Productivity, infrastructure and human capital in the Spanish regions*

Josep Lluís Carrion-i-Silvestre† Laura Surdeanu‡

May 3, 2016

Abstract

We revisit the cointegration relation among output, physical capital, human capital, public capital and labor for 17 Spanish regions observed over the period 1964-2011. Our approach is based on the estimation of a panel data model where cross-section dependence is allowed among the members of the panel. The paper emphasizes the idea that common factors capturing, for instance, the total factor productivity, should be accounted for when estimating the parameters. We use several proposals to estimate the long-run relation among these variables, which render consistent and efficient estimates of the parameters.

JEL Classification: C23, H54, O40, R11

Keywords: Panel data, cross-section dependence, total factor productivity, human capital, public capital

*We thank the editor, Prof. J. P. Elhorst, and four referees for helpful comments and suggestions. The authors acknowledge the financial support from the Spanish Ministerio de Ciencia y Tecnología grant ECO2014-58991-C3-1-R.

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

‡Department of Econometrics, Statistics and Applied Economics. University of Barcelona. Av. Diagonal, 690. 08034 Barcelona. E-mail: lsurdeanu@gmail.com

1 Introduction

The estimation of production functions that relate the output of a firm, region or country to different combinations of factors of production – usually physical capital and labor – has devoted lot of interest in empirical economics – see Aschauer (1989), Munnell (1990), García-Milá and McGuire (1992), Holtz-Eakin (1994), Baltagi and Pinnoi (1995), and García-Milá *et al.* (1996) for the US, Merriman (1990) for Japan, Berndt and Hansson (1992) for Sweden, Dalamagas (1995) for Greece, Evans and Karras (1994) for a sample of industrialized countries, Otto and Voss (1996) for Australia, and Wylie (1996) for Canada. These studies estimate production functions including not only physical capital and labor as inputs, but also human and public capitals as productive factors.

Early studies of production function estimation employ time series data, focusing on an individual region or country. For example, for the case of the aggregated Spanish economy, Serrano (1997) finds no evidence of cointegration, whereas Sosvilla-Rivero and Alonso (2005) achieve the converse conclusion. The contradictory results indicate that the empirical evidence from the time series analysis is mixed, being one plausible explanation the low power of the univariate unit root and cointegration tests. Fortunately, recent analyses show that the power of unit root and cointegration test statistics can be improved when both the time series and cross-section dimensions are combined in a panel data framework – see, for example, Serrano (1996), Bajo and Díaz (2005) and Márquez *et al.* (2011).

One critical problem with the panel data studies for the Spanish regions mentioned above is the assumption of cross-section independence. This is an unrealistic and far too restrictive assumption from an empirical point of view, 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 of the parameters 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 *et al.* (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).

Cross-section dependence is more a recurrent than a rare characteristic that is present in macroeconomic time series of different units. There are diverse sources of cross-section dependence that can be expected to affect the units of a panel data set. 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 units – the so-called cross-cointegration concept, as defined in Banerjee *et al.* (2005). This implies that cross-section dependence is more the rule than the exception. Bai and Ng (2002, 2004) recognize early on this problem and lay 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 (strong) – i.e., the sort of cross-section dependence that affects all units of the panel data.

As pointed out in Banerjee *et al.* (2010), the empirical work on the estimation of production functions in panel data using the common factor technique is relatively limited. Two examples related to our study are Costantini and Destefanis (2009) and Banerjee and Carrion-i-Silvestre (2011). Costantini and Destefanis (2009) analyze the production function for the Italian regions 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. 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 1964-2011.

It is usually assumed that the application of non-stationary panel data techniques will enhance the statistical inference about the stochastic properties of the variables, especially if T is small. Practitioners have started to apply panel data unit root tests with the hope that taking into account both the time and cross-section dimensions of the panel data will lead to improvements of the statistical inference – see Breitung and Pesaran (2008) and Banerjee and Wagner (2009) for recent overviews of the literature. However, this desirable situation might not be achieved if features like cross-section dependence is not considered. Westerlund and Breitung (2013) stress the importance of several issues that can be found when applying panel data unit root tests that, if not accounted for, can ruin the statistical inference.

To the best of our knowledge, none of the existing studies for the Spanish economy take into consideration the (strong) cross-section dependence among the members of the panel when estimating production functions. This paper is based on the estimation of a Cobb and Douglas (1928) production function and gives a novel empirical evidence for the Spanish regional case. Further, we consider the presence of structural instabilities due to the existence of structural breaks. In this regard, the cointegrating relations are estimated allowing for the presence of one structural break, which defines a flexible framework where output elasticities, marginal products to private and public capitals and returns to education can change through time. Finally, it is worth mentioning that the panel cointegration estimation techniques that we apply are designed to capture the presence of pervasive dependence among the units of the panel. However, it is possible that the units are also affected by local dependence, which implies that the dependence is not spread widely as the cross-section dimension of the panel increases. This situation gives rise to the so-called weak dependence, being the spatial dependence a particular case of weak dependence. The analysis that is conducted in this paper also covers the issue of spatial dependence, a form of weak dependence that is typically found when working with regional data.

The structure of this paper is as follows. Section 2 presents the model for panel data and the data used in this study. The results of the panel data cointegration analysis are presented in Section 3, where the estimation of the production function is reported using different estimation procedures. Finally, the paper concludes with Section 4.

2 Model specification

The model specification is given by the modified Cobb-Douglas production function used in Bajo and Díaz (2005):

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

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 private capital $(K_{i,t})$, public capital $(G_{i,t})$, human capital $(H_{i,t})$ and labor $(L_{i,t})$. The variable $A_{i,t}$ reflects total factor productivity (TFP), which is the part of the output not explained by the observable inputs. The production function can be expressed in per worker terms:

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

TFP represents the unobservable part of the production function and usually reflects the technological progress of the respective country or region. If technology is defined as 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, 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 potential stochastic trend 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 the conclusion 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 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 it 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 analysis and decompose the TFP into an unobserved common factor component $F_t'\lambda_i$ – where F_t is a $(r\times 1)$-vector of unobserved common factors, λ_i is a $(r\times 1)$-vector of loadings – and an idiosyncratic error component $e_{i,t}$. The common factor approach captures the effect of common shocks that affect the countries or regions, making it a desirable way to model strong cross-section dependence. Taking into account these considerations and following Costantini and Destefanis (2009) and Banerjee *et al.* (2010), the TFP is modeled through the common factor specification given by:

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

Assuming a Cobb-Douglas function for $F(K_{i,t}, G_{i,t}, H_{i,t}, L_{i,t}) = K_{i,t}^{\alpha} G_{i,t}^{\beta} H_{i,t}^{\delta} L_{i,t}^{\gamma}$ in Equation (1), we have $f(K_{i,t}/L_{i,t}, G_{i,t}/L_{i,t}, H_{i,t}/L_{i,t}, L_{i,t}) = (K_{i,t}/L_{i,t})^{\alpha} (G_{i,t}/L_{i,t})^{\beta} (H_{i,t}/L_{i,t})^{\delta} L_{i,t}^{(\alpha+\beta+\delta+\gamma-1)}$, and taking the natural logarithm of the variables from Equations (2) and (3), we obtain the model:

$$y_{i,t} = a_{i,t} + (\alpha+\beta+\delta+\gamma-1)\, l_{i,t} + \alpha k_{i,t} + \beta g_{i,t} + \delta h_{i,t} \tag{4}$$

$$a_{i,t} = \mu_i + F_t'\lambda_i + e_{i,t}, \tag{5}$$

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 $h_{i,t} = \ln(H_{i,t}/L_{i,t})$. Note that the model can be written in a single-equation form as:

$$y_{i,t} = \mu_i + \zeta l_{i,t} + \alpha k_{i,t} + \beta g_{i,t} + \delta h_{i,t} + F_t'\lambda_i + e_{i,t}, \tag{6}$$

with $\zeta = (\alpha+\beta+\delta+\gamma-1)$. Following the existing contributions in the literature, $g_{i,t}$ is defined considering the productive public capital[1] and $h_{i,t}$ is measured as the average number of schooling years – see Serrano (1996).

The data employed in our study contains annual observations for the $N = 17$ Spanish regions (Autonomous Communities) observed over the $T = 48$ year period from 1964 to 2011.[2] The di-

[1]Productive public capital includes road and highways, ports, airports, railroads, water and sewer systems, public electric and gas utilities, and telecommunications.

[2]The Spanish regions are: Andalucía, Aragón, Asturias, Baleares, Canarias, Cantabria, Castilla y León, Castilla-La Mancha, Catalunya, Comunidad Valenciana, Extremadura, Galicia, Madrid, Murcia, Navarra, País Vasco and La

mensions of this panel data setup are similar to the ones that we can find in regional economic analysis, in general, where the statistical information is more scarce compared to the country basis studies. However, in this paper we use some panel data techniques that have shown good performance when applied to panel data setups with these dimensions – see, for instance, the simulation results in Pesaran (2007) and Kapetanios *et al.* (2011) for the test statistics that they propose.

The output $Y_{i,t}$ – measured by Gross Value Added – the stock of private capital $K_{i,t}$ and the stock of productive public capital $G_{i,t}$ are measured at 2008 constant prices, and are obtained from the BD.MORES database (December 2015) of the Spanish Ministry of Finance and Public Administrations.[3] The variable $H_{i,t}$ is the stock of human capital, measured as an average years of schooling, from the Stock de Capital Humano database, IVIE, and Serrano (1996). Finally, $L_{i,t}$ is labor, measured as the employed population of region i in the year t, which is obtained from the Stock de Capital Humano database, IVIE. The visual inspection of the variables that are used in this paper reveals, first, a clear trending pattern and, second, the comovement (cross-section dependence) that seems to be present in their evolution – see Figure 1.

3 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 weak cross-section dependence (WCD) statistic of Pesaran (2004, 2015) to test for the presence of cross-section dependence. Although initially Pesaran (2004) proposed the WCD statistic to test the null hypothesis of cross-section independence, Pesaran (2015) shows that the implicit null hypothesis of the WCD statistic is that the cross-sectional exponent of the vector of variables $y_t = (y_{1,t}, y_{2,t}, \ldots, y_{NT})'$ is

Rioja.

[3] At the moment of writing this paper (February 2016), December 2015 is the last update of the BD.Mores data base.

$\alpha_{BKP} < (2-\varepsilon)/4$ as $N \to \infty$, such that $T = \kappa N^{\varepsilon}$, for some $0 \leq \varepsilon \leq 1$, and a finite $\kappa > 0$ – for expositional purposes, we consider the vector of the logarithm of labor productivity y_t. Bailey *et al.* (2015) interpret α_{BKP} as the parameter that quantifies the degree of cross-sectional dependence, and is defined as the exponent of N that gives the maximum number of $y_{i,t}$ units that are pair-wise correlated. The values of α_{BKP} in the range $[0, 1/2)$ correspond to different degrees of weak cross-sectional dependence. In this case, the average of the pair-wise correlation coefficients tends to zero very fast. Values of α_{BKP} in the range $[1/2, 3/4)$ can be interpreted as an indicator of moderate degrees of cross-section dependence. Values of α_{BKP} in the range $[3/4, 1)$ point to the presence of quite strong cross-section dependence, for which the average of the pair-wise correlation coefficients tends to zero rather slowly. Finally, the average of the pair-wise correlation coefficients tends to a constant only if $\alpha_{BKP} = 1$. Consequently, the use of common factor models to capture strong dependence will be adequate if α_{BKP} is close or equal to unity – i.e., $\alpha_{BKP} \in [3/4, 1]$.

The values of the WCD test statistic reported in Table 1 indicate that we can easily reject the null hypothesis of weak cross-section dependence in favor of strong cross-section dependence for all variables – under the null hypothesis the WCD statistic converges to a standard normal distribution. As pointed out in Pesaran (2015) and Bailey *et al.* (2016), the large values of the WCD tests can be an indication that strong dependence is present, which can be captured by the means of an approximate common factor model. This conclusion is reinforced if we compute the α_{BKP} degree of cross-section dependence, which takes high values in all cases – Table 1 shows that α_{BKP} is larger than 0.9 in all cases. Therefore, the presence of cross-section dependence has to be taken into account when performing the panel data order of integration and cointegration analyses below.

3.1 Panel data order of integration analysis

Given the presence of cross-section dependence among the units of the panels, we proceed with the computation of the Bai and Ng (2004), Moon and Perron (2004) and Pesaran (2007) panel data unit root test statistics and the panel stationarity test in Bai and Ng (2005), using a linear time trend as the deterministic component in all cases. One feature that share these proposals is that they are valid when the units of the panel are affected by the presence of strong cross-section dependence, which is captured through the specification of an approximate common factor model. Thus, they cover one of the issues raised in Westerlund and Breitung (2013). However, the way in which the common factors appear in the model make these approaches to differ among them – see the discussion below. Finally, it is worth mentioning that these proposals also differ depending on the procedure that is used to estimate the deterministic specification, something that has been shown to be relevant by Westerlund and Breitung (2013). In this regard, Westerlund and Breitung (2013) evidence that using OLS detrending can reduce the empirical power of the panel unit root tests, whereas the use of, for instance, Maximum Likelihood (ML) estimates under the null hypothesis of unit root can give good results. All test statistics that we apply here are based on the use of OLS detrending, with the exception of the Bai and Ng (2004) proposal, which are based on ML detrending. Therefore, our analysis covers also another important issue raised in Westerlund and Breitung (2013).

The approximations that we apply here differ depending on the procedure that is used to estimate the common factors. Whereas Bai and Ng (2004) and Moon and Perron (2004) estimate the common factors using principal components analysis, the approximation in Pesaran (2007) uses the cross-section averages of the observable variables to proxy the common factors. It is worth mentioning that 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), the proposals in 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. These features have to be taken into account when interpreting the outcomes of the different statistical procedures.

Let us first focus on the results obtained using Pesaran's (2007) statistics. Table 1 presents the CIPS test statistic – the truncated version of the statistic produces identical results – which leads to conclude that, except for $y_{i,t}$ and $h_{i,t}$, the idiosyncratic component of the variables that we consider in the paper is I(1).[4] The evidence drawn from the computation of the panel data unit root test statistics in Moon and Perron (2004) reveals that the null hypothesis of unit root in the idiosyncratic component is only rejected at the 5% level of significance for $y_{i,t}$ – regardless of the test statistic that is used – and for $h_{i,t}$ when the t_b statistic is used – throughout the paper, the estimated number of common factors $(\hat{r})$ is obtained using the panel IC_{p2} information criterion in Bai and Ng (2002) with a maximum of four common factors. However, we cannot conclude anything about the order of integration of the common factors from the application of these statistics. A more informative picture is obtained from Bai and Ng's (2004) approach, provided that separate inference can be conducted on the idiosyncratic and the common factor components of the observable variables.

Table 1 summarizes the results from the application of the approach in Bai and Ng (2004), reporting the panel augmented Dickey-Fuller (ADF) statistic for the idiosyncratic component $\left(ADF_{\hat{e}}^{\tau}\right)$ of each variable and the MQ test statistics on the estimated common factors.[5] Except for $y_{i,t}$, the $ADF_{\hat{e}}^{\tau}$ test statistic does not reject the null hypothesis of panel unit root at the 5% level of significance for the idiosyncratic component. The MQ test statistics find that there is, at least, one I(1) non-stationary common factor affecting the variables under consideration – i.e., $\hat{r}_1 \geq 1$. These elements indicate that there is strong evidence that the five variables that are used in the estimation of the production function are I(1) non-stationary processes.

We complement the analysis of the stochastic properties following the proposal in Bai and Ng (2005), who test the null hypothesis of I(0) against the alternative hypothesis of I(1) considering the common factor model described in Bai and Ng (2004). The confirmatory analysis is carried

[4]The order of the autoregressive correction has been selected for each individual using the BIC information criterion with a maximum of five lags. This strategy has led us to compute the critical values as described in Pesaran (2007), but where this automatic selection of the order of the autoregressive correction has been used. The critical value at the 5% level of significance for $N = 17$ and $T = 48$ is -2.81 for both the untruncated and truncated CIPS test statistics.

[5]Following Ng and Perron (2001), the maximum number of lags that are used to compute the ADF statistic is set at $T^{1/3}$.

out computing the stationarity KPSS test statistic on the idiosyncratic and the common factor components. In all cases, the KPSS statistic of the estimated common factors for each variable reveals that there is, at least, one I(1) common factor affecting each variable, which reinforces the conclusions above – in order to save space, we do not report the values of these statistics, but they are available upon request. As shown in Bai and Ng (2005), the presence of I(1) non-stationary common factors prevents the computation of a pooled panel stationarity test for the idiosyncratic disturbance terms – pooling the individual KPSS test statistics would require all common factors to be I(0). Notwithstanding, the main conclusion that can be drawn is that the results that have been obtained using the stationarity test statistics in Bai and Ng (2005) are in accordance with the ones based on the panel data unit root test statistics.

To sum up, after analyzing the results from several types of panel data unit root and stationarity statistics, we can conclude that the variables can be characterized as I(1) stochastic processes, so that we can proceed with the panel data cointegration analysis.

3.2 Testing for panel data cointegration

This section tests the presence of panel cointegration using different proposals in the literature that consider the presence of cross-section dependence and structural breaks. Proceeding in this way, the analysis aims at obtaining robust conclusions about the existence of a long-run relationship among the variables involved in the production function that has been specified.

3.2.1 Panel cointegration without structural breaks

Let us first focus on the Banerjee and Carrion-i-Silvestre (2015) approach where the common factors are estimated using principal components. The panel IC_{p2} information criterion selects two common factors which are characterized as I(1) stochastic processes – see Table 2. The panel ADF statistic computed using the idiosyncratic disturbance terms (Z_c test statistic) leads to the rejection of the null hypothesis of spurious regression so we conclude that, once the presence of common factors is accounted for, there is a long-run relation among the variables that define the production function. This implies that the observable economic variables do not cointegrate alone – i.e., they take part of a cointegration relation that includes the presence of global stochastic trends. This result is in line with the theoretical arguments that claim that the TFP is an I(1) stochastic process.

Table 2 reports the panel Durbin-Hausman (DH) cointegration test statistics of Westerlund (2008). Both the DH_g and DH_p panel data statistics do not reject the null hypothesis of no cointegration at the 5% level of significance, although these statistics are designed under the assumption that the common factors have to be I(0). The later has been shown to be a problematic assumption, provided the evidence of I(1) common factors as mentioned above. The results from the $CADFC_P$ panel cointegration statistic in Banerjee and Carrion-i-Silvestre (2011) appear in Table 2, where the common factors are approximated by the cross-section averages of the observable variables as in Pesaran (2006).[6] As it can be seen, the $CADFC_P$ statistic leads to reject the null hypothesis of no cointegration at the 5% level of significance, which reinforces the conclusions that have been obtained so far.

The panel data test statistics that have been computed indicate that in general the variables involved in the production function define a cointegrating relationship. The evidence drawn by the panel statistics in Westerlund (2008) depends on the assumption that the common factors are I(0), a requirement that is not met in our case. All test proposed in Banerjee and Carrion-i-Silvestre (2011, 2015) are able to reject the null hypothesis of no cointegration with overwhelming evidence.

3.2.2 Panel cointegration with a structural break

In order to account for the presence of parameter instabilities, we have proceeded to compute the panel data cointegration tests designed in Banerjee and Carrion-i-Silvestre (2015) considering the effect of one structural break. Two different model specifications have been essayed depending on whether the structural break only affects the level of the model – Model 1 in Banerjee and Carrion-

[6] The order of the autoregressive correction has been selected for each individual using the BIC information criterion with a maximum of five lags. This strategy has led us to compute the critical values as described in Banerjee and Carrion-i-Silvestre (2011), but where this automatic selection of the order of the autoregressive correction has been used. The critical value at the 5% and 10% levels of significance for $N = 17$ and $T = 48$ are -2.33 and -2.23, respectively, for both the untruncated and truncated test statistics.

i-Silvestre (2015) – or both the level and the slope parameters of the model – Model 4 in Banerjee and Carrion-i-Silvestre (2015). To be specific, Model 4 implies the estimation of the extended version of the specification given in (6):

$$\begin{aligned} y_{i,t} &= \mu_{i,0} + \alpha k_{i,t} + \beta g_{i,t} + \delta h_{i,t} + \zeta l_{i,t} \\ &\quad + \left(\mu_{i,1} + \alpha_1 k_{i,t} + \beta_1 g_{i,t} + \delta_1 h_{i,t} + \zeta_1 l_{i,t}\right) DU_t + F_t' \lambda_i + e_{i,t}, \end{aligned} \tag{7}$$

where DU_t is a dummy variable defined as $DU_t = 1$ if $t > T_b$, and 0 otherwise, with T_b the break date – Model 1 imposes $\alpha_1 = \beta_1 = \delta_1 = \zeta_1 = 0$ in (7). The computation of the panel data cointegration statistic using Models 1 and 4 reveals that the null hypothesis of no cointegration is strongly rejected – see Table 2. The procedure detects the presence of one or two non-stationary I(1) common factors, depending on the model specification. As it can be seen, the consideration of parameter instabilities in the model does not change the conclusion that has been obtained so far, i.e., that there exists a cointegration relationship among the variables that define the production function that has been specified.

3.3 Estimation of the production function

The estimation of the panel production function is conducted in two stages. First, the analysis focuses on the production function that assumes constant parameters, covering the issues of strong and weak cross-section dependence. Second, the study concentrates on the specification that considers the effect of one structural break. This increases the flexibility of the model specification and permits the computation of elasticities and other related measures that change through time.

3.3.1 Estimation of the production function without structural breaks

There are few theoretical proposals in the literature that allow the estimation of panel cointegrating relations with common factors that capture pervasive cross-section dependence. First, we apply the continuously-updated and fully-modified (CupFM) and the continuously-updated and biascorrected (CupBC) estimators proposed in Bai *et al.* (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 in an iterative fashion. Both estimation procedures render consistent and efficient estimates of the cointegrating vector regardless of whether we have I(0) and/or I(1) common factors, but they differ in terms of when the endogeneity correction is applied – in principle, CupFM should have better properties, since it applies the endogeneity correction in each iteration, while the CupBC does so once convergence is achieved. Second, we also use the pooled common correlated effects (CCEP) estimator in Kapetanios *et al.* (2011), which 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 observable variables of the model.

Although both approaches lead to consistent estimates of the parameters if the assumptions in which they rely on are met, the proposal of Bai *et al.* (2009) uses an efficient estimation procedure, which takes into account the possibility that there might be endogenous regressors. On the contrary, the estimator in Kapetanios *et al.* (2011) assumes that the stochastic regressors are weakly exogenous, a situation that might not hold in our case – note that the definition of the variables in per worker terms implies that the employment appears both on the left (in the denominator of the dependent variable) and the right of the model equation, which casts doubts on the exogeneity assumption of the regressors. This suggests that more stress have to be given to the estimates of the parameters that consider endogenous regressors.[7]

Table 3 reports the estimation of the Cobb-Douglas production function in Equation (6).[8] As it can be seen, there are important differences among the parameter estimates depending on the estimation technique that is used. The parameters obtained using the CCE estimation procedure are in general smaller than the ones provided by the Cup-based estimation techniques. Note that none of the CCE parameter estimates are statistically significant at the 5% level of significance – the parameters for the private capital and the labor are statistically significant at the 10% level.

[7]We thank Chihwa Kao and Takashi Yamagata for providing the Gauss code.

[8]As suggested in Bai *et al.* (2009), the computations of the CupFM and CupBC estimators base on the use of the Fejer kernel with a bandwidth of 10 lags.

Although this might be surprising at first sight, it should be borne in mind that the CCE assumes that the stochastic regressors are exogenous, an assumption that might be undermining the analysis. Consequently, in what follows we will rely on the efficient parameter estimates that deliver the Cup-based estimation techniques.

The estimated coefficients represent the elasticity of output with respect to physical capital, public capital and human capital – the elasticity of output with respect to labor (γ) can be recovered recalling that $\zeta = (\alpha+\beta+\delta+\gamma-1)$ in Equation (6). Panel A of Table 3 presents the estimation of the parameters for the specification that does not include structural breaks. First, note that all Cup-based estimates are statistically significant. The coefficient of $k_{i,t}$ indicates that the elasticity of output with respect to the physical capital is 0.5 (CupBC) and 0.54 (CupFM), values that are in the middle of the range of values that are commonly found in the empirical literature for the Spanish regions.[9] The coefficient of $g_{i,t}$ shows that the elasticity of output with respect to the productive public capital is 0.082 (CupBC) and 0.095 (CupFM),[10] whereas the elasticity of output with respect to the human capital is 0.136 (CupFM) and 0.209 (CupBC).[11]

Although these values are, in words of Boscá *et al.* (2011), reasonable for the Spanish economy, before proceeding with further analyses we should check whether the approximated common factor model captures the cross-section dependence that is present among the regions. This requirement is needed in order to ensure that the estimation of the parameters is efficient.

Spatial dependence So far, we assumed that the cross-section dependence among the Spanish regions is captured through the specification of a model of unobserved common factors. The use of a common factor model aims at capturing the existence of strong dependence among the units of a panel data set, a feature that appears when the dependence affect all units in the panel and its

[9] 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 *et al.* (2011) is 0.31. Note that specification of the variables, the model, the data and estimation techniques differ from one study to another.

[10] The estimate of Bajo and Díaz (2005) is 0.09, and the one by Márquez *et al.* (2011) is 0.10. Note that the estimated values fall inside the “reasonable estimates” [0.05, 0.1] interval that define Boscá *et al.* (2011) for the Spanish economy – these authors argue that this interval is in accordance with the accounting information of the Spanish Economy.

[11] These estimates are similar than those of Serrano (1996), who obtained a value of 0.216, and of Bajo and Díaz (2005), who obtained a value of 0.14.

effect does not vanish as more units are added. The econometric techniques that have been applied consider this form of strong dependence when estimating the parameters of the model in order to get consistent and efficient estimates. However, it is possible that the Spanish regions reflect forms of local dependence that are spatial in nature – i.e., that the regions might be affected by the presence of weak dependence. Spatial dependence assumes that the structure of the cross-section dependence is related to location and distance among units, being popular specifications the spatial autoregressive model, the spatial moving average model and the spatial error component model. These structures define particular cases of weak cross-section dependence situations, where the dependence exists only among adjacent observations. 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 – see Chudik *et al.* (2011) and Banerjee and Carrion-i-Silvestre (2011) for the discussion about the distinction between weak and strong dependence.

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 *et al.* (2010) and Pesaran and Tosetti (2011), and for each idiosyncratic disturbance term, we specify the following spatial error model:

$$\tilde{e}^*_{i,t} = \rho \sum_{j=1}^{N} w_{i,j} \tilde{e}^*_{j,t} + v_{i,t}, \tag{8}$$

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

$$L = -\left(\frac{NT}{2}\right) \ln(\sigma^2_v) + T \ln |I_N - \rho W| - \frac{1}{2\sigma^2_v} \sum_{t=1}^{T} \left(\tilde{e}^*_t - \rho W \tilde{e}^*_t\right)' \left(\tilde{e}^*_t - \rho W \tilde{e}^*_t\right), \tag{9}$$

where $\tilde{e}^*_t = \left(\tilde{e}^*_{1,t}, \tilde{e}^*_{2,t}, \ldots, \tilde{e}^*_{N,t}\right)'$ denotes the idiosyncratic disturbance terms that are estimated using either the CupFM or CupBC estimators. Since Baleares and Canarias are islands, they have no neighbors and we eliminate their data for this analysis, leaving a panel data set of $N = 15$ regions and $T = 48$ years.

The results of the ML estimation of ρ in Equation (8) are presented in Table 3, which reveal that the estimation of ρ is positive and statistically significant, regardless of the estimation technique that is used. The detection of spatial autocorrelation among the disturbance terms leads us to conclude that, besides the existence of strong cross-section dependence, there is weak cross-section dependence among the idiosyncratic errors. Consequently, the estimation of the parameters would not be fully efficient, since the idiosyncratic errors are correlated. In order to address this issue, we have estimated the production function in Equation (6) using the spatial filtered variables. If y defines the $(T \times N)$-matrix of the logarithm of the output per worker and W the $(N \times N)$-matrix of weights, the spatial filtered $(T \times N)$-matrix y^* is computed as $y^* = y - \rho y W'$ – the same transformation is applied to the k, g, h and l $(T \times N)$-matrices.[12]

The estimation results using the spatial filtered variables are collected in Table 3, where the CupFM or CupBC-based spatial autoregressive parameter is used depending on the case. The estimated parameters that are obtained using the spatial filtered variables are statistically significant, with values that are similar to the ones estimated with the original variables – the elasticities for the private and human capitals are slightly smaller, whereas the elasticity of the public capital and the parameter of the labor are slightly bigger. Consequently and although the presence of uncorrelated idiosyncratic errors violates one of the assumptions in Bai *et al.* (2009), these results might indicate that the weak cross-section dependence does not affect the estimation procedures used in the previous section in an important way.

Finally, it would be interesting to look at the effect that the estimated common component has on each region – the common component for each region i is defined by $F_t'\lambda_i$ in Equation (5), $i = 1, \ldots, N$. Unfortunately, give an economic interpretation to this component is not straightforward, since the estimated common factors and loadings are identified up to a rotation of the true unknown

[12]The W matrix has been normalized so that the sum of the elements of each row equals one. To avoid the exclusion of the islands, we have assigned zeros on the corresponding rows and columns of W.

common factors and loadings – see Bai *et al.* (2009), Proposition 5. Figure 2 provides the estimated common component for each region (upper-left picture), and the mean and standard deviation (upper-right picture) of the common component – the CupBC-based estimates are almost identical. As it can be seen, the effect that the common component has on each region is heterogeneous, with an average effect that does not follows a monotonic pattern – the mean of the common component experiences a high increase up to mid 70s, but then decreases to evolve around zero from early 80s on. One interesting feature is that the standard deviation tends to decrease along the period that has been analyzed, which shows that the strong dependence component that has been affecting the Spanish regions has become more homogeneous from late 80s on. This might be related to the analysis of economic convergence across regions, where the common component would be capturing the convergence process that have experienced the Spanish regions – the σ-convergence definition in Barro and Sala-i-Martin (1992).

Marginal product of the private and public capitals and returns to education From an economic point of view, the estimation of the production function that has been conducted allows us to compute the marginal products of the inputs. The marginal product of the private and public capitals can be obtained as

$$MP_{K,t} = \varepsilon_{Y,K}\frac{Y_t}{K_t}; \quad MP_{G,t} = \varepsilon_{Y,G}\frac{Y_t}{G_t}, \tag{10}$$

where $\varepsilon_{Y,K}$ and $\varepsilon_{Y,G}$ denote the elasticity of output to private and public capitals. In the case of human capital, the literature has also investigated the impact of an additional year of education on output, i.e., the return to education. This is captured through the estimation of the semi-elasticity of output with respect to human capital, which is given by – see López-Bazo and Moreno (2004):

$$R_{H,t} = \frac{\partial \ln Y_t}{\partial H_t} = \varepsilon_{Y,H}\frac{1}{H_t}, \tag{11}$$

where $\varepsilon_{Y,H}$ denotes the elasticity of output to human capital. In what follows, we use the elasticities estimates that are obtained using the spatial filtered variables – results available upon request show that similar values are obtained if we use the estimates that only accounts for the strong cross-section dependence.

Table 4 collects the marginal products and return to human capital for the CupFM and CupBC estimates considering a representative region, computed using the cross-section average of the Spanish regions – for instance, for the output, $Y_t = N^{-1}\sum_{i=1}^{N} Y_{i,t}$. Regardless of the estimator, it can be seen that these measures decrease during the period that is studied. It is worth noticing that the marginal product of the public capital is always above the marginal product of the private capital, although a convergence process has taken place. The large gap between the marginal products, especially from the beginning of the period up to mid 90s, would indicate that the public capital was under-provided, relative to the endowment of private capital – see Bajo and Díaz (2005). The dramatic decrease of the marginal product of the public capital can be due to the infrastructure scarcity that suffered the Spanish economy during the 60s and 70s – see the discussion below. The average values of the marginal products for the 1964-2011 period are 0.164 (CupFM) and 0.149 (CupBC) for the private capital, and 0.329 (CupFM) and 0.298 (CupBC) for the public capital, values that are in accordance with previous analyses of the Spanish regional case – Bajo and Díaz (2005) estimated marginal products of 0.156 and 0.231 for the private and public capitals, respectively, for the 1965-1995 period. The estimated returns to education that have been obtained – 0.018 (CupFM) and 0.027 (CupBC) on average for the 1964-2011 period – are considerably smaller than others reported in the literature for the Spanish economy. In this regard, López-Bazo and Moreno (2004) using a shorter time period and an approximation based on the estimation of cost functions, report values for the returns to education that go from 0.131 (for the year 1980) to 0.069 (for the year 1995). Similar results are obtained in Serrano (1996), although his approach does not base on a long-run equilibrium relationship. Notwithstanding, our estimates are in accordance with existing evidence at international level – for instance, De La Fuente and Doménech (2006) estimate the returns to education at 0.03 using a sample of OECD countries for the period 1960-1990.

Table 5 presents the marginal products and returns to education computed using the time average for each region. In this case, Y_t K_t, G_t and H_t in Equations (10) and (11) are replaced by Y_i K_i, G_i and H_i, with, for instance, for the output $Y_i = \sum_{t=1964}^{2011} Y_{i,t}$, $i = 1, \ldots, N$. The marginal product of the public capital is larger than the private capital one in all regions, which reflects the under-provision of public infrastructures, relative to the endowment of private capital. As for the returns to education, the CupFM-based estimates (around 0.016) are smaller than the CupBC-based ones (around 0.024), and both of them clearly below the estimates reported in López-Bazo and Moreno (2004) – 0.093 for the average for Spain. These discrepancies might be due to the use of a different time period, although it might be the case that the specification of a model that includes more determinants of the output would be reducing the estimation bias in which other analyses might be incurring.

Finally, we should highlight the negative and highly significant coefficient for $l_{i,t}$, which implies that the constant returns to scale hypothesis of the observable inputs cannot be accepted. In this regard, the negative sign of $\hat{\zeta}$ indicates diminishing returns to scale on the observable productive factors that have been considered in the model.

3.3.2 Estimation of the production function with a structural break

This section extends the previous analysis to accommodate one structural break using the general specification given by Model 4 in Equation (7), where the structural break can affect all parameters of the model. The estimation of the common break date (T_b) is conducted through the minimization of the sum of squared residuals of (7) over all possible break dates – following the convention in the literature, the admissible T_b is defined in the close set given by $T_b \in [0.15T, 0.85T]$.

Panel B of Table 3 presents the parameter estimates for the specification that allows for one structural break, which has been estimated at 1998. Before proceeding with the analysis, it is worth noticing that the estimation of the spatial autoregressive model for the idiosyncratic residuals that has been described in Section 3.3.1 indicates that the spatial autoregressive parameter ρ is not statistically significant, regardless of the specification used. In this case, the approximate common factor model captures the cross-section dependence that exists among the Spanish regions in a satisfactory way. This robustness analysis reinforces the validity of the optimal estimates of the production function that accommodates the presence of one structural break.

As it can be seen, there are some elasticities that experience a change in its value after the euro was launched, although results and the economic interpretation depend on the estimation procedure that is used. Let us first focus now on the CupFM-based estimates reported in Table 3. The elasticities of the private and public capitals are positive and statistically significant – note that the elasticity of the public capital falls inside the [0.05, 0.10] reasonable interval suggested in Boscá *et al.* (2011) – whereas the negative and statistically significant coefficient for the labor suggest the existence of diminishing returns to scale. Interestingly, the elasticity of the human capital is not statistically significant. Except for the labor, the structural break affects all coefficients. The private and public capital elasticities decrease after the structural break has taken place, although the overall magnitudes continue to be positive. During the second subperiod the estimated elasticity for the human capital is positive and statistically significant.

Table 3 presents the estimation results of the restricted model specification, where the variables for which the parameters that are not statistically significant at the 5% level have been removed from the model. Similar results are obtained for the estimated parameters, although now the elasticity of the public capital after 1998 is negative (with a small value of $0.077 - 0.081 = -0.004$). As above, the elasticity of the human capital is positive and statistically significant. The economic implications of these results are quite interesting and in accordance with some previous analyses for the Spanish economy. In this regard, Boscá *et al.* (2011) and De La Fuente (2010) stress the idea that the reason behind the high values of the public capital elasticity might be the infrastructure scarcity that is found in developing economies. Early stages of economic development require a level of infrastructures that will impact productivity in a positive way, but once economic development is achieved, the effects of the infrastructures would be smaller becoming either non-significant or negative in some cases – see Holtz-Eakin (1994) and Holtz-Eakin and Schwartz (1995) for the US regions. This defines the so-called "saturation effect" of infrastructures. Consequently, the fact that high public capital elasticity values are found for the Spanish regions might be due to infrastructure scarcity during the 60s and 70s. Notwithstanding, investment efforts of local, regional and national governments during the 80s have reduced (economic and social) infrastructure scarcity – see Boscá *et al.* (2011). This agrees with Mas *et al.* (1996) infrastructures saturation effect analysis, who reestimate a production function with different subsamples of increasing length (recursive estimation) and found that the recursive estimate of public capital elasticity was experiencing a monotonic decrease – going from 0.1404 for the 1964-1973 subsample to 0.0771 when the whole period (1964-1991) was used. Note that this can also be interpreted as evidence of structural breaks affecting the model specification, something that is covered by the framework that has been used in this section. The estimate of public capital elasticity in Mas *et al.* (1996) for the 1964-1991 period (0.0771) is quite similar to the estimate that is obtained in this paper for the 1964-1998 period (0.077). The fact that this elasticity is -0.004 for the 1999-2011 subperiod evidences the saturation effect of the infrastructures for the Spanish regions.

The picture is slightly different for the CupBC-based estimates. As above, the coefficient of the labor is not affected by the structural break, whereas the elasticity of the human capital – δ in (6) – is statistically significant at the 10% level – it is significant at the 5% level if we consider the sign and use the right tail of the distribution when performing the statistical inference. Similarly, the coefficient that captures the effect of the structural break on the public capital – β_1 in (6) – is significant at the 5% level if we consider the sign and use the left tail of the distribution when carrying out the statistical inference. Taking into account these features, we have proceeded to estimate the restricted model that excludes the effect of the structural break on the labor. The CupBC-based estimates are statistically significant,[13] with sign and magnitude that are similar to the CupFM-based ones. As with the CupFM-based estimates, the elasticities of the private and public capitals decrease after the structural break. The novelty now is that the elasticity of the human capital is significant throughout the whole period, but it experiences an important increase after the struc-

[13] The coefficient of the human capita is statistically significant at the 5.9% level of significance if we base on a two-tailed inference analysis, although is significant at the 5% level if we consider the sign and use the right tail of the distribution.

tural break, reaching a values that is comparable to the CupFM-based estimate. Interestingly, the elasticity of the public capital is still positive after the structural break, although it shows a small value ($0.061 - 0.059 = 0.002$). Strictly speaking, this contradicts the economic interpretation that has been obtained using the CupFM-based estimates (-0.004 for the 1999-2011 period), although in both cases the qualitative conclusion is coherent, i.e., the elasticity of the output with respect to the stock of infrastructure is really small, pointing to the presence of a saturation effect.

Figure 2 depicts the estimated common component – $F_t'\lambda_i$ in Equation (5) – for each region (lower-left picture), and its mean and standard deviation (lower-right picture) – as above, we only report the results for the CupFM-based estimates since the ones using the CupBC are similar. The effect of the common component on the Spanish regions is quite heterogeneous during the 1964-1998 subperiod, although after the structural break we observe a complete different pattern. On average, the effect of the common component shows an increasing trend, although the evolution is not smooth. Further, the effects of the common component turns out to be quite homogeneous across regions after 1998. This feature is evidenced by the evolution of the standard deviation, which shows a huge decrease during the 60s, then experiences a mild increase up to 1998 and, finally, falls to values close to zero after the structural break. This suggests that the incorporation of Spain to the EMU in 1999 has increased the homogeneity of the strong cross-section dependence effects across the Spanish regions. As above, this can also be related to the σ-convergence process that might have experienced the Spanish regions, which would be accelerated in the EMU era.

Marginal product of the private and public capitals and returns to education Table 4 summarizes the results of computing the marginal products of private and public capitals, and the returns to education considering a representative region, using the restricted specification of Model 4. The qualitative conclusion that arises is that the marginal products have decreased throughout the period studied. Up to late 80s, the marginal product of public capital was above the private capital one. However, after the incorporation of Spain at the European Economic Community in 1986, the marginal product of private capital has been placed above the public capital one. The structural break estimated in 1998 has implied a decrease of the marginal products, achieving small or even negative values in the case of the public capital. The effects of the human capital on productivity depend on the estimates that are used. If we focus on the CupFM estimates, returns to education are significantly different from zero in the 1999-2011 subperiod, and show a decreasing pattern (from 0.024 to 0.02). If we focus on the CupBC estimates, the effect of human capital on productivity is statistically significant for the whole 1964-2011 period. During 1964-1998 subperiod the returns to education decrease, reaching a value close to zero (from 0.013 to 0.006). The European and Monetary Union launch has implied an increase of returns to education, although they have experienced a mild decrease (from 0.023 to 0.019). These estimates are clearly lower than the one obtained in López-Bazo and Moreno (2004) for the average of Spain – which is 0.093 for the 1980-1995 period – but in accordance with other values reported in the economic literature. As mentioned above, de la Fuente and Doménech (2006) estimate the returns to education at 0.03 using a sample of OECD countries for the period 1960-1990.

Finally, Table 5 presents the marginal products and returns to education computed using the time average for each region. For the first subperiod, the marginal product of public capital is above the private capital one in 13 (CupFM) and 11 (CupBC) out of 17 Spanish regions, indicating under-provision of public capital, relative to the endowment of private capital.[14] During the 1999-2011 subperiod, the marginal product of public capital is negative in all cases. In general, we observe heterogeneity on the marginal products, when they are statistically significant. On the contrary, returns to education are quite homogeneous across regions – around 2% in 1999-2011, regardless of the estimator that is used.

4 Conclusions

This papers reexamines the evidence of cointegration among the output, physical capital, human capital, public capital, and labor. We consider annual data for seventeen Spanish regions observed

[14]The exceptions are Aragón, Castilla-León, Castilla-La Mancha, Extremadura, Galicia and Madrid.

over the period 1964-2011. 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 can be characterized as I(1) non-stationary stochastic processes. Therefore, the application of panel data cointegration techniques are required to obtain consistent estimates of the parameters of interest. The paper takes advantage of the recently developed non-stationary panel data analysis methodology that permits cross-section dependence across the units of the panel.

The results reveal evidence of panel data cointegration among the variables of the model up to the presence of I(1) non-stationary common factors. Consequently, the observable economic variables alone do not generate an equilibrium relationship. Thus, we need to consider the otherwise expected global stochastic trends that define the TFP. We estimate the Spanish regional production function using Bai *et al.* (2009) and Kapetanios *et al.* (2011) panel data cointegration estimators, considering the possibility of parameter instabilities – due to the existence of structural breaks – and cross-section dependence. The results indicate that physical capital, human capital, public capital (all in per capita terms) affect the Spanish productivity, although the sign and magnitude depend on the period that is analyzed. Finally, the negative coefficient that has been obtained for the labor indicates the existence of decreasing returns to scale on the observable inputs.

Andrews, D. W. K. (2005) Cross-section regression with common shocks, *Econometrica*, 73, 1551-1585.

Aschauer, D. (1989) Is public expenditure productive?, *Journal of Monetary Economics*, 23, 177-200.

Bai, J., Kao, C. & Ng, S. (2009) Panel cointegration with global stochastic trends, *Journal of Econometrics*, 149, 82-99.

Bai, J,. & Ng, S. (2002) Determining the number of factors in approximate factor models, *Econometrica*, 70, 191-221.

Bai, J,. & Ng, S. (2004) A PANIC attack on unit roots and cointegration. *Econometrica*, 72, 1127-1177.

Bai, J,. & Ng, S. (2005) A new look at panel testing of stationarity and the PPP hypothesis. In: *Identification and Inference for Econometric Models, Essays in Honor of Thomas J. Rothenberg*. Editors, Andrews D. W. K., Stock J. H. & Rothenberg T. J. Cambridge University Press.

Bai, J,. & Ng, S. (2010) Panel unit root tests with cross-section dependence: a further investigation, *Econometric Theory*, 26, 1088-1114.

Bailey, N., Holly, S. & Pesaran, M. H. (2016) A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence, *Journal of Applied Econometrics*, 31, 249-280.

Bailey, N., Kapetanios, G. & Pesaran, M. H. (2015) Exponent of cross-sectional dependence, estimation and inference, *Journal of Applied Econometrics*, forthcoming.

Bajo-Rubio, O. & Díaz-Roldán, C. (2005) Optimal endowments of public capital: an empirical analysis for the Spanish regions, *Regional Studies*, 39, 297-304.

Baltagi, B. H. & Pinnoi, N. (1995) Public capital stock and state productivity growth: further evidence from an error component model, *Empirical Economics*, 20, 351-359.

Banerjee, A. & Carrion-i-Silvestre, J. L. (2011) Testing for panel cointegration using common correlated effects estimators, *Working Paper*, University of Barcelona.

Banerjee, A. & Carrion-i-Silvestre, J. L. (2015) Cointegration in panel data with breaks and cross-section dependence, *Journal of Applied Econometrics*, 30, 1-23.

Banerjee, A., Marcellino, M. & Osbat, C. (2005) Testing for PPP, should we use panel methods?, *Empirical Economics*, 30, 77-91.

Banerjee, A., Eberhardt, M. & Reade, J. J. (2010) Panel estimation for worriers, *Working Paper*, University of Birmingham.

Banerjee, A. & Wagner, M. (2009) Panel methods to test for unit roots and cointegration. In *Palgrave Handbook of Econometrics, Vol. 2, Applied Econometrics*, Mills T. C. & Patterson, K. (eds). Palgrave Macmillan.

Barro, R. & Sala-i-Martin, X. (1992) Convergence, *Journal of Political Economy*, 100, 223-251.

Berndt, E. R. & Hanson, B. (1992) Measuring the contribution of public infrastructure capital in Sweden, *Scandinavian Journal of Economics*, 94, 151-168.

Boscá, J. E., Escribá, J. & Murgui, M. J. (2011) La efectividad de la inversión en infraestructuras públicas: una panorámica para la economía española y sus regiones, *Investigaciones Regionales*, 20, 195-217.

Breitung, J. & Pesaran, M. H. (2008) Unit roots and cointegration in panels. In *The Econometrics of Panel Data*, Matyas, L. & Sevestre, P. (eds), Klüver Academic Press.

Carrion-i-Silvestre, J. L. & Surdeanu, L. (2011) Panel cointegration rank testing with cross-section dependence, *Studies in Nonlinear Dynamics & Econometrics*, 15, article 4.

Chudik, A., Pesaran, M. H. & Tosetti, E. (2011) Weak and strong cross section dependence and estimation of large panels, *Econometrics Journal*, 14, C45-C90.

Cobb, C. W. & Douglas, P. H. (1928) A theory of production, *American Economic Review*, 18, 139-165.

Costantini, M. & Destefanis, S. (2009) Cointegration analysis for cross-sectionally dependent panels: the case of regional production functions, *Economic Modelling*, 26, 320-327.

Dalamagas, B. (1995) A reconsideration of the public sector's contribution to growth, *Empirical Economics*, 20, 385-414.

De La Fuente, A. & Doménech, R. (2006) Human capital in growth regressions: how much difference does data quality make?, *Journal of the European Economic Association*, 1, 1-36.

Evans, P. & Karras, G. (1994) Is government capital productive? evidence from a panel of seven countries, *Journal of Macroeconomics*, 16, 271-279.

García-Milá, T. & McGuire, T. (1992) The contribution of publicly provided inputs to states economics, *Regional Science and Urban Economics*, 22, 229-41.

García-Milá, T., McGuire, T. & Porter, R. (1996) The effect of public capital in state-level production functions reconsidered, *The Review of Economics and Statistics*, 78, 177-180.

Holly, S., Pesaran, M. H. & Yamagata, T. (2010) A spatio-temporal model of house prices in the USA, *Journal of Econometrics*, 158, 160-173.

Holtz-Eakin, D. (1994) Public sector capital and the productivity puzzle, *Review of Economics and Statistics*, 76, 12-21.

Holtz-Eakin, D. & Schwartz, A. (1995) Infrastructure in a structural model of economic growth, *Regional Science and Urban Economics*, 25, 131-151.

Kapetanios, G., Pesaran, M. H. & Yamagata, T. (2011) Panels with non-stationary multifactor error structures, *Journal of Econometrics*, 160, 326-348.

López-Bazo, E. & Moreno, R. (2004) Profitability of investments in education: evidence from Spanish regions, *Regional Studies*, 46, 1333-1346.

Márquez, M. A., Ramajo, J. & Hewings, G. J. D. (2011) Public capital and regional economic growth, a SVAR approach for the Spanish regions, *Investigaciones Regionales*, 21, 199-223.

Mas, M., Maudos, F., Pérez, F. & Uriel, E. (1996) Infrastructures and productivity in the Spanish regions, *Regional Studies*, 30, 641-649.

Merriman, D. (1990) Public capital and regional output, *Regional Science and Urban Economics*, 20, 437-458.

Moon, H. R. & Perron, B. (2004) Testing for a unit root in panels with dynamic factors, *Journal of Econometrics*, 122, 81-126.

Munnell, A. (1990) How does public infrastructure affect regional economic performance?, *New England Economic Review*, Federal Reserve Bank of Boston, September-October, 11-32.

Ng, S. & Perron, P. (2001) Lag length selection and the construction of unit root tests with good size and power, *Econometrica*, 69, 1519-1554.

Otto, G. & Voss, G. (1996) Public capital and private production in Australia, *Southern Economic Journal*, 62, 723-39.

Pesaran, M. H. (2004) General diagnostic tests for cross section dependence in panels, *Working Paper* 435. Cambridge University.

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

Pesaran, M. H. (2007) A simple panel unit root test in the presence of cross section dependence, *Journal of Applied Econometrics*, 22, 265-312.

Pesaran, M. H. (2015) Testing weak cross-sectional dependence in large panels, *Econometric Reviews*, 34, 1089-1117.

Serrano, L. (1996) Indicadores de capital humano y productividad, *Revista de Economía Aplicada*, 10, 177-190.

Serrano, L. (1997) Productividad del trabajo y capital humano en la economía española, *Moneda y Crédito*, 205, 79-101.

Sosvilla-Rivero, S. & Alonso, J. (2005) Estimación de una función de producción MRW para la economía española, 1910-1995, *Investigaciones Economicas,* 29, 609-624.

Westerlund, J. (2008) Panel cointegration tests of the fisher hypothesis, *Journal of Applied Econometrics*, 23, 193-233.

Westerlund, J. & Breitung, J. (2013) Lessons from a decade of IPS and LLC, *Econometric Reviews*, 32, 547-591.

Wylie, P. (1996) Infrastructure and Canadian economic growth, 1946-91, *Canadian Journal of Economics*, 29, S350-S355.

Table 1: Cross-section independence and panel data unit root tests

	Statistic	$y_{i,t}$	$k_{i,t}$	$g_{i,t}$	$h_{i,t}$	$l_{i,t}$
Pesaran (2015)	WCD	19.738	32.927	35.090	12.896	37.197
Bailey et al. (2015)	Exponent of CSD (α_{BKP})	0.990	1.004	1.004	0.906	1.004
	$s.e.(\alpha_{BKP})$	0.023	0.036	0.063	0.036	0.052
Pesaran (2007)	CIPS	-2.934	-2.549	-2.003	-2.952	-2.342
Moon and Perron (2004)	t_a	-3.467	0.229	-1.217	-1.589	-1.519
		(0.000)	(0.590)	(0.112)	(0.056)	(0.064)
	t_b	-3.268	0.218	-1.190	-1.684	-1.537
		(0.001)	(0.586)	(0.117)	(0.046)	(0.062)
Bai and Ng (2004)	$ADF_{\hat{e}}^{\tau}$	-1.681	0.739	-0.615	1.592	0.438
		(0.046)	(0.770)	(0.269)	(0.944)	(0.669)
	$\hat{r}$	1	2	2	1	1
	$\hat{r}_1$	1	2	2	1	1
	MQ_f^{τ}	-4.582	-13.103	-17.348	-1.915	-4.324
	MQ_c^{τ}	-2.639	-8.911	-10.140	1.444	-1.760

Note: p-values between parentheses. The estimation of the number of common factors (r) is obtained using the panel IC_{p2} information criterion in Bai and Ng (2002) with a maximum of four common factors

Table 2: Panel data cointegration analysis

	No structural breaks				One structural break					
	Statistic	$\hat{r}$	$\hat{r}_1^{NP}$	$\hat{r}_1^{P}$	Model	Z_c	$\hat{r}$	$\hat{r}_1^{NP}$	$\hat{r}_1^{P}$	$\hat{T}_b$
Z_c	-3.767	2	2	2	1	-4.177	2	2	2	1998
DH_p	-0.401	2			4	-7.798	1	1	1	1978
	(0.656)									
DH_g	-0.093	2								
	(0.537)									
$CADFC_P$	-2.394									

Note: p-values between parentheses. The estimation of the number of common factors (r) is obtained using the panel IC_{p2} information criterion in Bai and Ng (2002) with a maximum of four common factors. For the no structural breaks model, the 5% critical value for Z_c statistic is -1.645. For the $CADFC_P$ statistic, the 5% critical value is -2.33. In Model 1, presented by Banerjee and Carrion-i-Silvestre (2015), the structural break only affects the level of the model. In Model 4, illustrated by the same authors, the structural break affects both the level and the slope parameters of the model. The 5% critical value for the Z_c statistic for both Models 1 and 4 specifications is -2.219 (see Table III in Banerjee and Carrion-i-Silvestre (2015)).

Table 3: Estimation of the Spanish regional Cobb-Douglas production function

Estimation of the production function

	Panel A: No structural breaks					Panel B: One structural break			
				Spatial filtered vars					
	CCE	CupFM	CupBC	CupFM	CupBC	CupFM		CupBC	
						Model 4	Model 4r	Model 4	Model 4r
$k_{i,t}$	0.284	0.539	0.496	0.517	0.471	0.698	0.674	0.662	0.648
	(1.812)	(28.357)	(26.684)	(24.946)	(23.308)	(26.428)	(29.316)	(25.152)	(26.280)
$g_{i,t}$	0.074	0.095	0.082	0.104	0.094	0.082	0.077	0.065	0.061
	(1.079)	(7.126)	(6.191)	(7.151)	(6.538)	(5.610)	(5.779)	(4.322)	(4.105)
$h_{i,t}$	-0.007	0.136	0.209	0.116	0.179	-0.035		0.050	0.053
	(-0.092)	(5.383)	(8.259)	(4.187)	(6.474)	(-1.189)		(1.668)	(1.887)
$l_{i,t}$	-0.382	-0.281	-0.292	-0.274	-0.286	-0.385	-0.402	-0.383	-0.409
	(-1.824)	(-21.454)	(-22.444)	(-18.911)	(-19.693)	(-16.470)	(-24.051)	(-16.607)	(-22.453)
$k_{i,t} * DU_t$						-0.390	-0.382	-0.391	-0.364
						(-10.194)	(-10.694)	(-10.193)	(-9.904)
$g_{i,t} * DU_t$						-0.063	-0.081	-0.046	-0.059
						(-2.061)	(-2.765)	(-1.520)	(-2.028)
$h_{i,t} * DU_t$						0.228	0.233	0.144	0.172
						(3.315)	(5.083)	(2.084)	(3.668)
$l_{i,t} * DU_t$						-0.005		-0.043	
						(-0.150)		(-1.267)	
$\hat{r}$		3	3	3	3	3	3	3	3
IC_{p2}		-2.854	-2.860	-2.847	-2.854	-2.979	-2.984	-2.991	-2.989
$\hat{T}_b$						1998	1998	1998	1998

Spatial dependence of the idiosyncratic disturbance terms

	Panel A: No structural breaks					Panel B: One structural break			
	CCE	CupFM	CupBC	CupFM	CupBC	CupFM		CupBC	
						Model 4	Model 4r	Model 4	Model 4r
$\hat{\rho}$		0.125	0.114			-0.056	-0.054	-0.054	-0.054
		(2.614)	(2.373)			(-1.104)	(-1.084)	(-1.064)	(-1.065)

Note: t-ratio statistics between parentheses. DU_t denotes the dummy variable defined as $DU_t = 1$ for $t > \hat{T}_b$ and 0 otherwise. $\hat{r}$ denotes the estimated number of common factors, and IC_{p2} the information criteria defined in Bai and Ng (2002).

Table 4: Marginal product of private and public capitals and returns to education for the average of the Spanish regions

	No breaks						One structural break (Model 4r)					
	Private capital		Public capital		Human capital		Private capital		Public capital		Human capital	
Year	CupFM	CupBC	CupFM	CupBC	CupFM	CupBC	CupFM	CupBC	CupFM	CupBC	CupFM	CupBC
1964	0.187	0.170	0.576	0.521	0.029	0.045	0.244	0.234	0.427	0.338	0.000	0.013
1965	0.186	0.169	0.545	0.493	0.029	0.044	0.242	0.233	0.403	0.320	0.000	0.013
1966	0.184	0.168	0.515	0.465	0.029	0.044	0.240	0.231	0.381	0.302	0.000	0.013
1967	0.182	0.166	0.490	0.443	0.028	0.044	0.238	0.229	0.363	0.288	0.000	0.013
1968	0.184	0.167	0.491	0.444	0.028	0.043	0.239	0.230	0.363	0.288	0.000	0.013
1969	0.185	0.169	0.487	0.441	0.027	0.042	0.241	0.232	0.361	0.286	0.000	0.012
1970	0.185	0.169	0.481	0.434	0.026	0.041	0.241	0.232	0.356	0.282	0.000	0.012
1971	0.186	0.170	0.470	0.425	0.026	0.040	0.243	0.233	0.348	0.276	0.000	0.012
1972	0.190	0.173	0.474	0.429	0.025	0.039	0.248	0.238	0.351	0.278	0.000	0.012
1973	0.193	0.175	0.482	0.436	0.025	0.038	0.251	0.241	0.357	0.283	0.000	0.011
1974	0.188	0.172	0.473	0.428	0.024	0.037	0.246	0.236	0.350	0.278	0.000	0.011
1975	0.186	0.170	0.464	0.419	0.024	0.037	0.243	0.233	0.343	0.272	0.000	0.011
1976	0.182	0.166	0.453	0.409	0.023	0.036	0.237	0.228	0.335	0.266	0.000	0.011
1977	0.178	0.162	0.437	0.395	0.023	0.035	0.232	0.223	0.324	0.256	0.000	0.010
1978	0.172	0.157	0.427	0.386	0.022	0.034	0.225	0.216	0.316	0.251	0.000	0.010
1979	0.168	0.153	0.420	0.379	0.021	0.033	0.219	0.210	0.311	0.246	0.000	0.010
1980	0.163	0.148	0.408	0.369	0.021	0.032	0.212	0.204	0.302	0.240	0.000	0.009
1981	0.159	0.145	0.396	0.358	0.020	0.031	0.207	0.199	0.293	0.232	0.000	0.009
1982	0.156	0.142	0.377	0.341	0.019	0.030	0.204	0.196	0.279	0.221	0.000	0.009
1983	0.157	0.143	0.368	0.333	0.019	0.029	0.204	0.196	0.273	0.216	0.000	0.009
1984	0.159	0.145	0.363	0.328	0.018	0.028	0.208	0.200	0.269	0.213	0.000	0.008
1985	0.160	0.146	0.348	0.314	0.018	0.027	0.209	0.201	0.257	0.204	0.000	0.008
1986	0.157	0.143	0.325	0.293	0.017	0.026	0.205	0.197	0.240	0.190	0.000	0.008
1987	0.160	0.146	0.319	0.288	0.016	0.025	0.209	0.201	0.236	0.187	0.000	0.007
1988	0.163	0.149	0.313	0.283	0.016	0.024	0.213	0.205	0.232	0.183	0.000	0.007
1989	0.163	0.149	0.296	0.268	0.015	0.023	0.213	0.204	0.219	0.174	0.000	0.007
1990	0.162	0.148	0.278	0.252	0.015	0.023	0.212	0.204	0.206	0.163	0.000	0.007
1991	0.161	0.147	0.262	0.237	0.014	0.022	0.210	0.202	0.194	0.154	0.000	0.007
1992	0.159	0.145	0.248	0.224	0.014	0.022	0.207	0.199	0.184	0.146	0.000	0.006
1993	0.154	0.141	0.231	0.209	0.014	0.021	0.201	0.193	0.171	0.136	0.000	0.006
1994	0.155	0.141	0.223	0.202	0.013	0.021	0.202	0.194	0.165	0.131	0.000	0.006
1995	0.155	0.142	0.217	0.196	0.013	0.020	0.203	0.195	0.161	0.127	0.000	0.006
1996	0.155	0.141	0.213	0.193	0.013	0.019	0.202	0.194	0.158	0.125	0.000	0.006
1997	0.156	0.142	0.212	0.192	0.012	0.019	0.203	0.196	0.157	0.124	0.000	0.006
1998	0.157	0.143	0.210	0.190	0.012	0.019	0.205	0.197	0.156	0.123	0.000	0.006
1999	0.158	0.144	0.209	0.189	0.012	0.018	0.089	0.087	-0.008	0.004	0.024	0.023
2000	0.158	0.144	0.208	0.188	0.012	0.018	0.089	0.087	-0.008	0.004	0.023	0.023
2001	0.158	0.144	0.206	0.187	0.011	0.018	0.089	0.087	-0.008	0.004	0.023	0.022
2002	0.156	0.142	0.203	0.183	0.011	0.017	0.088	0.086	-0.008	0.004	0.023	0.022
2003	0.155	0.141	0.198	0.179	0.011	0.017	0.087	0.085	-0.008	0.004	0.022	0.021
2004	0.153	0.139	0.196	0.177	0.011	0.017	0.086	0.084	-0.008	0.004	0.022	0.021
2005	0.151	0.137	0.195	0.176	0.011	0.016	0.085	0.083	-0.008	0.004	0.021	0.021
2006	0.150	0.136	0.195	0.176	0.011	0.016	0.085	0.082	-0.007	0.004	0.021	0.020
2007	0.148	0.135	0.194	0.175	0.010	0.016	0.084	0.081	-0.007	0.004	0.021	0.020
2008	0.144	0.131	0.189	0.171	0.010	0.016	0.081	0.079	-0.007	0.004	0.021	0.020
2009	0.136	0.124	0.175	0.158	0.010	0.016	0.077	0.075	-0.007	0.003	0.021	0.020
2010	0.134	0.122	0.169	0.153	0.010	0.016	0.076	0.074	-0.007	0.003	0.020	0.020
2011	0.133	0.121	0.168	0.152	0.010	0.015	0.075	0.073	-0.006	0.003	0.020	0.019

Note: Model 4r denotes Model 4 restricted

Table 5: Marginal product of private and public capitals and returns to education for the Spanish regions (average across time)

	No breaks						One structural break (Model 4r)											
							Private capital				Public capital				Human capital			
	Private capital		Public capital		Human capital		CupFM		CupBC		CupFM		CupBC		CupFM		CupBC	
	CupFM	CupBC	CupFM	CupBC	CupFM	CupBC	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
AND	0.146	0.133	0.280	0.253	0.017	0.026	0.197	0.075	0.190	0.073	0.251	-0.007	0.199	0.004	0.000	0.023	0.009	0.022
ARA	0.168	0.153	0.215	0.194	0.015	0.024	0.224	0.091	0.215	0.088	0.184	-0.006	0.146	0.003	0.000	0.021	0.008	0.021
AST	0.159	0.145	0.269	0.243	0.016	0.024	0.214	0.082	0.206	0.080	0.282	-0.006	0.224	0.003	0.000	0.021	0.009	0.021
BAL	0.151	0.138	0.567	0.513	0.016	0.024	0.209	0.074	0.201	0.072	0.532	-0.014	0.422	0.007	0.000	0.022	0.009	0.021
CAN	0.143	0.130	0.309	0.280	0.016	0.025	0.204	0.065	0.196	0.063	0.271	-0.008	0.215	0.004	0.000	0.023	0.009	0.022
CANT	0.146	0.133	0.289	0.261	0.015	0.023	0.194	0.079	0.187	0.076	0.294	-0.006	0.233	0.003	0.000	0.021	0.008	0.021
CYL	0.161	0.147	0.194	0.175	0.016	0.024	0.215	0.086	0.207	0.083	0.172	-0.005	0.136	0.003	0.000	0.022	0.009	0.021
CLM	0.181	0.165	0.197	0.178	0.017	0.026	0.247	0.092	0.238	0.089	0.165	-0.006	0.131	0.003	0.000	0.023	0.010	0.022
CAT	0.177	0.161	0.498	0.450	0.015	0.023	0.242	0.089	0.232	0.087	0.446	-0.013	0.354	0.007	0.000	0.022	0.008	0.021
VAL	0.158	0.144	0.356	0.322	0.016	0.024	0.217	0.078	0.209	0.076	0.324	-0.009	0.256	0.005	0.000	0.021	0.009	0.021
EXT	0.133	0.121	0.162	0.146	0.017	0.026	0.168	0.082	0.161	0.080	0.138	-0.005	0.109	0.002	0.000	0.022	0.010	0.021
GAL	0.171	0.155	0.290	0.262	0.017	0.026	0.230	0.088	0.221	0.086	0.276	-0.007	0.219	0.003	0.000	0.022	0.010	0.021
MAD	0.230	0.210	0.424	0.383	0.013	0.021	0.322	0.110	0.310	0.107	0.368	-0.012	0.291	0.006	0.000	0.020	0.007	0.019
MUR	0.160	0.146	0.340	0.307	0.016	0.025	0.221	0.079	0.213	0.076	0.307	-0.009	0.243	0.004	0.000	0.023	0.009	0.022
NAV	0.174	0.159	0.284	0.257	0.014	0.022	0.238	0.088	0.229	0.085	0.260	-0.007	0.206	0.004	0.000	0.021	0.008	0.020
PV	0.195	0.178	0.385	0.348	0.014	0.022	0.257	0.107	0.247	0.104	0.351	-0.010	0.278	0.005	0.000	0.020	0.008	0.020
RIO	0.157	0.143	0.298	0.269	0.016	0.024	0.214	0.080	0.205	0.078	0.264	-0.008	0.209	0.004	0.000	0.022	0.009	0.021

Note: Model 4r denotes Model 4 restricted, 1st indicates the subperiod 1964-1998 and 2nd the subperiod 1999-2011