned}
$$

since from Bai and Ng (2004), the terms $T^{-1/2}\left\|\sum_{j=2}^{t}v_j\right\| = O_p\left(C_{NT}^{-1}\right)$, $\|d_i\| = O_p\left(C_{NT}^{-1}\right)$ and $T^{-1/2}\left\|\sum_{j=2}^{t}\tilde{f}_j\right\| = O_p\left(1\right)$. Note that we can define $b_1 = s/\lambda$ so that $0 < b_1 < 1$, which in turn implies that

$$
\begin{aligned}
T^{-1/2}\sum_{j=2}^{[sT]}\tilde{z}_{i,j} &\Rightarrow \sigma_i\sqrt{\lambda}W_i\left(b_1\right)-\sigma_i b_1\sqrt{\lambda}W_i\left(1\right) \\
&= \sigma_i\sqrt{\lambda}\left(W_i\left(b_1\right)-b_1W_i\left(1\right)\right)\equiv\sigma_i\sqrt{\lambda}V_i\left(b_1\right),
\end{aligned}
$$

given the properties of Brownian motions. On the other hand, for $t > T_b$

$$
\begin{aligned}
T^{-1/2}\sum_{j=2}^{[sT]}\Delta x_{i,j}^{d}\left(\Delta x_i^{d\prime}\Delta x_i^{d}\right)^{-1}\Delta x_i^{d\prime}\Delta\tilde{e}_i &= \frac{T_b}{T}\frac{1}{\lambda}\left(T^{-1/2}\left(\tilde{e}_{i,T_b}-\tilde{e}_{i,1}\right)\right) \\
&\quad +\frac{s-T_b}{T}\frac{1}{1-\lambda}\left(T^{-1/2}\left(\tilde{e}_{i,T}-\tilde{e}_{i,T_b}\right)\right)+o_p\left(1\right) \\
&\Rightarrow \sigma_i\left(W_i\left(\lambda\right)+\frac{s-\lambda}{1-\lambda}\left(W_i\left(1\right)-W_i\left(\lambda\right)\right)\right),
\end{aligned}
$$

so that

$$
\begin{aligned}
T^{-1/2}\sum_{j=2}^{[sT]}\tilde{z}_{i,j} &= T^{-1/2}\tilde{e}_{i,t}-\frac{s}{T}T^{-1/2}\tilde{e}_{i,T}+O_p\left(C_{NT}^{-1}\right) \\
&\Rightarrow \sigma_i\left(W_i\left(s\right)-W_i\left(\lambda\right)-\frac{s-\lambda}{1-\lambda}\left(W_i\left(1\right)-W_i\left(\lambda\right)\right)\right).
\end{aligned}
$$

As before, we can define $b_2 = \left(s-\lambda\right)/\left(1-\lambda\right)$ so that $0 < b_2 < 1$, which in turn implies that

$$
T^{-1/2}\sum_{j=2}^{t}\tilde{z}_{i,j}\Rightarrow\sigma_i\sqrt{1-\lambda}\left(W_i\left(b_2\right)-b_2W_i\left(1\right)\right)\equiv\sigma_i\sqrt{1-\lambda}V_i\left(b_2\right).
$$

Using similar developments as in the previous proof, the numerator of the DF statistic converges to $T^{-1}\sum_{t=2}^{T}\tilde{e}_{i,t-1}\Delta\tilde{e}_{i,t} \to^p -\sigma_i^2/2$, while the denominator is

$$\begin{aligned} T^{-2}\sum_{t=2}^{T}\tilde{e}_{i,t-1}^2 &= T^{-2}\sum_{t=2}^{T_b+1}\tilde{e}_{i,t-1}^2 + T^{-2}\sum_{t=T_b+2}^{T}\tilde{e}_{i,t-1}^2 \\ &\Rightarrow \sigma_i^2\left(\lambda^2\int_0^1 V_i(b_1)^2\,db_1 + (1-\lambda)^2\int_0^1 V_i(b_2)^2\,db_2\right), \end{aligned}$$

with $V(b_1)$ and $V(b_2)$ two independent Brownian bridges. Therefore, the limiting distribution of the DF statistic is

$$t_{\tilde{e}_i}^{\gamma}(\lambda_i) \Rightarrow -\frac{1}{2}\left(\lambda^2\int_0^1 V_i(b_1)^2\,db_1 + (1-\lambda)^2\int_0^1 V_i(b_2)^2\,db_2\right)^{-1/2}.$$

Note that this limiting distribution is symmetric around $\lambda = 0.5$ since in this case we can interchange λ^2 and $(1-\lambda)^2$ and obtain the same distribution. As before, the same limiting distribution is found for the ADF statistic.

A.4 Proof of statement *(2)* of Theorem 2

Let us now deal with the unit root hypothesis testing when there is $r = 1$ common factor and no change in trend. The model in first differences defines an idempotent matrix M_i that is unit-dependent. At first sight this goes against the definition of a common factor since we assume that this element is common to all units and, hence, cannot depend on i. Nevertheless, it is shown below that the elements that depend on i vanish asymptotically. Thus, note that

$$\begin{aligned} \sum_{j=2}^{t}\tilde{f}_j &= \sum_{j=2}^{t}\left[M_i\Delta\tilde{F}\right]_t \\ &= \tilde{F}_t - (x_{i,t}-x_{i,1})'\left(\Delta x_i'\Delta x_i\right)^{-1}\Delta x_i'\Delta\tilde{F}, \end{aligned} \qquad (34)$$

since we define $\tilde{F}_1 = 0$. Note that the first element of (34) is

$$\tilde{F}_t = H(F_t - F_1) + V_t,$$

since $\Delta\tilde{F}_t = H\ \Delta F_t + v_t$ and $V_t = \sum_{j=2}^{t} v_j$. The detrended estimated factor will remove F_1:

$$\tilde{F}_t^d = H\ F_t^d + V_t^d,$$

and it can be shown that

$$T^{-1/2}\tilde{F}_t^d = H\ T^{-1/2}F_t^d + O_p\left(C_{NT}^{-1}\right),$$

since $T^{-1/2}V_t^d = O_p\left(C_{NT}^{-1}\right)$ – see Bai and Ng (2004), Lemma B.2. The second term in (34) is $T^{-1/2}(x_{i,t}-x_{i,1})'(\Delta x_i'\Delta x_i)^{-1}\Delta x_i'\Delta\tilde{F} = o_p(1)$, since $T^{-1}\Delta x_i'\Delta x_i$ converges to the matrix of covariance of Δx_i and $T^{-1}\Delta x_i'\Delta\tilde{F} = o_p(1)$ by assumption. Since

$$\begin{aligned} T^{-1/2}\tilde{F}_t^d &\Rightarrow H\ W_w^d(s) \\ T^{-2}\sum_{t=2}^{T}\tilde{F}_{t-1}^d\tilde{F}_{t-1}^{d\prime} &\Rightarrow H^2\ \sigma_w^2\int_0^1 W_w^d(s)^2\,ds \\ T^{-1}\sum_{t=2}^{T}\tilde{F}_{t-1}^d\Delta\tilde{F}_t &\Rightarrow H^2\ \sigma_w^2\int_0^1 W_w^d(s)\,dW(s)\,, \end{aligned}$$

the DF statistic converges to

$$\begin{aligned} t_{\tilde{F}}^d &= \frac{T^{-1}\sum_{t=2}^{T}\tilde{F}_{t-1}^d\Delta\tilde{F}_t}{\left(\tilde{\sigma}_u^2 T^{-2}\sum_{t=2}^{T}\left(\tilde{F}_{t-1}^d\right)^2\right)^{1/2}} \\ &\Rightarrow \frac{\int_0^1 W_w^d(s)\,dW(s)}{\left(\int_0^1 W_w^d(s)^2\,ds\right)^{1/2}}, \end{aligned} \tag{35}$$

where $W_w^d(s)$ denotes the detrended Brownian motion and $\tilde{\sigma}_w^2 \xrightarrow{p} H^2\ \sigma_w^2$. The ADF statistic has the same limiting distribution since the order of the autoregressive correction is selected such that $k \to \infty$ and $k^3/\min[N,T] \to 0$.

The limiting distribution of the ADF statistic when there is a change in trend is derived by Perron (1989). The limiting distributions of the test statistics that are used when there is more than one common factor ($r > 1$) but no break – see statement *(3)* of Theorem 2 – are the same as the ones derived in Bai and Ng (2004). These steps may be followed routinely to derive the distributions given in *(3)* for the case where the break is unknown.

A.5 Non strictly-exogenous regressors

Following developments in Bai and Carrion-i-Silvestre (2005) we can show that the same results are obtained when stochastic regressors are non-strictly exogenous. Thus, the model given by (11) and (12) with non-strictly exogenous regressors can be expressed as

$$y_{i,t} = x_{i,t}'\delta_i + \Delta x_{i,t}'A_i(L) + F_t\lambda_i + \xi_{i,t},$$

where $A_i(L)$ denotes the $(k\times 1)$-vector of lead and lag polynomials. Previous derivations concerning idiosyncratic disturbance term still hold but replacing $\Delta\tilde{e}_{i,t}$ with $\Delta\tilde{\xi}_{i,t}$. Now we define $\Delta x_{i,t}^d = \left(\Delta x_{i,t}', \Delta^2 x_{i,t}'\right)'$. Note that $T^{-1/2}(\Delta x_{i,t} - \Delta x_{i,1}) = T^{-1/2}O_p(1) \to 0$, $T^{-1}\Delta x_i^{d\prime}\Delta x_i^d \xrightarrow{p} Q_{\Delta x_i^d\Delta x_i^d}$, the covariance matrix of $\Delta x_i^{d\prime}\Delta x_i^d$, and $T^{-1}\Delta x_i^{d\prime}\Delta\tilde{\xi}_i \to 0$, so that we can see that $T^{-1/2}\sum_{j=2}^{t}\left[P_i\Delta\tilde{\xi}_i\right]_j \to 0$. Then,

$$T^{-1/2}\sum_{j=2}^{t}\tilde{z}_{i,j} = T^{-1/2}\tilde{\xi}_{i,t} + o_p(1)\,,$$

and

$$T^{-1/2}\sum_{j=2}^{t} z_{i,j} = T^{-1/2}\xi_{i,t} + o_p(1),$$

which indicates that the presence of (non-strictly) stochastic regressors does not have any effect on the partial sum processes once endogeneity has been taken into account and, hence, the rest of the proof follows the one above for strictly exogenous regressors.

A.6 Proof of Theorem 4

The proof of this Theorem follows Zivot and Andrews (1992), Gregory and Hansen (1996), and Perron (1997). The panel test statistic can be expressed as a composite functional:

$$\inf_{\lambda\in\Lambda}\left(\frac{N^{-1/2}Z_{\hat{t}_{NT}}^{\tilde{e}\gamma}(\lambda) - \Theta_2^{\tilde{e}\gamma}(\lambda)\sqrt{N}}{\sqrt{\Psi_2^{\tilde{e}\gamma}(\lambda)}}\right) = g\left(T^{-1}\sum_{t=2}^{T}\tilde{e}_{i,t-1}\Delta\tilde{e}_{i,t}, T^{-2}\sum_{t=2}^{T}\tilde{e}_{i,t-1}^2, \tilde{\sigma}_i, \Theta_2^{\tilde{e}\gamma}(\lambda), \Psi_2^{\tilde{e}\gamma}(\lambda)\right),$$

where

$$g = h^{**}\left[h^*\left[h\left[m_1(\lambda), m_2(\lambda), m_3(\lambda)\right], \Theta_2^{\tilde{e}\gamma}(\lambda), \Psi_2^{\tilde{e}\gamma}(\lambda)\right]\right],$$

with $h^{**}(m) = \inf_{\lambda\in\Lambda} m(\lambda)$ for any real function $m = m(\cdot)$ on Λ, $h^*(m) = N^{-1}\sum_{i=1}^{N} m_i(\lambda)$ for any real function $m_i(\lambda)$ on Λ, $i = 1, \ldots, N$. Furthermore, for any real functions $n_{1,i}(\lambda)$, $n_{2,i}(\lambda)$ and $n_{3,i}(\lambda)$ on Λ, $h\left[n_{1,i}(\lambda), n_{2,i}(\lambda), n_{3,i}(\lambda)\right] = n_{1,i}(\lambda)^{-1/2} n_{2,i}(\lambda)/n_{3,i}(\lambda)$, where $n_{1,i}(\lambda) = T^{-2}\sum_{t=2}^{T}\tilde{e}_{i,t-1}^2$, $n_{2,i}(\lambda) = T^{-1}\sum_{t=2}^{T}\tilde{e}_{i,t-1}\Delta\tilde{e}_{i,t}$ and $n_{3,i}(\lambda) = \tilde{\sigma}_i$.

The weak convergence joint result for $n_{1,i}(\lambda)$ and $n_{2,i}(\lambda)$ has been obtained in the proof of statement *(1.c)* of Theorem 2 above, while $n_{3,i}(\lambda) \to^p \sigma_i$. Continuity of h is proved in Zivot and Andrews (1992), so that assuming that the random variable $h\left[n_{1,i}(\lambda), n_{2,i}(\lambda), n_{3,i}(\lambda)\right]$ has finite second moments, h^* is also continuous on Λ since this only implies the mean functional of the standardized continuous h function on Λ. Using the Lindberg-Levy Central Limit Theorem, h^* converges to the standard normal distribution as $N \to \infty$ for any fixed $\lambda \in \Lambda$.

Finally, Zivot and Andrews (1992) establish continuity of $h^{**}(m)$ for all real functions m on Λ. Therefore, the continuity of g follows from the continuity of a composition of continuous functions, so that the Central Mapping Theorem (CMT) can be used to obtain the result of Theorem 4.

B Data Appendix

Sources: Eurostat, COMEXT.

import prices - monthly indexes of import unit values (calculated to be based on local currency) for imports originating outside the euro area.

foreign prices - monthly indexes of import unit values (calculated to be based on US dollars) from imports originating outside the euro area into the euro zone.

exchange rates - index of monthly average exchange rate of local currency against the US dollar.

All variables are in logs.

SITC code - Industry

0 - Food and live animals chiefly for food

1 - Beverages and tobacco

2 - Crude materials, inedible, except fuels

3 - Mineral fuels, lubricants and related materials

4 - Animal and vegetable oils, fats and waxes

5 - Chemicals and related products, n.e.s.

6 - Manufactured goods classified chiefly by materials

7 - Machines, transport equipment

8 - Manufactured goods n.e.c.

CM data set 1989-2001 - series for 1989:1-2001:3 for Belgium+Luxembourg, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal and Spain. Series for 1996:1-2001:3 for Austria and Finland.

- "new" data set 1995-2005 - 1995:1-2005:3 for 10 out of 11 countries of the CM data set (Belgium+Luxembourg excluded, Austria and Finland start 1996:1, Portugal and Austria stop 2004:12)

- full panel - reduced version of 1995-2005 data set, trimmed in order to obtain a balanced panel. Covers 1996:1-2004:12 for all 10 countries.

- full panel for CM 1989-2001 sample - 9 countries: Austria and Finland excluded, due to short series. Series Ireland SITC 4 and Portugal SITC 4 also excluded due to missing values.

Table 1: Asymptotic moments for the normalized bias and t-ratio test statistics

	Model 1				Model 2				Model 3			
	$Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$	
$m-1$	Θ_1	Ψ_1	Θ_2	Ψ_2	Θ_1	Ψ_1	Θ_2	Ψ_2	Θ_1	Ψ_1	Θ_2	Ψ_2
1	-25.124	73.605	-3.558	0.388	-31.702	80.102	-4.003	0.341	-36.102	98.290	-4.276	0.366
2	-30.807	89.178	-3.943	0.392	-37.262	97.782	-4.343	0.355	-41.353	113.560	-4.581	0.374
3	-36.241	99.942	-4.285	0.373	-42.352	112.792	-4.637	0.369	-46.254	124.446	-4.853	0.364
4	-41.323	113.847	-4.580	0.373	-47.420	127.582	-4.912	0.368	-51.393	136.173	-5.124	0.364
5	-46.457	121.902	-4.865	0.365	-51.847	136.375	-5.145	0.362	-56.221	148.416	-5.366	0.366
6	-51.609	142.541	-5.131	0.384	-56.491	152.524	-5.375	0.378	-60.893	159.531	-5.593	0.365
7	-56.732	151.879	-5.389	0.372	-61.259	163.744	-5.606	0.375	-65.777	172.601	-5.820	0.369

	Model 4				Model 5				Model 6			
	$Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$	
$m-1$	Θ_1	Ψ_1	Θ_2	Ψ_2	Θ_1	Ψ_1	Θ_2	Ψ_2	Θ_1	Ψ_1	Θ_2	Ψ_2
1	-28.682	91.014	-3.798	0.431	-36.915	106.592	-4.324	0.393	-45.094	139.700	-4.783	0.418
2	-38.757	123.284	-4.427	0.436	-45.797	136.480	-4.821	0.408	-58.158	175.030	-5.453	0.415
3	-48.118	149.200	-4.944	0.431	-54.411	161.488	-5.271	0.415	-70.768	217.036	-6.037	0.432
4	-56.713	173.081	-5.380	0.430	-63.063	184.648	-5.687	0.410	-83.254	256.429	-6.573	0.441
5	-65.513	206.886	-5.798	0.447	-71.671	210.886	-6.081	0.417	-95.459	284.133	-7.065	0.435
6	-73.589	221.307	-6.163	0.427	-79.723	240.506	-6.425	0.434	-106.892	318.951	-7.498	0.443
7	-81.754	240.575	-6.513	0.423	-88.079	251.068	-6.771	0.417	-118.597	357.847	-7.923	0.455

Table 2: Response surfaces for the automatic lag length selection method ($k_{max} = 5$)

	Model 1				Model 2			
	$Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$	
	Θ_1	Ψ_1	Θ_2	Ψ_2	Θ_1	Ψ_1	Θ_2	Ψ_2
$\hat{\beta}_{0,0}$	0.41	56.823	-3.218	-19.638	0.372	71.034	-3.778	-26.654
$\hat{\beta}_{0,1}$	10.777	2079.863	-34.87	-97.193	1.676	1730.194	-42.359	-392.725
$\hat{\beta}_{0,2}$	-284.429		737.622	-3103.602	97.645	40207.55	1018.228	4124.832
$\hat{\beta}_{0,3}$	4332.145		-11377.84				-13147.4	
$\hat{\beta}_{1,0}$	-0.004	18.14	-0.442	-6.027	0.005	13.145	-0.351	-5.628
$\hat{\beta}_{1,1}$	-2.036		1.628	-68.79		1293.969	3.225	
$\hat{\beta}_{1,2}$	55.887	28710.63		1511.876		-18644.32	-36.265	
$\hat{\beta}_{1,3}$								
$\hat{\beta}_{2,0}$		-0.748	0.017	0.081	-0.001		0.01	0.064
$\hat{\beta}_{2,1}$	0.165	205.976	-0.114	0.967		48.061	-0.161	-5.218
$\hat{\beta}_{2,2}$		-5962.404			5.98			140.98
$\hat{\beta}_{2,3}$								

	Model 3				Model 4			
	$Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$	
	Θ_1	Ψ_1	Θ_2	Ψ_2	Θ_1	Ψ_1	Θ_2	Ψ_2
$\hat{\beta}_{0,0}$	0.389	72.251	-4.061	-31.033	0.517	70.453	-3.286	-19.519
$\hat{\beta}_{0,1}$	5.779	7427.681	-43.941	-465.591	1.919		-26.176	-101.832
$\hat{\beta}_{0,2}$	-225.895	-177465.4	921.364	1330.639		44801.21	166.728	-2334.409
$\hat{\beta}_{0,3}$	5584.734	2808044	-13082.02					
$\hat{\beta}_{1,0}$		19.721	-0.335	-5.637	-0.02	26.003	-0.649	-9.78
$\hat{\beta}_{1,1}$	-0.798		3.616			2162.096	3.806	-26.883
$\hat{\beta}_{1,2}$	56.865	37740.3	-30.174		72.559		-45.143	
$\hat{\beta}_{1,3}$								
$\hat{\beta}_{2,0}$	0.001	-0.737	0.009	0.059	0.001		0.025	0.086
$\hat{\beta}_{2,1}$	0.093	190.91	-0.194	-6.067	0.176	275.749	-0.227	-8.473
$\hat{\beta}_{2,2}$		-5491.499		146.45		-8513.873		292.56
$\hat{\beta}_{2,3}$								

	Model 5				Model 6			
	$Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$		$Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$	
	Θ_1	Ψ_1	Θ_2	Ψ_2	Θ_1	Ψ_1	Θ_2	Ψ_2
$\hat{\beta}_{0,0}$	0.399	109.977	-3.875	-27.694	0.424	87.103	-4.071	-31.407
$\hat{\beta}_{0,1}$		-8193.521	-35.047	-296.345		5713.206	-41.846	-243.518
$\hat{\beta}_{0,2}$	119.632	607421	681.665	1996.116	147.021	-286832.2	938.021	-12550.07
$\hat{\beta}_{0,3}$		-9915995	-8374.721			7973939	-14997.98	240521.8
$\hat{\beta}_{1,0}$	0.011	7.772	-0.549	-9.262	0.011	19.269	-0.509	-8.675
$\hat{\beta}_{1,1}$	0.937	6841.871	4.83	-9.079	0.858	4385.407	5.806	-66.601
$\hat{\beta}_{1,2}$		-256154.3	-74.577	-540.416		-58727.81	-111.692	3590.481
$\hat{\beta}_{1,3}$		3915350					1563.76	-69517.66
$\hat{\beta}_{2,0}$	-0.001	1.661	0.018	0.048	-0.001	1.239	0.014	
$\hat{\beta}_{2,1}$			-0.235	-8.318			-0.178	-7.49
$\hat{\beta}_{2,2}$	13.846			283.209	15.245		-3.999	271.886
$\hat{\beta}_{2,3}$								

Table 3: Mean and variance for the $ADF_{\tilde{e}}^{c}$, $ADF_{\tilde{e}}^{\tau}$ and $ADF_{\tilde{e}}^{\gamma}$ statistics

$ADF_{\tilde{e}}^{c}(i)$ statistic			$ADF_{\tilde{e}}^{\tau}(i)$ statistic		
T	$\Theta_2^{e^c}$	$\Psi_2^{e^c}$	T	$\Theta_2^{e^\tau}$	$\Psi_2^{e^\tau}$
50	-0.418	0.991	50	-1.549	0.367
100	-0.419	0.980	100	-1.541	0.353
250	-0.424	0.955	250	-1.538	0.346
500	-0.418	0.959	500	-1.536	0.346
1000	-0.424	0.964	1000	-1.535	0.341

$ADF_{\tilde{e}}^{\gamma}(i)$ statistic

T	λ	$\Theta_2^{e^\gamma}$	$\Psi_2^{e^\gamma}$	T	λ	$\Theta_2^{e^\gamma}$	$\Psi_2^{e^\gamma}$	T	λ	$\Theta_2^{e^\gamma}$	$\Psi_2^{e^\gamma}$
50	0.1	-1.684	0.423	100	0.1	-1.680	0.405	250	0.1	-1.682	0.399
	0.2	-1.829	0.450		0.2	-1.816	0.415		0.2	-1.810	0.394
	0.3	-1.932	0.414		0.3	-1.920	0.402		0.3	-1.904	0.378
	0.4	-2.013	0.398		0.4	-1.981	0.368		0.4	-1.957	0.354
	0.5	-2.022	0.383		0.5	-1.998	0.358		0.5	-1.967	0.330
	0.6	-2.011	0.404		0.6	-1.975	0.368		0.6	-1.961	0.349
	0.7	-1.940	0.425		0.7	-1.913	0.390		0.7	-1.913	0.385
	0.8	-1.834	0.447		0.8	-1.817	0.423		0.8	-1.808	0.402
	0.9	-1.681	0.423		0.9	-1.676	0.397		0.9	-1.682	0.400
500	0.1	-1.688	0.395	1,000	0.1	-1.676	0.389				
	0.2	-1.812	0.395		0.2	-1.809	0.392				
	0.3	-1.900	0.369		0.3	-1.902	0.371				
	0.4	-1.954	0.343		0.4	-1.950	0.338				
	0.5	-1.967	0.330		0.5	-1.972	0.339				
	0.6	-1.955	0.344		0.6	-1.953	0.346				
	0.7	-1.898	0.369		0.7	-1.900	0.365				
	0.8	-1.800	0.396		0.8	-1.799	0.390				
	0.9	-1.678	0.392		0.9	-1.691	0.392				

Table 4: Asymptotic critical values for the MQ tests

	$\lambda = 0.1$			$\lambda = 0.2$			$\lambda = 0.3$		
r	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	-32.163	-23.629	-19.865	-34.858	-26.091	-22.144	-36.123	-27.562	-23.619
2	-43.372	-34.321	-30.056	-46.436	-37.139	-32.688	-46.773	-37.778	-33.492
3	-53.648	-44.378	-39.748	-55.828	-46.232	-41.766	-57.136	-47.511	-42.775
4	-63.359	-53.470	-48.595	-65.206	-55.582	-50.645	-65.570	-55.883	-51.370
5	-73.691	-62.796	-57.434	-74.601	-64.165	-59.199	-75.573	-64.731	-59.919
6	-81.346	-71.238	-65.663	-83.575	-72.562	-67.309	-83.921	-73.247	-67.908
	$\lambda = 0.4$			$\lambda = 0.5$			$\lambda = 0.6$		
r	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	-36.635	-28.147	-24.140	-36.775	-28.226	-24.419	-36.805	-28.178	-24.176
2	-47.134	-38.391	-34.282	-48.148	-38.907	-34.553	-47.611	-38.587	-34.246
3	-57.176	-47.642	-43.088	-56.753	-47.715	-43.333	-57.230	-47.865	-43.200
4	-67.481	-56.958	-52.039	-65.752	-56.418	-51.708	-67.094	-56.599	-51.785
5	-75.603	-65.386	-60.204	-75.378	-65.302	-60.251	-75.182	-64.986	-60.057
6	-84.718	-73.703	-68.372	-83.902	-73.746	-68.222	-84.059	-73.136	-67.973
	$\lambda = 0.7$			$\lambda = 0.8$			$\lambda = 0.9$		
r	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	-36.302	-27.751	-23.890	-35.249	-26.722	-22.713	-32.918	-24.712	-20.896
2	-47.383	-38.223	-34.045	-46.572	-37.227	-33.085	-43.959	-35.248	-31.190
3	-56.908	-47.282	-42.693	-55.960	-46.442	-41.998	-54.568	-45.183	-40.623
4	-66.869	-56.270	-51.337	-65.833	-55.750	-50.890	-63.920	-53.985	-49.399
5	-75.074	-64.828	-59.867	-74.046	-64.430	-59.290	-74.177	-63.063	-57.839
6	-85.434	-73.646	-68.332	-83.244	-72.857	-67.721	-82.664	-71.518	-66.449

Table 5: Critical values for the $Z^{e}_{\hat{t}_{NT}}(\hat{\lambda})$, $ADF_{\tilde{F}}(\hat{\lambda})$ and $MQ(q,\hat{\lambda})$ statistics

Panel A: $Z^{e}_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$ statistic

Constant with or without change in level

T	1%	2.5%	5%	10%
50	-2.926	-2.517	-2.219	-1.901
100	-2.824	-2.402	-2.113	-1.759
250	-2.560	-2.250	-1.985	-1.619

Time trend with or without change in level

T	1%	2.5%	5%	10%
50	-2.900	-2.537	-2.120	-1.822
100	-2.924	-2.538	-2.240	-1.835
250	-2.619	-2.269	-1.931	-1.506

Time trend with one change in trend

T	1%	2.5%	5%	10%
50	-3.679	-3.389	-3.097	-2.714
100	-3.826	-3.467	-3.147	-2.804
250	-3.740	-3.373	-3.134	-2.794

Panel B: Common factor statistics

$ADF_{\tilde{F}}\left(\hat{\lambda}\right)$: Time trend with one change in trend

T	1%	2.5%	5%	10%
50	-4.779	-4.306	-4.008	-3.679
100	-4.549	-4.243	-3.930	-3.602
250	-4.474	-4.136	-3.873	-3.594

$MQ\left(q,\hat{\lambda}\right)$

T	r	1%	2.5%	5%	10%
50	1	-31.046	-27.569	-24.828	-21.669
	2	-38.827	-35.362	-32.792	-29.925
	3	-44.744	-42.436	-39.703	-36.641
	4	-47.752	-46.476	-44.865	-42.381
	5	-48.756	-48.305	-47.472	-46.119
	6	-48.890	-48.746	-48.444	-47.879
100	1	-34.474	-30.234	-26.833	-23.102
	2	-44.748	-40.147	-36.464	-32.729
	3	-53.423	-49.142	-45.879	-41.862
	4	-61.972	-57.307	-53.251	-49.284
	5	-69.033	-64.937	-61.099	-56.747
	6	-74.663	-70.434	-67.183	-63.437
250	1	-32.985	-28.983	-25.697	-22.843
	2	-46.953	-41.768	-38.103	-33.778
	3	-52.827	-48.542	-45.066	-41.136
	4	-59.494	-56.474	-53.392	-49.240
	5	-70.495	-66.474	-62.404	-57.440
	6	-78.589	-73.456	-68.748	-64.459

Table 6: Empirical size of the normalized bias and t-ratio tests (nominal size = 5%)

$Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$ statistic

N	T	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
20	50	0.039	0.046	0.043	0.033	0.054	0.045
	100	0.055	0.049	0.053	0.059	0.048	0.050
	250	0.050	0.053	0.046	0.052	0.056	0.059
40	50	0.040	0.049	0.046	0.030	0.044	0.056
	100	0.047	0.047	0.057	0.066	0.051	0.047
	250	0.056	0.061	0.047	0.044	0.046	0.055

$Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$ statistic

N	T	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
20	50	0.044	0.045	0.049	0.047	0.050	0.045
	100	0.050	0.050	0.045	0.046	0.043	0.053
	250	0.043	0.047	0.043	0.040	0.049	0.053
40	50	0.045	0.051	0.055	0.048	0.041	0.052
	100	0.041	0.047	0.047	0.044	0.046	0.043
	250	0.048	0.053	0.046	0.032	0.045	0.048

Table 7: Empirical power of the normalised bias statistic (nominal size = 5%)

				$Z_{\hat{\rho}_{NT}}\left(\hat{\lambda}\right)$ statistic			
				DGP: Model 1			
N	T	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
20	50	**0.455**	0.312	0.216	0.361	0.223	0.183
	100	**1**	0.998	0.989	1	0.931	0.980
	250	**1**	1	1	1	1	1
40	50	**0.676**	0.467	0.320	0.577	0.310	0.269
	100	**1**	1	1	1	0.998	1
	250	**1**	1	1	1	1	1
				DGP: Model 2			
N	T	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
20	50	0.001	**0.306**	0.219	0.004	0.211	0.185
	100	0	**1**	0.988	0.001	0.935	0.983
	250	0	**1**	1	0.000	1	1
40	50	0	**1**	0.334	0.001	0.309	0.261
	100	0	**1**	1	0.000	0.998	1
	250	0	**1**	1	0	1	1
				DGP: Model 3			
N	T	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
20	50	0	0.016	**0.110**	0	0.015	0.088
	100	0	0.089	**0.907**	0.001	0.121	0.861
	250	0	0.932	**1**	0	0.998	1
40	50	0	0.010	**0.125**	0	0.005	0.129
	100	0	0.085	**0.995**	0	0.159	0.992
	250	0	0.787	**1**	0	0.997	1
				DGP: Model 4			
N	T	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
20	50	0.864	0.389	0.527	**0.987**	0.478	0.413
	100	1	1	1	**1**	1	1
	250	1	1	1	**1**	1	1
40	50	0.996	0.754	0.671	**1**	0.781	0.687
	100	1	1	1	**1**	1	1
	250	1	1	1	**1**	1	1
				DGP: Model 5			
N	T	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
20	50	0.093	0.356	0.330	0.187	**0.485**	0.578
	100	0.236	0.999	1	0.283	**1**	1
	250	0.044	1	1	0.113	**1**	1
40	50	0.089	0.657	0.667	0.233	**0.714**	0.743
	100	0.305	1	1	0.515	**1**	1
	250	0.037	1	1	0.105	**1**	1
				DGP: Model 6			
N	T	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
20	50	0.002	0.151	0.311	0.018	0.077	**0.546**
	100	0.009	0.990	1	0.054	0.914	**1**
	250	0.001	0.997	1	0.022	0.998	**1**
40	50	0	0.328	0.606	0.005	0.244	**0.785**
	100	0	1	1	0.021	0.994	**1**
	250	0	1	1	0.003	1	**1**

Table 8: Empirical power of the pseudo t-ratio statistic (nominal size = 5%)

				$Z_{\hat{t}_{NT}}\left(\hat{\lambda}\right)$ statistic			
				DGP: Model 1			
N	T	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
20	50	**1**	1	1	1	1	1
	100	**1**	1	1	1	1	1
	250	**1**	1	1	1	1	1
40	50	**1**	1	1	1	1	1
	100	**1**	1	1	1	1	1
	250	**1**	1	1	1	1	1
				DGP: Model 2			
N	T	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
20	50	0.957	**1**	1	0.995	1	1
	100	0.403	**1**	1	0.675	1	1
	250	0.034	**1**	1	0.073	1	1
40	50	1	**1**	1	1	1	1
	100	0.647	**1**	1	0.908	1	1
	250	0.026	**1**	1	0.070	1	1
				DGP: Model 3			
N	T	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
20	50	0.303	1	**1**	0.765	1	1
	100	0.018	1	**1**	0.221	1	1
	250	0.001	1	**1**	0.009	1	1
40	50	0.497	1	**1**	0.958	1	1
	100	0.014	1	**1**	0.324	1	1
	250	0	1	**1**	0.003	1	1
				DGP: Model 4			
N	T	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
20	50	1	1	1	**1**	1	1
	100	1	1	1	**1**	1	1
	250	1	1	1	**1**	1	1
40	50	1	1	1	**1**	1	1
	100	1	1	1	**1**	1	1
	250	1	1	1	**1**	1	1
				DGP: Model 5			
N	T	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
20	50	1	1	1	1	**1**	1
	100	0.917	1	1	0.981	**1**	1
	250	0.219	1	1	0.413	**1**	1
40	50	1	1	1	1	**1**	1
	100	0.992	1	1	1	**1**	1
	250	0.308	1	1	0.594	**1**	1
				DGP: Model 6			
N	T	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
20	50	0.976	1	1	0.999	1	**1**
	100	0.287	1	1	0.741	1	**1**
	250	0.009	1	1	0.083	1	**1**
40	50	0.999	1	1	1	1	**1**
	100	0.229	1	1	0.901	1	**1**
	250	0.003	1	1	0.060	1	**1**

Table 9: Empirical size and power of the ADF statistic for the idiosyncratic and factor components. Model 3 with one common factor, with break under the alternative hypothesis at $\lambda = 0.5$ and assumed unknown

T	σ_F^2	α	ρ_i	$Z^e_{\hat{t}_{NT}}$	$ADF^d_{\hat{F}}$	ρ_i	$Z^e_{\hat{t}_{NT}}$	$ADF^d_{\hat{F}}$	ρ_i	$Z^e_{\hat{t}_{NT}}$	$ADF^d_{\hat{F}}$
50	0.5	1	1	0.031	0.036	0.99	0.041	0.016	0.95	0.174	0.005
100	0.5	1	1	0.035	0.039	0.99	0.028	0.042	0.95	0.348	0.054
250	0.5	1	1	0.037	0.034	0.99	0.047	0.048	0.95	1	0.060
50	0.5	0.9	1	0.025	0.067	0.99	0.041	0.039	0.95	0.193	0.001
100	0.5	0.9	1	0.031	0.145	0.99	0.012	0.133	0.95	0.738	0.012
250	0.5	0.9	1	0.037	0.630	0.99	0.041	0.583	0.95	1	0.661
50	0.5	0.8	1	0.024	0.148	0.99	0.053	0.032	0.95	0.171	0.002
100	0.5	0.8	1	0.030	0.466	0.99	0.093	0.103	0.95	0.315	0.397
250	0.5	0.8	1	0.038	0.992	0.99	0.044	0.981	0.95	1	0.997
50	1	1	1	0.028	0.058	0.99	0.073	0.002	0.95	0.210	0.001
100	1	1	1	0.023	0.054	0.99	0.086	0.012	0.95	0.338	0.047
250	1	1	1	0.036	0.039	0.99	0.043	0.043	0.95	1	0.038
50	1	0.9	1	0.026	0.065	0.99	0.015	0.058	0.95	0.145	0.017
100	1	0.9	1	0.020	0.123	0.99	0.053	0.078	0.95	0.311	0.116
250	1	0.9	1	0.023	0.628	0.99	0.048	0.620	0.95	1	0.628
50	1	0.8	1	0.031	0.134	0.99	0.043	0.051	0.95	0.152	0.019
100	1	0.8	1	0.020	0.468	0.99	0.005	0.401	0.95	0.417	0.323
250	1	0.8	1	0.025	0.999	0.99	0.042	0.991	0.95	1	0.888
50	10	1	1	0.043	0.037	0.99	0.076	0.008	0.95	0.163	0.009
100	10	1	1	0.040	0.043	0.99	0.027	0.040	0.95	0.238	0.036
250	10	1	1	0.032	0.044	0.99	0.043	0.051	0.95	1	0.055
50	10	0.9	1	0.039	0.052	0.99	0.053	0.007	0.95	0.140	0.010
100	10	0.9	1	0.026	0.123	0.99	0.020	0.115	0.95	0.262	0.140
250	10	0.9	1	0.032	0.638	0.99	0.030	0.637	0.95	1	0.639
50	10	0.8	1	0.031	0.075	0.99	0.038	0.060	0.95	0.064	0.060
100	10	0.8	1	0.040	0.396	0.99	0.016	0.410	0.95	0.221	0.393
250	10	0.8	1	0.032	0.996	0.99	0.036	0.972	0.95	1	0.968

Table 10: Empirical size and power of the ADF statistic for the idiosyncratic and factor components. Model 6 with one common factor, with break under the alternative hypothesis at $\lambda = 0.5$ and assumed unknown

T	σ_F^2	α	ρ_i	$Z^e_{\hat{t}_{NT}}$	$ADF^d_{\hat{F}}$	ρ_i	$Z^e_{\hat{t}_{NT}}$	$ADF^d_{\hat{F}}$	ρ_i	$Z^e_{\hat{t}_{NT}}$	$ADF^d_{\hat{F}}$
50	0.5	1	1	0.027	0.049	0.99	0.072	0.001	0.95	0.138	0.003
100	0.5	1	1	0.041	0.043	0.99	0.026	0.003	0.95	0.267	0.009
250	0.5	1	1	0.041	0.040	0.99	0.037	0.005	0.95	1	0.043
50	0.5	0.9	1	0.037	0.081	0.99	0.071	0.001	0.95	0.115	0.004
100	0.5	0.9	1	0.027	0.127	0.99	0.023	0.004	0.95	0.253	0.018
250	0.5	0.9	1	0.023	0.610	0.99	0.029	0.107	0.95	1	0.326
50	0.5	0.8	1	0.032	0.144	0.99	0.052	0.002	0.95	0.105	0.006
100	0.5	0.8	1	0.021	0.418	0.99	0.015	0.020	0.95	0.245	0.055
250	0.5	0.8	1	0.047	0.995	0.99	0.029	0.148	0.95	1	0.509
50	1	1	1	0.035	0.055	0.99	0.056	0.005	0.95	0.107	0.010
100	1	1	1	0.037	0.044	0.99	0.027	0.008	0.95	0.256	0.009
250	1	1	1	0.035	0.040	0.99	0.036	0.014	0.95	1	0.026
50	1	0.9	1	0.030	0.057	0.99	0.078	0.006	0.95	0.106	0.002
100	1	0.9	1	0.033	0.129	0.99	0.012	0.006	0.95	0.242	0.023
250	1	0.9	1	0.051	0.659	0.99	0.029	0.084	0.95	1	0.328
50	1	0.8	1	0.032	0.145	0.99	0.059	0.007	0.95	0.092	0.008
100	1	0.8	1	0.029	0.457	0.99	0.018	0.018	0.95	0.259	0.063
250	1	0.8	1	0.039	0.995	0.99	0.030	0.128	0.95	1	0.487
50	10	1	1	0.114	0.014	0.99	0.072	0.002	0.95	0.097	0.001
100	10	1	1	0.065	0.029	0.99	0.023	0.002	0.95	0.172	0.006
250	10	1	1	0.058	0.042	0.99	0.039	0.010	0.95	0.997	0.018
50	10	0.9	1	0.105	0.032	0.99	0.056	0.001	0.95	0.090	0.003
100	10	0.9	1	0.080	0.076	0.99	0.015	0.006	0.95	0.156	0.012
250	10	0.9	1	0.056	0.632	0.99	0.031	0.076	0.95	1	0.283
50	10	0.8	1	0.121	0.047	0.99	0.071	0.003	0.95	0.080	0.005
100	10	0.8	1	0.117	0.293	0.99	0.030	0.022	0.95	0.131	0.040
250	10	0.8	1	0.064	0.934	0.99	0.026	0.153	0.95	1	0.407

Table 11: Empirical size and power of the ADF statistic for the idiosyncratic and MQ test for factor components. Model 3, three common factors, with break under the alternative hypothesis at $\lambda = 0.5$ assumed unknown, and $\rho_i = 1$

T	σ_F^2	α	ρ_i	$Z_{\hat{t}_{NT}}^e$	$MQ\,(0)$	$MQ\,(1)$	$MQ\,(2)$	$MQ\,(3)$	$MQ\,(4)$	$MQ\,(5)$	$MQ\,(6)$
50	0.5	1	1	0.044	0	0.000	0.025	0.974	0	0	0
100	0.5	1	1	0.039	0.003	0.010	0.052	0.935	0	0	0
250	0.5	1	1	0.038	0.002	0.008	0.098	0.892	0	0	0
50	0.5	0.9	1	0.031	0.005	0.004	0.036	0.955	0	0	0
100	0.5	0.9	1	0.023	0.004	0.014	0.073	0.909	0	0	0
250	0.5	0.9	1	0.044	0.392	0.051	0.278	0.279	0	0	0
50	0.5	0.8	1	0.013	0.005	0.003	0.041	0.951	0	0	0
100	0.5	0.8	1	0.029	0.091	0.046	0.161	0.702	0	0	0
250	0.5	0.8	1	0.029	0.990	0.002	0.004	0.004	0	0	0
50	1	1	1	0.043	0	0.001	0.022	0.974	0.003	0	0
100	1	1	1	0.033	0	0	0.033	0.967	0	0	0
250	1	1	1	0.037	0.002	0.013	0.104	0.881	0	0	0
50	1	0.9	1	0.022	0	0.003	0.036	0.961	0	0	0
100	1	0.9	1	0.030	0.011	0.014	0.066	0.909	0	0	0
250	1	0.9	1	0.042	0.436	0.056	0.271	0.237	0	0	0
50	1	0.8	1	0.024	0.002	0.009	0.050	0.934	0.005	0	0
100	1	0.8	1	0.023	0.125	0.062	0.135	0.678	0	0	0
250	1	0.8	1	0.035	0.995	0	0.004	0.001	0	0	0
50	10	1	1	0.063	0	0.003	0.013	0.306	0.342	0.195	0.141
100	10	1	1	0.045	0	0.001	0.031	0.765	0.170	0.027	0.006
250	10	1	1	0.038	0.001	0.010	0.099	0.884	0.006	0	0
50	10	0.9	1	0.045	0.001	0.001	0.015	0.246	0.288	0.231	0.218
100	10	0.9	1	0.038	0.004	0.015	0.071	0.690	0.180	0.029	0.011
250	10	0.9	1	0.045	0.470	0.077	0.236	0.214	0.003	0	0
50	10	0.8	1	0.024	0.002	0.006	0.020	0.228	0.307	0.218	0.219
100	10	0.8	1	0.025	0.111	0.044	0.148	0.559	0.124	0.013	0.001
250	10	0.8	1	0.031	1	0	0	0	0	0	0

Table 12: Empirical size and power of the ADF statistic for the idiosyncratic and MQ test for factor components. Model 3, three common factors, with break under the alternative hypothesis at $\lambda = 0.5$ assumed unknown, and $\rho_i = 0.99$

T	σ_F^2	α	ρ_i	$Z^e_{\hat{t}_{NT}}$	$MQ\,(0)$	$MQ\,(1)$	$MQ\,(2)$	$MQ\,(3)$	$MQ\,(4)$	$MQ\,(5)$	$MQ\,(6)$
50	0.5	1	0.99	0.023	0.002	0.001	0.019	0.976	0.002	0	0
100	0.5	1	0.99	0.004	0.002	0.005	0.038	0.955	0	0	0
250	0.5	1	0.99	0.052	0.003	0.017	0.125	0.855	0	0	0
50	0.5	0.9	0.99	0.046	0.001	0.002	0.011	0.351	0.635	0	0
100	0.5	0.9	0.99	0.011	0.005	0.007	0.078	0.910	0	0	0
250	0.5	0.9	0.99	0.043	0.407	0.096	0.247	0.250	0	0	0
50	0.5	0.8	0.99	0.013	0.003	0.009	0.033	0.759	0.196	0	0
100	0.5	0.8	0.99	0.004	0.049	0.057	0.134	0.76	0	0	0
250	0.5	0.8	0.99	0.044	0.980	0.014	0.005	0.001	0	0	0
50	1	1	0.99	0.049	0	0.002	0.017	0.635	0.346	0	0
100	1	1	0.99	0.013	0.001	0.005	0.051	0.943	0	0	0
250	1	1	0.99	0.031	0	0.009	0.115	0.876	0	0	0
50	1	0.9	0.99	0.032	0.002	0.003	0.019	0.621	0.354	0.001	0
100	1	0.9	0.99	0.005	0.004	0.013	0.074	0.909	0	0	0
250	1	0.9	0.99	0.036	0.433	0.069	0.266	0.232	0	0	0
50	1	0.8	0.99	0.015	0.002	0.012	0.043	0.844	0.099	0	0
100	1	0.8	0.99	0.010	0.066	0.061	0.155	0.718	0	0	0
250	1	0.8	0.99	0.038	0.985	0.013	0.002	0	0	0	0
50	10	1	0.99	0.035	0.001	0.002	0.008	0.306	0.324	0.204	0.155
100	10	1	0.99	0.009	0.003	0.004	0.056	0.793	0.128	0.016	0
250	10	1	0.99	0.049	0.004	0.012	0.115	0.851	0.018	0	0
50	10	0.9	0.99	0.019	0	0.001	0.011	0.274	0.307	0.231	0.176
100	10	0.9	0.99	0.012	0.010	0.012	0.075	0.712	0.153	0.033	0.005
250	10	0.9	0.99	0.024	0.472	0.057	0.247	0.223	0.001	0	0
50	10	0.8	0.99	0.033	0.001	0.004	0.014	0.193	0.288	0.262	0.238
100	10	0.8	0.99	0.013	0.13	0.04	0.144	0.543	0.125	0.015	0.003
250	10	0.8	0.99	0.028	1	0	0	0	0	0	0

Table 13: Empirical size and power of the ADF statistic for the idiosyncratic and MQ test for factor components. Model 3, three common factors, with break under the alternative hypothesis at $\lambda = 0.5$ assumed unknown, and $\rho_i = 0.95$

T	σ_F^2	α	ρ_i	$Z_{\hat{t}_{NT}}^e$	$MQ\,(0)$	$MQ\,(1)$	$MQ\,(2)$	$MQ\,(3)$	$MQ\,(4)$	$MQ\,(5)$	$MQ\,(6)$
50	0.5	1	0.95	0.077	0	0.003	0.024	0.917	0.056	0	0
100	0.5	1	0.95	0.297	0	0.002	0.046	0.952	0	0	0
250	0.5	1	0.95	1	0.002	0.006	0.107	0.885	0	0	0
50	0.5	0.9	0.95	0.062	0.001	0.003	0.026	0.955	0.015	0	0
100	0.5	0.9	0.95	0.268	0.004	0.011	0.101	0.884	0	0	0
250	0.5	0.9	0.95	1	0.450	0.072	0.257	0.221	0	0	0
50	0.5	0.8	0.95	0.050	0.001	0.003	0.038	0.956	0.002	0	0
100	0.5	0.8	0.95	0.241	0.081	0.068	0.161	0.688	0.002	0	0
250	0.5	0.8	0.95	1	0.997	0.001	0.001	0.001	0	0	0
50	1	1	0.95	0.068	0	0.003	0.014	0.980	0.003	0	0
100	1	1	0.95	0.282	0.001	0.008	0.039	0.952	0	0	0
250	1	1	0.95	1	0.003	0.01	0.120	0.867	0	0	0
50	1	0.9	0.95	0.060	0	0.005	0.017	0.965	0.013	0	0
100	1	0.9	0.95	0.227	0.014	0.017	0.058	0.911	0	0	0
250	1	0.9	0.95	1	0.483	0.077	0.250	0.190	0	0	0
50	1	0.8	0.95	0.039	0.004	0.008	0.030	0.881	0.077	0	0
100	1	0.8	0.95	0.215	0.088	0.06	0.133	0.712	0.007	0	0
250	1	0.8	0.95	1	0.996	0.001	0.002	0.001	0	0	0
50	10	1	0.95	0.073	0.002	0.001	0.023	0.352	0.297	0.193	0.132
100	10	1	0.95	0.167	0	0.003	0.032	0.769	0.165	0.026	0.005
250	10	1	0.95	0.997	0.005	0.018	0.119	0.849	0.009	0	0
50	10	0.9	0.95	0.056	0	0	0.004	0.21	0.273	0.241	0.272
100	10	0.9	0.95	0.138	0.007	0.009	0.068	0.763	0.130	0.018	0.005
250	10	0.9	0.95	1	0.467	0.057	0.25	0.223	0.003	0	0
50	10	0.8	0.95	0.038	0.001	0.004	0.012	0.374	0.309	0.170	0.130
100	10	0.8	0.95	0.113	0.095	0.050	0.140	0.551	0.132	0.031	0.001
250	10	0.8	0.95	0.998	0.999	0	0	0	0.001	0	0

Table 14: Empirical size and power of the ADF statistic for the idiosyncratic and MQ test for factor components. Model 6, three common factors, with break under the alternative hypothesis at $\lambda = 0.5$ assumed unknown, and $\rho_i = 1$

T	σ_F^2	α	ρ_i	$Z^e_{\hat{t}_{NT}}$	$MQ\,(0)$	$MQ\,(1)$	$MQ\,(2)$	$MQ\,(3)$	$MQ\,(4)$	$MQ\,(5)$	$MQ\,(6)$
50	0.5	1	1	0.070	0.002	0.001	0.022	0.975	0	0	0
100	0.5	1	1	0.050	0.001	0.004	0.039	0.956	0	0	0
250	0.5	1	1	0.048	0.002	0.008	0.108	0.882	0	0	0
50	0.5	0.9	1	0.030	0	0.005	0.018	0.977	0	0	0
100	0.5	0.9	1	0.036	0.003	0.016	0.072	0.909	0	0	0
250	0.5	0.9	1	0.030	0.375	0.063	0.298	0.264	0	0	0
50	0.5	0.8	1	0.034	0.006	0.012	0.039	0.943	0	0	0
100	0.5	0.8	1	0.018	0.083	0.042	0.168	0.707	0	0	0
250	0.5	0.8	1	0.026	0.986	0.005	0.008	0.001	0	0	0
50	1	1	1	0.080	0	0.002	0.022	0.974	0.002	0	0
100	1	1	1	0.043	0	0.007	0.041	0.952	0	0	0
250	1	1	1	0.042	0.001	0.013	0.128	0.858	0	0	0
50	1	0.9	1	0.058	0.002	0.001	0.022	0.964	0.011	0	0
100	1	0.9	1	0.048	0.015	0.018	0.077	0.890	0	0	0
250	1	0.9	1	0.034	0.412	0.047	0.271	0.270	0	0	0
50	1	0.8	1	0.044	0.004	0.005	0.045	0.943	0.003	0	0
100	1	0.8	1	0.037	0.115	0.053	0.154	0.678	0	0	0
250	1	0.8	1	0.032	0.998	0.001	0.001	0	0	0	0
50	10	1	1	0.254	0	0	0.009	0.210	0.266	0.229	0.286
100	10	1	1	0.251	0	0.004	0.033	0.670	0.233	0.050	0.010
250	10	1	1	0.113	0.001	0.008	0.093	0.857	0.041	0	0
50	10	0.9	1	0.225	0	0.002	0.005	0.116	0.214	0.248	0.415
100	10	0.9	1	0.243	0.003	0.007	0.058	0.687	0.189	0.047	0.009
250	10	0.9	1	0.166	0.451	0.077	0.224	0.241	0.007	0	0
50	10	0.8	1	0.203	0.001	0.005	0.017	0.242	0.272	0.208	0.255
100	10	0.8	1	0.245	0.093	0.029	0.141	0.463	0.194	0.064	0.016
250	10	0.8	1	0.114	0.999	0	0.001	0	0	0	0

Table 15: Empirical size and power of the ADF statistic for the idiosyncratic and MQ test for factor components. Model 6, three common factors, with break under the alternative hypothesis at $\lambda = 0.5$ assumed unknown, and $\rho_i = 0.99$

T	σ_F^2	α	ρ_i	$Z_{\hat{t}_{NT}}^e$	$MQ\,(0)$	$MQ\,(1)$	$MQ\,(2)$	$MQ\,(3)$	$MQ\,(4)$	$MQ\,(5)$	$MQ\,(6)$
50	0.5	1	0.99	0.077	0	0	0.001	0.060	0.525	0.176	0.238
100	0.5	1	0.99	0.033	0	0.001	0.007	0.107	0.781	0.092	0.012
250	0.5	1	0.99	0.070	0.002	0.007	0.040	0.228	0.723	0	0
50	0.5	0.9	0.99	0.056	0	0	0.008	0.144	0.389	0.189	0.270
100	0.5	0.9	0.99	0.021	0	0	0.009	0.108	0.770	0.101	0.012
250	0.5	0.9	0.99	0.041	0.036	0.168	0.172	0.281	0.343	0	0
50	0.5	0.8	0.99	0.045	0.001	0	0.005	0.048	0.517	0.179	0.250
100	0.5	0.8	0.99	0.025	0.002	0.022	0.037	0.175	0.664	0.087	0.013
250	0.5	0.8	0.99	0.029	0.106	0.742	0.123	0.019	0.010	0	0
50	1	1	0.99	0.072	0	0	0	0.025	0.601	0.155	0.219
100	1	1	0.99	0.029	0	0.001	0.001	0.071	0.851	0.070	0.006
250	1	1	0.99	0.055	0	0.006	0.038	0.196	0.76	0	0
50	1	0.9	0.99	0.050	0	0.001	0	0.028	0.584	0.168	0.219
100	1	0.9	0.99	0.040	0	0.002	0.010	0.085	0.825	0.077	0.001
250	1	0.9	0.99	0.045	0.042	0.155	0.164	0.303	0.336	0	0
50	1	0.8	0.99	0.050	0	0	0.007	0.039	0.601	0.152	0.201
100	1	0.8	0.99	0.016	0.001	0.019	0.019	0.165	0.720	0.066	0.010
250	1	0.8	0.99	0.049	0.086	0.608	0.232	0.061	0.013	0	0
50	10	1	0.99	0.212	0	0.001	0	0.012	0.643	0.231	0.113
100	10	1	0.99	0.079	0	0	0.001	0.039	0.800	0.127	0.033
250	10	1	0.99	0.071	0	0.003	0.036	0.174	0.779	0.008	0
50	10	0.9	0.99	0.188	0	0.001	0.001	0.009	0.480	0.298	0.211
100	10	0.9	0.99	0.054	0.001	0.004	0.011	0.062	0.808	0.102	0.012
250	10	0.9	0.99	0.061	0.220	0.08	0.207	0.196	0.291	0.006	0
50	10	0.8	0.99	0.148	0	0.001	0	0.021	0.482	0.273	0.223
100	10	0.8	0.99	0.070	0.03	0.01	0.03	0.134	0.719	0.062	0.015
250	10	0.8	0.99	0.024	0.829	0.127	0.039	0.005	0	0	0

Table 16: Empirical size and power of the ADF statistic for the idiosyncratic and MQ test for factor components. Model 6, three common factors, with break under the alternative hypothesis at $\lambda = 0.5$ assumed unknown, and $\rho_i = 0.95$

T	σ_F^2	α	ρ_i	$Z^e_{\hat{t}_{NT}}$	$MQ\,(0)$	$MQ\,(1)$	$MQ\,(2)$	$MQ\,(3)$	$MQ\,(4)$	$MQ\,(5)$	$MQ\,(6)$
50	0.5	1	0.95	0.138	0	0	0.004	0.11	0.487	0.184	0.215
100	0.5	1	0.95	0.272	0	0.003	0.009	0.129	0.744	0.106	0.009
250	0.5	1	0.95	1	0.002	0.012	0.062	0.515	0.409	0	0
50	0.5	0.9	0.95	0.09	0	0	0	0.081	0.538	0.156	0.225
100	0.5	0.9	0.95	0.233	0.002	0.005	0.016	0.193	0.693	0.084	0.007
250	0.5	0.9	0.95	0.999	0.219	0.176	0.174	0.257	0.174	0	0
50	0.5	0.8	0.95	0.096	0	0.002	0.009	0.106	0.479	0.167	0.237
100	0.5	0.8	0.95	0.188	0.017	0.034	0.067	0.233	0.552	0.092	0.005
250	0.5	0.8	0.95	1	0.495	0.493	0.009	0.002	0.001	0	0
50	1	1	0.95	0.118	0	0	0.003	0.041	0.505	0.193	0.258
100	1	1	0.95	0.243	0	0	0.006	0.151	0.768	0.074	0.001
250	1	1	0.95	1	0	0.004	0.072	0.455	0.469	0	0
50	1	0.9	0.95	0.115	0	0.001	0.002	0.037	0.622	0.153	0.185
100	1	0.9	0.95	0.226	0.002	0.006	0.018	0.166	0.730	0.073	0.005
250	1	0.9	0.95	1	0.200	0.158	0.198	0.264	0.180	0	0
50	1	0.8	0.95	0.072	0	0.002	0.004	0.056	0.624	0.154	0.160
100	1	0.8	0.95	0.186	0.020	0.026	0.030	0.199	0.654	0.067	0.004
250	1	0.8	0.95	1	0.454	0.539	0.006	0.001	0	0	0
50	10	1	0.95	0.217	0	0	0.002	0.009	0.518	0.267	0.204
100	10	1	0.95	0.205	0	0	0.005	0.065	0.792	0.116	0.022
250	10	1	0.95	0.972	0.001	0.005	0.039	0.346	0.600	0.009	0
50	10	0.9	0.95	0.211	0	0.001	0.003	0.019	0.603	0.251	0.123
100	10	0.9	0.95	0.213	0.001	0	0.010	0.072	0.792	0.11	0.015
250	10	0.9	0.95	0.979	0.260	0.08	0.190	0.233	0.226	0.011	0
50	10	0.8	0.95	0.204	0.001	0.001	0.005	0.018	0.437	0.297	0.241
100	10	0.8	0.95	0.172	0.027	0.013	0.029	0.146	0.661	0.103	0.021
250	10	0.8	0.95	0.965	0.881	0.107	0.008	0.004	0	0	0

Table 17: Panel cointegration tests for the pass-through in the euro area

	Cross-section independent		Cross-section dependent						
	Panel test	Panel test	Homogeneous break dates				Heterogeneous break dates		
Model	t-ratio	norm. bias	Idiosyn.	$\hat{T}_b$	$\hat{r}$	$\hat{r}_1$	Idiosyn.	$\hat{r}$	$\hat{r}_1$
1	-22.596	-51.548	-5.009	1999.02	6	1	-2.744	6	2
2	-26.500	-59.188	-6.409	2002.11	6	1	-2.933	6	1
3	-25.018	-60.867	-5.816	2002.03	6	1			
4	-22.523	-51.507	-5.201	2000.03	6	1	-3.060	6	3
5	-24.677	-56.949	-6.189	2002.08	6	1	-2.239	6	2
6	-23.498	-57.660	-6.353	2002.10	6	1			