UNIVERSITÀ DEGLI STUDI DI FERRARA
DIPARTIMENTO DI ECONOMIA E MANAGEMENT
Via Voltapaletto, 11 – 44121 Ferrara

Quaderno DEM 6/2015

March 2015

Weak and Strong cross-sectional dependence: a panel data analysis of international technology diffusion

Cem Ertur – Antonio Musolesi

Quaderni DEM, volume 4 ISSN 2281-9673

Weak and Strong cross-sectional dependence: a panel data analysis of international technology diffusion

CEM ERTUR† and ANTONIO MUSOLESI‡

February 25, 2015

Abstract

This paper provides an econometric examination of geographic R&D spillovers among countries by focusing on the issue of cross-sectional dependence, and in particular on the different ways – weak and strong – it may affect the model. A preliminary analysis based on the estimation of the exponent of cross-sectional correlation proposed by Bailey et al. (2013), a, provides a very clear-cut result with an estimate of a very close to unity, not only indicating the presence of strong cross-sectional correlation but also being consistent with the factor literature typically assuming that $a = 1$. Moreover, second generation unit roots tests suggest that while the unobserved idiosyncratic component of the variables under study may be stationary, the unobserved common factors appear to be nonstationary. Consequently, a factor structure appears to be preferable to a spatial error model and in particular the Correlated Common Effects approach is employed since, among other things, it is still valid in the more general case of nonstationary common factors. Finally, comparing the results with those obtained with a spatial model gives some insights on the possible bias occurring when allowing only for weak correlation while strong correlation is present in the data.

JEL classification: C23; C5; F0; O3.

Keywords: panel data; cross-sectional correlation; spatial models; factor models; unit root; international technology diffusion; geography.

†*LEO, University of Orleans UMR 6221 CNRS Faculté de Droit d'Economie et de Gestion, Rue de Blois - B.P. 6739 45067 Orléans Cedex 2 –France*
(email: cem.ertur@univ-orleans.fr)

‡*Department of Economics and Management (DEM), University of Ferrara, and SEEDS, Via Voltapaletto 11, 44100 Ferrara - Italy*
(antonio.musolesi@unife.it)

1 Introduction

Since the seminal paper by Coe and Helpman (1995), henceforth CH (recently revisited by Coe et al., 2009), there has been an increasing interest in international technology diffusion. CH test the predictions of models of innovation and growth (Grossman and Helpman, 1991) in which total factor productivity (TFP) is an increasing function of cumulative research and development (R&D). In particular, CH analyse the role of international trade. By assuming that some intermediate inputs are traded internationally, whereas others are not, they relate TFP to both domestic and foreign R&D and construct the foreign R&D capital stock as the import-share weighted average of the domestic R&D capital stocks of the trading partners. The influence of this approach is based on its plausibility with respect to endogenous growth theory (Keller, 2004) and its versatility in allowing for the consideration of alternative channels of international technology diffusion, such as foreign direct investment (FDI) (Lichtenberg and Van Pottelsberghe, 2001), bilateral technological proximity and patent citations between countries (Lee, 2006), language skills or geographical proximity (Keller, 2002).

The present paper aims to contribute to the empirical literature on R&D spillovers among countries by focusing on the issue of cross-sectional dependence. The rationale is that cross-country correlation from a variety of sources can plausibly be present in CH-type specifications; however, this correlation complicates standard estimation and inference.

Cross-sectional dependence can be introduced as a result of a finite number of *unobservable (and/or observed) common factors* that may have different effects on TFP across countries. Such factors might include, for instance, aggregate technological shocks, national policies aimed at raising the level of technology or oil price shocks that may influence TFP through their effects on production costs. The heterogeneous effect of these factors may be the result, for instance, of country-specific technological constraints. A model with multifactor error structure can be estimated adopting the correlated common effects (CCE) approach developed by Pesaran (2006), which has been further developed and proved to be valid in a variety of situations (Chudik et al., 2011; Pesaran and Tosetti, 2011; Kapetanios et al., 2011). Cross-sectional correlation can be alternatively posed as a result of spatial spillover effects and can be modeled by adopting a spatial panel econometric framework allowing for spatially correlated disturbances. The estimation can be performed using Lee and Yu's (2010) quasi-maximum likelihood (QML) estimator. Some remarks are in order.

First, the unobserved common factors approach and the spatial error model are related to the recently developed concepts of weak and strong cross-sectional dependence. Although the literature does not provide a unique definition of "weak" and "strong" dependence (see, e.g., Chudik et al., 2011, and Sarafidis, 2009), it is interesting to note that the spatial models, under a standard set of regularity conditions, entail weak cross-sectional correlation regardless of the definition adopted (Breitung and Pesaran, 2008; Pesaran and Tosetti, 2011; Sarafidis and Wansbeek, 2012), whereas the type of dependence arising from the factor model depends on the adopted definition of weak/strong dependence and the limiting properties of averaged factor loadings (Sarafidis and Wansbeek, 2012). A related concept is that of strong and weak factors recently proposed by Chudik et al. (2011). The authors demonstrate that there is a direct relationship between the concept of weak/strong factors and their conception of weak/strong dependence. They demonstrate that a process that is the sum of a finite number of common factors and an idiosyncratic error term is cross-sectionally strongly dependent at a given point in time if at least one of those common factors is strong. Specifically, the CCE approach explicitly introduces a finite number of *strong* factors, entailing strong dependence, but does not explicitly introduce *weak* factors.

Second, the most general model would obviously allow for both forms of dependence – weak and strong – as suggested by some recent papers (Pesaran and Tosetti, 2011; Chudick et al., 2011; Bresson and Hsiao, 2011; Bailey et al. 2013b). However, the CCE approach has been shown to be valid even in this case. In particular, Pesaran and Tosetti (2011) prove that the CCE estimator provides consistent estimates of the slope coefficients and their standard errors under a generalized data generating process (DGP) with an error term that is the sum of a multifactor structure and a spatial process, i.e. when both forms of cross-correlation – weak and strong – characterize the DGP. Moreover, Chudik et al. (2011) extended the CCE approach by allowing for the presence of both a limited number of strong factors and a large number of weak or semi-strong factors and then show that, even under this extended framework, the CCE method still provides consistent estimates of the slope coefficients.

Third, while both factor and spatial models allow for cross-section correlation, the motivations underlying such models differ importantly. In the first approach, the unobserved factors are viewed as nuisance variables introduced to allow for cross-sectional dependence and to capture information in a parsimonious way, whereas the main focus is on the estimation and inference of the slope parameters. Moreover, this set-up introduces endogeneity due to unobservables, whereby the explanatory variables are allowed to be correlated with the factors. Bai (2009) and Sarafidis and Wansbeek (2012) provide many examples of circumstances under which this may occur, such as production and cost function specifications. This can also be relevant in the CH specification, as stressed by Keller (2004, p.763). The spatial models, instead, aim to model interactions among cross-sections, such as spillover effects. Such cross-section interactions will produce differentiated impact coefficients computed from the reduced form of the spatial model (see e.g., Debarsy and Ertur, 2010). Moreover, spatial models have been shown to be relevant in many contexts which are related to this paper such as neoclassical and endogeneous growth models (Ertur and Koch, 2007, 2011).

We argue that, when analysing international R&D spillovers, there are neither theoretical reasons nor well-established empirical evidence allowing for an a priori choice between the spatial approach and the factor model. If the data exhibit only weak correlation, this could be an indication of the necessity to model spatial interactions while if strong correlation is present, this would suggest that relevant variables have not been accounted for in the original formulation or that both forms of correlation may affect the model. In this paper, we first focus on testing and measuring cross section correlation. This appears very relevant from both a theoretical and a policy oriented perspective. According to the obtained results, we finally estimate the model with the most suitable approach and provide some new results.

The remainder of the paper is organized as follows: section 2 describes the baseline model. Section 3 extends the benchmark specification by allowing for cross-sectional dependence. Section 4 presents the results. Finally, section 5 concludes.

2 Baseline econometric specification

The baseline econometric model is an extended version of that adopted by CH, as modified by Coe et al. (2009) by including human capital on the right-hand side of the equation:

$$f_{it} = \exp(\alpha_i + e_{it}) \left(S_{it}^d\right)^\theta \left(S_{it}^f\right)^\gamma H_{it}^\delta \tag{1}$$

where f_{it} is the total factor productivity of country $i = 1, ..., N$ at time $t = 1, ..., T$; α_i are individual fixed effects that take into account unobserved time-invariant characteristics, which are allowed to be freely correlated with both R&D capital stocks (domestic, S_{it}^d, and foreign, S_{it}^f) and human capital (H_{it}); and e_{it} is the error term. The foreign capital stock S_{it}^f is defined as the weighted arithmetic mean of S_{jt}^d for $j \neq i$:

$$S_{it}^f = \sum_{j \neq i} \omega_{ij} S_{jt}^d \tag{2}$$

where w_{ij} represents the weighting scheme. The model is then linearised by taking logs:

$$\log f_{it} = \alpha_i + \theta \log S_{it}^d + \gamma \log \sum_{j \neq i} \omega_{ij} S_{jt}^d + \delta \log H_{it} + e_{it} \tag{3}$$

It is interesting to note that this simple empirical specification can be derived from an endogenous growth model (see, e.g., Keller, 2004, p. 762). However, as noted by Lichtenberg and Van Pottelsberghe (2001, p. 490), *International technological spillovers have no widely accepted measures"*. According to Keller (2004), the main channels of technology diffusion are trade, FDI and language skills. For instance, Coe et al. (2009) and Lichtenberg and Van Pottelsberghe (1998) use alternative definitions of w_{ij} based on imports, Lichtenberg and Van Pottelsberghe (2001) focus on FDI, and Musolesi (2007) adopts a weighting scheme that takes language skills into account. More recently, Spoalore and Warcziag (2009) suggest genetic distance as a barrier to the diffusion of development.

In this paper, we focus on geographical proximity as a channel for technology diffusion for many reasons. First, it is theoretically consistent. Keller (2002) and Eaton and Kortum (2002) show that international technology diffusion is related to geographical distance due to transport costs or geographical barriers. Second, the geographic localisation of international technology diffusion can have economically relevant implications. Specifically, it can affect the process of convergence across countries (Grossman and Helpman, 1991), the agglomeration that takes place in an economy (Krugman and Venables, 1995) and the long-term effectiveness of macroeconomic policies aimed at technological progress (Keller, 2002). Third, there have been far fewer studies on geographic international R&D spillovers than on spillovers via other channels, such as trade or FDI, in spite of the theoretical consistency and empirical relevance of geography. Finally, and perhaps most importantly in the context of this paper, which focuses on the methodological issue of cross-sectional dependence, traditional channels of international technology diffusion might create reverse causality problems when included in econometric specifications. For instance, a country's international trade, FDI or patent activity may depend on its technological level and, in turn, might be endogenous with respect to TFP (see, e.g., Hong and Sun, 2011). In contrast, geographical distance is generally considered exogenous, *"Global technology spillovers favor income convergence, and local spillovers tend to lead to divergence, no matter through which channel technology diffuses... An advantage of this is that geography is arguably exogenous in this process"* (Keller, 2004, p.772). Moreover, geographical distance may be considered an exogenous proxy for certain endogenous measures of socioeconomic, institutional, cultural or language-based similarities that might enhance the diffusion of technology. Following Keller (2002), we propose a specification of foreign R&D that incorporates the notion that the impact of foreign R&D is a decreasing function of geographical distance from foreign economies. Therefore, the foreign R&D capital stock for each country i is obtained by weighting the domestic R&D capital stocks of every other country $j \neq i$ in the sample using an exponential distance decay function, $\omega_{ij} = \exp(-\varphi d_{ij})$, such that

$$S^f_{it} = \sum_{j \neq i} \exp(-\varphi d_{ij}) S^d_{jt} \tag{4}$$

where d_{ij} represents the geographic distance between country i and country j. Finally, to construct the stock of human capital, we use the average number of years of schooling in the population over 25 years old. Following Hall and Jones (1999), this parameter is converted into a measure of human capital stock through the following formula:

$$H_{it} = \exp\left[g\left(Edu_{it}\right)\right] \tag{5}$$

where Edu_{it} is the average number of years of schooling and the function $g\left(Edu_{it}\right)$ reflects the efficiency of a unit of labor with Edu years of schooling relative to one with no schooling. Following Psacharopoulos (1994) and Caselli (2005), it is assumed that $g\left(Edu_{it}\right)$ is piecewise linear, which implies a log-(piecewise)linear relationship between H and Edu.[1] Combining equations (4) and (5):

$$\log f_{it} = \alpha_i + \theta \log S^d_{it} + \gamma \log \sum_{j \neq i} \exp(-\varphi d_{ij}) S^d_{jt} + \delta \log H_{it} + e_{it} \tag{6}$$

If there are positive geographical spillovers (if foreign R&D enhances domestic productivity, $\gamma > 0$), then a positive value of φ indicates that the impact of such spillovers decreases non-linearly with

[1] with slope 0.134 for $0 < Edu \leq 4$, 0.101 for $4 < Edu \leq 8$, and 0.068 for $Edu > 8$.

distance, whereas a negative value of φ suggests that the benefits of foreign R&D are increasing with distance. Finally, $\varphi = 0$ indicates that the impact of spillovers does not depend on the distance separating two countries. To allow the impact of foreign R&D capital to differ between the G7 countries and the others, the *benchmark* specification we adopt is a simple variant of equation (6) that has been widely used in the literature:

$$\log f_{it} = \alpha_i + \theta \log S_{it}^d + \gamma_{G7} \mathbf{1}_{G7} \log \sum_{j \neq i} \exp(-\varphi_{G7} d_{ij}) S_{jt}^d + \tag{7}$$

$$+ \gamma_{NOG7} \mathbf{1}_{NOG7} \log \sum_{j \neq i} \exp(-\varphi_{NOG7} d_{ij}) S_{jt}^d + \delta \log H_{it} + e_{it}$$

$$\text{with:} \mathbf{1}_{G7} = \begin{cases} 1 & \text{if country } \in \text{ G7 group} \\ 0 & \text{if country } \notin \text{ G7 group} \end{cases}, \text{ and: } \mathbf{1}_{NOG7} = 1 - \mathbf{1}_{G7}$$

In the following, for ease of exposition, we define:

Equation (7) can then be expressed as:

$$y_{it} = \alpha_i + \beta' \mathbf{x}_{it} + e_{it} \tag{8}$$

where:

$$y_{it} = \log f_{it}$$

$$\mathbf{x}_{it} = [\log(S_{it}^d), \mathbf{1}_{G7} \log \sum_{j \neq i} \exp(-\varphi_{G7} d_{ij}) S_{jt}^d, \mathbf{1}_{NOG7} \log \sum_{j \neq i} \exp(-\varphi_{NOG7} d_{ij}) S_{jt}^d, \log(H_{it})]'$$

$$\beta = [\theta, \gamma_{G7}, \gamma_{NOG7}, \delta]'$$

Because one of our main objectives is the comparison of the results with previous studies on international R&D spillovers, which do not consider the issue of cross-sectional correlation, our main source is the CH data set, which has been widely used in the literature (see Table 1). This data set is a balanced panel of 21 OECD countries plus Israel observed over the period 1971-90. Our measures of TFP and domestic R&D capital stock come from this data source. The average number of years of schooling used to construct our measure of human capital is taken from Barro and Lee (2001), as in Coe et al. (2009). Finally, the distance between two countries is calculated as the spherical distance between capitals.

3 Strong and weak cross-sectional correlation

In order to introduce cross-sectional correlation in the benchmark specification defined by equation (8), following Sarafidis (2009) and Sarafidis and Wansbeek (2012), a *general* error structure can be considered:

$$e_{it} = (\varrho_i \odot \mathbf{w}_i)' \xi_t + \varepsilon_{it} = \sum_{j=1}^{m} \varrho_{ij} w_{ij} \xi_{jt} + \varepsilon_{it} \tag{9}$$

where $\odot$ denotes the Hadamard product, $\xi_t = (\xi_{1t}, \xi_{2t}, ..., \xi_{mt})'$ is a $m \times 1$ vector of unobserved common factors, $\varrho_i = (\varrho_{i1}, \varrho_{i2}, ..., \varrho_{im})'$ is a $m \times 1$ vector of factor loadings, $\mathbf{w}_i = (w_{i1}, w_{i2}, ..., w_{im})'$ is a $m \times 1$ vector of deterministic bounded weights and $\varepsilon_{it} \sim i.i.d.\,(0, \sigma_\varepsilon^2)$. Equation (9) allows us to regard the unobserved components, ξ_t as shocks, the impact of which are a combination of heterogeneous factor loadings (ϱ_i) with a weight scheme ($\mathbf{w}_i$).

Setting $\mathbf{w}_i = \iota$, ι as a vector of ones, equation (9) boils down to the multifactor error structure:

$$e_{it} = \varrho_i^{'}\xi_t + \varepsilon_{it} = \sum_{j=1}^{m} \varrho_{ij}\xi_{jt} + \varepsilon_{it} \tag{10}$$

With $m = N$, the N common factor model (Chudick et al., 2011) is obtained:

$$e_{it} = \sum_{j=1}^{N} \varrho_{ij}\xi_{jt} + \varepsilon_{it}. \tag{11}$$

In matrix form, stacking over individuals, we obtain the following:

$$\boldsymbol{e}_t = \mathbf{P}\xi_t + \boldsymbol{\varepsilon}_t \tag{12}$$

where $\xi_t = \xi_t = (\xi_{1t}, \xi_{2t}, ..., \xi_{Nt})'$ is a $N \times 1$ vector of unobserved factors, $\mathbf{P}$ is a $N \times N$ matrix of associated factor loadings with typical element $\{\varrho_{ij}\}$ and $\boldsymbol{\varepsilon}_t = i.i.d.\,(0, \sigma_\varepsilon^2 \mathbf{I}_N)$.

Imposing appropriate zero restrictions on $\mathbf{w}_i$, at least $w_{ij} = 0$ for $i = j$, homogeneity restrictions on $\varrho_i = \lambda, \forall i$, and setting $\xi_{jt} = e_{jt}$, for $j = 1, ..., N$, with $m = N$, equation (9) boils down to a spatial autoregressive error specification:

$$e_{it} = \lambda \sum_{j=1}^{N} w_{ij} e_{jt} + \varepsilon_{it} \tag{13}$$

that can be rewritten in matrix form as follows:[2]

$$\boldsymbol{e}_t = \lambda \mathbf{W}_N \boldsymbol{e}_t + \boldsymbol{\varepsilon}_t \tag{14}$$

where $\boldsymbol{W}_N$ is defined in the spatial econometrics litterature as an $N \times N$ interaction or spatial weight matrix. It is most of the time not derived from theory, but exogeneously given as to reflect the interaction pattern connecting individuals, which is considered to be time invariant.[3] Its elements are non stochastic, non negative and finite.[4] Under the invertibility condition of $(\boldsymbol{I}_N - \lambda \boldsymbol{W}_N)$, we obtain:

$$\boldsymbol{e}_t = \mathbf{R}_N \boldsymbol{\varepsilon}_t \tag{15}$$

[2]It is worth noting that equation (9) also contains the spatial moving average process, i.e., $e_{it} = \lambda \sum_{j \neq i} w_{ij}\varepsilon_{jt} + \varepsilon_{it}$, and the spatial error component process, $e_{it} = \lambda \sum_{j \neq i} w_{ij}\varsigma_{jt} + \varepsilon_{it}$ where $E(\varsigma_{jt}) = 0$, $V(\varsigma_{jt}) = \sigma_\varsigma^2$ and $E(\varsigma_{jt}\varepsilon_{it}) = 0$ as special cases.

[3]An exception is Lee and Yu (2012).

[4]An exception where it is theoretically defined is for instance Behrens et al. (2010).

where $\mathbf{R}_N = (\boldsymbol{I}_N - \lambda \boldsymbol{W}_N)^{-1}$. Note that $\boldsymbol{W}_N$ and $\mathbf{R}_N$ must satisfy some regularity conditions given for instance by Lee (2004) for QML estimation.[5] Those conditions mainly require that $\boldsymbol{W}_N$ and $\mathbf{R}_N$ be uniformly bounded in row and column sums both at the true value of the λ parameter and uniformly in λ in a compact parameter space Λ. The true value of the λ parameter being in the interior of Λ.[6]

It may be useful to understand how these two approaches are related to the concepts of weak and strong cross-sectional dependence recently developed in the literature. Forni and Lippi (2001) introduce the notion of an idiosyncratic process to characterise a weak form of dependence that involves both time series and cross-sectional dimensions. More recently, Chudick et al. (2011) (henceforth CPT) propose a new and more widely applicable definition. They consider the asymptotic behavior of weighted averages at each point in time and define a process $\{z_{it}\}$ to be cross-sectionally weakly dependent at a given point in time if its weighted average at that time, conditional on the information set available in the previous period, $\Im_{t-1}$, converges to its expectation in quadratic mean, as the cross-sectional dimension is increased without bounds for all weights, w, that satisfy certain granularity conditions ensuring that the weights are not dominated by a few individuals,[7] that is, $\lim_{N\to\infty} Var\left(\sum_{i=1}^{N} w_{ij} z_{it} \mid \Im_{t-1}\right) = 0$. Another definition has been recently proposed by Sarafidis (2009) (henceforth, SARA), who defines a process $\{z_{it}\}$ to be cross-sectionally weakly dependent if $\sum_{j\neq i} |Cov\left(z_{it}, z_{js}|F_{ij}\right)| < \infty$, where F_{ij} denotes the conditioning set of all time-invariant characteristics of individuals i and j.[8]

Spatial error models satisfy, under a standard set of regularity conditions, weak dependence under both definitions. For example, the standard uniform boundedness conditions given by Lee (2004) are sufficient but not necessary to guarantee weak dependence.[9]

Conversely, the factor approach entails strong dependence under both definitions unless further restrictions are imposed on the factor loadings. To see this relation, consider the single factor error process $e_{it} = \varrho_i \xi_t + \varepsilon_{it}$. With N and T both large and noticing that $\sigma_{ij,t} = cov\left(e_{it}, e_{jt}|F_{ij}\right) = \varrho_i \varrho_j \sigma_\xi^2 \neq 0$; therefore, $\sum_{j\neq i} |\sigma_{ij,t}|$ is unbounded, thus entailing strong correlation under SARA. A related concept is that of strong and weak factors (Chudik et al. 2011). Let b be a constant in the range $0 \leq b \leq 1$, and consider the condition $\lim_{N\to\infty} N^{-b} \sum_{i=1}^{N} |\varrho_i| = K < \infty$. According to Chudik et al. (2011), the strong and weak factors correspond to $b = 1$ and $b = 0$, respectively. For $b \in (0, 1)$, the factor ξ_t is said to be semi-strong $(1/2 \leq b < 1)$ or semi-weak $(0 < b < 1/2)$. Thus, $b = 0$ implies that the factor affects only a fixed number of cross-sectional units, whereas $b < 1$ means that the subset of cross-sectional units affected by the factor grows more slowly than N at a rate depending on b. Under CPT, if there exists at least one strong factor, the underlying

[5] Those assumptions are all originated in Kelejian and Prucha (1998, 1999, 2001).

[6] A slightly different set is given for instance by Kelejian and Prucha (2010) for GMM estimation along with a detailed discussion of the parameter space.

[7] The definition of idiosyncratic process advanced by Forni and Lippi (2001) and the definition by CPT of weak dependence differ in the way the weights used to construct weighted averages are defined.

[8] Robinson (2011) and Robinson and Thawornkaiwong (2012) give a very similar definition.

[9] See Sarafidis and Wansbeek (2012) for a more detailed discussion.

process is strongly cross-sectionally dependent; otherwise, it is cross-sectionally weakly dependent. As also noted by Chudik et al. (2011), the CCE approach by Pesaran (2006) explicitly introduces a finite number of *strong* factors according to their definition of *strong* and *weak* factors. Thus, it entails strong dependence under both SARA and CPT.

3.1 Strong correlation: errors with multifactor structure

The empirical setup adopted in this paper builds on the framework originally proposed by Pesaran (2006) and further developed and studied more recently (Chudik et al. 2011; Pesaran and Tosetti, 2011; Kapetanios et al. 2011). Such a framework has a number of appealing features. It is sufficiently general to render a variety of panel data models as special cases, it allows correlated common factors, it does not require specifying the number of factors, and it is computationally very simple.[10]

Let us consider the following DGP:

$$y_{it} = \alpha_i^{'}\mathbf{d}_t + \beta_i^{'}\mathbf{x}_{it} + e_{it} \tag{16}$$

where $\mathbf{d}_t$ is a $l \times 1$ vector of observed common effects, $\alpha_i^{'}$ is the associated vector of parameters and $\mathbf{x}_{it}$ is a 4×1 vector of explanatory variables. The 'one-way' specification is simply obtained by setting $\mathbf{d}_t = 1$. The slope coefficients $\beta_i' = [\theta, \gamma_{G7}, \gamma_{NOG7}, \delta]$ can be assumed to be fixed and homogeneous across countries, $\beta_i' = \beta' \; \forall \; i$, or assumed to follow a random coefficients specification: $\beta_i = \beta + \mathbf{v}_i$, $\mathbf{v}_i \sim IID\left(\mathbf{0}, \Theta_v\right)$. The errors e_{it} are assumed to have a multifactor structure as in equation (10):

$$e_{it} = \varrho_i^{'}\xi_t + \varepsilon_{it} \tag{17}$$

where ξ_t is a $m \times 1$ vector of unobserved common factors with country-specific factor loadings ϱ_i. Combining (16) with (17), we thus obtain the following:

$$y_{it} = \alpha_i^{'}\mathbf{d}_t + \beta_i^{'}\mathbf{x}_{it} + \varrho_i^{'}\xi_t + \varepsilon_{it} \tag{18}$$

where the idiosyncratic errors, ε_{it}, are assumed to be independently distributed over $(\mathbf{d}_t, \mathbf{x}_{it})$, whereas the unobserved factors, ξ_t, can be correlated with $(\mathbf{d}_t, \mathbf{x}_{it})$. This correlation is allowed by

[10] In the macro panel data literature, the standard approach to addressing cross-sectional correlation has been to adopt a seemingly unrelated regressions (SURE) framework and estimate that system of equations by generalised least squares. The SURE approach, however, is not applicable if the panel has a large cross-sectional dimension because it involves nuisance parameters that increase at a quadratic rate as the cross-sectional dimension of the panel is allowed to rise. Moreover, an often questionable assumption behind this approach is that the errors are uncorrelated with the regressors. This has led to the consideration of unobserved factor models. Coakley et al. (2002) propose a principal component approach requiring, however, that the unobserved factors be uncorrelated with the explanatory variables to be consistent. Other estimators based on principal component analysis have been proposed by Kapetanios and Pesaran (2007) and Bai (2009). For a fixed time dimension, however, both are inconsistent under serial correlation or heteroskedasticity (see, also, Sarafidis and Wansbeek, 2012).

modeling the explanatory variables as linear functions of the observed common factors $\mathbf{d}_t$ and the unobserved common factors ξ_t:

$$\mathbf{x}_{it} = \mathbf{A}_i'\mathbf{d}_t + \mathbf{\Gamma}_i'\xi_t + \mathbf{v}_{it} \tag{19}$$

where $\mathbf{A}_i$ and $\mathbf{\Gamma}_i$ are $l \times 4$ and $m \times 4$ factor loadings matrices and $\mathbf{v}_{it} = (v_{i1t}, v_{i2t}, , v_{i3t}, v_{i4t})'$. Combining (18) and (19), we finally obtain a system of equations simultaneously explaining TFP, R&D (domestic and foreign) and human capital:

$$\underset{5\times1}{\mathbf{z}_{it}} = \begin{pmatrix} y_{it} \\ \mathbf{x}_{it} \end{pmatrix} = \begin{pmatrix} \log\left(f_{it}\right) \\ \log(S^d_{it}) \\ \mathbf{1}_{G7}\log\sum_{j\neq i}\exp(-\varphi_{G7}d_{ij})S^d_{jt} \\ \mathbf{1}_{NOG7}\log\sum_{j\neq i}\exp(-\varphi_{NOG7}d_{ij})S^d_{jt} \\ \log(H_{it}) \end{pmatrix} = \underset{5\times l}{\mathbf{B}_i'}\underset{l\times1}{\mathbf{d}_t} + \underset{5\times m}{\mathbf{C}_i'}\underset{m\times1}{\xi_t} + \underset{5\times1}{\mathbf{u}_{it}}, \tag{20}$$

where:

$$\mathbf{u}_{it} = \begin{pmatrix} \mathbf{1} & \beta_i' \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix}\begin{pmatrix} \varepsilon_{it} \\ \mathbf{v}_{it} \end{pmatrix} = \begin{pmatrix} \varepsilon_{it} + \beta_i'\mathbf{v}_{it} \\ \mathbf{v}_{it} \end{pmatrix},$$

$$\mathbf{B}_i = \begin{pmatrix} \alpha_i & \mathbf{A}_i \end{pmatrix}\begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \beta_i & \mathbf{I}_k \end{pmatrix},$$

$$\mathbf{C}_i = \begin{pmatrix} \varrho_i & \mathbf{\Gamma}_i \end{pmatrix}\begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \beta_i & \mathbf{I}_k \end{pmatrix},$$

where $\mathbf{I}_k$ is an identity matrix of order k. In our specific case, $k = 4$.

Under the restrictive assumption of homogeneous factor loadings and homogeneous slope parameters, it is possible to consistently estimate the slope coefficients by ordinary least squares (OLS): one may adopt either a two-way fixed-effects model or a first difference specification augmented with time dummies (Eberhardt and Bond, 2009). To consistently generalise the model to heterogeneous factor loadings, we adopt the Pesaran CCE (2006) approach, which solves the identification problem by augmenting the regression with proxies for the unobserved factors. Pesaran suggests using $[\mathbf{d}_t' \quad \overline{\mathbf{z}}_{wt}']$ as observable proxies for the unobserved factors, and $\overline{\mathbf{z}}_{wt}$ indicates the cross-sectional average: $\overline{\mathbf{z}}_{wt} = \sum_{j=1}^{N} w_j\mathbf{z}_{jt}$, w_j are weights equal to $1/N$. The individual slopes β_i' or their mean can be consistently estimated by regressing y_{it} on $\mathbf{x}_{it}, \mathbf{d}_t$ and $\overline{\mathbf{z}}_{wt}$. This type of estimator is referred to as a common correlated effect estimator. In particular, Pesaran (2006) proposes two estimators of the individual coefficients' mean, β : the CCE pooled estimator (CCEP) and the Mean Group estimator known as CCEMG, which is obtained by averaging the country-specific estimates following Pesaran and Smith (1995), which also allows the slope parameters to differ across cross-sections. As an alternative to the CCEMG, Eberhardt and Bond (2009) and Eberhardt and Teal (2010) propose the Augmented Mean Group estimator, where the Mean Group group-specific regressions are augmented with a preliminary OLS estimate of a "common dynamic process". Note, however, that whereas in Eberhardt and Teal' s (2010) production function framework, such a common dynamic process represents the estimated cross-sectional average of the unobservable TFP (the "residual"), in our empirical setup, where the dependent variable is TFP itself, the common dynamic process does not seem to have a straightforward interpretation. In this paper, we focus on the CCEP estimator (and assume $\beta_i' = \beta' \ \forall \ i$).

Some remarks are in order. First, and very important, this set-up introduces endogeneity, whereby the $\mathbf{x}_{it}$ are correlated with the unobservable e_{it} via the correlation between ξ_t and $\mathbf{x}_{it}$. As noted by Kapetanios et al. (2011), standard approaches that neglect common factors fail to identify β_i'; instead, they yield an estimate of:

$$\kappa_i' = \beta_i' + \varrho_i' \mathbf{\Gamma}_i'^{-1}. \tag{21}$$

The estimation bias is thus function of the factors loadings ϱ_i' and $\mathbf{\Gamma}_i'$ only.[11] Second, specifying a factor-loadings matrix $\mathbf{C}_i$ of the kind presented above permits a variety of situations, as each variable is allowed to be affected in a specific way by each factor because the typical element of such a matrix, say c_{imj}, measures the country-specific effect (eventually being zero) of the m^{th} common factor on the j^{th} variable. For example, it may allow some of the unobserved common factors driving the evolution of TFP to also drive the variation in R&D and human capital stocks. Such factors may be linked to oil price shocks or global policies aimed at raising the level of technology, for example. It could also allow other factors to specifically affect only one variable in the system.[12]

It is finally worth recalling some recent results concerning the validity of the CCE approach when the underlying DGP is also characterised by weak factors or spatial error correlation. Chudik et al. (2011) also extended the CCE approach by allowing for the presence of both a limited number of strong factors and a large number of weak or semi-strong factors and then show that, even under this extended framework, the CCE method still provides consistent estimates of the slope coefficients. Pesaran and Tosetti (2011) prove that the CCE estimator provides consistent estimates of the slope coefficients and their standard errors under the more general case of a multi-factor error structure and spatial error correlation (see, also, Bresson and Hsiao, 2011, for further simulation results), i.e. when both forms of cross-correlation – weak and strong – characterise the DGP:

$$e_{it} = \varrho_i' \xi_t + \lambda \sum_{j \neq i} w_{ij} e_{jt} + \varepsilon_{it}. \tag{22}$$

[11] To see how this may occur, let us rewrite the model for y_{it} as in Kapetanios et al. (2011) equation (52): abstracting from $\mathbf{d}_t$, assuming that k (the number of regressors)= m(the number of common unobserved factors) and that $\mathbf{\Gamma}_i$ is invertible, we can write the following: $y_{it} = \beta_i' \mathbf{x}_{it} + \varrho_i' \mathbf{\Gamma}_i'^{-1} (\mathbf{x}_{it} - \mathbf{v}_{it}) + \varepsilon_{it} = \kappa_i' \mathbf{x}_{it} + \varkappa_{it}$, where $\kappa_i' = \beta_i' + \varrho_i' \mathbf{\Gamma}_i'^{-1}$ and $\varkappa_{it} = \varepsilon_{it} - \varrho_i' \mathbf{\Gamma}_i'^{-1} \mathbf{v}_{it}$. Therefore, applying least squares to such an equation consistently estimates κ_i' rather than β_i'.

[12] To make this feature more apparent, Eberhardt and Teal (2010) adopt a scalar notation and replace equation (19) with $x_{kit} = \pi_{ki}' \mathbf{d}_{kt} + \delta_{ki}' \mathbf{g}_{kt} + \vartheta_{1ki} \xi_{1kt} + ... + \vartheta_{lki} \xi_{lkt} + \omega_{kit}$, where $k = 1, ..., 3$ and $\mathbf{g}_{kt}$ are common factors that are specific to each regressor.

This appears as a very appealing feature since it could be that both forms of dependence are present in the data as shown for instance by Bailey et al. (2013b).

3.2 Weak correlation: errors with spatial autocorrelation

We consider the following panel model with spatially autocorrelated error terms:

$$\begin{aligned} y_{it} &= \alpha_i + \beta' \mathbf{x}_{it} + e_{it} \\ e_{it} &= \lambda \sum_{j \neq i} w_{ij} e_{jt} + \varepsilon_{it} \end{aligned} \tag{23}$$

where λ is the spatial autoregressive parameter. For $\lambda = 0$, equation (21) simply reduces to the baseline *a-spatial* specification (8). To obtain a better understanding of such a spatial process, it is useful to examine the so-called *reduced form.* In matrix form, stacking over all individuals for time period t, we have the following:

$$\boldsymbol{y}_t = \alpha + \boldsymbol{X}_t \beta + \mathbf{e}_t \tag{24}$$
$$\boldsymbol{e}_t = \lambda \mathbf{W}_N \boldsymbol{e}_t + \boldsymbol{\varepsilon}_t \qquad t = 1, ..., T$$

where $\boldsymbol{y}_t$ represents the $N \times 1$ vector of log TFP, $\boldsymbol{X}_t$ is the $N \times 4$ matrix of explanatory variables and $\boldsymbol{W}_N$ is an $N \times N$ row-normalised interaction matrix.[13] Under the invertibility condition of $(\boldsymbol{I}_N - \varrho \boldsymbol{W}_N)$, equation (24) can be rewritten in its *reduced form* representation::

$$\boldsymbol{y}_t = \alpha + \boldsymbol{X}_t \beta + (\boldsymbol{I}_N - \lambda \boldsymbol{W}_N)^{-1} \boldsymbol{\varepsilon}_t \tag{25}$$

where $(\boldsymbol{I}_N - \lambda \boldsymbol{W}_N)^{-1}$ is the so-called *global* spatial multiplier. This reduced form has the following important implications. First, in the conditional mean, the total factor productivity in country i will only be affected by the domestic R&D capital stock or human capital stock in the same country i and not by those in any other country j, exactly as in the standard *a-spatial* panel data model. *Therefore there are no spatial spillover effects in this model.*[14] Second, and more specifically, one can easily see that a random shock due to unobservable factors (i.e., a shock in the disturbances) in a specific country i not only affects TFP in country i, but it also has an impact on TFP in all the countries of the sample through the inverse spatial transformation $(\boldsymbol{I}_N - \lambda \boldsymbol{W}_N)^{-1}$: this is the so-called *global spatial diffusion process* of a random shock, which can be expressed as follows:[15]

$$\boldsymbol{\Xi}_y^{\varepsilon} \equiv \frac{\partial \boldsymbol{y}_t}{\partial \boldsymbol{\varepsilon}_t'} \equiv (\boldsymbol{I}_N - \lambda \boldsymbol{W}_N)^{-1} = \boldsymbol{I}_N + \lambda \boldsymbol{W}_N + \lambda^2 \boldsymbol{W}_N^2 + \lambda^3 \boldsymbol{W}_N^3 + ... \tag{26}$$

[13]According to Lee and Yu (2010), it allows us to consider the parameter space for λ to be a compact subset of $(-1, 1)$. Row-normalization also facilitates the interpretation of the results but is not theoretically necessary (Kelejian and Prucha, 2010).

[14]In contrast, such spatial spillovers exist in the spatially lagged endogenous variable model, sometimes referred to as the mixed regressive spatial autoregressive model, or the SAR model in the spatial econometric literature.

[15]Note that the use of the term diffusion only refers here to the spatial dimension, not to the space-time dimension, and therefore might be misleading.

where Ξ_y^ε is a $N \times N$ matrix of the partial derivatives of $\boldsymbol{y}_t$ with respect to the disturbance $\boldsymbol{\varepsilon}_t$. This matrix is in general full and not symmetric whatever the sparsity and structure of the interaction matrix $\mathbf{W}_n$.

Therefore, in this model, *a random shock in a given country will spill over the entire sample* and will have a *global* impact. It should be clear that the impact of a random shock hitting a given country is by no means "local" in such a model. The diagonal elements of this matrix represent the *direct impacts* of random shocks affecting each of the countries of the sample, including *feedback* effects, which are inherently heterogeneous in the presence of spatial autocorrelation due to differentiated interaction terms in the $\boldsymbol{W}_N$ matrix.[16] This type of heterogeneity is called *interactive heterogeneity*, in opposition to standard individual heterogeneity in panel data models (Debarsy and Ertur, 2010). The off-diagonal elements of this matrix represent the *indirect impacts* of random shocks.[17] Considering column j of the impact matrix Ξ_y^ε, a random shock in a given country j will differently affect each of the countries of the sample. It represents the emission side of the spatial diffusion process. Considering row i, random shocks in each of all the countries of the sample will each differently affect country i even if they are common or identical. It represents the reception side of the spatial diffusion process. Using obvious notations, we have the following:

$$\frac{\partial \mathbf{y}_{t,i}}{\partial \varepsilon_{t,i}} \equiv (\Xi_{\mathbf{y}}^\varepsilon)_{t,ii} \qquad \text{and} \qquad \frac{\partial \mathbf{y}_{t,i}}{\partial \varepsilon_{t,j}} \equiv (\Xi_{\mathbf{y}}^\varepsilon)_{t,ij} \tag{27}$$

The magnitude of those direct and indirect impacts of random shocks will depend on (1) the degree of interaction between countries, which is governed by the $\boldsymbol{W}_N$ matrix, and (2) the parameter λ, measuring the strength of interactions or cross-sectional correlation between countries. Note that feedback effects are at best of second order and may die out rather quickly, as can be easily seen in equation (26), whereas indirect impacts, although presumably small in magnitude, should not be a priori neglected.

A possible method for estimating spatial panel econometric models consists of using the direct ML approach. However, the direct ML approach may suffer from the incidental parameter problem discussed in Neyman and Scott (1948), who illustrate the inconsistency of the variance parameter for the "a-spatial" linear panel data model when the time dimension is finite. Focusing on spatial panel models, Lee and Yu (2010) show that the direct ML provides consistent estimates of regressor coefficients. However, the direct ML provides inconsistent estimates of the variance parameter when T is finite. Thus, Lee and Yu (2010) propose a consistent QML approach based on a data transformation that eliminates the individual fixed effects. They also demonstrate that except for the variance parameter, the estimates of the direct approach are identical to the corresponding estimates of the transformation approach.

[16]More specifically, the own derivative for country i includes the *feedback* effects, where country i affects country j and country j also affects country i, and longer paths that might go from country i to j to k and back to i.

[17]LeSage and Pace (2009) present a comprehensive analysis of those effects along with some useful summary measures in the cross-section setting. Their extension to our panel data setting is straightforward. See, also, Kelejian et al. (2006, 2008) for other applications.

4 Results

4.1 Preliminary estimation results: the baseline specification

Table 2 summarises the results obtained by estimating the benchmark specifications presented above. Column (i) shows the estimated parameters from the specification in equation (6), where the model is estimated using Nonlinear Least Squares as in Keller (2002). The output elasticity of domestic R&D capital stock, θ, is estimated to be 0.067 and is statistically significant. This result is in line with the related empirical literature, such as Coe and Helpman (1995), Coe et al. (2009), Lichtenberg and Van Pottelsberghe (2001) and Keller (2002). The estimated coefficient of human capital (δ) is highly significant and of the same order of magnitude as that found by Coe et al. (2009). The inclusion of human capital is relevant not only because it affects productivity and the ability of firms to absorb information but also because it is potentially correlated with R&D; hence, estimating the model without human capital should bias the coefficient associated with R&D upward. In some previous studies (Barrio-Castro et al., 2002; Frantzen, 2000; Engelbrecht, 1997), this bias has been estimated to be approximately 20% to 30%. Next, we focus on the output elasticity of foreign R&D capital stock incorporated into the geographical technology transfer channel and how the effectiveness of such spillovers decreases with distance (i.e., we focus on the parameters γ and φ). The output elasticity of foreign R&D capital stock incorporated into the geographical technology transfer channel (γ) is estimated to be 0.042 and is significant at the 1% level. In other words, we find evidence of positive (but small in magnitude) geographical spillovers across countries. It is interesting to compare this result with those obtained using alternative technology transfer channels. There is a large body of literature focusing on trade and FDI that generally finds a larger point estimate but presents conflicting results regarding statistical significance (Table 1). Conversely, analyses of technology diffusion via language skills are rare. Musolesi (2007) finds a significant and quite high estimate (approximately 0.2) for the coefficient associated with foreign R&D incorporated into language skills. The positive estimate of φ suggests that the impact of such spillovers decreases with distance. This result is consistent with Bottazzi and Peri (2003), who find that R&D spillovers are small in magnitude and highly localised in European regions. Finally, we turn to the estimates of the benchmark specification in equation (8), which allows the output elasticity with respect to foreign R&D to differ between large and small countries (Table 2, column ii), and focus on the parameters $\gamma_{G7}, \gamma_{NOG7}, \varphi_{G7}$ and φ_{NOG7}. Such a specification will be extended in the following sections to accommodate cross-sectional dependence. Clearly, both elasticities are significant, and the effect of foreign R&D on TFP is much higher for G7 than for non-G7 countries ($\hat{\gamma}_{G7} = 0.170, \hat{\gamma}_{NOG7} = 0.026$). This result is similar to that of Lichtenberg and Van Pottelsberghe (2001), who focus on FDI spillovers. We also find that the effectiveness of such spillovers decreases with distance more quickly for G7 than for non-G7 countries ($\hat{\varphi}_{G7} > \hat{\varphi}_{NOG7}$). In other words, the spillovers are more localised for G7 countries than for smaller countries. These results suggest the existence of substantial differences between richer and poorer countries in terms of how effective they are in adopting foreign technology. Richer countries are, according to our results, better at adopting foreign technology than poorer countries. This pattern can be seen as consistent with the existence of a minimum level of absorptive capacity allowing a country to benefit from foreign technology (see, e.g., Xu, 2000) and theories describing how technology that is invented in frontier countries is less appropriate for poorer countries (e.g., Basu and Weil, 1998).

4.2 Testing for cross-sectional correlation

A widely adopted test, likely due to its useful small-sample properties, is the CD test developed by Pesaran (2004). An interesting feature of this test is that, as shown by Pesaran (2012a), the implicit null hypothesis of the CD test is, in the most common cases, weak cross-sectional correlation. This assumption makes the test more appealing from an applied perspective because when estimating a model, only strong cross-sectional correlation may pose serious problems. More precisely, let us define ϵ as a measure of the degree to which T expands relative to N, as defined by $T = O\left(N^{\epsilon}\right)$ for $0 < \epsilon \leq 1$ and a being the exponent of cross-sectional correlation introduced in Bailey et al. (2013), which can take any value in the range $[0, 1]$, with 1 indicating the highest degree of cross-sectional dependence, while $a < 0.5$ and $a > 0.5$ correspond to the cases of weak and strong cross-sectional correlation discussed in CPT, respectively. Pesaran (2012a) shows that the implicit null of the CD test is given by $0 \leq a < (2 - \epsilon)/4$. Thus, for ϵ close to zero (T almost fixed as $N \to \infty$), such a null hypothesis is $0 \leq a < 1/2$, whereas in the case where $\epsilon = 1$ (N and $T \to \infty$ at the same rate, as is roughly the case of the data used in this paper), the implicit null of the CD test is given by $0 \leq a < 1/4$. The CD test has been performed on the residuals of the benchmark specification (equation (8)) (Table 3). We also performed the CD test for all the variables taken separately. The result of this test is a strong rejection of the null hypothesis in all cases, suggesting that the exponent of cross-sectional correlation, a, is in the range $[1/4, 1]$.[18]

In order to discriminate between these two typologies of correlation and to obtain a measure of the degree of such correlation, we adopt the method recently proposed by Bailey et al. (2013a) and compute the bias corrected version of a for all the variables under study. With the exception of Bailey et al. (2013a), this paper is, to the best of our knowledge, the first one providing fresh empirical evidence on a for some relevant macroeconomic variables. Following Bailey et al. (2013a) the Holm's approach has been preferred over the Bonferroni procedure. The exponent of cross sectional correlation a is estimated at about 0.999 for all the variables. This result concerns the version of the estimator which is robust to serial correlation in the factors and weak cross-sectional dependence in the error terms. Moreover the non-robust version of the estimator provides very similar results with a ranging from 0.998 to 0.999. This is a very clear-cut result not only indicating

[18] The full correlation matrix of residuals is available upon request; the average absolute value of the its off-diagonal elements is 0.470. Table 3 also reports the results of other tests based on the pair-wise correlation coefficients and precisely provides the tests by Frees (1995) and Friedman (1937). Both are discussed in Sarafidis and Wansbeek (2012). Both tests strongly reject the null hypothesis of cross-sectional independence.

the presence of strong cross-sectional correlation but also being consistent with the factor literature typically assuming $a = 1$ (Stock and Watson, 2002; Bai and Ng, 2002). A first implication of this result is that in our empirical framework almost all the existing cross sectional correlation can be modeled with common factors while other spatial linkages plays a very marginal role. The second implication is that we do not encounter the problems arising when the assumption that $a = 1$ fails and which are discussed in Chudick et al. (2011), Kapetanios and Marcellino (2010) and Onatski (2012).

4.3 Testing for unit roots

We start by considering some first-generation tests, notably the test proposed by Im, Pesaran and Shin (2003) (IPS) and the Fisher-type tests introduced by Maddala and Wu (1999) and further developed by Choi (2001). These are ADF-type tests where the non-stationarity null hypothesis that the coefficient associated with the autoregressive term is zero for all cross-sections is tested against the alternative that such a coefficient is negative for some cross-sections and is zero for others (see also Pesaran, 2012b). We then move to second-generation tests allowing for cross-sectional dependence. We both consider augmented ADF-type specifications (Pesaran, 2007; Pesaran et al., 2013) and tests decomposing the panel into deterministic, common and idiosyncratic components (Bai and Ng, 2004; Moon and Perron, 2004; Choi, 2006). We report here the main conclusions of our analysis. While most of the previous works find evidence of nonstationary variables by applying first-generation tests (see among others, Coe et al. 2009), we provide a more nuanced and thorough picture. First, focusing on first-generation tests, we documented that when the number of lags of the autoregressive component of heterogeneous ADF-type specifications is estimated in a model selection framework, the null of nonstationarity for all the cross-sectional units is rejected, suggesting that for some countries the variables are, in fact, stationary. Secondly, adopting second generation tests allows us to provide further and interesting insights. On one hand, indeed, the use of augmented ADF-type specifications (Pesaran, 2007; Pesaran et al., 2013), which are built ruling out the possibility of the factors having unit roots, confirms the previous findings and goes towards the rejection of the null hypothesis. On the other hand, the adoption of tests decomposing the panel into deterministic, common and idiosyncratic components (Bai and Ng, 2004; Moon and Perron, 2004; Choi, 2006) suggest that while the unobserved idiosyncratic component of the variables under study is stationary, the unobserved common factors seem to be be nonstationary. The main results focusing on second generation tests (Bai and Ng, 2004; Moon and Perron, 2004; Pesaran, 2007; Pesaran et al., 2013) are provided in tables 4, 5 and 6 while a more detailed discussion and all the results are available in the online appendix 1.

These results are very relevant for an empirical perspective and are related to the recent work by Kapetanios et al. (2011). They partitioned the vector of observed common factors as $\mathbf{d}_t = \left(\mathbf{d}_{1t}^{'}, \mathbf{d}_{2t}^{'}\right)^{'}$ where $\mathbf{d}_{1t}$ is an $l_1 \times 1$ vector of determinstic components and $\mathbf{d}_{2t}$ is an $l_2 \times 1$ vector of unit root observed common factors, with $l_1 + l_2 = l$ and then suppose that the $(l_2 + m) \times 1$

vector of stocastic common effects $\mathbf{h}_t = \left(\mathbf{d}_{2t}', \xi_t'\right)'$ follows a multivariate unit root process. Both analytical results and a simulation study indicate that the CCE approach is still valid when the unobserved factors are allowed to follow unit root processes.

4.4 CCEP estimation results

Next, we adopt the CCEP approach. Our estimates (in Table 2) are obtained using alternative definitions of the vector $\mathbf{d}_t$ of observed common factors. The results are structured as follows. In column (iii), as in Mastromarco et al. (2012), a linear time trend is used as an observed common factor, such that $\mathbf{d}_t = (\iota, t)'$. In column (iv), more flexibility is allowed by adding a squared time trend, such as $\mathbf{d}_t = (\iota, t, t^2)'$. Finally, in columns (v) and (vi), the oil price, denoted p, is added to the previous two specifications such that $\mathbf{d}_t = (\iota, t, p)'$ and $\mathbf{d}_t = (\iota, t, t^2, p)'$, respectively.

The results provide a robust picture. The coefficient associated with domestic R&D is always very close to zero, ranging from -.0043 to 0.026, and is never significant at standard levels. In other words, both the estimated coefficient of domestic R&D and its significance level have decreased substantially with respect to the benchmark specification. Conversely, adopting the CCEP method increases the effect of external R&D. For G7 countries, the effect of foreign R&D on TFP ranges from 0.22 to 0.46, which is statistically significant at least at 5% in all cases, whereas for the other countries, the effect ranges from 0.12 to 0.34 but is only statistically significant for the specification in column (vi). Finally, human capital is found to be insignificant at standard levels, although the estimated coefficient ranges between 0.19 and 0.37.[19]

These results are meaningful with respect to both the magnitude of the estimated coefficients and their significance levels. The very low (and not significant) estimate of the coefficient associated with domestic R&D complements Eberhardt et al. (2012), who estimate a knowledge capital production function *à la* Griliches - i.e. a production function augmented with domestic R&D – at industry level, and find that when unobserved common factors are introduced, the effect of R&D is close to zero and no longer significant. The high value of the estimated coefficient for foreign R&D is another result what seems interesting to us. Indeed, our results indicate not only that unobserved common factors matter for enhancing productivity growth, and that their inclusion in

[19] It is worth to note that the foreign RD is a weighted average of domestic R&D, i.e. $\mathbf{1}_{G7} \log \sum_{j \neq i} \exp(-\varphi_{G7} d_{ij}) S_{jt}^d$ and $\mathbf{1}_{NOG7} \log \sum_{j \neq i} \exp(-\varphi_{NOG7} d_{ij}) S_{jt}^d$, for G7 and non-G7 countries, respectively and that the CCE procedure inroduces in the regression another average of domestic R&D, $\overline{\log S_t^d}$. One may wonder if and how these quantities are correlated and if this correlation may affect the results. A closer look to the data indicates that the correlation between the observed variables, $\mathbf{z}_{it}$, and the cross-sectional averages, $\overline{\mathbf{z}_t}$, is not very high, ranging from 0.44 to 0.80, and in particular the correlation coefficient between $\overline{\log S_t^d}$ and the foreign R&D is quite low: 0.44 and 0.62 for G7 and non-G7 countries, respectively. These correlations are of the same order of magnitude than the correlations between foreign R&D stocks and the others cross-sectionally averaged variables. The correlation of cross-sectional averaged variables among themselves is, instead, very high in all cases ranging between 0.97 and 0.99. Such variables also present a high degree of correlation with the trend and the squared trend. Finally, it can be noticed that the oil price presents a low degree of correlation with both $\mathbf{z}_{it}$ and $\overline{\mathbf{z}_t}$ (detailed results are available upon request).

the model decreases the effect of domestic R&D, but also suggest that all types of spillovers, both observed (foreign R&D) and unobserved (common factors), may play a key role in explaining TFP growth.

Concerning the inference provided by these estimates, it may be interesting to recall the Monte Carlo simulations by Pesaran (2006) made under the assumption that the DGP is characterised by unobserved common factors. For $N = T = 20$, these simulations indicate that, whereas the naïve estimators (i.e., the estimators that do not account for cross-sectional correlation, such as the LSDV) are oversized but have high power, the CCE estimators have the correct size but have low power. Moreover, Pesaran and Tosetti (2011) (see Bresson and Hsiao, 2011, for additional results) provide interesting simulations under the assumption that the error is generated by a spatial autoregressive model or is a mixture of a spatial process and a multifactor model. In all cases (for $T = N = 20$), the CCE estimators have better size than any others (including the ML spatial error estimator) but low power compared to the alternative estimators.[20]

These simulation results may have relevant implications in our context, especially for the coefficient associated with foreign R&D for the countries that do not belong to the G7 group. The point estimate of this coefficient is high, taking values of 0.12, 0.16, 0.21 and 0.34 (from column (iii) to (vi)), and the corresponding t values, which are also relatively high, range from 0.93, 1.11, 1.44 to 2.27 respectively. This finding might suggest that in some cases, our inference erroneously indicates a non-significant effect of foreign R&D for the non-G7 countries. Finally, it is also interesting to note that both the magnitude of the estimated parameter, γ_{NOG7}, and the corresponding significance level increase with the number of observed common factors introduced in the CCEP framework.

Finally, comparing the results with those obtained with a spatial panel data model with spatially autocorrelated errors, as in equation (23), may provide some interesting insights on the possible bias occurring when allowing only for weak correlation while strong correlation is present in the data. Therefore, considering an interaction matrix with typical element $w_{ij} = \exp(-\phi d_{ij})/\sum_j \exp(-\phi d_{ij})$ to model interactions between countries i and j, a spatial model has been estimated by the QML approach proposed by Lee and Yu (2010), and the estimation results are presented in Table 7. We also consider adding the trend, the trend squared and the oil price as in the CCEP estimation. Again, the results are not sensitive to the different specifications for the observed common factors. Introducing such observed common variables makes decrease the estimated parameters for all the explanatory variables. In particular, both the foreign R&D for non-G7 countries and the human capital become not significant at standard levels. This is the same result we obtain from the a-spatial baseline specification when such observed common factors are introduced (results available upon request). However, in contrast to the CCEP estimation results, it appears immediately that the estimated coefficient for domestic R&D is about 0.04 and remains highly significant. In summary, when using a spatial model instead of a model

[20]These simulations also show that the CCE estimators are superior to all competitors with respect to bias when the errors are a mixture of a spatial and a multifactor process.

with multifactor error structure, we obtain a higher effect of domestic R&D and a lower effect of foreign R&D for G7 countries. This is a relevant result both in terms of econometric modelling and with respect to the policy implications. This result may be due to the bias induced by strong cross-section correlation, which is not taken into account in the spatial econometric approach.[21]

5 Conclusion

This paper provides an analysis of international technology diffusion by accounting for the role of cross-country correlation when estimating the econometric specification. Theoretical consistency, empirical relevance and exogeneity arguments have allowed us to focus on geographical proximity as a channel of technology diffusion.

Cross-sectional dependence can be taken into account using two alternative approaches: unobserved common factors and spatial models. The motivations underlying these two approaches are completely different. The first approach is related to the notion of strong cross-sectional dependence and is based on a parsimonious way to capture information in a data rich environment using a small number of unobserved factors, which are allowed to be freely correlated with the conditioning variables. The second approach is related to weak cross-sectional dependence. It is explicitly oriented towards modeling cross-sectional interactions and capturing spatial spillovers. There are no theoretical or empirical reasons to a priori favor one of them and the type of cross-sectional dependence should be tested before estimating the most suitable model.

A preliminary analysis based on the CD test proposed by Pesaran (2004, 2012a) and on the estimation of the exponent of cross section correlation, proposed by Bailey et al. (2013a), provides a very clear-cut result in favor of strong correlation. Indeed, first, the CD test strongly rejects the null. Using the results by Pesaran (2012a) and given the size of our sample (N and T of the same order), this result suggests a situation in which $1/4 \leq a \leq 1$, where a is the exponent of cross-sectional correlation introduced in Bailey et al. (2013a), which can take any value in the range $[0, 1]$, with 1 indicating the highest degree of cross-sectional dependence, and where $a < 0.5$ and $a > 0.5$ correspond to the cases of weak and strong cross-sectional correlation discussed in Chudick et al. (2011), respectively. Given this, we focus on the estimation of a and obtain an estimate of a very close to unity, not only indicating the presence of strong cross-sectional correlation but also being consistent with the factor literature typically assuming that $a = 1$. Moreover, before moving to the estimation, we also study the order of integration of the variables of interest using several tests, most of which allow for cross-sectional dependence (Pesaran, 2007; Pesaran et al., 2013; Bai and Ng, 2004; Moon and Perron, 2004; Choi, 2006). While most of the previous works find evidence of nonstationary variables by applying first-generation tests (see among others, Coe et al. 2009), we provide a more nuanced and thorough picture, in the end suggesting that while the unobserved idiosyncratic component of the variables under study is stationary, the unobserved

[21] The results are very robust to the choice of the exponential decay parameter of the spatial autoregressive component ϕ, which take three different values (1, 5 and 10).

common factors seem to be nonstationary. This result is very relevant for an empirical perspective and is related to the recent work by Kapetanios et al. (2011) showing that the CCE approach is still valid when the unobserved factors are allowed to follow unit root processes. Finally, the model is estimated with the CCE approach and the results are compared with those obtained with a spatial model providing some interesting insights on the possible bias occurring when allowing only for weak correlation while strong correlation is present in the data.

References

[1] Bai, J. (2003). 'Inferential theory for factor models of large dimensions', *Econometrica*, 71, 135–17

[2] Bai, J. and Ng, S. (2004). 'A Panic Attack on Unit Roots and Cointegration', *Econometrica*, 72, 1127–1177.

[3] Bai, J. (2009). 'Panel data models with interactive fixed effects', *Econometrica*, 77, 1229-1279.

[4] Baltagi, B.H., Bresson, G. and Pirotte, A. (2007), 'Panel Unit Root Tests and Spatial Dependence', *Journal of Applied Econometrics*, 22, 339–360.

[5] Bailey, N., G. Kapetanios, and Pesaran, M.H. (2013a) 'Exponent of Cross-sectional Dependence: Estimation and Inference', mimeo.

[6] Bailey, N., Holly, S. and Pesaran, M. H. (2013b). A two stage approach to spatiotemporal analysis with strong and weak cross-sectional dependence. Center for Applied Financial Economics (CAFE) research paper no. 14.01.

[7] Barrio-Castro, T.; Lopez Bazo, E. and Serrano Domingo, G. (2002) 'New Evidence on International R&D Spillovers, Human Capital and Productivity in the OECD', *Economics Letters*, 77, 41-45.

[8] Barro, R.J. and Lee, J. (2001). 'International Data on Educational Attainment: Updates and Implications', *Oxford Economic Papers*, 53, 541-563.

[9] Basu, S. and Weil, D. (1998). 'Appropriate Technology and Growth', *Quarterly Journal of Economics*, 113, 1025-54.

[10] Behrens, K., Ertur, C. and Koch, W. (2012). 'Dual gravity: using spatial econometrics to control for multilateral resistance', *Journal of Applied Econometrics*, 2012, 27, 773-794.

[11] Bottazzi, L., and Peri, G. (2003). 'Innovation, Demand, and Knowledge Spillovers: Evidence from European Patent Data', *European Economic Review*, 47, 687–710.

[12] Bresson, G. and Hsiao, C. (2011), 'A functional connectivity approach for modeling cross-sectional dependence with an application to the estimation of hedonic housing prices in Paris', *Advances in Statistical Analysis*, 95, 501-529.

[13] Caselli, F. (2005). 'Accounting for Cross-Country Income Differences', *Handbook of Economic Growth* (Aghion and Durlauf, eds.), North-Holland.

[14] Choi, I. (2001). 'Unit Root Tests for Panel Data', *Journal of International Money and Finance*, 20, 249-272.

[15] Choi, I. (2006). 'Combination Unit Root Tests for Cross-Sectionally Correlated Panels'. In: D. Corbae, S. N. Durlauf, and B. E. Hansen (eds.): *Econometric Theory and Practice: Frontiers of Analysis and Applied Research: Essays in Honor of Peter C. B. Phillips*. Cambridge University Press, Chapt. 11, 311–333.

[16] Chudik, A., Pesaran, H. and & Tosetti, E. (2011). 'Weak and strong cross-section dependence and estimation of large panels', *Econometrics Journal*, 14, 45-90.

[17] Coakley, J., Fuertes, A.M. and Smith, R. (2002. 'A principal components approach to cross-section dependence in panels', Birkbeck College Discussion Paper 01/2002.

[18] Coe, D.T. and Helpman, E. (1995). 'International R&D spillovers', *European Economic Review*, 39, 859–887.

[19] Coe, D.T., Helpman, E. and Hoffmaister, A. (2009). 'International R&D spillovers and institutions', *European Economic Review*, 53, 723-741,

[20] Debarsy, N. and Ertur, C. (2010). 'Testing for Spatial Autocorrelation in a Fixed Effects Panel Data Model', *Regional Science and Urban Economics*, 40, 453-470

[21] Eaton, J. and Kortum, S. (2002). 'Technology, Geography, and Trade', *Econometrica*, 70, 1741-1779.

[22] Eberhardt, M. and Helmers, C. and Strauss, H. (2012). 'Do spillovers matter when estimating private returns to R&D?' *The Review of Economics and Statistics*, (In Press.).

[23] Eberhardt, M. and Bond, S. (2009). 'Cross-section dependence in nonstationary panel models: a novel estimator', MPRA Paper 17692, University Library of Munich, Germany.

[24] Eberhardt, M. and Teal. F. (2010). 'Productivity Analysis in Global Manufacturing Production', Economics Series Working Papers 515, Department of Economics, University of Oxford

[25] Engelbrecht, H.J. (1997). 'International R&D spillovers, human capital and productivity in OECD economies: an empirical investigation', *European Economic Review*, 41, 1479–1488.

[26] Ertur, C. and Koch, W. (2007). 'Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence', *Journal of Applied Econometrics*, 22, 1033-1062.

[27] Ertur, C. and Koch, W. (2011). 'A contribution to the theory and empirics of Schumpeterian growth with worldwide interactions', *Journal of Economic Growth*, 6, 215-255.

[28] Forni, M., and Lippi, M. (2001). 'The Generalized Factor Model: Representation Theory', *Econometric Theory*, 17, 1113-1141.

[29] Frantzen, D. (2000). 'R&D, human capital and international technology spillovers: a cross-country analysis', *Scandinavian Journal of Economics*, 102, 57–75.

[30] Frees, E. W. (1995). ‘Assessing cross-sectional correlations in panel data’, *Journal of Econometrics*, 69, 393-414.

[31] Friedman, M. (1937). ‘The use of ranks to avoid the assumption of normality implicit in the analysis of variance’, *Journal of the American Statistical Association*, 32, 675-701.

[32] Grossman, G., and Helpman, E. (1991). ‘Innovation and growth in the global economy’, Cambridge, MA, MIT Press.

[33] Hall, R.E. and Jones, C.I. (1999). ‘Why Do Some Countries Produce So Much More Output Per Worker Than Others?’, *The Quarterly Journal of Economics*, 114, 83-116.

[34] Hong, E. and Sun, L. (2011). ‘Foreign Direct Investment and Total Factor Productivity in China: A Spatial Dynamic Panel Analysis’, *Oxford Bulletin of Economics and Statistics*, 73, 6, 771-791.

[35] Im, K.S., Pesaran, M.H. and Shin, Y. (2003). ‘Testing for Unit Roots in Heterogenous Panels’, *Journal of Econometrics,* 115, 53–74.

[36] Kao, C. Chiang, M.H. and Chen, B. (1999). ‘International R&D Spillovers: An Application of Estimation and Inference in Panel Cointegration’, *Oxford Bulletin of Economics and Statistics*, 61, 691-709.

[37] Kapetanios, G. and M. Marcellino (2010). Factor-GMM estimation with large sets of possibly weak instruments. Computational Statistics and Data Analysis 54 (11), 2655.2675.

[38] Kapetanios, G. and Pesaran, M.H. (2007). ‘Alternative approaches to estimation and inference in large multifactor panels: small sample results with an application to modelling of asset returns. In: Phillips, G., Tzavalis, E. (Eds.), The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis. Cambridge University Press, Cambridge.

[39] Kapetanios, G., Pesaran, M.H. and Yamagata, T. (2011). ‘Panels with Nonstationary Multifactor Error Structures’, *Journal of Econometrics*, 160, 326–348.

[40] Kelejian, H. H. and Prucha, I. R. (1998). ’A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances’, *Journal of Real Estate Finance and Economics*, 1998, 17, 99-121.

[41] Kelejian, H. H. and Prucha, I. R. (1999). ’A generalized moments estimator for the autoregressive parameter in a spatial model’, *International Economic Review*, 1999, 40, 509-533.

[42] Kelejian, H. H. and Prucha, I. R. (2001). ’On the Asymptotic Distribution of The Moran I Test Statistic with Applications’, *Journal of Econometrics*, 2001, 104, 219-257.

[43] Kelejian, H. H. and Prucha, I. R. (2010). 'Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances', *Journal of Econometrics*, 157, 53-67.

[44] Keller, W. (2002). 'Geographic Localization of International Technology Diffusion', *American Economic Review*, 92, 120-42.

[45] Keller, W. (2004). 'International Technology Diffusion', *Journal of Economic Literature*, 42, 752-782.

[46] Krugman, P. and Venables, A. (1995). 'Globalization et the Inequality of Nations', *Quarterly Journal of Economics*, 110, 857-880.

[47] Lee, G. (2006). 'The effectiveness of international knowledge spillover channels, *European Economic Review*, 50, 2075–2088.

[48] Lee, L., and Yu, J. (2010). 'Estimation of Spatial Autoregressive Panel Data Models With Fixed Effects', *Journal of Econometrics,* 154, 165-185.

[49] Lee, L.-f. and Yu, J. (2012). 'QML estimation of spatial dynamic panel data models with time varying spatial weights matrices', *Spatial Economic Analysis*, 2012, 7, 31-74.

[50] Levin, A., Lin, C. and Chu, C.J. (2002). 'Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties', *Journal of Econometrics*, 108, 1–24.

[51] Lichtenberg, F.R. and Van Pottelsberghe de la Potterie, B.P. (1998). 'International R&D spillovers: A comment', *European Economic Review*, 42, 1483-1491.

[52] Lichtenberg, F.R. and Van Pottelsberghe de la Potterie, B.P. (2001). 'Does foreign direct investment transfer technology across borders?', *The Review of Economics and Statistics*, 83, 490–497.

[53] Maddala, G.S. and Wu, S. (1999). A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test, *Oxford Bulletin of Economics and Statistics*, 61, 631–652.

[54] Mastromarco, C., Serlenga L. and Shin, Y. (2012). 'Globalisation and technological convergence in the EU', *Journal of Productivity Analysis*, in press.

[55] Moon, R. and Perron, B. (2004), 'Testing for Unit Root in Panels with Dynamic Factors', *Journal of Econometrics*, 122, 81-126.

[56] Musolesi, A. (2007). 'Basic stocks of knowledge and productivity: further evidence from the hierarchical Bayes estimator', *Economics Letters*, 95, 54-59.

[57] Onatski, A. (2012). 'Asymptotics of the principal components estimator of large factor models with weakly influential factors', *Journal of Econometrics,* 168, 244.258.

[58] Neyman, J., and Scott, E.L. (1948). 'Consistent estimates based on partially consistent observations', *Econometrica*, 16, 1 32.

[59] Pesaran, M.H. (2004), General diagnostic tests for cross section dependence in panels, Cambridge Working Papers in Economics No. 0435, Faculty of Economics, University of Cambridge.

[60] Pesaran, M.H. (2006), 'Estimation and inference in large heterogeneous panels with a multifactor error structure', *Econometrica*, 74, 967-1012.

[61] Pesaran, M.H. (2007). 'A Simple Panel Unit Root Test in the Presence of Cross Section Dependence', *Journal of Applied Econometrics*, 22, 265–312.

[62] Pesaran, M. H. (2012a). 'Testing weak cross-sectional dependence in large panels', CESifo Working Paper no. 3800.

[63] Pesaran, M. H. (2012b). 'On the interpretation of panel unit root tests', *Economics Letters,* 116, 545-546.

[64] Pesaran, M. H. and Smith, R. (1995). 'Estimating long-run relationships from dynamic heterogeneous panels', *Journal of Econometrics,* 68, 79-113.

[65] Pesaran, M. H. and Tosetti, E. (2011). 'Large panels with common factors and spatial correlation', *Journal of Econometrics*, 161, 182-202.

[66] Pesaran, M. H., V. Smith and T. Yamagata (2013). 'Panel unit root tests in the presence of a multifactor error structure', *Journal of Econometrics,* 175, 94-115.

[67] Psacharopulos, G. (1994). 'Returns to investment in education: A global update', *World Development*, 22, 1325-1343.

[68] Robinson, P.M. (2011). 'Asymptotic theory for nonparametric regression with spatial data', *Journal of Econometrics,* 165, 5-19.

[69] Robinson, P.M., and Thawornkaiwong, S. (2012). 'Statistical inference on regression with spatial dependence', *Journal of Econometrics* 167, 521-542.

[70] Sarafidis, V. (2009). 'GMM Estimation of Short Dynamic Panel Data Models with Error Cross-sectional Dependence', Mimeo.

[71] Sarafidis, V., and Wansbeek, T. (2012). 'Cross-sectional dependence in panel data analysis', *Econometric Reviews,* 31, 483-531.

[72] Spolaore, E. and Wacziarg, R. (2009). 'The Diffusion of Development', *The Quarterly Journal of Economics*, 124, 469-529.

[73] Stock, J. H. and M. W. Watson (2002). 'Macroeconomic forecasting using diffusion indexes', *Journal of Business and Economic Statistics,* 20, 147.162.

[74] Xu, B. (2000). 'Multinational Enterprises, Technology Diffusion, and Host Country Productivity Growth', *Journal of Development Economics*, 62, 477-93.

TABLE 1

Some previous studies on R&D international spillovers

Author	sample	Technology transfer	Method	Foreign R&D
Coe and Helpman (1995)	22 countries, 1971-90	trade	LSDV	.06-.092
Coe et al. (2009)	22 countries, 1971-90	trade	DOLS	.165-.186
	24 countries, 1971-2004		LSDV	.185-.206
			DOLS	.206-.213
Kao et al. (1999)	22 countries, 1971-90	trade	BC-OLS	.09-.125
			FM-OLS	.075-.103
			DOLS	.044NS-.056NS
Lichtenberg and Van Pottelsberghe (1997)	22 countries, 1971-90	trade	LSDV	.058-.276
Lichtenberg and Van Pottelsberghe (2001)	23 countries, 1971-90	trade	LSDV	.154
		FDI		-.06NS-.072
		trade	FD	.067
		FDI		.006NS-.039
Musolesi (2007)	13 countries, 1981-98	trade	HB	.09
		FDI		-0.01NS-.004NS
		language		.23
Lee (2006)	16 countries, 1981-2000	trade	DOLS	-.02NS-.17
		FDI		-.017NS-.034
		patents		.157-.183
Keller (2002)	14 countries, 1970-95	geographic distance	NLS	.843
		language		.565
Engelbrecht (2002)	21 countries, 1971-85	trade	LSDV	.220-.305
Barrio-Castro et al. (2002)	21 countries, 1971-85	trade	LSDV	.094-.225
	21 countries, 1966-95		LSDV	0.016-0.106
			DOLS	.092-.141

Notes:

LSDV: Least Square Dummy Variable ; DOLS: Dynamic Ordinary Least Square; BC-OLS: Bias Corrected OLS;

FM-OLS: Fully Modified OLS; FD: First Difference; HB: Hierarchical Bayes; NLS: Non Linear Least Square; NS: not significant.

TABLE 2

Estimation results. Benchmark NLS estimates and CCEP

	Benchmark NLS		*CCEP*			
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
β	0.067***	0.069***	0.026	0.020	0.010	-0.004
	(0.0115)	(0.0105)	(0.064)	(0.096)	(0.069)	(0.105)
γ	0.042***					
	(0.0125)					
φ	9.649***					
	(0.0125)					
γ_G7		0.170***	0.317**	0.624***	0.216**	0.464***
		(0.0183)	(0.159)	(0.156)	(0.095)	(0.165)
φ_G7		15.217***				
		(0.0272)				
γ_NOG7		0.0267**	0.122	0.161	0.213	0.337**
		(0.0116)	(0.131)	(0.145)	(0.148)	(0.148)
φ_NOG7		7.042***				
		(0.1801)				
δ	0.538***	0.375***	0.305	0.372	0.194	0.222
	(0.1224)	(0.1135)	(0.372)	(0.421)	(0.540)	(0.410)
$\mathbf{d}_t$	$(\iota)'$	$(\iota)'$	$(\iota, t)'$	$(\iota, t, t^2)'$	$(\iota, t, p)'$	$(\iota, t, t^2, p)'$

Notes:

NLS: Nonlinear Least Squares. CCEP: Correlated Common Effect Pooled.

***, **, *: significant at 1%, 5%, 10%, respectively.

Standard error between brackets.

CCEP: standard error based on Newey-West type variance estimator of eq. (74) in Pesaran (2006).

CCEP: the parameters $\varphi_G7$and $\varphi_NOG7$have been settled to 15.22 and 7.04, respectively.

TABLE 3

Tests of cross sectional independence

Pair-wise correlation coefficients		
CD	Friedman	Frees
4.710	40.696	5.235
(0.000)	(0.0061)	
		[0.470]

Notes:

p-value between brackets. Average absolute value of the off-diagonal elements between square brackets.

Critical value from Frees' Q distribution: alpha = 0.01 : 0.2468.

TABLE 4

Moon and Perron test

	r	kernel	$\log f$	$\log S^d$	$\log S^f$	$\log H$
m*(IC2)			3	4	4	4
m*(BIC3)			2	4	4	4
t_a^*	1	QS	-2.6604(0.0039)	-1.7862 (0.0370)	-0.7206 (0.2356)	-0.2279(0.4099)
	1	B	-2.9104(0.0018)	-1.7852 (0.0371)	-0.7889(0.2151)	-0.2382 (0.4059)
	2	QS	-4.2229(1.2057e-005)	-1.6096 (0.0537)	-0.7984 (0.2123)	-4.5750 (2.3815e-006)
	2	B	-4.4621 (4.0585e-006)	-1.6856 (0.0459)	-0.6783 (0.2488)	-4.5750 (2.3815e-006)
	3	QS	-4.3942 (5.5589e-006)	-0.7444 (0.2283)	-4.2315 (1.1605e-005)	-3.5113 (2.2296e-004)
	3	B	-4.5508 (2.6726e-006)	-0.7703 (0.2206)	-3.7149 (1.0164e-004)	-3.6976 (1.0883e-004)
	4	QS	-4.9478 (3.7529e-007)	-3.6826 (1.1544e-004)	-0.4179 (0.3380)	-3.6500 (1.3112e-004)
	4	B	-5.0309 (2.4409e-007)	-4.6880 (1.3792e-006)	-0.5859 (0.2790)	-4.0252 (2.8465e-005)
t_b^*	1	QS	-2.6563 (0.0040)	-1.1354 (0.1281)	-0.6198 (0.2677)	-0.0747 (0.4702)
	1	B	-2.7187(0.0033)	-1.0977 (0.1362)	-0.7003 (0.2419)	-0.0834 (0.4668)
	2	QS	-4.6476 (1.6787e-006)	-0.9969 (0.1594)	-0.8512 (0.1973)	-5.5644 (1.3150e-008)
	2	B	-4.4694 (3.9228e-006)	-1.1119 (0.1331)	-0.7572 (0.2245)	-4.6502 (1.6583e-006)
	3	QS	-4.6777 (1.4502e-006)	-0.3774 (0.3530)	-5.9500 (1.3410e-009)	-6.6819 (1.1796e-011)
	3	B	-4.6046 (2.0666e-006)	-0.4235 (0.3360)	-3.5446 (1.9659e-004)	-6.3602 (1.0077e-010)
	4	QS	-6.3634 (9.8699e-011)	-4.6224 (1.8964e-006)	-0.3629 (0.3583)	-6.2396 (2.1941e-010)
	4	B	-5.1844 (1.0838e-007)	-4.7808 (8.7295e-007)	-0.3591 (0.3598)	-7.2367 (2.2992e-013)

Notes:

p-value between brackets. QS: Quadratic spectral kernel. B: Bartlett kernel. r is the number of factors.

Here we impose that the maximum number of factors is 4 as in Pesaran et al. (2013). In almost all cases the criteria suggest that the number of unobserved factors, r, equals the maximum number we allowed, 4. This is the same result by Pesaran (2007) and by Pesaran et al. (2013) (see also Gutierrez, 2006).

This suggests that the number of factors could be even higher than 4. However, given the possibility that the criteria over estimate the number of factors and the number of observations available, Pesaran et al. (2013) do not allow that the maximum number of factors could be greater than 4.

We provide our main results doing the same thing but, as a robustness check available upon request, we also perform the test for all the possible values of r in the range $1-15$. The null hypotesis is generally not rejected for r in the range 1-9 while it is always rejected for $10 \leq r \leq 15$.

TABLE 5

Bai and Ng test

		Idiosyncratic component		Common factors	
	Number of factors			Nonstationary factors $\widehat{r}_1$	
	$\widehat{r}_{BIC3}$	P_e^c	Z_e^c	MQ_f	MQ_c
		p-value			
Model with intercept					
$\log f$	2	0.210	0.202	2	2
$\log S^d$	4	0	8.76E-10	4	4
$\log S^f$	4	0	8,75E-10	4	4
$\log H$	4	0	0	4	4
Model with intercept and trend					
$\log f$	2			2	2
$\log S^d$	4			4	4
$\log S^f$	4			4	4
$\log H$	4			4	4

Notes:

$\widehat{r}_{BIC3}$ss the estimated number of common factors, based on BIC3 criterion.

Here again we impose that the maximum number of factors is 4.

For the idiosyncratic components, only pooled unit root tests are reported.

Individual ADF t statistics based on de-factored components are available upon request.

The pooled test for the idiosyncratic component is valid only for the intercept case.

In the linear trend case, the limiting distribution of the pooled test is not a DF type distribution.

For the common factors components, the estimated number $\widehat{r}_1$of independent stochastic trends is reported (5% level).

We provide available upon request the results of the test for all values of r in the range 1-15.

They strongly confirm the finding of nonstationary common factors and possibly stationary idiosyncratic components.

TABLE 6

CIPS and CSB tests - Pesaran (2007) and Pesaran et al. (2013)

	$\log f$		$\log S^d$		$\log S^f$		$\log H$	
	$CIPS(\widehat{p})$	$CSB(\widehat{p})$	$CIPS(\widehat{p})$	$CSB(\widehat{p})$	$CIPS(\widehat{p})$	$CSB(\widehat{p})$	$xCIPS(\widehat{p})$	$CSB(\widehat{p})$
$r = 1$	-2.146	0.146*	-2.152	0.192*	-1.935	0.161*	-2.193	0.131*
$r = 2$								
$\overline{x}_t = (\overline{\log S^d})$	-3.001*	0.102*	-2.074	0.167*	-1.511	0.123*	-1.913	0.177*
$\overline{x}_t = (\overline{\log S^f})$	-2.671	0.105*	-2.071	0.156*	-1.453	0.048	-2.27	0.143*
$\overline{x}_t = (\overline{\log H})$	-2.307	0.13*	-2.649	0.158*	-1.651	0.133*	-2.504	0.091*
$r = 3$								
$\overline{x}_t = (\overline{\log S^d}, \log S^f)$	-2.821	0,092*	-1.8	0.132*	-3.128	0.039	-1.693	0.126*
$\overline{x}_t = (\overline{\log S^d}, \overline{\log H})$	-3.119*	0.093*	-2.068	0.11*	-1.973	0.071*	-2.716	0.102*
$\overline{x}_t = (\overline{\log S^f}, \overline{\log H})$	-3.377*	0.09*	-1.986	0.125*	-1.144	0.035	-3.555*	0.1067*
$r = 4$								
$\overline{x}_t = (\overline{\log S^d}, \log S^f, \log H)$	-2.614	0.081*	-2.896	0.086*	-2.992*	0.0244	-5.621*	0.0954*

Notes:

Pesaran et al. (2013) set the lag order, $\widehat{p} = \left[4\,(T/100)^{1/4}\right]$. Tthis gives $\widehat{p} = 2.6$. For $p = 3$we had computational problems. For $p = 2$we had problems in few cases.

To provide comparable results for all possible combinations of regressors we set $p = 1$. Results for $p = 0$and $p = 2$are available upon request.

The variables under the heading $\overline{x}_t$indicate the regressors used for cross-section augmentation in addition to $\overline{y}_t$. In the case where $r = 1$no additional regressors are used (Pesaran, 2007).

For the selected lag order, the 5% critical values for the CIPS are -2.67, -2.83, -2.88, -2.97 for r=1, 2, 3, 4, respectively.

The 5% critical values for the CSB are 0.104, 0.087, 0.071, 0.057 for r=1, 2, 3, 4, respectively. * Denotes rejection at 5% level.

According to Pesaran et al. (2013), for small T, the CSB test has higher power than that of CIPS, and should thus be preferred in such cases.

TABLE 7

Estimation results. Spatial autoregressive error model (QML)

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)	(xi)	(xii)
ϕ	1	5	10	1	5	10	1	5	10	1	5	10
β	0.078***	0.074***	0.074***	0.045***	0.036***	0.036***	0.044***	0.036***	0.035***	0.044***	0.036***	0.035***
	(8.481)	(8.067)	(7.871)	(4.381)	(3.369)	(3.304)	(4.352)	(3.318)	(3.251)	(4.364)	(3.313)	(3.246)
γ_G7	0.183***	0.180***	0.175***	0.122***	0.116***	0.110***	0.122***	0.115***	0.109***	0.122***	0.115***	0.109***
	(10.527)	(10.358)	(10.032)	(6.378)	(5.940)	(5.623)	(6.343)	(5.868)	(5.547)	(6.355)	(5.850)	(5.533)
γ_NOG7	0.020**	0.032***	0.029***	-0.007	-0.001	-0.003	-0.007	-0.002	-0.004	-0.007	-0.002	-0.004
	(1.979)	(3.123)	(2.792)	(-0.668)	(-0.128)	(-0.324)	(-0.699)	(-0.178)	(-0.378)	(-0.695)	(-0.187)	(-0.386)
δ	0.010	0.182	0.235**	-0.159	-0.024	0.006	-0.151	-0.009	0.022	-0.154	-0.007	0.025
	(0.086)	(1.599)	(2.061)	(-1.404)	(-0.208)	(0.053)	(-1.32)	(-0.082)	(0.189)	(-1.34)	(-0.061)	(0.213)
λ	0.4569***	0.254***	0.207***	0.406***	0.220***	0.175***	0.408***	0.218***	0.173***	0.411***	0.215***	0.171***
	(9.106)	(5.535)	(4.875)	(7.584)	(4.700)	(4.064)	(7.640)	(4.653)	(4.014)	(7.726)	(4.582)	(3.965)
$\mathbf{d}_t$	$(\iota)'$	$(\iota)'$	$(\iota)'$	$(\iota,t)'$	$(\iota,t)'$	$(\iota,t)'$	$(\iota,t,t^2)'$	$(\iota,t,t^2)'$	$(\iota,t,t^2)'$	$(\iota,t,t^2,p)'$	$(\iota,t,t^2,p)'$	$(\iota,t,t^2,p)'$

Notes:

***, **, *: significant at 1%, 5%, 10%, respectively.

Asymptotic t statistics between brackets

The parameters $\varphi_G7$and $\varphi_NOG7$have been settled to 15.22 and 7.04, respectively.

The specifications with $d_t = (\iota,t,p)'$have been excluded to save space. The results are very close to those obtained with $d_t = (\iota,t,t^2)'$and $(\iota,t,t^2,p)'$.

These estimates are available upon request.

Appendix: Panel Unit Root Tests

We first present the results obtained using the test proposed by Im, Pesaran and Shin (2003) (IPS) and the Fisher-type tests introduced by Maddala and Wu (1999) and further developed by Choi (2001). These tests allow the same degree of heterogeneity: let define δ as the coefficient associated with the autoregressive term in the ADF type regression; Levin, Lin and Chu (2002), among others, propose a test assuming that this coefficient is the same for all cross sections. As noted by Maddala and Wu (1999), whereas the null unit root hypothesis ($\delta = 0$) is appropriate in some empirical applications, the alternative ($\delta < 0$) seems to be too strong to hold in any relevant case. Im, Pesaran and Shin (2003); Maddala and Wu (1999); and Choi (2001) relax the assumption that $\delta_1 = \delta_2 \ldots = \delta_N$ under the alternative, allowing under such alternative some of the individual series to have a unit root, say $\delta_i < 0$ for $i = 1, ..., N1$ and $\delta_i = 0$ for $i = N1 + 1, ..., N$ with $0 < N1 < N$ (see also Pesaran, 2012). Both tests combine the information obtained from the N independent individual tests and (at least when linear trends are included in the deterministic component and the errors are serially correlated) both tests obtain their asymptotic properties by first sending T to infinity and then N to infinity, $(T, N) \rightarrow_{\text{seq}} \infty$.[22]

In performing the tests, we make the following choices: i) because the series are clearly trended, linear time trends have been included in the deterministic component, and ii) the selection of the lag order of the autoregressive components p has to be performed carefully because it is well known that ADF-type tests are highly sensitive to this choice. There is, of course, a delicate balance between choosing a p that is sufficiently large to allow for serially uncorrelated residuals and, simultaneously, sufficiently small such that the model is not overparameterised. Moreover, this choice is crucial because if k is overestimated this may decrease the power of the test while if it is underestimated, this may invalidate the asymptotic distribution of the test (see also Westerlund and Breitung, 2009, for a more detailed discussion of this issue).Therefore, the order of the (individual) AR components has been chosen using alternative criteria (AIC, SBC, HQIC) subject to a maximum lag of 3.

The IPS test is based on combining individual ADF t statistics. The reported standardised statistic – the W_{t-bar} – has an asymptotically standard normal distribution and has been shown to perform well even in small samples. The results in **table A1** have three major implications. First, the test is highly sensitive to the number of lags of the AR component, k. Second, when the number of lags k is chosen with AIC, BIC or HQIC, there is strong evidence against the null hypothesis for all variables. This, thus suggests that some of the individual series in the panel are, in fact, stationary.

Next, we use the Fisher-type tests (Fisher, 1932) provided by Maddala and Wu (1999) and Choi (2001) based on combining the p-values of the N independent test statistics. Two statistics are provided here, labeled P and Z. These values differ in whether they use the inverse chi-square or the inverse normal distribution of the p-values.[23] The Fisher-type statistics (in **table A1**) fully confirm the IPS tests.

Recent work has demonstrated the importance of accounting for cross-sectional correlation when testing the

[22]Although the sequential limit results may appear to be more restrictive than the joint limit results obtained by sending T and N to infinity simultaneously, it has been shown that the sequential and joint limit results are identical under additional moment conditions (Phillips and Moon, 1999). As a practical matter, this means that in both cases, a reasonably large number of time periods and cross-sections are required to implement these tests.

[23]Choi's (2001) simulation results suggest the use of the Z statistic, which offers the best trade-off between size and power. With the aim of comparing Fisher-type tests to the IPS test, Choi (2001) finds that the Fisher tests are more powerful than the IPS test in finite samples, and Maddala and Wu (1999) confirm this finding even when the errors are cross-correlated.

unit root hypothesis. Pesaran's (2007) simulations show that tests assuming cross-sectional independence tend to over-reject the null hypothesis if cross-sectional correlation is present. Baltagi et al. (2007) find that when spatial autoregression is present, first-generation tests become oversised, but the tests explicitly allowing for cross-sectional dependence yield a lower frequency of type I errors. As noted by Pesaran (2007), subtracting the cross-sectional averages from the series before applying the panel unit root test can mitigate the impact of cross-sectional dependence even if cross-sectional demeaning could not work in general in conditions under which the pairwise cross-sectional errors' covariances differ across individuals.[24] Moreover, while weak cross section correlation can be dealt by a simple correction of the tests, the presence of strong cross section correlation is more problematic, making the tests statistics becoming divergent (Westerlund and Breitung, 2009). Since we clearly documented the presence of strong correlation in our data, it is of fundamental importance applying the so-called second generation unit root tests (Bai and Ng, 2002, 2004; Moon and Perron, 2004; Pesaran, 2007, Pesaran et al. 2013).

Bai and Ng (2004) propose decomposing the panel into deterministic, common and idiosyncratic components, i.e.

$$y_{it} = D_{it} + \zeta_i^{'} \mathbf{f}_t + v_{it},$$

where D_{it} is the deterministic component with individual effects and eventually individual trends, $\zeta_i^{'} \mathbf{f}_t$ the common component, with r unobserved factors, and v_{it} the idiosyncratic component. Such a decomposition allows considering factors as objects of interest and understanding not only if the data are stationary or not but also if the eventual nonstationarity derives from nonstationary common component, nonstationary idiosyncratic component or nonstationarity of both components. More precisely, Bai and Ng (2004) also assume that

$$(\mathbf{I} - L)\,\mathbf{f}_t = \mathbf{C}\,(L)\,u_t,$$
$$(1 - \rho_i)\,v_{it} = \mathbf{B}_i\,(L)\,\epsilon_{it}$$

where $\mathbf{C}\,(L) = \sum_{j=0}^{\infty} C_j L^j$ and $\mathbf{B}_i\,(L) = \sum_{j=0}^{\infty} B_{ij} L^j$. The idiosyncratic component is $I(1)$ if $\rho_i = 1$ and is stationary if $\rho_i < 1$. There are r_0 stationary common factors and r_1 common stochastic trends, so that $r_0 + r_1 = r$, the total number of factors; the rank of $\mathbf{C}\,(1)$ is r_1. The goal of Bai and Ng (2004) is to determine r_1 and test if $\rho_i = 1$ when neither $\mathbf{f}_t$ nor v_{it} is observed. This approach is known as the PANIC (panel analysis of nonstationarity in idiosyncratic nd common components) approach. A preliminary issue that arises is to determine how many common factors, r, are necessary to capture the existing cross sectional correlation. To this end, we will employ the information criteria suggested by Bai and Ng (2002). They have been built with a similar spirit than the AIC and BIC criteria for time series, involving a trade-off between some measure of fit and a penalty for complexity. As the number of factors increase, the fit must improve but the penalty also increases. We compute all the criteria but following a relevant literature (Bai and Ng, 2002; Moon and Perron, 2007; Hurlin, 2010), we pay particular attention to the IC2 and BIC3 criteria which are expected to minimize the risk of overestimating the number of factors.[25] These criteria are applied to factors estimated by principal components on first differences

[24] We have performed both the IPS test and the Fisher-type tests on the demeaned series, and the (non-reported) results are fully consistent with results obtained without demeaning (reported in table 2).

[25] According to Bai and Ng (2002), the IC2 selects the true number of factors and dominate the other criteria. The BIC3 has been shown (Bai and Ng, 2002) to perform better than the others when $\min(T, N) \leq 20$ and T and N are roughly of the same size; this result holds even if the BIC3 does not satisfy the conditions for consistency when either N or T dominates the other exponentially.

(Bai and Ng, 2004). Recent literature suggests that a small number if unobserved common factors is sufficient to explain most of the variations in many macroeconomic variables (see e.g. Stock and Watson, 2002; Pesaran et al., 2013; Moon and Perron, 2007 and Hurlin, 2010). We start our analysis by following such a literature and apply the above discussed criteria by imposing that the maximum number of factors is 4 as in Pesaran et al. (2013). In almost all cases the criteria suggest that the number of unobserved factors, r, equals the maximum number we allowed, 4. This is the same result by Pesaran (2007) and by Pesaran et al. (2013) and it is not surprising given our sample sizes (see also Gutierrez, 2006). This suggests that the number of factors could be even higher than 4. However, given the possibility that the criteria over estimate the number of factors and the number of observations available, Pesaran et al. (2013) do not allow that the maximum number of factors could be greater than 4. We provide our main results doing the same thing but, as a robustness check available upon request, we also perform the procedure by Bai and Ng (2004) to test the stationarity in the common component, to identify the number of nonstationary common factors (if they exists) and to test the stationarity in the idiosyncratic component for different values of r in the range $1 - 15$ (**table A2**).

It is interesting to note that the nonstationarity of the idiosyncratic components can be tested without knowing if the factors are stationary, and vice versa. The only thing it is needed to know is the total number of factors, r. This is why, given the considerable uncertainity that surrounds the number of factors, we perform the tests for a large range of possible values of r. In order to test the nonstationarity of idiosyncratic components, Bai and Ng (2004) proceed pooling individual ADF t statistics obtained on de-factored components. Pooling, however, requires cross-sectional independence of the idiosyncratic components. Since the idiosyncratic components in a factor model can only be weakly correlated across units by construction while the factors involve strong correlation, it appears that the pooled tests based on de-factored components are likely to satisfy the required cross-sectional independence assumption. The two Fisher type statistics proposed by Bai and Ng (2004), and denoted P_e^c and Z_e^c provide strong evidence towards the rejection of the null hypothesis of nonstationarity of the idiosyncratic components for all the variables. For domestic R&D, foreign R&D and human capital the null is rejected irrespective of the value of r, the total number of common factors, while for TFP, the null is rejected in many cases. The rejection of the non stationarity of the idiosyncratic component does not imply that the series are stationary, since some of the common factors may be non-stationary. We have already tried to determine the total number of factors using information criteria on first differences and the next task is thus to determine how many of these factors are nonstationary. For this purpose, we follow Bai and Ng (2004) and proceed as follows. For $r = 1$, we use a standard ADF test, its rejection indicates that the unique common factor is stationary, while for $r > 1$, we look at the MQ_f and MQ_c statistics. The limiting distributions of these statistics are nonstandard, and critical values are reported in Bai and Ng (2004) up to 6 factors. The results provide a very clear-cut picture: for all the variables, whatever the test used, the number of nonstationary common factors, r_1, is always equal to the total number of common factors, r. This is the same result found by Hurlin (2010). The application of the PANIC approach by Bai and Ng (2004) suggest thus that the variables are nonstationary and that this property is the result of multiple nonstationary common factors combined with stationary idiosyncratic components.

Next, to further investigate the order of integration of the variables of interest, we follow Moon and Perron (2004), who also allow for r unobserved common factors but propose expressing the panel in an autoregressive form of the type

$$y_{it} = D_{it} + y_{it}^0,$$
$$y_{it}^0 = \rho_i y_{it-1}^0 + u_{it} \ ,$$
$$u_{it} = \zeta_i^{'} \mathbf{f}_t + v_{it}.$$

As in Bai and Ng (2004), data are first defactored and then panel unit root test statistics based on de-factored data are proposed. Moon and Perron (2004), however, consider the factors to be nuisance parameters and the unit root test is only based on the estimated idiosyncratic components. This is a relevant difference with respect to Bai and Ng (2004). The proposed test statistic uses defactored data obtained by projecting the data onto the space orthogonal to the factor loadings. The authors derive two modified t statistics - denoted t_a and t_b - which have a Gaussian distribution under the null hypothesis, and propose the implementation of feasible statistics - t_a^* and t_b^* - based on the estimation of long-run variances. To assess the robustness of the results to the choice of the kernel function used to estimate the long-run variances, we compute t_a^* and t_b^* with both quadratic spectral and Bartlett kernels. In Moon and Perron (2004), the above mentioned information criteria to detect the number of common factors are applied on residuals (rather than on first-differences). Such criteria provide the same results we obtained using PANIC tending to select the maximum number of factors which is allowed, that is 4. In almost all cases (except for foreign R&D), these tests strongly reject the unit root hypothesis (table A3). As for PANIC, we then perform the tests for all the possible values of r in the range 1-15 . for TFP, domestic R&D and human capital, in almost all cases these tests strongly reject the unit root hypothesis of the idiosyncratic components, while for foreign R&D the results of the tests dipend crucially to r. The null hypotesis is often not rejected for r in the range 5-9 while it is always rejected for $10 \leq r \leq 15$.[26]

Another test we perform is that proposed by Choi (2006), who uses a two-way error-component model rather than a factor model. Such a model differs from that proposed by Moon and Perron (2004) mainly because each cross-sectional unit is influenced homogeneously by a single factor, so that the component $\zeta_i^{'} \mathbf{f}_t$ is replaced with f_t, and $\zeta_i = \zeta$ for all i. To perform this test, the common component and cross-sectional correlations are eliminated by GLS detrending (Ellitot et al., 1996) and cross-sectional demeaning. Three Fisher-type statistics - denoted P_m, Z, L^* - are obtained by combining p-values from the ADF test applied to each (detrended and demeaned) individual time series. From an applied perspective, such an approach can be viewed as complementary to Moon and Perron (2004). Indeed, Gutierrez (2006) has shown through Monte Carlo simulation that Moon and Perron's tests have a better size than Choi's when the common factor influences the cross-sectional units heterogeneously; however, Choi's test performs well under the more restrictive assumption that the cross-sectional units are homogeneously influenced by the common factor, and in a few cases, it outperforms Moon and Perron's test in terms of power. Additionally, for such a test (in **table A4**), the choice of the lags is crucial, and when such a choice is made with the AIC (or other non-reported criteria), it clearly goes toward the rejection of the null hypothesis.

Finally, we implement the tests which have been proposed by Pesaran (2007) and Pesaran et al. (2013). Instead of basing the unit roots tests on deviations from the estimated factors, they augment standard ADF regressions with cross sectional averages. In the case of a single unobserved common factor, Pesaran (2007) suggests augmenting the standard (individual) ADF regression with the cross-sectional average of first differences ($\Delta \bar{y}_t = N^{-1} \sum_{i=1}^{N} \Delta y_{it} = \bar{y}_t - \bar{y}_{t-1}$) and lagged levels ($\bar{y}_{t-1}$) of the individual series, which are $\sqrt{N}$- consistent

[26] Results available from the authors upon request.

estimators for the rescaled factors $\bar{\zeta} f$ and $\bar{\zeta} \sum_{j=0}^{t-1} f_j$, respectively, where $\bar{\zeta} = N^{-1} \sum_{i=1}^{N} \zeta_i$. This expression gives the cross-sectionally augmented Dickey-Fuller (CADF) statistics; the individual CADF statistics are used to develop a modified version of the IPS test named CIPS. However, Monte Carlo experiments show that Pesaran's CIPS test has desirable small sample properties in the presence of a single unobserved common factor but show size distortions if the number of common factors exceeds unity. Recently, Pesaran et al. (2013) extend Pesaran's CIPS test to the case of a multifactor error structure. They propose to utilise the information contained in a number of k additional variables, x_{it} , that are assumed to share the common factors of the series of interest, y_{it}. In particular, they propose two tests. The first test, CIPS, is an extension of the test proposed in Pesaran (2007) and is based on the average of t-ratios from ADF regressions augmented by the cross section averages of the dependent variable as well as k additional regressors. The second test, CSB, exploits cross-sectional augmentation for the Sargan–Bhargava test. It is worth noting that the perspective of these tests is quite different with respect to that of Bai and Ng (2004). Indeed, while Bai and Ng (2004) consider whether the source of non-stationarity is due to the common factors and/or the idiosyncratic components, neither of which are observed directly, Pesaran et al. (2013) aim to test for the presence of a unit root in the y_{it} process, which is observed. In doing so, they adopt an auto-regressive specification augmented with common factors and the unit root test if performed by testing that the auto-regressive component of the specification expressed in first difference, δ_i, is 0 for all i against the alternative which can be expressed as $\delta_i = 0$ for some countries but $\delta_i < 0$ for some others. In such a set up, they rule out the possibility of the factors having unit roots since otherwise all series in the panel could be $I(1)$ irrespective of whether $\delta_i = 0$ or not. To deal with the uncertainty that surrounds the value of r, we follow Pesaran et al. (2013) and consider the application of the CIPS and CSB tests allowing the number of factors, $r = k+1$, to take any value between 1 and 4 and present the results of these tests for all possible combinations of regressors. When $k = 3$, we implicitly assume that the four observed variables used in the econometric analysis, f_{it}, S_{it}^{d}, S_{it}^{f}, H_{it} share the same common factors. While Pesaran et al. (2013) set the lag order, $\widehat{p} = \left[4\left(T/100\right)^{1/4}\right]$, in our case, this rule gives $\widehat{p} = 2.6$. However, for $p = 3$ we encounter computational problems in many cases and even for $p = 2$ we had the same problems in few cases. To provide comparable results for all possible combinations of regressors we thus set $p = 1$, while results for $p = 0$ and $p = 2$ are available upon request. The results are summarised in **Table A5**. The test outcomes are as follows. When the CSB test is used, the null hypothesis of a panel unit root is rejected in almost all cases for all the variables under investigation. When the CIPS test is used, the results are mixed and crucially depend on the variables which are used to augment the ADF regression. Rejection of the null hypothesis is more likely to appear when $r \geq 3$. It is worth noting that according to Pesaran et al. (2013), for small T, the CSB test has higher power than that of CIPS, and should thus be preferred in such cases.

References

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

[2] Bai, J. and Ng, S. (2004). 'A Panic Attack on Unit Roots and Cointegration', *Econometrica*, 72, 1127–1177.

[3] Baltagi, B.H., Bresson, G. and Pirotte, A. (2007), 'Panel Unit Root Tests and Spatial Dependence', *Journal of Applied Econometrics*, 22, 339–360.

[4] Breitung, J. and Pesaran, M.H. (2008). 'Unit roots and cointegration in panels', In: Matyas, L., Sevestre, P. (Eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, Springer, Berlin.

[5] Choi, I. (2001). 'Unit Root Tests for Panel Data', *Journal of International Money and Finance*, 20, 249-272.

[6] Choi, I. (2006). 'Combination Unit Root Tests for Cross-Sectionally Correlated Panels'. In: D. Corbae, S. N. Durlauf, and B. E. Hansen (eds.): *Econometric Theory and Practice: Frontiers of Analysis and Applied Research: Essays in Honor of Peter C. B. Phillips*. Cambridge University Press, Chapt. 11, 311–333.

[7] Elliott, G., Rothenberg, T. and Stock, J. (1996). 'Efficient Tests for an Autoregressive Unit Root', *Econometrica*, 64, 813–836.

[8] Fisher, R. A. (1932). *Statistical Methods for Research Workers*, Oliver & Boyd, Edinburgh, 4th Edition.

[9] Gutierrez, L. (2006). 'Panel Unit-root Tests for Cross-sectionally Correlated Panels: A Monte Carlo Comparison', *Oxford Bulletin of Economics and Statistics*, 68, 519-540.

[10] Hurlin, C. (2010). 'What would Nelson and Plosser find had they used panel unit root tests?', *Applied Economics*, 42, 1515-1531.

[11] Levin, A., Lin, C. and Chu, C.J. (2002). 'Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties', *Journal of Econometrics*, 108, 1–24.

[12] Maddala, G.S. and Wu, S. (1999). A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test, *Oxford Bulletin of Economics and Statistics*, 61, 631–652.

[13] Moon, R. and Perron, B. (2004), 'Testing for Unit Root in Panels with Dynamic Factors', *Journal of Econometrics*, 122, 81-126.

[14] Moon, R. and Perron, B. (2007), 'An empirical analysis of nonstationarity in a panel of interest rates with factors', *Journal of Applied Econometrics*, 22, 383-400.

[15] Pesaran, M.H. (2007). 'A Simple Panel Unit Root Test in the Presence of Cross Section Dependence', *Journal of Applied Econometrics*, 22, 265–312.

[16] Pesaran, M. H., V. Smith and T. Yamagata (2013). 'Panel unit root tests in the presence of a multifactor error structure', *Journal of Econometrics,* 175, 94-115.

[17] Stock, J. H. and M. W. Watson (2002). 'Macroeconomic forecasting using diffusion indexes', *Journal of Business and Economic Statistics,* 20, 147.162.

[18] Westerlund, J. and J. Breitung (2009). 'Myths and facts about panel unit root tests', Working Papers in Economics No 380, University of Gothenburg.

TABLE A1

IPS (W_{-t-bar}) and Fisher-type statistics (P and Z)

lag order (k)	Test	$\log f$	$\log S^d$	$\log S^f$	$\log H$
1	W_{-t-bar}	-2.14861 (0.0158)	-1.75052 (0.0400)	-2.33921 (0.0097)	-1.55144 (0.0604)
1	P	64.6462 (0.0229)	70.2988 (0.0071)	71.8629 (0.0050)	60.6723 (0.0483)
1	Z	-2.31288 (0.0104)	-1.58178 (0.0569)	-2.47066 (0.0067)	-1.74831 (0.0402)
2	W_{-t-bar}	0.55338 (0.7100)	0.16992 (0.5675)	-1.15137 (0.1248)	-2.36996 (0.0089)
2	P	29.0843 (0.9593)	49.1626 (0.2740)	47.2918 (0.3397)	65.1072 (0.0209)
2	Z	1.31609 (0.9059)	1.30367 (0.9038)	-0.63744 (0.2619)	-1.75297 (0.0398)
3	W_{-t-bar}	-0.25439 (0.3996)	0.21641 (0.5857)	2.09292 (0.9818)	-2.17194 (0.0149)
3	P	46.5330 (0.3685)	39.1391 (0.6797)	15.3795 (1.0000)	66.5139 (0.0158)
3	Z	0.81728 (0.7931)	1.40485 (0.9200)	3.56924 (0.9998)	-1.33588 (0.0908)
aic	W_{-t-bar}	-3.32945 (0.0004)	-1.84356 (0.0326)	-3.58654 (0.0002)	-2.97393 (0.0015)
aic	P	79.1326 (0.0009)	71.8741 (0.0050)	79.3389 (0.0009)	79.8767 (0.0008)
aic	Z	-3.14469 (0.0008)	-1.1403 (0.1271)	-3.58923 (0.0002)	-2.60986 (0.0045)
		$\bar{k}$= 1.00	$\bar{k}$ =1.55	$\bar{k}$ =1.45	$\bar{k}$ =1.36
bic	W_{-t-bar}	-2.57982 (0.0049)	-1.81427 (0.0348)	-3.68351 (0.0001)	-2.91527 (0.0018)
bic	P	70.0827 (0.0074)	73.6950 (0.0033)	80.9023 (0.0006)	79.4714 (0.0008)
bic	Z	-2.41172 (0.0079)	-1.18474 (0.1181)	-3.68534 (0.0001)	-2.58006 (0.0049)
		$\bar{k}$= 0.77	$\bar{k}$ =1.27	$\bar{k}$= 1.32	$\bar{k}$= 1.32
hqic	W_{-t-bar}	-3.32945 (0.0004)	-1.51511 (0.0649)	-3.58654 (0.0002)	-2.97393 (0.0015)
hqic	P	79.1326 (0.0009)	70.5376 (0.0067)	79.3389 (0.0009)	79.8767 (0.0008)
hqic	Z	-3.14469 (0.0008)	-0.77733 (0.2185)	-3.58923 (0.0002)	-2.60986 (0.0045)
		$\bar{k}$= 1.00	$\bar{k}$ =1.41	$\bar{k}$ =1.45	$\bar{k}$ =1.36

Notes:

p-value between brackets.

TABLE A2

Bai and Ng test

		Idiosyncratic component		Common factors	
	Number of factors			Nonstationary factors $\widehat{r}_1$	
	$\widehat{r}_{BIC3}$	P_e^c	Z_e^c	MQ_f	MQ_c
		p-value			
Model with intercept					
$\log f$	2	0.210	0.202	2	2
$\log S^d$	4	0	8.76E-10	4	4
$\log S^f$	4	0	8,75E-10	4	4
$\log H$	4	0	0	4	4
Model with intercept and trend					
$\log f$	2			2	2
$\log S^d$	4			4	4
$\log S^f$	4			4	4
$\log H$	4			4	4

Notes:

$\widehat{r}_{BIC3}$ss the estimated number of common factors, based on BIC3 criterion.

Here again we impose that the maximum number of factors is 4.

For the idiosyncratic components, only pooled unit root tests are reported.

Individual ADF t statistics based on de-factored components are available upon request.

The pooled test for the idiosyncratic component is valid only for the intercept case.

In the linear trend case, the limiting distribution of the pooled test is not a DF type distribution.

For the common factors components, the estimated number $\widehat{r_1}$of independent stochastic trends is reported (5% level).

We provide available upon request the results of the test for all values of r in the range 1-15.

They strongly confirm the finding of nonstationary common factors and possibly stationary idiosyncratic components.

TABLE A3

Moon and Perron test

	r	kernel	$\log f$	$\log S^d$	$\log S^f$	$\log H$
m*(IC2)			3	4	4	4
m*(BIC3)			2	4	4	4
t_a^*	1	QS	-2.6604(0.0039)	-1.7862 (0.0370)	-0.7206 (0.2356)	-0.2279(0.4099)
	1	B	-2.9104(0.0018)	-1.7852 (0.0371)	-0.7889(0.2151)	-0.2382 (0.4059)
	2	QS	-4.2229(1.2057e-005)	-1.6096 (0.0537)	-0.7984 (0.2123)	-4.5750 (2.3815e-006)
	2	B	-4.4621 (4.0585e-006)	-1.6856 (0.0459)	-0.6783 (0.2488)	-4.5750 (2.3815e-006)
	3	QS	-4.3942 (5.5589e-006)	-0.7444 (0.2283)	-4.2315 (1.1605e-005)	-3.5113 (2.2296e-004)
	3	B	-4.5508 (2.6726e-006)	-0.7703 (0.2206)	-3.7149 (1.0164e-004)	-3.6976 (1.0883e-004)
	4	QS	-4.9478 (3.7529e-007)	-3.6826 (1.1544e-004)	-0.4179 (0.3380)	-3.6500 (1.3112e-004)
	4	B	-5.0309 (2.4409e-007)	-4.6880 (1.3792e-006)	-0.5859 (0.2790)	-4.0252 (2.8465e-005)
t_b^*	1	QS	-2.6563 (0.0040)	-1.1354 (0.1281)	-0.6198 (0.2677)	-0.0747 (0.4702)
	1	B	-2.7187(0.0033)	-1.0977 (0.1362)	-0.7003 (0.2419)	-0.0834 (0.4668)
	2	QS	-4.6476 (1.6787e-006)	-0.9969 (0.1594)	-0.8512 (0.1973)	-5.5644 (1.3150e-008)
	2	B	-4.4694 (3.9228e-006)	-1.1119 (0.1331)	-0.7572 (0.2245)	-4.6502 (1.6583e-006)
	3	QS	-4.6777 (1.4502e-006)	-0.3774 (0.3530)	-5.9500 (1.3410e-009)	-6.6819 (1.1796e-011)
	3	B	-4.6046 (2.0666e-006)	-0.4235 (0.3360)	-3.5446 (1.9659e-004)	-6.3602 (1.0077e-010)
	4	QS	-6.3634 (9.8699e-011)	-4.6224 (1.8964e-006)	-0.3629 (0.3583)	-6.2396 (2.1941e-010)
	4	B	-5.1844 (1.0838e-007)	-4.7808 (8.7295e-007)	-0.3591 (0.3598)	-7.2367 (2.2992e-013)

Notes:

p-value between brackets. QS: Quadratic spectral kernel. B: Bartlett kernel. $r$is the number of factors.

Here we impose that the maximum number of factors is 4 as in Pesaran et al. (2013). In almost all cases the criteria suggest that the number of unobserved factors, r, equals the maximum number we allowed, 4. This is the same result by Pesaran (2007) and by Pesaran et al. (2013) (see also Gutierrez, 2006).

This suggests that the number of factors could be even higher than 4. However, given the possibility that the criteria over estimate the number of factors and the number of observations available, Pesaran et al. (2013) do not allow that the maximum number of factors could be greater than 4.

We provide our main results doing the same thing but, as a robustness check available upon request, we also perform the test for all the possible values of $r$in the range $1-15$.

The null hypotesis is generally not rejected for $r$in the range 1-9 while it is always rejected for $10 \leq r \leq 15$.

TABLE A4

Choi (2006) Fisher-type statistics (P_m, Z, L^*)

lag order (k)	Test	$\log f$	$\log S^d$	$\log S^f$	$\log H$
1	P_m	1.9495 (0.0256)	7.1271 (5.1237e-013)	10.8056 (0)	0.5401(0.2946)
1	Z	-2.4621 (0.0069)	-3.8111 (6.9186e-005)	-6.7776 (6.1086e-012)	-1.3161 (0.0941)
1	L^*	-2.3866 (0.0085)	-4.6426 (1.7204e-006)	-7.5693 (1.8768e-014)	-1.2233 (0.1106)
2	P_m	-1.3238 (0.9072)	1.5406 (0.0617)	7.3856 (7.5939e-014)	0.3634 (0.3581)
2	Z	0.3686 (0.6438)	0.5268 (0.7008)	-5.4012 (3.3094e-008)	-1.0541 (0.1459)
2	L^*	0.3375 (0.6321)	1.1224 (0.8691)	-5.5847 (1.1708e-008)	-0.9689 (0.1663)
3	P_m	-0.1229 (0.5489)	-0.7096 (0.7610)	-0.6469 (0.7411)	-0.1871 (0.5742)
3	Z	0.0251 (0.5100)	1.1386 (0.8726)	-0.5273 (0.2990)	-0.3876 (0.3492)
3	L^*	-0.0452 (0.4820)	1.4084 (0.9205)	-0.4800 (0.3156)	-0.2621 (0.3966)
aic	P_m	2.6412 (0.0041)	5.9129 (1.6806e-009)	12.2388 (0)	1.9548 (0.0253)
aic	Z	-2.5990 (0.0047)	-2.6092 (0.0045)	-7.7401 (4.9682e-015)	-1.2459 (0.1064)
aic	L^*	-2.5389 (0.0056)	-2.7818 (0.0027)	-8.5914 (4.2945e-018)	-1.1490 (0.1253)

Notes:

p-value between brackets.

TABLE A5

CIPS and CSB tests - Pesaran (2007) and Pesaran et al. (2013)

	$\log f$		$\log S^d$		$\log S^f$		$\log H$	
	$CIPS(\widehat{p})$	$CSB(\widehat{p})$	$CIPS(\widehat{p})$	$CSB(\widehat{p})$	$CIPS(\widehat{p})$	$CSB(\widehat{p})$	$xCIPS(\widehat{p})$	$CSB(\widehat{p})$
$r = 1$	-2.146	0.146*	-2.152	0.192*	-1.935	0.161*	-2.193	0.131*
$r = 2$								
$\overline{x}_t = (\overline{\log S^d})$	-3.001*	0.102*	-2.074	0.167*	-1.511	0.123*	-1.913	0.177*
$\overline{x}_t = (\overline{\log S^f})$	-2.671	0.105*	-2.071	0.156*	-1.453	0.048	-2.27	0.143*
$\overline{x}_t = (\overline{\log H})$	-2.307	0.13*	-2.649	0.158*	-1.651	0.133*	-2.504	0.091*
$r = 3$								
$\overline{x}_t = (\overline{\log S^d}, \log S^f)$	-2.821	0,092*	-1.8	0.132*	-3.128	0.039	-1.693	0.126*
$\overline{x}_t = (\overline{\log S^d}, \overline{\log H})$	-3.119*	0.093*	-2.068	0.11*	-1.973	0.071*	-2.716	0.102*
$\overline{x}_t = (\overline{\log S^f}, \overline{\log H})$	-3.377*	0.09*	-1.986	0.125*	-1.144	0.035	-3.555*	0.1067*
$r = 4$								
$\overline{x}_t = (\overline{\log S^d}, \log S^f, \log H)$	-2.614	0.081*	-2.896	0.086*	-2.992*	0.0244	-5.621*	0.0954*

Notes:

Pesaran et al. (2013) set the lag order, $\widehat{p} = \left[4\,(T/100)^{1/4}\right]$. Tthis gives $\widehat{p} = 2.6$. For $p = 3$we had computational problems. For $p = 2$we had problems in few cases.

To provide comparable results for all possible combinations of regressors we set $p = 1$. Results for $p = 0$and $p = 2$are available upon request.

The variables under the heading $\overline{x}_t$indicate the regressors used for cross-section augmentation in addition to $\overline{y}_t$. In the case where $r = 1$no additional regressors are used (Pesaran, 2007).

For the selected lag order, the 5% critical values for the CIPS are -2.67, -2.83, -2.88, -2.97 for r=1, 2, 3, 4, respectively.

The 5% critical values for the CSB are 0.104, 0.087, 0.071, 0.057 for r=1, 2, 3, 4, respectively. * Denotes rejection at 5% level.

According to Pesaran et al. (2013), for small T, the CSB test has higher power than that of CIPS, and should thus be preferred in such cases.