Estimation of production functions: The Spanish regional case

Josep Lluís Carrion-i-Silvestre* Laura Surdeanu†

March 1, 2012

Abstract

In this study, we revisit the cointegration relation between output, physical capital, human capital, public capital and labor for 17 Spanish regions observed over the period 1964-2000. The novelty of our approach is that we allow for cross-section dependence between the members of the panel. To see if the variables are cointegrated or not, we employ two different techniques at the panel level. More exactly, we compare the statistics from the single equation method of Banerjee and Carrion-i-Silvestre (2011) with those from the VAR framework of Carrion-i-Silvestre and Surdeanu (2011). Moreover, using the VAR method, we identify at least one common cointegrating relation among output, physical capital, human capital, public capital and labor. Finally, we use several estimators to estimate the long-run relation between these variables.

JEL Classification: C12, C22

Keywords: Panel data, cross-section dependence,

*AQR-IREA Research Group. Department of Econometrics, Statistics and Spanish Economy. University of Barcelona. Av. Diagonal, 690. 08034 Barcelona. Tel: +34 934021826. Fax: +34 934021821. E-mail: carrion@ub.edu

†Department of Econometrics, Statistics and Spanish Economy. University of Barcelona. Av. Diagonal, 690. 08034 Barcelona. E-mail: lsurdesu7@docd2.ub.edu

1 Introduction

Ever since Aschauer (1989) added the stock of public capital to the Cobb-Douglas production function, the interest in this field rose considerably. The production function relates the output of a firm, region or country to different combinations of factors of production, usually physical capital and labor. It is one of the key concepts in economics, making it attractive in the empirical work.

Early studies of production function employ time series data, focusing on an individual region or country. For example, for the case of the aggregated Spanish economy, Serrano (1997) uses annual data observed over the period 1964-1991 and finds no evidence of cointegration between gross value added, human capital, private physical capital and labor. Sosvilla-Rivero and Alonso (2005) obtain a different result and find that in Spain, gross domestic product, physical capital, human capital and labor define a cointegration relationship. They employ annual data that is observed over the period 1910-1995. The contradictory results indicate that the empirical evidence from the time-series studies is mixed. One plausible explanation is the low power of the univariate unit root and cointegration tests that were used in these studies.

More recent studies show that the power of unit root and cointegration test statistics can be improved if we consider panel data techniques. Thus, another category of studies that estimate long-run production functions using panel data tools. Some examples for Spanish data are Serrano (1996), Bajo and Díaz (2005) and Márquez, Ramajo and Hewings (2011). Serrano (1996) employs a panel of regional data observed over the years 1980-1991 and simply avoids the risk of a spurious regression working with a model that relates the gross value added, human capital, physical capital and labor in first difference.[1] Bajo and Díaz (2005) go one step further and add public capital to the production function. The authors investigate the relation between gross domestic product, private capital, public capital, human capital and labor using data for the 17 Spanish regions over 1965-1995. They find that these variables are cointegrated. Recently, Márquez, Ramajo and Hewings (2011) investigate the relation between gross value added, public capital, private capital and labor for the 17 Spanish regions observed over the period 1972-2000. The authors find that indeed, there is cointegration between these variables.

One critical problem with the earlier panel data studies for the Spanish regions is the assumption of cross-section independence that they make. This is an unrealistic and far too restrictive assumption, especially since regions are so closely related to each other. If the independence assumption is violated then we might expect to have, on the one hand, biased and inconsistent estimates and, on the other hand, spurious statistical inference – see Andrews (2005). More specifically, in the case of non-stationary panel data, the unaccounted cross-section dependence might lead to conclude that panel data is actually stationary when in fact it might be non-stationary – see Banerjee, Marcellino and Osbat (2005). Similarly, the panel data cointegration test statistics might indicate than there are more cointegrating relations than there exist – see Carrion-i-Silvestre and Surdeanu (2011). Consequently, accounting for the presence of cross-section dependence is crucial to draw meaningful conclusions from the analysis.

Cross-section dependence is more a recurrent than a rare characteristic that is present in macroeconomic time series of different units – i.e., countries, regions or sectors. There is different sources of cross-section dependence that can be expected to affect the units of a panel data set.

[1] Note that this leads to the estimation of a short-run relationship among the variables since the long-run one would require to use cointegration techniques.

For instance, cross-section dependence is usually caused by the presence of common shocks (oil price shocks or financial crises) or the existence of local productivity spillover effects. Further, the economic literature on output stochastic convergence implies the existence of a long-run relation (cointegration relation) among the different economies, so that the use of macroeconomic variables such as the output or production should account for the presence of this long-run relation across the cross-section – the so-called cross cointegration concept, as defined in Banerjee, Marcellino and Osbat (2005). This implies that cross-section dependence is more the rule than the exception. Therefore, in country or regional level studies is practically impossible to ignore the effect of cross-section dependence in the analysis of the models that are to be estimated. Bai and Ng (2002, 2004) recognized early on this problem and laid down the foundation of the theoretical panel framework with common factors.The use of common factor models is particularly useful to capture the presence of cross-section that is pervasive or strong, i.e., the sort of cross-section dependence that affects all units of the panel data.

However, as Banerjee, Eberhardt and Reade (2010) mention, the empirical work on the estimation of production function in panel data using the common factor technique is relatively limited. The most relevant example to our study is the work by Costantini and Destefanis (2009). They analyze the production function for the Italian regions over the 1970-2003 period and find that the regional value added, physical capital and human capital augmented labor are cointegrated. They also find that ignoring the cross-section dependence biases upward the estimates for the returns to scale. Note that the authors use the single equation framework while our paper focus on both the single equation and vector autoregressive (VAR) frameworks. The advantage of a VAR model is knowing exactly how many cointegration relations or, conversely, how many stochastic trends exist among the units of the panel. To the best of our knowledge, none of the existing studies for the Spanish economy take into consideration the cross-section dependence among the members of the panel in a VAR model.

In this paper, we reexamine the cointegration relation among the output, physical capital, human capital, public capital and labor for the 17 Spanish regions observed over the period 1964-2000. We model the cross-section dependence through the specification of a common factor model. In order to analyze the order of integration of the variables in our model we apply the panel data unit root test statistics in Bai and Ng (2004), Moon and Perron (2004) and Pesaran (2007) and the panel data stationarity test statistics in Hadri (2000). All these test statistics account for the existence of cross-section dependence in different ways. In general, the application of these statistics leads to the same qualitative conclusion, i.e., that all panel data sets are characterized as non-stationary panels. We then use the panel cointegration statistics recently proposed in Carrion-i-Silvestre and Surdeanu (2011) using a VAR framework and in Banerjee and Carrion-i-Silvestre (2006, 2011) for the single equation framework. All the cointegration statistics allow for cross-section dependence through the use of common factors. By using the vector autoregressive model we are able to determine the exact number of cointegrating vectors that exist in the model. Finally, we compute the panel data estimators proposed in Pesaran (2006), Bai, Kao and Ng (2009) and Kapetanios, Pesaran and Yamagata (2011) to estimate the long-run production function for the Spanish regions.

The structure of this paper is as follows. Section 2 presents the model for panel data and the data used in this study. In the Section 3 we present the econometric methodology while the results are presented in Section 4. Finally, the paper concludes with Section 5.

2 Specification of the model and the data

This section presents the production function and the data used in this study. We use a modified Cobb-Douglas production function used also by Bajo and Díaz (2005) that has the following form:

$$Y_{i,t} = A_{i,t}\, F\left(K_{i,t}, G_{i,t}, H_{i,t}, L_{i,t}\right), \tag{1}$$

where $i = 1,\ldots,N$ represents the cross-section dimension and $t = 1,\ldots,T$ represents the time-series dimension. The variable $Y_{i,t}$ is the output that depends on the private capital $(K_{i,t})$, the public capital $(G_{i,t})$, the human capital $(H_{i,t})$ and the labor $(L_{i,t})$. The variable $A_{i,t}$ is the total factor productivity (TFP), which is the part of the output not explained by the inputs. Next, we express the production function in per worker terms, obtaining:

$$Y_{i,t}/L_{i,t} = A_{i,t}/L_{i,t}\; f\left(K_{i,t}/L_{i,t}, G_{i,t}/L_{i,t}, H_{i,t}/L_{i,t}\right). \tag{2}$$

As it is well known, TFP represents the unobservable part of the production function and usually reflects the technological progress of the respective country or region. Further, if the technology represent the cumulation of the innovations and progress efforts made by economic agents, we should expect the TFP to be a non-stationary stochastic process. However, since the TFP cannot be measured directly, the empirical researchers estimate it as the residual of the estimated production function. Although intuitive, this approach causes serious econometric and interpretation problems. First, if not appropriately accounted for, the non-stationarity nature of the TFP would imply that the estimation of the production function is, in fact, a spurious regression. Therefore, panel data cointegration test statistics would lead to conclude that the variables involved in the production function are not cointegrated. Second, the issue that part of the technology that is available is common to all the economies implies a source of cross-section dependence, which needs to be accounted for in order to obtain meaningful conclusions of the panel cointegration test statistics. As can be seen, the specification of a common factor model can capture this unobservable variable that is difficult to approximate.

We take advantage of the recent developments in the field of non-stationary panel data and decompose the TFP into an unobserved common factor component $F_t'\lambda_i$ and an idiosyncratic component $e_{i,t}$. The common factor approach allows us to capture the effect of common shocks that affect the countries or regions, making it a desirable way to model the cross-section dependence. Therefore, following Costantini and Destefanis (2009) and Banerjee et al. (2010), TFP is modeled through the common factor specification given by:

$$A_{i,t}/L_{i,t} = e^{D_{i,t}+F_t'\lambda_i+e_{i,t}}, \tag{3}$$

where $D_{i,t}$ denotes the deterministic component being either a constant $(D_{i,t} = \mu_i)$ or a linear time trend $(D_{i,t} = \mu_i + \delta_i t)$. Assuming a Cobb-Douglas function and taking the natural logarithm of the variables from (2) and (3), we obtain the single equation model:

$$\begin{aligned} y_{i,t} &= a_{i,t} + (\alpha+\beta+\gamma+\delta-1)\, l_{i,t} + \alpha k_{i,t} + \beta g_{i,t} + \delta h_{i,t} \\ a_{i,t} &= D_{i,t} + F_t'\lambda_i + e_{i,t}, \end{aligned} \tag{4}$$

where $y_{i,t} = \ln(Y_{i,t}/L_{i,t})$, $a_{i,t} = \ln(A_{i,t}/L_{i,t})$, $l_{i,t} = \ln L_{i,t}$, $k_{i,t} = \ln(K_{i,t}/L_{i,t})$, $g_{i,t} = \ln(G_{i,t}/L_{i,t})$ and $h_{i,t} = \ln(H_{i,t}/L_{i,t})$. Following the existing contributions in the literature, we have essayed two alternative measures for public and human capital. First, we have used the total public capital ($g_{i,t}$) or the productive public capital ($gp_{i,t} = \ln(Gp_{i,t}/L_{i,t})$). Second, the human capital has been proxied by the rate of employes with at least a secundary school studies over the total number of employes ($h_{i,t}$) and the average number of schooling years ($hs_{i,t} = \ln(Hs_{i,t}/L_{i,t})$) – see Serrano (1996). The use of these variables defines up to four different model specifications depending on whether total or productive public capital is used and on whether we use $h_{i,t}$ or $hs_{i,t}$ to proxy the human capital. For a detailed description of the variables and the sources, see the appendix. These are the variables that we will be using throughout the rest of the paper.

The data employed in our study contains annual observations for the $N = 17$ Spanish regions observed over the $T = 37$ year period from 1964 to 2000. We collect the data from the BD.MORES database provided by the Spanish Ministry of Economy and Finance and from the Instituto Valenciano de Investigaciones Económicas (IVIE). The Spanish regions are: Andalucía, Aragón, Asturias, Baleares, Canarias, Cantabria, Castilla y León, Castilla-La Mancha, Catalonia, Comunidad Valenciana, Extremadura, Galicia, Madrid, Murcia, Navarra, País Vasco and La Rioja. The picture of the variables can be found in Figure 1, which evidences, first, the clear trend pattern shown by the four variables of the model and, second, the comovement (cross-section dependence) that seem to be present in their evolution.

3 Econometric methodology

In this section we describe the tools that are used throughout the paper in order to analyze our dataset. The order in which we present the econometric procedures is the one that will be followed when applying them in the empirical estimation of the regional Spanish production function. Since the validity of the panel data unit root, stationarity and cointegration test statistics requires to assess whether the units in the panel data set are cross-section dependent, we first start the discussion describing Pesaran's (2004) CD test statistic that test the null hypothesis of cross-section independence against the alternative hypothesis of cross-section dependence. It should be bear in mind that in our case we are analyzing macroeconomic time series of highly economic integrated regions, provided that the regions belong to the same economy – see Figure 1. Therefore we can expect the presence of cross-section dependence among the units of the panel. Second, we present the panel unit root and stationarity tests that control for the presence of cross-section dependence in different ways. To be specific, we apply the panel data unit root tests in Bai and Ng (2004), Moon and Perron (2004) and Pesaran (2007), and the panel stationarity tests in Hadri (2000). Finally, we summarize recent developments in panel cointegration testing and estimation that take into consideration the cross-section dependence. In this regard, we first focus on the system-based approach in Carrion-i-Silvestre and Surdeanu (2011) and in the single-equation-based procedures proposed in Banerjee and Carrion-i-Silvestre (2006, 2011) to test whether there exist a cointegration relationship among the variables of the model. Second, we proceed to estimate the cointegration relationships using the proposals in Bai, Kao and Ng (2009) and Kapetanios, Pesaran and Yamagata (2011).

3.1 Cross-section dependence

In this subsection we test the null hypothesis of cross-section independence against the alternative hypothesis of cross-section dependence using the approach suggested in Pesaran (2004). For notational convenience, throughout this and the next section, we will use $y_{i,t}$ as the variable of interest, although the same applies for the other variables of the system – i.e., $k_{i,t}$, $g_{i,t}$, $gp_{i,t}$, $h_{i,t}$ and $hs_{i,t}$. The test statistic is based on the average of pair-wise Pearson's correlation coefficients $\hat{p}_j$, $j = 1,2,\ldots,n$, $n = N(N-1)/2$, of the residuals $\varepsilon_{i,t}$ obtained from the following augmented Dickey-Fuller (ADF) type regression equation:

$$\Delta y_{i,t} = \mu_i + \delta_i t + \alpha_{i,0} y_{i,t-1} + \sum_{j=1}^{p_i} \alpha_{i,j} \Delta y_{i,t-j} + \varepsilon_{i,t}, \tag{5}$$

$i = 1,\ldots,N$. Pesaran's CD test is based on averaging all pair-wise correlation coefficients ($\hat{\rho}_{i,j}$) of the Ordinary Least Squares (OLS) estimated residuals $\hat{\varepsilon}_{i,t}$ in (5):

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{i,j} \right), \tag{6}$$

with $i = 1,\ldots,N-1$ and $j = i+1,\ldots,N$. Under the null hypothesis of cross-section independence, the CD statistic of Pesaran (2004) converges to the standard normal distribution. Pesaran (2004) shows that the CD statistic has the correct size and satisfactory power even in small samples, making it attractive in the empirical research.

3.2 Panel data unit root and stationarity test statistics

The panel data unit root tests that are applied in this paper are those of Bai and Ng (2004), Moon and Perron (2004) and Pesaran (2007), which that take into account the presence of cross-section dependence by specifying a model of common factors. Of the three approaches the most general one is the one in Bai and Ng (2004), provided that it allows to test the order of integration of the idiosyncratic and common components in a separate way. For exposition convenience, we first present Pesaran (2007) approach, then Moon and Perron (2004) and, finally, we discuss the proposal in Bai and Ng (2004). In addition and as a confirmatory analysis, we also compute the panel data stationarity test statistics in Hadri (2000).

3.2.1 Pesaran (2007) panel data unit root test statistic

The approach in Pesaran (2007) assumes that the cross-section dependence is driven by one unobservable stationary common factor, which can be proxied using cross-section averages of the units that define the panel data set. For the case of uncorrelated residuals, the starting regression has the following form:

$$\Delta y_{i,t} = \mu_i + \delta_i t + \alpha_i y_{i,t-1} + \lambda_i f_t + \varepsilon_{i,t},$$

where $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$, f_t denotes the unobserved common factor and $\varepsilon_{i,t}$ is the idiosyncratic error. The common factor f_t can be proxied by the cross-section mean of $y_{i,t}$ (or $\bar{y}_t = N^{-1} \sum_{i=1}^{N} y_{i,t}$) and its lagged values ($\bar{y}_{t-1}, \bar{y}_{t-1}\ldots$). Pesaran (2007) notes that $\bar{y}_t$ and $\bar{y}_{t-1}$ (or $\bar{y}_{t-1}$ and $\Delta\bar{y}_t$) are sufficient for eliminating the effect of the common factor. Therefore, after substituting the proxies we obtain the modified cross-sectionally ADF (CADF) regression:

$$\Delta y_{i,t} = \mu_i + \delta_i t + \alpha_{i,0} y_{i,t-1} + \xi_i \bar{y}_{t-1} + \eta_{i,0} \Delta\bar{y}_t + e_{i,t}.$$

In the case of correlated disturbance terms, the CADF regression equation is given by:

$$\Delta y_{i,t} = \mu_i + \delta_i t + \alpha_{i,0} y_{i,t-1} + \sum_{j=1}^{p} \alpha_{i,j} \Delta y_{i,t-j} + \xi_i \bar{y}_{t-1} + \sum_{j=0}^{p} \eta_{i,j} \Delta \bar{y}_{t-j} + e_{i,t}. \tag{7}$$

One of the panel unit root statistics proposed by Pesaran (2007) consists of the average of the individual CADF statistics:

$$CIPS(N,T) = N^{-1} \sum_{i=1}^{N} CADF_i,$$

where $CADF_i$ is the cross-sectionally augmented Dickey–Fuller statistic for the i-th cross-section unit given by the t-ratio of the OLS estimate of $\alpha_{i,0}$ in (7).

Pesaran (2007) also proposes a truncated version of $CIPS$ test, denoted $CIPS^*$. This statistic is useful when T is small, usually between 10 and 20. The author presents the simulated critical values of $CIPS$ and $CIPS^*$ in his paper. If the value of the $CIPS$ and $CIPS^*$ statistics is smaller than the critical value, then the null hypothesis of a panel unit root is rejected. Although Pesaran (2007)'s framework allows for only one stationary common factor, Smith, Pesaran and Yamagata (2010) show that it is also valid when there is either more than one common factor and/or the common factors are stationary or non-stationary.

3.2.2 Moon and Perron (2004) panel data unit root test statistics

Moon and Perron (2004) propose two statistics that test for the presence of a unit root while accounting for cross-sectional dependence among the units of the panel. Their approach is based on an approximate common factor model and tests for the unit root in the defactored series. The authors consider an autoregressive process in which the error term follows a factor structure:

$$\begin{aligned} y_{i,t} &= \mu_i + \delta_i t + y^0_{i,t} \\ y^0_{i,t} &= \rho_i y^0_{i,t-1} + u_{i,t} \\ u_{i,t} &= F'_t \lambda_i + e_{i,t}. \end{aligned}$$

Moon and Perron (2004) first transform the model in order to eliminate the common factors and obtain a defactored data that has no cross-sectional dependence. In the second step, they construct the panel unit root tests t_a and t_b using the defactored data. Then the null hypothesis $H_0 : \rho_i = 1$ for all i against the alternative hypothesis $H_1 : \rho_i < 1$ for some i, $i = 1, \ldots, N$, is tested, using the following pooled test statistics:

$$\begin{aligned} t_a &= \frac{T\sqrt{N}\left(\hat{\rho}^*_{pool} - 1\right)}{\sqrt{2\hat{\phi}^4_e / \hat{\omega}^4_e}} \\ t_b &= T\sqrt{N}\left(\hat{\rho}^*_{pool} - 1\right)\sqrt{\frac{1}{NT^2} tr\left(Z_{-1} Q Z'_{-1}\right)}\left(\frac{\hat{\omega}_e}{\hat{\phi}^2_e}\right), \end{aligned}$$

where, under the null hypothesis of panel unit root, both test statistics converge to the standard normal distribution. The terms $\hat{\rho}^*_{pool}$, $\hat{\phi}_e$, $\hat{\omega}_e$, Z_{-1} and Q are defined in Moon and Perron (2004).

The null hypothesis of a unit root is rejected if the value of the t_a or t_b statistics is smaller than the critical value drawn from the standard normal distribution. Moon and Perron (2004) also show that estimating the common factors by principal components lead to feasible statistics with the same limiting distribution as if they were observable.

3.2.3 Bai and Ng (2004) panel data unit root test statistics

Bai and Ng (2004) decompose the observable variable $y_{i,t}$ into a deterministic component $D_{i,t}$, a common component $\lambda_i' F_t$ and an idiosyncratic component $e_{i,t}$:

$$\begin{aligned} y_{i,t} &= D_{i,t} + \lambda_i' F_t + e_{i,t} && (8)\\ (1-L)\,F_{j,t} &= C_j\,(L)\,w_{j,t}; \quad j = 1\ldots,r && (9)\\ (1-\rho_i L)\,e_{i,t} &= H_i\,(L)\,\varepsilon_{i,t}, && (10) \end{aligned}$$

where $D_{i,t}$ denotes the deterministic part of the model – either a constant or a linear time trend – F_t is a $(r \times 1)$-vector that accounts for the common factors that are present in the panel, and $e_{i,t}$ is the idiosyncratic disturbance term, which is assumed to be cross-section independent. The vector of loading parameters λ_i measures the effect that the common factors have on the i-th time series. Unobserved common factors and idiosyncratic disturbance terms are estimated using principal components on the first difference model. The estimation of the number of common factors is obtained using the panel BIC information criterion in Bai and Ng (2002).

Once both the idiosyncratic and common components have been estimated, we can proceed to test their order of integration using unit root tests. On the one hand, it is possible to test whether there are stationary or non-stationary common factors (F_t) using the ADF (for the one common factor case) or the MQ test statistics in Bai and Ng (2004) (for the general case where there are more than one common factor) – either in its parametric ($MQ_f^j(m)$) and/or non-parametric ($MQ_c^j(m)$) version, where $j = c$ for the model that includes a constant, $j = \tau$ for the model that includes a linear time trend and m denotes the number of stochastic trends under the null hypothesis. The critical values for up to six factors for the MQ tests can be found in Table 1 of Bai and Ng (2004), whereas the usual critical values of the Dickey-Fuller test can be used in the case of one common factor. Therefore, using these statistics we will be able to conclude how many (if any) of the r common factors that have been estimated are the stationary (r_0) and how many are non-stationary (r_1), so that $r = r_0 + r_1$. On the other hand, we can test the panel unit root hypothesis focusing on the idiosyncratic shocks ($e_{i,t}$). In this case, Bai and Ng (2004) propose to compute the usual ADF pseudo t-ratio statistic applied to the idiosyncratic component. If the model contains only an intercept, the pseudo t-ratio statistic is denoted as $ADF_{\hat{e}}^c$ and its asymptotic distribution coincides with the Dickey-Fuller distribution for the case of no constant. If the model has an intercept and a linear trend that the statistic is denoted as $ADF_{\hat{e}}^{\tau}$, which asymptotic distribution is function of a Brownian bridge.

As can be seen, this technique can determine the source of the non-stationarity. It is possible that the non-stationarity of the observed variables ($y_{i,t}$) is the result of the presence of I(1) common factors – or a combination of I(0) and I(1) common factors – which implies that the panel data set is non-stationary and that the source of non-stationarity is a common cause for all the units that define the panel. In this case, we should conclude that there are global permanent shocks affecting the whole panel. It could also be the case that source of non-stationarity of the panel is that the idiosyncratic disturbance terms are I(1) non-stationary processes, a fact that implies that shocks that affect only each time series – i.e., not the global shocks – have a permanent character.

The approach of Bai and Ng (2004) nests the ones in Moon and Perron (2004) and Pesaran (2007). As noted by Bai and Ng (2010), Moon and Perron (2004) and Pesaran (2007) control the presence of cross-section dependence allowing for common factors, although the common factors and idiosyncratic shocks are restricted to have the same order of integration. Therefore, it is not possible to cover situations in which one component (e.g., the common factors) is I(0) and the other component (for example, the idiosyncratic shocks) is I(1) and vice versa. In practical terms, the test statistics in Moon and Perron (2004) and Pesaran (2007) turn out to be statistical procedures to make inference only on the idiosyncratic shocks, where the dynamics of both the idiosyncratic and the common components are restricted to be the same.

3.2.4 Hadri (2000) panel data stationarity test statistics

The panel data stationarity test statistic in Hadri (2000) specifies the null hypothesis that the units in the panel data set are I(0) against the alternative hypothesis that there are some units that are I(1). The test is based on the OLS estimation of the following regression equation:

$$y_{i,t} = D_{i,t} + u_{i,t}, \qquad (11)$$

where $D_{i,t}$ denotes the deterministic component. The estimated residuals from (11) are used to define the partial sum processes $\hat{S}_{i,t} = \sum_{j=1}^{t} \hat{u}_{i,j}$ for each unit. Using this individual information, Hadri (2000) proposes a panel stationarity test:

$$LM^j = N^{-1} \sum_{i=1}^{N} \eta_i^j,$$

where $\eta_i^j = \hat{\omega}_i^{-2} T^{-2} \sum_{t=1}^{T} \hat{S}_{i,t}^2$, $i = 1, \ldots, N$, denotes the individual stationarity test statistic proposed in Kwiatkowski, Phillips, Schmidt and Shin (1992) – KPSS henceforth – where $j = c$ for the model that only includes a constant ($D_{i,t} = \mu_i$) and $j = \tau$ for the one that includes a linear time trend ($D_{i,t} = \mu_i + \delta_i t$), with $\hat{\omega}_i^2$ being a consistent estimate of the long-run variance of the error term $e_{i,t}$ – Carrion-i-Silvestre et al. (2005) suggest to estimate the long-run variance following the procedure described by Sul et al. (2005), using the Quadratic spectral kernel. At this stage, we should mention that it is possible to compute two different LM statistics, depending on whether the long-run variance is allowed to be heterogeneous across i (LM_{HET}^j) or homogeneous for all individuals (LM_{HOM}^j) – in the latter case we use $\hat{\omega}^2 = N^{-1} \sum_{i=1}^{N} \hat{\omega}_i^2$. After standardizing the LM statistic by its mean and variance and assuming that $u_{i,t}$ in (11) are cross-section independent, the authors derive the new test Z_k^j, $k = \{HOM, HET\}$, that has the following distribution under the null hypothesis of panel stationarity:

$$Z_k^j = \frac{\sqrt{N}(LM_k^j - \overline{\xi^j})}{\overline{\varsigma^j}} \Rightarrow N(0,1),$$

$j = \{c, \tau\}$, $k = \{HOM, HET\}$. The terms $\overline{\xi^j}$ and $\overline{\varsigma^j}$ are the cross-section average of the mean and the variance of the individual KPSS statistic defined in Hadri (2000), $j = \{c, \tau\}$. Finally, it should be mention that the test in Carrion-i-Silvestre et al. (2005) bases on Hadri's (2000) proposal but uses bootstrapped p-values following the lines given in Maddala and Wu (1999) to deal with the cross-section dependence among the time series in the panel. Note thus that the statistic does not account for the presence of cross-section dependence using a common factor model.

3.3 Panel data cointegration test statistics

3.3.1 Carrion-i-Silvestre and Surdeanu (2011) panel data cointegration test statistics

The first category of testing for cointegration in panel data is based on a system-based approach. As mentioned above, the main advantage of the system-based approach is allowing more than one cointegrating relation among the variables for each individual system. Let us define the vector $X_{i,t} = (y_{i,t}, k_{i,t}, g_{i,t}, h_{i,t})'$ that collects the observable variables of our model, for which we define the following VAR representation:

$$\begin{aligned} X_{i,t} &= D_{i,t} + \lambda_i F_t + e_{i,t} && (12) \\ \left(I_q - L\right) F_t &= C(L)\, w_t && (13) \\ (I_k - L)\, e_{i,t} &= H_i(L)\, \varepsilon_{i,t}, && (14) \end{aligned}$$

where $i = 1, \ldots, N$ and $t = 1, \ldots, T$. In this setup $D_{i,t}$ is defined as a $(k \times 1)$ vector that cointains the deterministic component of each of the variables in the vector $X_{i,t}$, i.e., $k = 4$ in our case. The term F_t is a $(q \times 1)$ vector of common factors, λ_i is a $(k \times q)$ matrix of factor loadings and $e_{i,t}$ is a $(k \times q)$ vector that collects the idiosyncratic stochastic term. The estimation of the unobservable common factors is made using the principal component approach suggested in Bai and Ng (2002, 2004). Once the effects of the common factors are removed, cointegration analysis is then performed focusing on both the idiosyncratic and common factor components. This gives us further insight on the cointegration analysis, since the inference on the cointegrating rank can be distorted if common factors are not accounted for in the model – see Carrion-i-Silvestre and Surdeanu (2011) for further details.

The determination of the number of stochastic trends in the system relies on a sequential testing procedure that starts assuming that the cointegrating rank is zero – i.e., there is $m = k$ stochastic trends – and, defining the multivariate MSB test statistic $MSB_{j,i}(m)$, $j = \{c, \tau\}$, we can proceed to test whether there is $m = k$ stochastic trends or less than k. Using the $MSB_{j,i}(m)$, $j = \{c, \tau\}$, test statistic we can estimate the number of stochastic trends for each individual system using the critical values in Carrion-i-Silvestre and Surdeanu (2011).

It is also possible to combine the individual information and define panel data cointegrating rank tests. Assuming the same number of stochastic trends m in all individual systems, Carrion-i-Silvestre and Surdeanu (2011) test the null hypothesis that all N individual systems have m stochastic trends against the alternative hypothesis that there are $m-1$ stochastic trends:

$$\left\{ \begin{array}{ll} H_0: & m \text{ stochastic trends} \quad \forall i = 1, \ldots, N \\ H_1: & m-1 \text{ stochastic trends} \quad \forall i = 1, \ldots, N. \end{array} \right. \qquad (15)$$

The first panel data statistic is based on the standardized mean of the individual statistics:

$$PMSB_j^Z(m) = \frac{\sqrt{N}(\overline{MSB}_j(m) - E(MSB_j(m)))}{\sqrt{Var(MSB_j(m))}},$$

where $\overline{MSB}_j(m) = N^{-1}\sum_{i=1}^{N} MSB_{j,i}(m)$, and $E(MSB_j(m))$ and $Var(MSB_j(m))$ are the mean and the variance of the $MSB_j(m)$ statistic given in Carrion-i-Silvestre and Surdeanu (2011). Under the null hypothesis of m stochastic trends $PMSB_j^Z(m) \Rightarrow N(0,1)$. The remaining tests are based on the combination of the p-values (φ_i) of the individual MSB statistic:

$$
\begin{aligned}
PMSB_j^F(m) &= -2\sum_{i=1}^{N} \ln \varphi_i \\
PMSB_j^C(m) &= \frac{-2\sum_{i=1}^{N} \ln \varphi_i - 2N}{\sqrt{4N}},
\end{aligned}
$$

where under the null hypothesis of m stochastic trends $PMSB_j^F(m) \Rightarrow \chi^2_{2N}$ and $PMSB_j^C(m) \Rightarrow N(0,1)$. In this paper we use another test, $PMSB_j^{CZ}$, in order to test for cointegration among the cross-sections of the panel. The $PMSB_j^{CZ}$ statistic, originally proposed by Choi (2001), is based on the p-values of the individual MSB tests and has following form:

$$
PMSB_j^{CZ}(m) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \Phi^{-1}(\hat{\varphi}_i),
$$

where $\Phi(\cdot)$ denotes the standard Normal cumulative distribution function. Although Carrion-i-Silvestre and Surdeanu (2011) do not prove it, they conjecture that the limiting distribution of this statistic is also standard normal, a claim that is supported by the Monte Carlo simulations. Like the previous panel cointegration tests, the null hypothesis of $PMSB_j^{CZ}$ is that of no panel cointegration.

3.3.2 Banerjee and Carrion-i-Silvestre (2006) panel data cointegration test statistics

Banerjee and Carrion-i-Silvestre (2006) deal with the following model specification:

$$
y_i = D_i + x_i\beta_i + F\lambda_i + e_i, \tag{16}
$$

where the common factors and factor loadings are estimated using principal components following the approach in Bai and Ng (2004). In order to do so, orthogonal projections on the first difference of (16) are taken:

$$
\begin{aligned}
M_i\Delta y_i &= M_i\Delta F\lambda_i + M_i\Delta e_i \\
y_i^* &= f\lambda_i + z_i,
\end{aligned} \tag{17}
$$

with $M_i = I_{T-1} - \Delta x_i^d\left(\Delta x_i^{d\prime}\Delta x_i^d\right)^{-1}\Delta x_i^{d\prime}$ being the idempotent matrix, $\Delta x_i^d = [\Delta D_i\ \Delta x_i]$ a matrix that contains the first difference of the deterministic component and the stochastic regressors, $y_i^* = M_i\Delta y_i$, $f = M_i\Delta F$ and $z_i = M_i\Delta e_i$.[2] The estimation of the common factors and factor loadings is done as in Bai and Ng (2004) using principal components. Specifically, the estimated principal components of $f = (f_2, f_3, \ldots, f_T)$, denoted as $\tilde{f}$, are $\sqrt{T-1}$ times the r eigenvectors corresponding to the first r largest eigenvalues of the $(T-1)\times(T-1)$ matrix $y^*y^{*\prime}$. Under the normalization $\tilde{f}\tilde{f}'/(T-1) = I_r$, the estimated loading matrix is $\tilde{\Lambda} = \tilde{f}'y^*/(T-1)$. Therefore, the estimated residuals are defined as:

$$
\tilde{z}_{i,t} = y_{i,t}^* - \tilde{f}_t'\tilde{\lambda}_i. \tag{18}
$$

[2]It should be understood that $\Delta x_i^d = \Delta x_i$ for Models 0 and 1 provided that $\Delta D_i = 0$.

Using these estimates, the idiosyncratic disturbance term is recovered and the common factors are computed through cumulation, i.e., $\tilde{e}^*_{i,t} = \sum_{j=2}^{t} \tilde{z}_{i,j}$ and $\tilde{F}_t = \sum_{j=2}^{t} \tilde{f}_j$. Then we proceed to the estimation of the ADF-type regression equation:

$$\Delta\tilde{e}^*_{i,t} = \alpha_{i,0}\tilde{e}^*_{i,t-1} + \sum_{j=1}^{p} \alpha_{i,j}\Delta\tilde{e}^*_{i,t-j} + w_{i,t}, \tag{19}$$

so that the null hypothesis of no cointegration can be tested using the pseudo t-ratio of $\alpha_{i,0}$ ($t_{\tilde{\alpha}_{i,0}}$). Banerjee and Carrion-i-Silvestre (2006) define the panel cointegration test statistic $Z_j = (N^{-1}\sum_{i=1}^{N} t_{\tilde{\alpha}_{i,0}} - \Theta^e_j)(\Psi^e_j/N)^{-1/2}$, where $j = c$ refers to the model that includes a constant and $j = \tau$ to the model that includes a linear time trend, with Θ^e_j and Ψ^e_j the mean and variance of the relevant functionals of Brownian motions.[3] As $T, N \to \infty$ the Z_j, $j = \{c, \tau\}$, test statistic converges in the limit under the null hypothesis of no panel cointegration to a standard normal distribution. If there is only one common factor, its order of integration can be tested using the ADF-type regression equation in (19) with $\tilde{e}^*_{i,t}$ replaced by $\tilde{F}_t$, while in the case where more than one factor is estimated, the number of stochastic trends among the common factors can be estimated using the MQ test statistics as in Bai and Ng (2004).[4]

3.3.3 Banerjee and Carrion-i-Silvestre (2011) panel data cointegration test statistics

Banerjee and Carrion-i-Silvestre (2011) propose a panel cointegration test based on the common correlated effects (CCE) estimation approach developed by Pesaran (2006). The idea behind the CCE estimation is relatively simple. Since the cross-section dependence is sometimes caused by unobservable common factors, Pesaran (2006) uses cross-section averages of the dependent and the explanatory variables as proxies for common factors. Banerjee and Carrion-i-Silvestre (2011) use the following model:

$$y_{i,t} = D_{i,t} + x^{'}_{i,t}\beta_i + \bar{z}^{'}_t\eta_i + e_{i,t},$$

where $\bar{z}^{'}_t = \left(\bar{y}_t, \bar{x}^{'}_t\right)^{'}$ is the vector of cross-section means of the dependent and explanatory variables. Following Pesaran (2006), Holly, Pesaran and Yamagata (2010) and Kapetanios, Pesaran and Yamagata (2011), Banerjee and Carrion-i-Silvestre (2011) use the pooled estimator:

$$\hat{\beta}_{CCEP} = \left(\sum_{i=1}^{N} x^{'}_i\bar{M}x_i\right)^{-1}\left(\sum_{i=1}^{N} x^{'}_i\bar{M}y_i\right),$$

where $x_i = (x_{i,1}, x_{i,2}, ..., x_{i,T})^{'}$, $y_i = (y_{i,1}, y_{i,2}, ..., y_{i,T})^{'}$ and the matrix $\bar{M}$ is defined in Holly et al. (2010). In the next step, Banerjee and Carrion-i-Silvestre (2011) define the variable $\tilde{y}_{i,t} = y_{i,t} - x^{'}_{i,t}\hat{\beta}_{CCEP}$ and then estimate the regression below using the OLS procedure:

$$\tilde{y}_{i,t} = D_{i,t} + e_{i,t}. \tag{20}$$

[3] Banerjee and Carrion-i-Silvestre (2006) approximate the moments of the limiting distribution of the statistics by means of Monte Carlo simulation, which are $(\Theta^e_c, \Theta^e_\tau) = (-0.424, -1.535)$ and $(\Psi^e_c, \Psi^e_\tau) = (0.964, 0.341)$.

[4] The limiting distribution of the ADF test statistic when there is one common factor is the one obtained in Dickey and Fuller (1979), so that the standard critical values for the ADF test statistic can be used in this case. The critical values for the MQ test can be found in Table I in Bai and Ng (2004).

Both individual ($t_{\hat{\alpha}_{i,0}}$) and panel cointegration ($CADFC_P$) test statistics are based on the OLS residuals $\hat{e}_{i,t}$ from (20). The individual cointegration test statistic is the pseudo t-ratio of the estimated parameter $\hat{\alpha}_{i,0}$ in the following regression:

$$\Delta \hat{e}_{i,t} = \alpha_{i,0}\hat{e}_{i,t-1} + \sum_{j=1}^{s} \alpha_{i,j}\Delta \hat{e}_{i,t-j} + \zeta_i \bar{\hat{e}}_{t-1} + \sum_{j=0}^{s} \theta_{i,j}\Delta \bar{\hat{e}}_{t-j} + \kappa_{i,t}.$$

Finally, the panel cointegration statistic is defined as:

$$CADFC_P = N^{-1} \sum_{i=1}^{N} t_{\hat{\alpha}_{i,0}}.$$

The critical values for both individual and panel cointegration test statistics are presented in Banerjee and Carrion-i-Silvestre (2011). The null hypothesis of no cointegration is rejected if the value of the corresponding test statistic is smaller than the critical value.

4 Empirical results

We start the empirical analysis by checking whether cross-section dependence exists among the variables of our model. Note that while it is convenient to think of cross-section independence as the ideal case, in real world this is not likely to hold in most situations. It should be natural to assume that the regions of Spain are dependent of each other. We employ the CD statistic of Pesaran (2004) and present the results of the statistic for each variable for different augmentation orders ($p = 0, 1, \ldots, 5$) at the top of Table 1. The values of the CD test statistic indicate that we can easily reject the null hypothesis of cross-section independence in favor of cross-section dependence for all variables regardless of the augmentation order that is used.

4.1 Panel data unit root and stationarity test statistics

Let us first focus on the results obtained using Pesaran's (2007) statistics. Table 2 presents the CIPS(p) test statistic for different augmentation orders ($p = 0, 1, \ldots, 5$). The 5% critical value of the statistic for the case with intercept and time trend is -2.72 – see Table II(c) in Pesaran (2007). The results indicate that, in almost all cases, the idiosyncratic component of the variables that we consider in the paper are non-stationary – the null hypothesis of unit root is marginally rejected for $y_{i,t}$ with $p = 0$ and $p = 1$ and for $k_{i,t}$ with $p = 1$. In general, these results suggest the non-stationarity of the idiosyncratic component of the variables in our model.

Since the cross-section dependence is accounted for through the specification of an approximate common factor model, it is important to estimate the number of common factors. It is interesting to note that the number of common factors (estimated by Bayesian information criteria (BIC) as in Bai and Ng (2002)) equals the maximum number of common factors permitted. This seems a rather typical problem encountered by Basher and Carrion-i-Silvestre (2007), Sul (2005) and Holly, Pesaran and Yamagata (2010), among others. One reasonable explication, sustained by Bai and Ng (2002) as well, is that for small number of cross-sections (less than 20), the number of common factors is difficult to estimate. We then determine the number of non-stationary factors using the three criteria (IPC_1, IPC_2 and IPC_3) proposed by Bai (2004) and the MQ test statistics by Bai and Ng (2004) while setting the maximum number of factors at 6. The IPC_1 and IPC_2 criteria yield 4 non-stationary factors, IPC_3 criteria suggests 3 non-stationary factors and the MQ test statistics indicate that all 6 factors are non-stationary. Therefore, for the rest of our analysis we calculate the statistics described previously for 3, 4 or 6 factors.

Next, we compute the Moon and Perron (2004) panel data unit root test statistics. As can be seen in Table 2, regardless of the number of factors considered, for the variables $g_{i,t}$ and $h_{i,t}$ the two statistics of Moon and Perron (2004) do not reject the null hypothesis of panel unit root at the 5% level of significance. For the rest of the variables, the results are somewhat mixed, depending on the number of factors taken into account – for more than half of the cases, the idiosyncratic component of these seven variables is I(1).

However, we cannot conclude anything about the order of integration of the common factors from the application of these statistics since we are focusing only on the idiosyncratic component – note that we wipe out the effect of the common factors, so that we are just focusing on the idiosyncratic disturbance terms. A more informative picture is thus obtained from Bai and Ng (2004) approach, provided that separate inference can be conducted on the idiosyncratic and the common factor components of the observable variable. Table 2 reports the ADF statistic for the idiosyncratic component of each variable and the MQ test statistics on the estimated common factors.[5] The null hypothesis of panel unit root cannot be rejected at the 5% level of significance for the idiosyncratic component of $y_{i,t}$, $h_{i,t}$ and $hs_{i,t}$ for any number of factors considered. The rest of the variables are I(1) for the case of 3 and 4 factors, and I(0) when 6 factors are taken into account. Besides, the application of the $MQ_f^\tau(m)$ and $MQ_c^\tau(m)$ statistics of Bai and Ng (2004) characterize the 6, 4 or 3 estimated common factors as stochastic trends – see the critical values for these statistics in Table I in Bai and Ng (2004). These two set of results leads to conclude that the seven observable variables are I(1). When we consider 3 or 4 factors, the source of non-stationarity is of global and idiosyncratic nature for all seven variables. However, when we consider 6 factors, the source of non-stationarity comes from a global nature for $k_{i,t}$, $g_{i,t}$, $l_{i,t}$ and $gp_{i,t}$ and of a global and idiosyncratic nature for $y_{i,t}$, $h_{i,t}$ and $hs_{i,t}$.

In the next step, we look at the stationarity of the variables. First, we employ the panel stationary test of Hadri (2000) assuming that the long-run variance is either homogeneous or heterogeneous. Since the variables of our model present a linear trend, we estimate a model where the deterministic term consists of both the constant and the linear time trend. The results of the panel stationary test of Hadri (2000) and bootstrapped critical values computed as in Carrion-i-Silvestre et al. (2005) are presented in Table 3. It is easy to see that with the exception of the calculated value for $l_{i,t}$, all values of the stationarity test of Hadri (2000) are greater than the 95% bootstrapped critical values. Therefore, we reject the null hypothesis of stationarity for six out of seven variables at the 5% level of significance and we imply that these six variables are I(1) regardless of the way in which the long-run variance is estimated. As noted in the previous section, while the stationarity test of Hadri (2000) allows for cross-section dependence, it does not accommodate for the common factors.

[5]Following Ng and Perron (1998), the maximum number of lags that is used to compute the ADF statistic is set at $T^{1/3}$.

4.2 Panel data cointegration tests

The model for the production function involves five observable variables that are driven by global and idiosyncratic stochastic trends. Since we consider two types of public capital and two types of human capital, we analyze four different combinations of variables. The first combination consists of $y_{i,t}$, $k_{i,t}$, $g_{i,t}$, $h_{i,t}$ and $l_{i,t}$ – hereafter, we denote this model specification as Combination 1. The variables that we test secondly are $y_{i,t}$, $k_{i,t}$, $gp_{i,t}$, $h_{i,t}$ and $l_{i,t}$ (Combination 2). The third combination consists of $y_{i,t}$, $k_{i,t}$, $g_{i,t}$, $hs_{i,t}$ (Combination 3) and $l_{i,t}$ and the last one consists of $y_{i,t}$, $k_{i,t}$, $gp_{i,t}$, $hs_{i,t}$ and $l_{i,t}$ (Combination 4). The non-stationarity of these variables implies that the estimation of the model that links these macroeconomic aggregates needs to restore on the use of the cointegration analysis. Thus, as a first stage, we should test whether cointegration is present among these variables, accounting for the feature that global stochastic trends are present. To this end, we first analyze how many cointegration relations exist among these variables applying the VAR approach devised in Carrion-i-Silvestre and Surdeanu (2011).

The individual MSB based statistic and its respective cointegration rank for the first combination of variables are shown in the upper part of Table 4. The most common selected rank for the individual Spanish regions is one, suggesting the existence of one cointegrating relation among the variables of the model. For almost half of the regions, the univariate statistic detects no cointegration at all. For two regions, namely Andalucía and La Rioja, the rank is two indicating two cointegrating relations. Overall, the results are mixed indicating the low power of the univariate statistic.

The bottom part of Table 4 presents the Carrion-i-Silvestre and Surdeanu (2011) $PMSB_{\tau}^{Z}$, $PMSB_{\tau}^{F}$ and $PMSB_{\tau}^{C}$ panel data statistics. Since the panel cointegrating test $PMSB_{\tau}^{CH}$ also used in this paper is based on the same MSB statistic as the previous three panel statistics, we present its results in the same table with the Carrion-i-Silvestre and Surdeanu (2011) statistics. The panel cointegration ranks are presented in the last column of Table 4. All panel data statistics strongly reject the null hypothesis of no cointegration at the 5% level of significance. Moreover, with the exception of $PMSB_{\tau}^{Z}$ test, all panel data cointegration test statistics indicate that the cointegration rank is two. This result implies the existence of two common cointegrating relations between output, physical capital, human capital, public capital (all in per worker terms) and labor. Table 5 presents both univariate and panel data cointegration statistics for the second combination of variables. At the univariate level, cointegrating rank is 0 for seven regions, 1 for six regions and 2 for four regions. The panel data results are similar to the previous combination of variables. More exactly, three panel data statistics detect two cointegration relations while only one statistic detects one cointegration relation. The results of the cointegration tests both for univariate and panel data for the third combination are presented in Table 6. The univariate statistic indicates the absence of any cointegrating relation for seven regions, one cointegrating relation for nine regions and two cointegrating relations for one region (Canarias). At the panel level, the results indicate the existence of one common cointegrating relation. Finally, the results from the univariate and panel data cointegration statistics for the last combination of variables are presented in Table 7. There are seven regions for which the univariate statistic does not detect any cointegration between the variables. The univariate cointegrating rank is one for nine regions and two for only one region (Galicia). The results from the panel cointegration tests indicate the existence of one cointegration relation between this combination of variables. Overall, we can infer that the results from the individual MSB based statistic are mixed. Approximately half of the time the test statistic detects no cointegration at all and this can be due to the low power of the univariate test. However, the panel cointegration statistics indicate with overwhelming evidence that there exist at least one common cointegrating relation, depending on the variables presented in the model.

The next part of the empirical analysis examines the results from the single-equation-based framework of Banerjee and Carrion-i-Silvestre (2006, 2011). Let us first focus on the results for the approach based on principal components. In this case, we have allowed a maximum of six common factors and the BIC information criterion in Bai and Ng (2002) selects two, three or four common factors depending on the combination of variables that we consider. In all cases, the estimated common factors are characterized as I(1) processes – see Table 8. The panel ADF test statistic computed using the idiosyncratic disturbance terms (Z_c test statistic) leads to the rejection of the null hypothesis of spurious regression for all four combinations of variables. Therefore, from the application of the test statistic in Banerjee and Carrion-i-Silvestre (2006) we conclude that there is a long-run relationship among the variables of all four combinations that we have considered once the presence of common factors are accounted for. Notice that this results imples that the observable economic variables of the model do not cointegrate alone, they take part of a cointegration relationship that includes the presence of non-stationary global stochastic trends. This result is in line with the theoretical arguments that claim that the TFP is a non-stationary stochastic process.

The results from panel cointegration test statistic in Banerjee and Carrion-i-Silvestre (2011) are presented in Table 9 for all four combinations of variables. At the individual level, we are able to reject the null hypothesis of no cointegration for only a few regions. Therefore, at the individual level, for the majority of Spanish regions there is not enough evidence that the variables cointegrate, regardless of the combination of variables used – the results from the individual statistics are not shown in order to save space but they are available upon request.

Let's turn our attention to panel statistic $CADFC_P$ presented for up to 10 lags. For $p = 0,5,6,7,8,9$, the panel statistic detects no cointegration at any acceptable levels of significance, regardless of the combination of variables used. For $p = 1,2,3$, the statistic is able to reject the null hypothesis of no cointegration at either the 5% or 10% level of significance for every combination. For the augmentation orders $p = 4$ and $p = 10$, we obtain mixed results. The statistic finds evidence in favor of no cointegration for the first two combinations and the opposite conclusion for the last two combinations of variables. Although these results might seem contradictory with the evidence provided by the Z_c test statistic, it should bear in mind that the $CADFC_P$ tends to show mild under-rejection size distortions problem when the common factors and the idiosyncratic component have different orders of integration, as this is the case if we rely on the previous test statistic – see Banerjee and Carrion-i-Silvestre (2011) for further details.

When we compare the results of single-equation-based and system-based cointegration analysis approaches, the results are similar at the individual level, provided that little evidence is found in favor of cointegration. However, if we compare the results at the panel level, we obtain mixed results. In general, Banerjee and Carrion-i-Silvestre (2011) statistic finds evidence of cointegration for lower lags and no cointegration for higher lags. On the other hand, the test statistics proposed in Banerjee and Carrion-i-Silvestre (2006) and Carrion-i-Silvestre and Surdeanu (2011) are able to reject the null hypothesis of no cointegration with overwhelming evidence for every combination of variables that is used.

4.3 The estimation of the long-run production function relationship

Once the presence of a long-run relationship among the different combination of variables that we have considered has been established, we proceed to estimate the panel cointegration relationship allowing for common factors. There are few theoretical proposals in the literature that fit our requirements. First, we apply the continuously-updated and fully-modified (CupFM) and the continuously-updated and bias-corrected (CupBC) estimators proposed in Bai, Kao and Ng (2009), which rely on the use of principal components to jointly estimate the cointegrating vector, the factor loadings and the common factors of the model specification.[6] Both estimation procedures render consistent estimates of the cointegrating vector regardless of whether we have I(0) and/or I(1) common factors. Second, we also use the pooled CCE estimator in Pesaran (2006) that, as established in Kapetanios, Pesaran and Yamagata (2011), produces a consistent estimator of the cointegrating vector. In this case, the common factors are proxied by the use of cross-section averages of the variables of the model. Finally, it is worth pointing out that it is possible to conduct statistical inference on the parameters estimated by any of these procedures.

Table 10 reports the estimated cointegrating vectors for the different combinations and estimation procedures. For the combinations 1 and 2, the number of factors estimated according to the information criteria in Bai, Kao and Ng (2009) is two. For the combinations 3 and 4, the number of factors estimated is three. The last estimator in Table 10 is the CCEP estimator of Pesaran (2006). When we look at the estimation results presented in Table 10, we see that the CupFM and CupBC estimators yield similar results and all the estimated parameters are statistically significant. The CCEP estimator, on the other hand, shows that only half the estimated parameters are statistically significant. Also, the estimated parameters of both forms of human capital ($h_{i,t}$ and $hs_{i,t}$) are not statistically significant, regardless of the combination of variables used.

The values for the estimated parameter of $k_{i,t}$ (log of physical capital per worker) range from 0.206 (CupBC) to 0.402 (CCEP). These results are similar with those obtained in other studies. For example, the values obtained by Serrano (1996) range from 0.38 to 0.45, those obtained by Bajo and Díaz (2005) range from 0.59 to 0.68 while that obtained by Márquez, Ramajo and Hewings (2011) is 0.31.[7] The estimated coefficient of the log of total public capital per worker $(g_{i,t})$ range from 0.118 (CCEP) to 0.173 (CupFM), while that of $gp_{i,t}$ range from 0.111 (CCEP) to 0.153 (CupFM). Similar results were obtained by Bajo and Díaz (2005), 0.09, and Márquez, Ramajo and Hewings (2011), 0.10. The estimates for the parameter of $h_{i,t}$ range from 0.213 (CupFM) to 0.272 (CupBC) while those for $hs_{i,t}$ range from 0.205 (CupFM) to 0.339 (CupBC). Again, the results are similar with those of Serrano (1996), who obtained a value of 0.216, or Bajo and Díaz (2005), who obtained a value of 0.14. What is somehow surprising is the negative and highly significant coefficient for $l_{i,t}$ in all three estimators, which indicates that the constant returns to scale assumption cannot be accepted – the negative sign indicates diminishing returns to scale on the factors that have been considered in the model.

[6]We thank Chihwa Kao, Takashi Yamagata and Mauro Costantini for providing the Gauss code.

[7]Note that specification of the variables, the model, the data and estimation techniques differ from one study to another.

4.4 Spatial dependence

Another way to deal with cross-section dependence is through the spatial econometrics approach. Until now we assumed that the cross-section dependence between the Spanish regions was captured through the (unobserved) common factors. Examples of such common shocks are oil price, stock market or technological shocks. However, it is possible that one Spanish region is affected by its neighbors – i.e., weak dependence as opposed to the strong dependence induced by the common factors. One obvious example is the labor mobility between the regions. Therefore, it makes sense to consider the tools developed by the spatial econometrics as a way to model the weak dependence that might be affecting the Spanish regions. The spatial dependence in econometric studies is carried out by defining a weight matrix, W, which indicates whether any pair of regions share a common border. If region i and j share a common border, then $W(i,j) = 1$ and zero otherwise. The testing for spatial dependence is typically done by maximum likelihood technique or generalized method of moments (Pesaran and Tosetti (2011)).

We follow Holly, Pesaran and Yamagata (2010) and we start the analysis with the following model:

$$\begin{aligned} y_{i,t} &= a_{i,t} + (\alpha+\beta+\gamma+\delta-1)\, l_{i,t} + \alpha k_{i,t} + \beta g_{i,t} + \delta h_{i,t} \\ a_{i,t} &= D_{i,t} + u_{i,t} \\ u_{i,t} &= F_t'\lambda_i + e_{i,t}, \end{aligned}$$

We decompose the term $\hat{u}_{i,t}$ into

$$\hat{u}_{i,t} = \sum_{j=1}^{r} \tilde{\lambda}_{i,j}\tilde{f}_{j,t} + \tilde{e}_{i,t},$$

where $\tilde{\lambda}_{i,j}$ are the factor loadings and $\tilde{f}_{j,t}$ denote the common factors, $j = 1,\dots,r$. We perform the OLS regression of $\hat{u}_{i,t}$ on the estimated factors and obtain the idiosyncratic components $\tilde{e}_{i,t}$. For each factor we compute:

$$\tilde{e}_{i,t} = \Gamma \sum_{j=1}^{N} w_{i,j}\tilde{e}_{j,t} + v_{i,t},$$

where Γ is the spatial autoregressive parameter, $w_{i,j}$ is the (i,j) element of the spatial weight matrix W and $v_{i,t} \sim iid\left(0,\sigma_v^2\right)$. We then calculate the log likelihood function:

$$L = -\left(\frac{NT}{2}\right)\ln(\sigma_v^2) + T\ln|I_N - \Gamma W| - \frac{1}{2\sigma_v^2}\sum_{t=1}^{T}(\tilde{e}_t - \Gamma W\tilde{e}_t)'(\tilde{e}_t - \Gamma W\tilde{e}_t),$$

where $\tilde{e}_t = (\tilde{e}_{1,t}, \tilde{e}_{2,t}, ..., \tilde{e}_{N,t})'$. Since Baleares and Canarias are islands, they have no neighbors and we eliminate their data for this analysis. Thus, we made the calculation considering $N = 15$ and $T = 37$. The maximum likelihood estimates of Γ are presented in Table 11. It is easy to see that the results are mixed and vary depending on the numbers of factors considered. When we consider 2 or 3 factors for all the combinations of variables or 6 factors for the combinations 1 and 2, the estimates are significant at the 5% level. This indicates that, even after controlling for the strong cross-section dependence, there exists spatial (weak) dependence between the Spanish regions.

5 Conclusions

This papers reexamines the evidence of cointegration between the output, physical capital, human capital, public capital, and labor. We consider annual data for $N = 17$ Spanish regions observed over the $T = 37$ year period from 1964 to 2000.

The empirical analyses that focus on the estimation of Spanish production functions usually assume cross-section independence, which is a restrictive assumption especially at the regional level. Our empirical analysis shows that the variables involved in the model are non-stationary, so that the application of panel data cointegration techniques are required to obtain a consistent estimate of the parameters of interest. The paper takes advantage of the recently developed non-stationary panel methodology, in both single-equation and system-equations based framework, that are general enough to permit the cross-section dependence across the units of the panel via common factors. The results reveal evidence of cointegration among the variables of the model up to the presence of non-stationary common factors. Consequently, the observable economic variables alone do not generate a long-run relationship, we need to consider the, otherwise, expected global stochastic common trends that defines the TFP of the regions. The procedures applied in the paper detect between one and two cointegration relation among output, physical capital, human capital, public capital (all in per worker terms) and labor. Finally, we estimate the Spanish regional production function using Pesaran (2006) and Bai, Kao and Ng (2009) panel data cointegration estimators.

A Appendix

The description of the variables that are used in the data base is the following:

- $Y_{i,t}$ = the output, measured by GVA at factor cost of region i in the year t at 1980 constant prices, from the BD.MORES data base, Spanish Ministry of Finance and Public Administrations.

- $K_{i,t}$ = the stock of private capital of region i in the year t at 1980 constant prices, from the Stock de Capital data base, IVIE.

- $G_{i,t}$ = the stock of total public capital of region i in the year t at 1980 constant prices, from the Stock de Capital data base, IVIE.

- $GP_{i,t}$ = the stock of productive public capital of region i in the year t at 1980 constant prices, from the Stock de Capital data base, IVIE.

- $H_{i,t}$ = the stock of human capital, measured as a share of the employed population with secondary and university education of region i in the year t, from the Stock de Capital Humano data base, IVIE.

- $HS_{i,t}$ = the stock of human capital, measured as an average years of schooling, from the Stock de Capital Humano data base, IVIE, and Serrano (1996)

References

[1] Andrews, D.W.K. (2005) Cross-Section Regression with Common Shocks, *Econometrica*, 73, 1551-1585.

[2] Aschauer, D. (1989) Is Public Expenditure Productive?, *Journal of Monetary Economics*, 23, 2, 177-200.

[3] Bai, J. (2004) Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data, *Journal of Econometrics*, 122, 137-183.

[4] Bai, J. (2009) Panel Data Models With Interactive Fixed Effects, *Econometrica*, 77, 1229-1279.

[5] Bai, J. and Kao C. (2006) On the Estimation and Inference of a Panel Cointegration Model with Cross-Sectional Dependence, *Panel Data Econometrics: Theoretical Contributions and Empirical Applications*, B. Baltagi (Ed.), 3-30.

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

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

[8] Bai, J. and Ng, S. (2010) Panel Unit Root Tests with Cross-Section Dependence: A Further Investigation, *Econometric Theory*, 26, 1088-1114.

[9] Bajo-Rubio, O. and Díaz-Roldán, C. (2005) Optimal Endowments of Public Capital: An Empirical Analysis for the Spanish Regions, *Regional Studies*, 39, 3, 297-304.

[10] Bajo-Rubio, O. and Sosvilla-Rivero, S. (1993) Does Public Capital Affect Private Sector Performance? An Analysis of the Spanish Case, 1964–88, *Economic Modelling*, 10, 179-185.

[11] Banerjee, A. and Carrion-i-Silvestre, J.L. (2006) Cointegration in Panel Data with Breaks and Cross-Section Dependence, ECB Working Paper 591.

[12] Banerjee, A. and Carrion-i-Silvestre, J.L. (2011) Testing for Panel Cointegration using Common Correlated Effects Estimators, Working Paper.

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

[14] Banerjee, A., Eberhardt, M. and Reade, J.J. (2010) Panel Estimation for Worriers, Working Paper, University of Birmingham.

[15] Basher, S.A. and Carrion-i-Silvestre, J.L (2007) Another Look at the Null of Stationary Real Exchange Rates: Panel Data with Structural Breaks and Cross-section Dependence, IREA Working Paper, University of Barcelona.

[16] Carrion-i-Silvestre, J.L., del Barrio, T. and López-Bazo, E. (2005) Breaking the Panels: An Application to GDP per capita, *Econometrics Journal*, 8, 159-175.

[17] Carrion-i-Silvestre, J.L. and Surdeanu, L. (2011) Panel Cointegration Rank Testing with Cross-Section Dependence, to appear in *Studies in Nonlinear Dynamics & Econometrics*, 15.

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

[19] Costantini, M. and Destefanis, S. (2009) Cointegration Analysis for Cross-Sectionally Dependent Panels: The Case of Regional Production Functions, *Economic Modelling*, 26, 320-327.

[20] Holly, S., Pesaran, M.H. and Yamagata, T. (2010) A Spatio-Temporal Model of House Prices in the USA, *Journal of Econometrics*, 158, 160-173.

[21] Kapetanios, G., Pesaran, M.H. and Yamagata, T. (2011) Panels with Non-stationary Multifactor Error Structures, *Journal of Econometrics*, 160, 326-348.

[22] LeSage, J.P. and Pace, R.K. (2009) Introduction to Spatial Econometrics, CRC Press, Boca Raton, FL.

[23] Márquez, M.A., Ramajo, J. and Hewings, G.J.D. (2011) Public Capital and Regional Economic Growth: a SVAR Approach for the Spanish Regions, Working Paper.

[24] Mas Ivars, M., Uriel Jiménez, E. and Pérez García, F. (2005) El Stock y los Servicios del Capital en España (1964-2002). Nueva metodología, Fundación BBVA.

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

[26] Pesaran, M.H. (2004) General Diagnostic Tests for Cross Section Dependence in Panels, Working Paper 435.

[27] Pesaran, M.H. (2006) Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure, *Econometrica*, 74, 967-1012.

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

[29] Serrano, L. (1996) Indicadores de Capital Humano y Productividad, *Revista de Economía Aplicada*, 10, 177-190.

[30] Serrano, L. (1997) Productividad del Trabajo y Capital Humano en la Economía Española, *Moneda y Crédito*, 205, 79-101.

[31] Sosvilla-Rivero, S. and Alonso, J. (2005) Estimación de una Función de Producción MRW para la Economía Española, 1910-1995, *Investigaciones Economicas*, Fundación SEPI, 29, 3, 609-624.

[32] Sul, D. (2005) New Panel Unit Root Tests under Cross Section Dependence for Practitioners, Working paper.

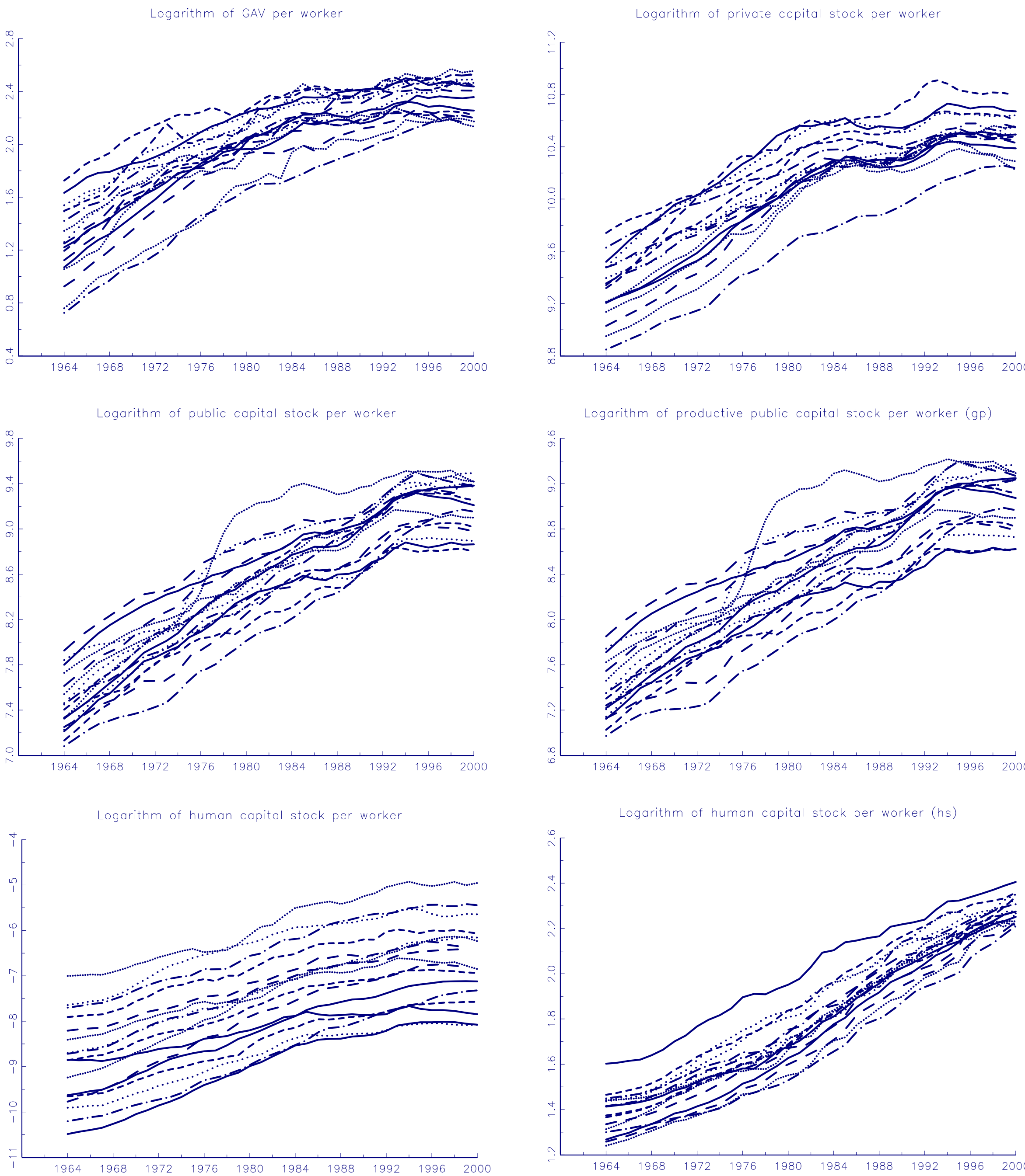

Figure 1. Time series variables of the seventeen Spanish regions

Table 1: Pesaran's (2004) cross-section dependence tests

	$y_{i,t}$	$k_{i,t}$	$g_{i,t}$	$h_{i,t}$	$l_{i,t}$	$gp_{i,t}$	$hs_{i,t}$
CD(0)	8.505	29.191	27.233	21.061	36.083	26.488	16.978
CD(1)	9.205	19.378	19.309	18.386	21.965	17.727	17.661
CD(2)	8.291	18.669	17.501	18.359	21.724	16.010	18.370
CD(3)	7.984	19.232	16.918	18.278	22.240	15.741	15.687
CD(4)	8.307	18.403	16.121	17.858	21.629	15.019	15.075
CD(5)	7.592	18.361	16.776	17.354	21.528	15.632	15.243

Table 2: Panel data unit root tests

Pesaran (2007)

	$y_{i,t}$	$k_{i,t}$	$g_{i,t}$	$h_{i,t}$	$l_{i,t}$	$gp_{i,t}$	$hs_{i,t}$
CADF(0)	-2.730	-2.687	-1.934	-2.293	-2.677	-1.844	-2.122
CADF(1)	-2.916	-2.817	-2.197	-2.553	-2.716	-2.065	-2.713
CADF(2)	-2.522	-2.407	-1.922	-2.123	-2.222	-1.809	-2.276
CADF(3)	-2.416	-2.306	-1.929	-2.065	-2.039	-1.846	-1.990
CADF(4)	-2.254	-2.088	-2.097	-1.646	-1.817	-2.075	-1.641
CADF(5)	-2.271	-2.170	-1.804	-1.614	-1.648	-1.805	-2.156

Moon and Perron (2004)

		$y_{i,t}$	$k_{i,t}$	$g_{i,t}$	$h_{i,t}$	$l_{i,t}$	$gp_{i,t}$	$hs_{i,t}$
6 factors	t_a	-1.267	-0.844	-0.146	-0.955	-1.414	-0.301	-4.707
	p-value	(0.103)	(0.199)	(0.442)	(0.170)	(0.079)	(0.382)	(0.000)
	t_b	-1.225	-0.743	-0.119	-0.818	-1.379	-0.253	-4.307
	p-value	(0.110)	(0.229)	(0.453)	(0.207)	(0.084)	(0.400)	(0.000)
4 factors	t_a	-1.621	-0.397	-0.793	-0.928	-1.719	-0.515	-2.470
	p-value	(0.052)	(0.346)	(0.214)	(0.177)	(0.043)	(0.303)	(0.007)
	t_b	-1.604	-0.354	-0.771	-0.837	-2.006	-0.478	-1.762
	p-value	(0.054)	(0.362)	(0.220)	(0.201)	(0.022)	(0.316)	(0.039)
3 factors	t_a	-1.761	-1.985	-1.073	-1.176	-2.015	-1.531	-1.539
	p-value	(0.039)	(0.024)	(0.142)	(0.120)	(0.022)	(0.063)	(0.062)
	t_b	-1.610	-1.886	-1.212	-1.082	-2.511	-1.759	-1.220
	p-value	(0.054)	(0.030)	(0.113)	(0.140)	(0.006)	(0.039)	(0.111)

Bai and Ng (2004)

		$y_{i,t}$	$k_{i,t}$	$g_{i,t}$	$h_{i,t}$	$l_{i,t}$	$gp_{i,t}$	$hs_{i,t}$
6 factors	$ADF_{\hat{e}}^{\tau}$	0.410	-1.770	-3.513	-0.219	-1.774	-2.707	2.210
	p-value	(0.659)	(0.038)	(0.000)	(0.413)	(0.038)	(0.003)	(0.986)
	$MQ_f^{\tau}(6)$	-31.907	-30.099	-27.514	-34.730	-30.349	-26.869	-31.790
	$MQ_c^{\tau}(6)$	-35.406	-27.884	-23.257	-33.837	-25.857	-24.339	-35.729
4 factors	$ADF_{\hat{e}}^{\tau}$	-0.733	-0.540	-0.026	-0.593	1.466	-0.313	0.199
	p-value	(0.232)	(0.294)	(0.490)	(0.276)	(0.929)	(0.377)	(0.579)
	$MQ_f^{\tau}(4)$	-27.996	-27.129	-13.887	-21.437	-24.762	-13.888	-5.911
	$MQ_c^{\tau}(4)$	-22.761	-26.854	-20.255	-20.019	-20.248	-19.498	-7.870
3 factors	$ADF_{\hat{e}}^{\tau}$	0.360	-1.221	0.544	-0.373	1.855	1.231	0.199
	p-value	(0.641)	(0.111)	(0.707)	(0.355)	(0.968)	(0.891)	(0.579)
	$MQ_f^{\tau}(3)$	-17.847	-24.356	-9.758	-20.501	-23.963	-8.660	-5.911
	$MQ_c^{\tau}(3)$	-15.914	-26.647	-13.875	-19.851	-18.138	-13.758	-7.870

Table 3: Hadri (2000) panel stationarity tests

Variable	Long-run variance	Z_j^{τ} Statistic	Bootstrapped critical values 90%	95%	99%
$y_{i,t}$	Homogeneous	17.359	6.894	8.911	12.403
	Heterogeneous	26.744	8.590	11.034	17.607
$k_{i,t}$	Homogeneous	38.494	7.444	9.836	15.575
	Heterogeneous	44.574	9.428	12.136	20.319
$g_{i,t}$	Homogeneous	21.396	5.737	7.207	11.470
	Heterogeneous	21.762	7.102	8.738	13.455
$h_{i,t}$	Homogeneous	11.421	6.895	8.791	13.027
	Heterogeneous	19.596	8.606	11.441	17.954
$l_{i,t}$	Homogeneous	5.456	5.890	7.363	10.961
	Heterogeneous	6.868	7.830	9.678	14.177
$gp_{i,t}$	Heterogeneous	17.534	5.520	6.998	10.596
	Heterogeneous	13.752	7.051	8.833	12.781
$hs_{i,t}$	Heterogeneous	9.220	6.019	7.663	11.580
	Heterogeneous	12.081	7.617	9.732	15.976

Table 4: Individual and panel data cointegration tests of Carrion-i-Silvestre and Surdeanu (2011). Results for the combination (1)

	Individual statistic					
Region	$m=5$	$m=4$	$m=3$	$m=2$	$m=1$	Rank
Andalucía	0.017**	0.021**	0.070	0.081	0.258	2
Aragón	0.029	0.036	0.048	0.060	0.257	0
Asturias	0.025**	0.034	0.037	0.038	0.162	1
Baleares	0.028	0.033	0.034	0.047	0.150	0
Canarias	0.036	0.058	0.070	0.111	0.120	0
Cantabria	0.026**	0.029	0.035	0.050	0.122	1
Castilla y León	0.024**	0.033	0.059	0.091	0.062	1
Castilla-La Mancha	0.016**	0.033	0.068	0.073	0.107	1
Cataluña	0.030	0.038	0.045	0.132	0.105	0
Comunidad Valenciana	0.026	0.027	0.056	0.055	0.198	0
Extremadura	0.017**	0.033	0.100	0.099	0.207	1
Galicia	0.021**	0.041	0.051	0.086	0.306	1
Madrid	0.029	0.029	0.044	0.070	0.193	0
Murcia	0.026	0.026	0.062	0.073	0.087	0
Navarra	0.019**	0.033	0.047	0.104	0.166	1
País Vasco	0.025**	0.037	0.048	0.078	0.061	1
La Rioja	0.021**	0.027**	0.050	0.063	0.057	2

	Panel statistics					
	$m=5$	$m=4$	$m=3$	$m=2$	$m=1$	Rank
$PMSB_\tau^Z$	-6.807**	-1.493	2.948	1.189	-0.611	1
$PMSB_\tau^F$	134.862**	61.071**	15.798	19.579	29.308	2
$PMSB_\tau^C$	12.231**	3.283**	-2.207	-1.749	-0.569	2
$PMSB_\tau^{CH}$	-7.434**	-2.170**	2.840	1.701	0.016	2

m represents the number of stochastic trends. ** denotes that the test is significant at the 5% level. The variables consisting of combination (1) are $y_{i,t}$, $k_{i,t}$, $g_{i,t}$, $h_{i,t}$ and $l_{i,t}$.

Table 5: Individual and panel data cointegration tests of Carrion-i-Silvestre and Surdeanu (2011). Results for the combination (2)

	Individual statistic					
Region	$m=5$	$m=4$	$m=3$	$m=2$	$m=1$	Rank
Andalucía	0.019**	0.023**	0.071	0.071	0.103	2
Aragón	0.028	0.037	0.062	0.067	0.226	0
Asturias	0.027	0.033	0.039	0.039	0.112	0
Baleares	0.029	0.032	0.036	0.051	0.055	0
Canarias	0.030	0.067	0.079	0.116	0.122	0
Cantabria	0.024**	0.028	0.034	0.104	0.120	1
Castilla y León	0.026	0.044	0.061	0.089	0.055	0
Castilla-La Mancha	0.014**	0.031	0.070	0.087	0.090	1
Cataluña	0.029	0.038	0.058	0.095	0.089	0
Comunidad Valenciana	0.026**	0.027**	0.036	0.074	0.167	2
Extremadura	0.018**	0.034	0.115	0.124	0.159	1
Galicia	0.022**	0.037	0.052	0.113	0.167	1
Madrid	0.027	0.027	0.048	0.076	0.210	0
Murcia	0.025**	0.025**	0.072	0.072	0.095	2
Navarra	0.018**	0.037	0.082	0.088	0.163	1
País Vasco	0.025**	0.037	0.053	0.066	0.064	1
La Rioja	0.020**	0.025**	0.050	0.068	0.078	2

	Panel statistics					
	$m=5$	$m=4$	$m=3$	$m=2$	$m=1$	Rank
$PMSB_{\tau}^{Z}$	-7.207**	-1.006	4.683	1.901	-1.624	1
$PMSB_{\tau}^{F}$	137.166**	70.158**	14.216	15.236	36.389	2
$PMSB_{\tau}^{C}$	12.511**	4.385**	-2.399	-2.275	0.290	2
$PMSB_{\tau}^{CH}$	-7.889**	-2.290**	3.849	2.444	-1.107	2

m represents the number of stochastic trends. ** denotes that the test is significant at the 5% level. The variables consisting of combination (2) are $y_{i,t}$, $k_{i,t}$, $gp_{i,t}$, $h_{i,t}$ and $l_{i,t}$.

Table 6: Individual and panel data cointegration tests of Carrion-i-Silvestre and Surdeanu (2011). Results for the combination (3)

	Individual statistic					
Region	$m=5$	$m=4$	$m=3$	$m=2$	$m=1$	Rank
Andalucía	0.021**	0.030	0.030	0.320	0.108	1
Aragón	0.025**	0.031	0.052	0.082	0.101	1
Asturias	0.025**	0.029	0.036	0.075	0.295	1
Baleares	0.031	0.033	0.062	0.071	0.282	0
Canarias	0.025**	0.027**	0.070	0.100	0.221	2
Cantabria	0.023**	0.035	0.066	0.101	0.101	1
Castilla y León	0.028	0.032	0.086	0.086	0.082	0
Castilla-La Mancha	0.033	0.037	0.037	0.118	0.112	0
Cataluña	0.032	0.047	0.064	0.106	0.091	0
Comunidad Valenciana	0.024**	0.031	0.045	0.088	0.371	1
Extremadura	0.021**	0.028	0.115	0.173	0.229	1
Galicia	0.030	0.028	0.053	0.053	0.146	0
Madrid	0.031	0.044	0.051	0.082	0.202	0
Murcia	0.022**	0.030	0.062	0.062	0.161	1
Navarra	0.018**	0.036	0.082	0.106	0.504	1
País Vasco	0.034	0.036	0.037	0.073	0.179	0
La Rioja	0.019**	0.033	0.045	0.079	0.081	1
	Panel statistics					
	$m=5$	$m=4$	$m=3$	$m=2$	$m=1$	Rank
$PMSB_{\tau}^{Z}$	-5.300**	-1.496	4.214	4.846	0.611	1
$PMSB_{\tau}^{F}$	120.704**	45.451	18.137	9.435	21.565	1
$PMSB_{\tau}^{C}$	10.514**	1.389	-1.924	-2.979	-1.508	1
$PMSB_{\tau}^{CH}$	-6.125**	-1.605	3.325	3.973	1.193	1

m represents the number of stochastic trends. ** denotes that the test is significant at the 5% level. The variables consisting of combination (3) are $y_{i,t}$, $k_{i,t}$, $g_{i,t}$, $hs_{i,t}$ and $l_{i,t}$.

Table 7: Individual and panel data cointegration tests of Carrion-i-Silvestre and Surdeanu (2011). Results for the combination (4)

	Individual statistic					
Region	$m=5$	$m=4$	$m=3$	$m=2$	$m=1$	Rank
Andalucía	0.018**	0.031	0.032	0.370	0.095	1
Aragón	0.025**	0.033	0.059	0.087	0.087	1
Asturias	0.025**	0.029	0.038	0.069	0.303	1
Baleares	0.030	0.034	0.065	0.077	0.256	0
Canarias	0.026	0.028	0.075	0.104	0.235	0
Cantabria	0.021**	0.036	0.082	0.097	0.204	1
Castilla y León	0.027	0.031	0.086	0.087	0.082	0
Castilla-La Mancha	0.032	0.037	0.038	0.116	0.099	0
Cataluña	0.032	0.042	0.069	0.116	0.075	0
Comunidad Valenciana	0.024**	0.032	0.037	0.096	0.346	1
Extremadura	0.021**	0.029	0.117	0.189	0.221	1
Galicia	0.025**	0.028**	0.048	0.050	0.257	2
Madrid	0.041	0.044	0.051	0.098	0.203	0
Murcia	0.018**	0.037	0.043	0.064	0.155	1
Navarra	0.018**	0.036	0.080	0.126	0.536	1
País Vasco	0.033	0.035	0.035	0.070	0.212	0
La Rioja	0.018**	0.031	0.044	0.083	0.066	1
	Panel statistics					
	$m=5$	$m=4$	$m=3$	$m=2$	$m=1$	Rank
$PMSB_{\tau}^{Z}$	-5.618**	-1.276	4.342	5.826	0.917	1
$PMSB_{\tau}^{F}$	127.145**	38.686	18.590	9.104	22.399	1
$PMSB_{\tau}^{C}$	11.295**	0.568	-1.869	-3.019	-1.407	1
$PMSB_{\tau}^{CH}$	-6.326**	-1.096	3.275	4.292	1.325	1

m represents the number of stochastic trends. ** denotes that the test is significant at the 5% level. The variables consisting of combination (4) are $y_{i,t}$, $k_{i,t}$, $gp_{i,t}$, $hs_{i,t}$ and $l_{i,t}$.

Table 8: Banerjee and Carrion-i-Silvestre (2006) panel data cointegration test statistic

	Combination 1					Combination 2			
	Test	$\hat{r}$	$\hat{r}_1^{NP}$	$\hat{r}_1^{P}$		Test	$\hat{r}$	$\hat{r}_1^{NP}$	$\hat{r}_1^{P}$
Z_c	-2.863	2	2	2	Z_c	-2.609	3	3	3
$MQ_c^c(2)$	-10.031				$MQ_c^c(3)$	-23.306			
$MQ_f^c(2)$	-8.447				$MQ_f^c(3)$	-24.115			
	Combination 3					Combination 4			
	Test	$\hat{r}$	$\hat{r}_1^{NP}$	$\hat{r}_1^{P}$		Test	$\hat{r}$	$\hat{r}_1^{NP}$	$\hat{r}_1^{P}$
Z_c	-2.578	4	4	4	Z_c	-2.797	3	3	3
$MQ_c^c(4)$	-21.995				$MQ_c^c(3)$	-19.798			
$MQ_f^c(4)$	-20.807				$MQ_f^c(3)$	-17.876			

Table 9: Panel data cointegration test statistic of Banerjee and Carrion-i-Silvestre (2011)

	$CADFC_P$			
p	(1)	(2)	(3)	(4)
0	-2.062	-2.057	-2.125	-2.121
1	-2.307*	-2.300*	-2.385**	-2.378**
2	-2.347**	-2.338**	-2.465**	-2.456**
3	-2.461**	-2.459**	-2.635**	-2.630**
4	-2.106	-2.122	-2.249*	-2.260*
5	-1.802	-1.823	-1.925	-1.940
6	-1.748	-1.768	-1.842	-1.857
7	-1.593	-1.621	-1.710	-1.735
8	-1.608	-1.650	-1.558	-1.583
9	-1.539	-1.621	-1.004	-1.025
10	-1.734	-1.804	-2.326**	-2.210*

** denotes that the test is significant at the 5% level and * denotes that the test is significant at the 10% level. p is the number of lags. The dependent variable is $y_{i,t}$. The exogenous variables for the combination (1) are $k_{i,t}$, $g_{i,t}$, $h_{i,t}$ and $l_{i,t}$. The exogenous variables for the combination (2) are $k_{i,t}$, $gp_{i,t}$, $h_{i,t}$ and $l_{i,t}$. The exogenous variables for the combination (3) are $k_{i,t}$, $g_{i,t}$, $hs_{i,t}$ and $l_{i,t}$. The exogenous variables for the combination (4) are $k_{i,t}$, $gp_{i,t}$, $hs_{i,t}$ and $l_{i,t}$.

Table 10: Estimates of the panel cointegrating vector

	CupFM				CupBC				CCEP			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
$k_{i,t}$	0.223	0.241	0.340	0.365	0.206	0.219	0.279	0.299	0.257	0.248	0.402	0.388
	(7.312)	(8.095)	(14.534)	(16.199)	(6.479)	(7.205)	(12.077)	(13.157)	(1.600)	(1.553)	(2.111)	(2.187)
$g_{i,t}$	0.173	-	0.149	-	0.164	-	0.145	-	0.120	-	0.118	-
	(7.326)		(8.929)		(6.752)		(8.701)		(3.760)		(2.741)	
$gp_{i,t}$	-	0.153	-	0.130	-	0.144	-	0.126	-	0.111	-	0.112
		(7.185)		(8.593)		(6.524)		(8.280)		(4.103)		(4.298)
$h_{i,t}$	0.218	0.213	-	-	0.272	0.271	-	-	0.084	0.089	-	-
	(10.939)	(10.298)			(13.839)	(13.672)			(1.144)	(1.132)		
$hs_{i,t}$	-	-	0.205	0.211	-	-	0.339	0.333	-	-	0.039	0.046
			(6.742)	(6.853)			(11.238)	(10.940)			(0.249)	(0.283)
$l_{i,t}$	-0.123	-0.133	-0.208	-0.209	-0.068	-0.064	-0.234	-0.233	-0.564	-0.619	-0.542	-0.619
	(-4.614)	(-4.991)	(-6.422)	(-6.368)	(-2.523)	(-2.426)	(-8.104)	(-7.804)	(-2.564)	(-2.865)	(-2.512)	(-3.028)

The numbers in parentheses are the t-statistics. The dependent variable is $y_{i,t}$. The exogenous variables for the combination (1) are $k_{i,t}$, $g_{i,t}$, $h_{i,t}$ and $l_{i,t}$. The exogenous variables for the combination (2) are $k_{i,t}$, $gp_{i,t}$, $h_{i,t}$ and $l_{i,t}$. The exogenous variables for the combination (3) are $k_{i,t}$, $g_{i,t}$, $hs_{i,t}$ and $l_{i,t}$. The exogenous variables for the combination (4) are $k_{i,t}$, $gp_{i,t}$, $hs_{i,t}$ and $l_{i,t}$. For the combinations (1) and (2), the number of factors estimated according to Bai, Kao and Ng (2009) is two. For the combinations (3) and (4), the number of factors estimated according to Bai, Kao and Ng (2009) is three.

Table 11: Spatial MLE estimates

Nr. of factors	(1)	(2)	(3)	(4)
1	-0.010	0.005	-0.051	-0.045
2	0.167**	0.176**	0.136**	0.143**
3	0.185**	0.183**	0.142**	0.151**
4	0.004	0.003	-0.045	-0.048
5	-0.020	-0.022	-0.041	-0.044
6	-0.220**	-0.218**	-0.038	-0.037

** denotes significance at the 5% level. The dependent variable is $y_{i,t}$. The exogenous variables for the combination (1) are $k_{i,t}$, $g_{i,t}$, $h_{i,t}$ and $l_{i,t}$. The exogenous variables for the combination (2) are $k_{i,t}$, $gp_{i,t}$, $h_{i,t}$ and $l_{i,t}$. The exogenous variables for the combination (3) are $k_{i,t}$, $g_{i,t}$, $hs_{i,t}$ and $l_{i,t}$. The exogenous variables for the combination (4) are $k_{i,t}$, $gp_{i,t}$, $hs_{i,t}$ and $l_{i,t}$.