The long-run relationship between R&D and regional knowledge: The case of France, Germany, Italy and Spain*

Marcos Sanso-Navarro[a,§], María Vera-Cabello[b]

[a]Departamento de Análisis Económico, Universidad de Zaragoza, Spain

[b]Centro Universitario de la Defensa de Zaragoza, Spain

June 2017

Abstract

This paper incorporates time-dependence into a regional knowledge production function framework. Within this setup, the long-run dynamic behaviour of R&D and knowledge has been analysed in four European countries - France, Germany, Italy and Spain - using unit root tests and cointegration techniques. We find that the regional stock of knowledge is cointegrated with R&D employment and external knowledge. Nonetheless, knowledge spillovers play a more important role in the generation and accumulation of new ideas. This suggests that innovation policies should aim at enhancing knowledge diffusion and regional absorptive capacity.

JEL classification: C33, O31, O47, R11

Key words: Knowledge production, heterogeneous panels, unit roots, cointegration, cross-sectional dependence

*The authors have benefited from the valuable comments of two associate editors, two anonymous referees and participants at the XLI Reunión de Estudios Regionales (Reus) and the XIX Encuentro de Economía Aplicada (Sevilla). This work was supported by Centro Universitario de la Defensa under Grant UZCUD2015-SOC-04; Gobierno de Aragón under Grant S16-ADETRE Research Group; and Universidad de Zaragoza under Grant JIUZ-2014-SOC-12.

§Corresponding author. Address: Departamento de Análisis Económico. Facultad de Economía y Empresa. Gran Vía 2. 50005 Zaragoza (Spain). Tel: (+34) 876 554 629; Fax: (+34) 976 761 996; e-mail: marcossn@unizar.es.

1 Introduction

Explicitly considering innovation as the driving force of economic growth, the Lisbon Agenda established the objective of making the European Union (EU) the most competitive and dynamic knowledge-based economy in the world. The importance of research and development (R&D), innovation and knowledge flows has subsequently been emphasised by the 'Europe 2020' strategy for smart, sustainable and inclusive growth. It has also been acknowledged that key incentives to foster innovative activities should be tackled at the regional level because that is where the transmission of knowledge takes place. This explains why the EU Cohesion Policy has increased the percentage of its total funds that is set aside for R&D and innovation from nearly 25% in the previous programming period (2007-2013) to 30% in the present one (2014-2020).

The estimation of knowledge production functions (Griliches, 1979) has been extensively used for the empirical analysis of innovation. At the regional level, this framework permits the study of how local inputs contribute to the generation of knowledge (Jaffe, 1989). The production of new ideas in a region not only depends on the amount of resources that it devotes to R&D but also on the stock of knowledge available in other regions, especially nearby ones. Therefore, knowledge production functions (KPFs, hereafter) are also used to measure the intensity and spatial extent of these knowledge spillovers that are of interest in the geography of innovation literature (Audrestch & Feldman, 2004).

Given the relevance of the determinants of regional innovation for policy-making, several studies have analysed this issue in European regions within a KPF setting[1]. Bottazi and Peri (2003) explored the influence of R&D and its externalities on the ability of regions to generate new ideas, finding that knowledge spillovers exist only within a distance of 300 kilometers. Using spatial econometric techniques, Moreno, Paci, and Usai (2005) dealt with the geographical distribution of innovation and the influence of spillovers in knowledge creation and diffusion. As well as the importance of regional factors, these authors provided evidence of spatial spillovers in R&D and patenting - constrained by national borders and within 250 kilometers - and of the role played by technological similarity in the diffusion of new ideas. In this line, Paci, Marrocu, and Usai (2014) analysed the effects of different dimensions of proximity on regional innovative capacity, concluding that technological closeness exerts the most important influence.

[1]See Guastella and van Oort (2015) for a comprehensive and up-to-date literature review on this topic.

Charlot, Crescenzi, and Musolesi (2015) adopted a semi-parametric setting - relaxing the assumptions of linearity, additivity and homogeneity - that included region-specific time trends. These authors found that knowledge generation in EU regions is characterized by the presence of nonlinearities, thresholds, complex interactions and shadow effects[2]. Guastella and van Oort (2015) considered the presence of spatial heterogeneity, concluding that the significance of knowledge spillovers decreases when geographical characteristics are accounted for. Miguélez and Moreno (2015) showed that regions with large absorptive capacity make the most of knowledge inflows from other regions.

This paper contributes to the literature on the determinants of knowledge generation in European regions by introducing time-dependence into the flow of new ideas. More specifically, we adopt a longitudinal perspective by considering that knowledge generation depends on the existing stock of ideas. Our main aim is to analyse the long-run dynamic behaviour of regional knowledge (output), R&D employment and external knowledge (inputs). This will be done using unit root tests and cointegration estimation techniques that exploit both the cross-sectional and temporal dimensions of the data. These methods will take into account both the presence of cross-sectional dependence - due to interactions between economic agents and institutions located in different regions - and spatial heterogeneity in the innovation process and its determinants. Constrained by the longitudinal perspective adopted and data availability, the present study will refer to the regions of four EU countries: France, Germany, Italy and Spain.

As well as uncovering the factors behind the creation of new ideas and how they flow between regions, KPFs serve as a testing ground to discriminate between alternative endogenous growth theories. Within this setup, the dynamics of the knowledge-generating sector and the long-run relationship between R&D and knowledge permits the testing of different 'ideas-based' growth models[3] through the study of the scale of the effect that the R&D sector exerts on knowledge creation. This distinction may serve as a basis for the design of regional innovation policies regarding R&D expenditures. Moreover, it is an empirical check of the assumptions made on the process of technological change in

[2] The complexity of the European innovation process at the regional level motivates the extension of the KPF empirical framework. In this study, and following Charlot et al. (2015), time-varying unobserved factors and heterogeneous effects of innovation inputs have been taken into account within a parametric setting.

[3] These models assign a prominent role to the non-rival nature of knowledge and its diffusion. The public good attributes of the knowledge produced by R&D expenditures - that generate intertemporal positive knowledge externalities - create a negative gap between private and social returns to R&D. Montmartin and Massard (2015) review the implications for R&D policies of the difference between the decentralised level of investment in R&D and the socially optimal level.

the models currently used to assess the impact of EU regional policies (Brandsma & Kancs, 2015; Brandsma, Kancs, Monfort, & Rillaers, 2015).

2 An 'ideas-based' growth framework for the R&D sector

The so-called 'ideas-based' endogenous growth models assign a key role to the mechanism through which the resources devoted to R&D turn into new knowledge. Nevertheless, there are alternative approaches within these theoretical models that establish different predictions about the scale of the effect that the R&D sector will have on the generation of technological knowledge and, hence, on long-run productivity growth. On the one hand, the first generation of R&D-based growth models (Romer, 1990; Grossman & Helpman, 1991; Aghion & Howitt, 1992) exhibits strong scale effects as a higher level of R&D employment is positively related to the growth rate of the knowledge stock. On the other hand, semi-endogenous growth models (Jones, 1995; Kortum, 1997; Segerstrom, 1998) avoid this strong scale effect by introducing diminishing returns in the R&D sector. As a consequence, an increase in the level of R&D resources will only have short-run effects, not affecting the steady-state growth rate.

The empirical setting presented in this section is based on the simple theoretical model proposed by Jones (2005), adopted later by Bottazi and Peri (2007). This framework begins by assuming that, in a given region, the final good is obtained from a neoclassical technology. That is, the production function in terms of output per worker (y_t) is specified as:

$$y_t = BA_t^{\sigma} k_t^{\theta}; \quad \sigma > 0; \quad 0 < \theta < 1, \tag{1}$$

where total factor productivity (TFP) depends on a term B that captures efficiency in production. This term is considered to be time-invariant as it is determined by slowly evolving regional factors such as the quality of institutions or agglomeration economies. σ denotes the elasticity of TFP to the available stock of scientific and technological knowledge A_t. It is also assumed that physical capital per worker, k_t, experiences decreasing marginal returns and satisfies the Inada conditions.

Taking natural logarithms on both sides of (1) and taking derivatives with respect to time, it is obtained that the growth rate of output per worker (g_{y_t}) is a linear combination - determined by their corresponding elasticities - of the growth rates of technological knowledge (g_{A_t}) and physical capital per worker (g_{k_t}):

$$g_{y_t} = \sigma g_{A_t} + \theta g_{k_t}. \tag{2}$$

Decreasing marginal returns of physical capital lead a region to converge to a balanced growth path (BGP, characterised by $g_y = g_k$) where the dynamics of technological knowledge determine those of labour productivity:

$$g_y = \frac{\sigma}{1-\theta} g_A. \tag{3}$$

This simple model introduces time-dependence into the flow of new ideas by considering that knowledge is generated by the workers in the R&D sector using their creativity and the available stock of knowledge as inputs. In the aggregate, the creation of new ideas can be reduced to a random noise. Neglecting this term, non-obsolete available knowledge, R&D resources and new ideas have a stable relationship that is reflected in the following KPF:

$$I_t = F(R\&D_t, A_t, \tilde{A}_t), \tag{4}$$

where I_t denotes the ideas generated in a given region in period t. Expression (4) implies that R&D has a strong contemporaneous effect on innovation. The influence of past resources devoted to R&D on innovation is exerted through the stock of useful knowledge. A_t reflects the knowledge generated up to $t-1$ that is available at the beginning of period t. Due to knowledge spillovers, a region also benefits from the ideas created in other regions. In this regard, $\tilde{A}_t$ reflects the external stock of ideas.

In order to make this theoretical framework operative, the number of new ideas generated in a region is considered to be proportional to the number of patents for which an application is filed by its residents ($Pat_t = \varkappa I_t$). This implies that, as is common practice in the related literature, patent statistics will allow us to proxy for the generation of new ideas. Furthermore, it will be assumed that the relationship in (4) is log-linear:

$$\ln(Pat_t) = \ln(\varkappa) + \lambda \ln(R\&D_t) + \phi \ln(A_t) + \xi \ln(\tilde{A}_t). \tag{5}$$

As well as considering that the available stock of knowledge increases with the development of new ideas, it will also be assumed that it is continually decreasing at a constant obsolescence rate δ:

$$A_{t+1} = Pat_t + (1-\delta)A_t. \tag{6}$$

Dividing both sides of this expression by A_t, taking natural logarithms and substituting into (5), the following long-run relationship - motivating the empirical analysis carried out in the present paper - is obtained:

$$\ln(g_{A_t} + \delta) - \ln(\varkappa) = (\phi - 1)\ln(A_t) + \lambda \ln(R\&D_t) + \xi \ln(\tilde{A}_t). \tag{7}$$

The left-hand side will become a region-specific stationary process if the stock of knowledge converges to a stochastic BGP. In this case, there will be a stationary long-run relationship between the resources devoted to R&D and the regional and external stocks of knowledge. If these variables are non-stationary, convergence to a BGP implies that there is a cointegration relationship between them. Standardising by the regional knowledge stock, the cointegration vector is $(-1, \mu, \gamma)$ and can be estimated from:

$$\ln(A_t) = \mu \ln(R\&D_t) + \gamma \ln(\tilde{A}_t) + \varepsilon_t, \tag{8}$$

with $\mu = \frac{\lambda}{1-\phi}$, $\gamma = \frac{\xi}{1-\phi}$ and where ε_t is the disturbance term.

If, on the contrary, the left-hand side of (7) is non-stationary, the stock of knowledge will not converge to a stochastic BGP. The growth rate of technological knowledge will increase with the levels of R&D resources and of the regional and external knowledge stocks. In this case, there will not be a cointegration relationship between these variables and the knowledge stock will diverge across regions.

This framework allows us to discriminate between alternative 'ideas-based' growth models through the study (i) of the stationarity of the regional knowledge growth rate, and (ii) of the presence of a cointegration relationship between the levels of R&D resources and of the regional and external stocks of knowledge. Unit root non-stationarity of the growth rate of the stock of knowledge in a region implies that it is determined by the amount of resources devoted to the research sector and the external and regional knowledge stocks. Therefore, the growth rates of regional knowledge and productivity increase with the level of R&D inputs. This can be interpreted as evidence that R&D has a strong scale effect and, hence, in favour of the first generation of 'ideas-based' growth models which hold that a constant level of R&D resources is enough to sustain long-run growth. Alternatively, the stationarity of the growth rate of the regional knowledge stock implies that, in the long run, the levels of R&D inputs and of external knowledge will determine the level of knowledge in a region and, as a consequence, that of its productivity. Therefore, a reallocation of resources to R&D will only affect the level of income. This can be interpreted as evidence in favour of semi-endogenous growth models which, due to diminishing returns in the research sector, are characterised by weak scale effects. In these models, long-run growth depends on the growth rate of R&D resources and the elasticities of the KPF. This implies that higher growth might be achieved with higher investments in education as long as they lead workers in the R&D sector to improve their performance and absorptive capacity to exploit knowledge spillovers.

3 Data sources, variable construction and sample description

The amount of resources that a region devotes to R&D activities has been measured by the total employment in all sectors (private and public, full-time equivalents), which has been extracted from EUROSTAT. The availability of this information at the regional level is limited until the mid-1990s in most European countries. As can be inferred from the empirical setting presented in the previous section, we are interested in the stochastic properties of the relevant variables in the 'ideas-based' economic growth models. For this reason, and given that the panel methods implemented in our analysis require both a long time dimension and a minimum number of units to obtain reliable results, our sample is made up of regions in France, Italy and Spain at the NUTS2 level and of NUTS1 regions in Germany[4]. The regions included in our sample are listed in Table 1. It is worth mentioning that, in order to work with a balanced panel for each country, some missing observations have been interpolated using the automatic procedure in the Windows version of the TRAMO-SEATS software.

[Insert Table 1 here]

[4] There is no information regarding employment in the R&D sector at the NUTS2 level for Germany. The reason is that there is no correspondence between NUTS2 regions (Regierungsbezirke) and the relevant administrative units in this country (Länder; NUTS1). It is worth noting that this does not affect the homogeneity and coherence of our empirical analysis. The regions that have been considered are the areas targeted by both national and supra-national institutions implementing innovation policies. Nevertheless, the results obtained in the present study, as well as in related ones, may be driven by the geographical scale at which the analysis is carried out.

The generation of ideas in a region has been proxied by the number of patent applications made by its residents to the European Patent Office (EPO). This information has been extracted from the OECD REGPAT database (January 2013 edition). Patent applications have been classified according to their priority year and calculated using fractional counting. In addition, each patent has been weighted by the factor $(1+\varphi_6)$, where φ_6 denotes the broadest quality index elaborated by Squicciani, Dernis, and Criscuolo (2013). This measure is calculated taking into account six dimensions - forward citations, family size, number of claims, generality index, backward citations and grant lag - and is intended to reflect the technological and economic value of patents as well as their impact on subsequent technological developments.

Following expression (6), the regional stock of scientific and technological knowledge has been constructed from the number of quality-weighted patent applications. To do so, and in order to establish an initial level of available knowledge, we assume that its accumulation is compatible with a BGP. Based on a perpetual inventory method, the initial level of knowledge for a given region i has been calculated as:

$$A_{i0} = \sum_{t=0}^{\infty} \frac{Pat_{i0}}{(1+\bar{g}_i)^{t+1}}(1-\delta)^t = \frac{Pat_{i0}}{(\bar{g}_i+\delta)}, \tag{9}$$

where, as in Bottazi and Peri (2007), δ has been set to 0.10 and $\bar{g}_i$ is the average annual growth rate of patent applications made by region i's residents during the first five years for which data are available. The stock of ideas in the other regions of the same country, which tries to capture the influence of knowledge spillovers, has been constructed as $\tilde{A}_{it} = \sum\limits_{i \neq j} A_{it}$.

[Insert Figure 1 here]

Figure 1 presents, for each country and region, box-plots of three variables (in natural logarithms[5]) during the period 1995-2010: R&D employment, patent applications and the stock of knowledge. Broadly speaking, German regions display a higher level of R&D employment than the regions in France and, especially, Italy and Spain. This is reflected in a higher number of patent applications to the EPO and, as a consequence, in a higher level of knowledge stocks. This shows that there is a high correlation between these variables at the regional level in the four countries that conform our sample. It can also be stated that Italy and Spain have a higher number of regions that are far

[5]Zero patents add a small constant before the logarithmic transformation.

from the technological frontier, defined as the stock of knowledge in the region with the highest level[6]. Moreover, these two countries also display high levels of both between- and within-regional variability. These results suggest that, although there exists a relationship between R&D employment and innovation, it is of a heterogeneous nature. This fact - that is already being taken into consideration by European regional innovation policies - will be controlled for by the methods we have implemented to analyse the long-run relationship between R&D efforts and knowledge.

[Insert Figure 2 here]

More in line with the main aim of our study, Figure 2 plots the temporal evolution of the three main variables of interest. It can also be observed that the variability between French regions is mainly determined by two innovation leaders (Île de France and Rhône-Alpes) - characterised by high levels of R&D employment and innovation results - and a lagging region (Corsica). Italy and Spain present more apparent differences across regions, with less innovative ones experiencing a higher volatility. The number of patent applications to the EPO has been affected by the crisis to a greater extent than the level of R&D employment. Finally, it is worth noting that the variables displayed in Figure 2 followed an upward trend during the period 1995-2010, especially R&D employment and the stock of knowledge, two important variables in our empirical framework.

4 Empirical analysis

In a review of the studies dealing with the geographical aspects of innovation, Autant-Bernard (2012) pointed out that, together with the temporal dimension, the possible presence of heterogeneity and spatial correlation should be taken into account when working with knowledge and innovation at the regional level. For this reason, we begin our empirical analysis by testing for the presence of weak cross-sectional dependence using the procedure developed by Pesaran (2015). Both the levels of (the natural logarithm of) R&D employment and the regional and external knowledge stocks as well as their first differences (growth rates) have been considered. The values obtained for the CD test statistic, which are reported in Table 2, show that the null hypothesis of the absence of cross-sectional dependence across regions can be rejected in all cases.

[6] In our sample, the technological frontier is represented by the stock of knowledge in Baden-Württemberg. This region hosts a cluster of technologically advanced firms and is among the most innovative regions in the EU.

[Insert Table 2 here]

Having provided evidence that the variables in our empirical framework display cross-sectional dependence, both in levels and first-differenced, the next step is to obtain a measure of its degree using the characterisation proposed by Bailey, Kapetanios, and Pesaran (2016). Point estimates for the $\mathring{\alpha}$ exponent of dependence and their 90% level confidence bands are also reported in Table 2. These figures show that the estimated exponents for the levels and growth rates of the regional and external knowledge stocks and R&D employment are indistinguishable from unity. This reflects that these variables have a strong dependence across the regions in the countries that form our sample which might be controlled for using a factor structure. The only exception is the growth rate of R&D employment in Germany, Italy and Spain where the estimated exponent of cross-sectional dependence tends to be smaller. In any case, the lower bound of their confidence interval is far from 0.5, the value that corresponds to the case of weak dependence.

It is difficult to obtain reliable inferences about the order of integration of a variable from short time series with a yearly frequency. This problem can be mitigated through the application of panel unit root tests that exploit both the cross-sectional and temporal dimensions of the data. However, and although panel unit root tests are a poweful alternative to univariate methods, they may be biased (size-distorted) in the presence of cross-sectional dependence. Baltagi, Bresson, and Pirotte (2007) show that, of the tests considered in their study, that proposed by Pesaran (2007) is the most robust to cross-sectional dependence of a spatial nature, which is commonly found when working with regional information. This data feature is controlled for by this method assuming the presence of a single common factor that, following the spirit of the Common Correlated Effects estimator (Pesaran, 2006; CCE), is proxied by the cross-sectional mean of the individual time series. An explanation for the good performance of the CCE estimator is provided by Pesaran and Tosetti (2011) who show that it eliminates the effects of all forms of correlations, irrespective of whether they are due to spatial and/or unobserved common factors. In this line, Breinlich, Ottaviano, and Temple (2014), consider the common factor structure to be a reasonable alternative to the spatial econometric approach where cross-sectional correlation is determined by location and the distance between units.

The unit root test developed by Pesaran (2007) for heterogeneous panels[7] is implemented by obtaining individual test statistics for each region in the panel and, then, calculating their country average ($CIPS$). As noted before, individual unit root test statistics are obtained from standard augmented Dickey-Fuller auxiliary regressions that include cross-sectional averages of lagged levels and first differences of the individual series ($CADF_i$). Resulting test statistics for each variable at country level are shown in Table 3. Although their magnitude depends on the number of augmentation lags introduced to mitigate size distortions (due to serial correlation in the error term), this does not prevent us from drawing some general conclusions. First, the unit root null hypothesis is rejected with more difficulty for the variables in levels. This implies that shocks to R&D employment and the stocks of technological knowledge have a persistent effect, so they can be considered to be non-stationary processes. Second, unit root non-stationarity is easily rejected for the growth rates of these variables. Table 3 also reports (in brackets) the p-values that result from the application of the weak cross-sectional dependence CD test statistic to the residuals of the augmented Dickey-Fuller auxiliary regressions from which the individual unit root test statistics are calculated. These figures show that, in general terms, the introduction of cross-sectional averages of the individual time series eliminates the correlation among the members of the panel. The exceptions are both the levels and growth rates of the external stock of knowledge in Germany and Spain and of R&D employment in the latter. This suggests that the assumption of a one-factor residual model may be very restrictive in these cases.

[Insert Table 3 here]

As explained in the theoretical section, we can empirically discriminate between alternative 'ideas-based' growth models based on the study of the stationarity of the knowledge stock growth rate and of the presence of a cointegration relationship between the levels of R&D employment and of the regional and external technological knowledge stocks. The $CIPS$ test statistic tends to reject the unit root null hypothesis for the growth rate of the regional stock of ideas. The main exceptions are found when three augmentation lags are included to deal with (possibly) serially correlated errors, which may be related to the adverse influence of these lags on the power of unit root tests. These results suggest that technological knowledge converges to a BGP and, in

[7]Omitted variables bias and unobserved heterogeneity are important issues in the estimation of KPFs. For this reason, the unit root tests and cointegration techniques for heterogeneous panels applied introduce regional fixed-effects to capture time-invariant unobserved regional features.

accordance with (7), that there exists a cointegration relationship between the regional levels of R&D employment and of the stocks of knowledge in these four countries. This long-run comovement implies that the regional stock of ideas is determined by the employment in the R&D sector and external technological knowledge. Moreover, it can be interpreted as evidence favourable to the predictions of semi-endogenous 'ideas-based' growth models that display weak scale effects from R&D resources to regional knowledge.

We are going to further investigate the presence of a long-run relationship among the relevant variables in our empirical setting by applying the residual-based test ($CADFC_p$) proposed by Banerjee and Carrion-i-Silvestre (2017) to expression (8). This procedure also controls for the dependence across the units that conform the panel using an unobserved common factor structure proxied by cross-sectional averages. In what follows, we are taking into account that knowledge diffusion is more effective among closer regions and that its intensity decreases with geographical distance (Dettori, Marrocu, & Paci, 2012). This will be done using two additional specifications for the external stock of knowledge. First, we will consider only the knowledge stock of the five nearest regions ($\tilde{\mathrm{A}}_{5\mathrm{nn}}$). Second, we will apply an inverse-distance weight to all the external stocks of knowledge ($\tilde{\mathrm{A}}_{\mathrm{dist}}$).

[Insert Table 4 here]

The resulting panel cointegration test statistics are displayed in Table 4. The null hypothesis of no cointegration cannot be rejected for French and German regions when the deterministic component consists only of a (region-specific) constant term. Under this specification, there is evidence of a cointegration relationship between R&D employment and knowledge in Italian regions when the stock of external knowledge is defined using inverse-distance weights. The $CADFC_p$ test corroborates the conclusions drawn from the panel unit root test in Spanish regions regardless of the number of augmentation lags included and the proxy for the influence of knowledge spillovers considered. In line with Charlot et al. (2015), the introduction of individual time trends increases the evidence favourable to the presence of a long-run relationship between R&D and knowledge in France and, especially, in Germany and Italy.

Expression (8) permits us to obtain long-run elasticities. We have implemented the mean group estimator of Pesaran and Smith (1995). The specification for the deterministic component in each country has been chosen according to the evidence provided by the panel cointegration test. Estimation results are reported in Table 5. It can be concluded that both R&D employment and external knowledge have a positive relationship with the stock of regional knowledge. Knowledge spillovers have a more significant and a higher estimated parameter than the employment in the R&D sector. More specifically, and for French regions, a 1% increase in regional R&D employment increases technological knowledge by 0.17% in two out of three specifications for the external stock of knowledge. A 1% increase of the external stock of knowledge is associated with a long-run increase in the regional knowledge stock of around 0.72%. The magnitude of this long-run elasticity, that reflects the importance of knowledge spillovers, does not critically depend on the way the external knowledge stock is calculated.

[Insert Table 5 here]

Long-run elasticities of the regional knowledge stock to R&D employment are positive but not statistically significant in Germany. However, the estimated parameters for the external stock of knowledge are statistically significant at the 1% significance level in all specifications. Similar results are obtained for the regions in the two Mediterranean countries. Although southern regions tend to display lower levels of R&D employment and, given the double role of R&D (Miguélez & Moreno, 2015), they are expected to have a lower absorptive capacity so their elasticities to external knowledge are higher than the regions in France and Germany. This may be reflecting that, when regions are further from the technological frontier, having greater opportunities to catch up, knowledge generation mainly relies on technological acquisition and imitation.

5 Conclusions

This paper studies the long-run relationship between R&D and knowledge - two key ingredients of the 'Europe 2020' strategy - at the regional level. After introducing time-dependence into the flow of new ideas within a knowledge production function framework, the empirical analysis is based on the application of panel unit root tests and cointegration techniques. Our main aim is to determine whether or not the stock of knowledge in a region is cointegrated with its level of R&D employment and the stock of knowledge in other regions. Trying to capture the complexity of the European regional innovation systems and the presence of cross-sectional dependence, the methods that have been implemented control for time-varying unobserved factors and heterogeneous effects.

The longitudinal perspective adopted and data availability have limited our sample to four countries: France, Germany, Italy and Spain. Therefore, our study might not be representative of all Europe. The number of patent applications has been used as a measure of innovation. Although patents have been weighted according to a quality index, they only represent a fraction of new knowledge. The reason is that patent applications do not reflect protected, process or organizational innovative activity. In addition, we have only considered knowledge spillovers through geographical proximity, neglecting other types of nonspatial proximity. These shortcomings should be kept in mind when interpreting our results.

The levels (growth rates) of R&D employment and of the regional and external knowledge stocks are non-stationary (stationary). Moreover, there exists a long-run cointegration relationship between these three variables. Within the simple theoretical model proposed by Jones (2005) and Bottazi and Peri (2007), these results may be interpreted as evidence favourable to semi-endogenous growth models. This implies that any policy supporting the level of resources devoted to R&D will only have transitory effects on the knowledge stock and, hence, productivity growth. These findings also provide empirical support to the assumptions made regarding technological change in the models used to assess the impact of regional policies in the EU (Brandsma & Kancs, 2015; Brandsma et al., 2015).

Estimated long-run elasticities from the cointegration relationship suggest that both the levels of R&D employment and external knowledge are positively related to the regional stock of knowledge. Nonetheless, knowledge spillovers generated in other regions exert a higher and more significant influence on the processes of generation and accumulation of new ideas. Southern regions, which tend to be far from the technological frontier, derive more benefits from knowledge spillovers. It is worth noting that these findings are robust to alternative ways of calculating the external stock of knowledge.

Our results show that knowledge spillovers are as important for regional innovative performance as the amount of resources devoted to R&D. Although the latter will determine the ability of a region to exploit the knowledge generated by other regions, we provide evidence supporting the shift in European innovation strategy from an almost exclusive focus on R&D to a broader set of dimensions. Our findings suggest that innovation policies should aim at enhancing knowledge diffusion and regional absorptive capacity. On the one hand, and with the purpose of fostering the transmission of knowledge, the mobility of skilled workers and the establishment of research networks should be encouraged. On the other hand, improvements in the educational level of the labour force will favour the capacity of regions to absorb external knowledge. Educational policy seems to be especially relevant in Southern countries, where the share of low-skilled workers is higher.

References

[1] Aghion, P., & Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, 60, 323-351.

[2] Audretsch, D. B., & Feldman, M. P. (2004). Knowledge spillovers and the geography of innovation. In V. Henderson & J. Thisse (Eds.), *Handbook of Urban and Regional Economics* (pp. 2713-2739). Princeton, NJ: Princeton University Press.

[3] Autant-Bernard, C. (2012). Spatial econometrics of innovation: Recent contributions and research perspectives. *Spatial Economic Analysis*, 7, 403-419.

[4] Bailey, N., Kapetanios, G., & Pesaran, M. H. (2016). Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics*, 31, 929-960.

[5] Baltagi, B. H., Bresson, G., & Pirotte, A. (2007). Panel unit root tests and spatial dependence. *Journal of Applied Econometrics*, 22, 339-360.

[6] Banerjee, A., & Carrion-i-Silvestre, J. L. (2017). Testing for panel cointegration using common correlated effects estimators. *Journal of Time Series Analysis*, 38, 610-636.

[7] Bottazzi, L., & Peri, G. (2003). Innovation and spillovers in regions: Evidence from European patent data. *European Economic Review*, 47, 687-710.

[8] Bottazzi, L., & Peri, G. (2007). The international dynamics of R&D and innovation in the long run and in the short run. *Economic Journal*, 117, 486-511.

[9] Brandsma, A., & Kancs, D. (2015). RHOMOLO: A dynamic general equilibrium modelling approach to the evaluation of the European Union's R&D policies. *Regional Studies*, 49, 1340-1359.

[10] Brandsma, A., Kancs, D., Monfort, P., & Rillaers, A. (2015). RHOMOLO: A dynamic spatial general equilibrium model for assessing the impact of cohesion policy. *Papers in Regional Science*, 94(S1), 197-221.

[11] Breinlich, H., Ottaviano, G. I. P., & Temple, J. R. W. (2014). Regional growth and regional decline. In P. Aghion & S. Durlauf (Eds.), *Handbook of Economic Growth* (pp. 683-779). The Netherlands: North-Holland.

[12] Charlot, S., Crescenzi, R., & Musolesi, A. (2015). Econometric modelling of the regional knowledge production function in Europe. *Journal of Economic Geography*, 15, 1227-1259.

[13] Dettori, B., Marrocu, E., & Paci, R. (2012). Total factor productivity, intangible assets and spatial dependence in the European regions. *Regional Studies*, 46, 1401-1416.

[14] Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *Bell Journal of Economics*, 10, 92-116.

[15] Grossman, G., & Helpman, E. (1991). *Innovation and growth in the global economy.* Cambridge, MA: MIT Press.

[16] Guastella, G., & van Oort, F. G. (2015). Regional heterogeneity and interregional research spillovers in European innovation: Modelling and policy implications. *Regional Studies*, 49, 1772-1787.

[17] Jaffe, A. B. (1989). Real effects of academic research. *American Economic Review*, 79, 957–70.

[18] Jones, C. I. (1995). Time series tests of endogenous growth models. *Quarterly Journal of Economics*, 110, 495-525.

[19] Jones, C. I. (2005). Growth and ideas. In P. Aghion & S. Durlauf (Eds.), *Handbook of Economic Growth* (pp. 1063-1111). The Netherlands: North-Holland.

[20] Kortum, S. (1997). Research, patenting, and technological change. *Econometrica*, 65, 1389-1419.

[21] Miguélez, E., & Moreno, R. (2015). Knowledge flows and the absorptive capacity of regions. *Research Policy*, 44, 833-848.

[22] Montmartin, B., & Massard, N. (2015). Is financial support for private R&D always justified? A discussion based on the literature on growth. *Journal of Economic Surveys*, 29, 479-505.

[23] Moreno, R., Paci, R., & Usai, S. (2005). Spatial spillovers and innovation activity in European regions. *Environment and Planning A*, 37, 1793-1812.

[24] Paci, R., Marrocu, E., & Usai, S. (2014). The complementary effects of proximity dimensions on knowledge spillovers. *Spatial Economic Analysis*, 9, 9-30.

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

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

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

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

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

[30] Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5), 71-102.

[31] Segerstrom, P. (1998). Endogenous growth without scale effects. *American Economic Review*, 80, 1077-1091.

[32] Squicciarini, M., Dernis, H., & Criscuolo, C. (2013). Measuring patent quality: Indicators of technological and economic value (Working Paper No. 2013/03). Paris: OECD Science, Technology and Industry.

Table 1: Sample description. NUTS classification, version 2013.

France (1991-2012)		Germany (1990-2012)		Italy (1994-2012)		Spain (1988-2012)	
FR10	Île de France	DE1	Baden-Württemberg	ITC1	Piemonte	ES11	Galicia
FR21	Champagne-Ardenne	DE2	Bayern	ITC2	Valle d'Aosta/Vallée d'Aoste	ES12	Principado de Asturias
FR22	Picardie	DE3	Berlin	ITC3	Liguria	ES13	Cantabria
FR23	Haute-Normandie	DE5	Bremen	ITC4	Lombardia	ES21	País Vasco
FR24	Centre	DE6	Hamburg	ITF1	Abruzzo	ES22	Comunidad Foral de Navarra
FR25	Basse-Normandie	DE7	Hessen	ITF3	Campania	ES23	La Rioja
FR26	Bourgogne	DE9	Niedersachsen	ITF4	Puglia	ES24	Aragón
FR30	Nord - Pas-de-Calais	DEA	Nordrhein-Westfalen	ITF5	Basilicata	ES30	Comunidad de Madrid
FR41	Lorraine	DEB	Rheinland-Pfalz	ITF6	Calabria	ES41	Castilla y León
FR42	Alsace	DEC	Saarland	ITG1	Sicilia	ES42	Castilla-La Mancha
FR43	Franche-Comté	DEF	Sachsen-Anhalt	ITG2	Sardegna	ES43	Extremadura
FR51	Pays de la Loire			ITH3	Veneto	ES51	Cataluña
FR52	Bretagne			ITH4	Friuli-Venezia Giulia	ES52	Comunidad Valenciana
FR53	Poitou-Charentes			ITH5	Emilia-Romagna	ES53	Illes Balears
FR61	Aquitaine			ITI1	Toscana	ES61	Andalucía
FR62	Midi-Pyrénées			ITI2	Umbria	ES62	Región de Murcia
FR63	Limousin			ITI3	Marche	ES70	Canarias
FR71	Rhône-Alpes			ITI4	Lazio		
FR72	Auvergne						
FR81	Languedoc-Roussillon						
FR82	Provence-Alpes-Côte d'Azur						
FR83	Corsica						

Table 2: Cross-sectional dependence testing for R&D employment and knowledge stocks.

	France					
	ln(R&D)	ln(A)	ln(Ã)	Δln(R&D)	Δln(A)	Δln(Ã)
CD test	64.69	54.61	67.88	36.73	16.82	65.47
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\mathring{\alpha}$ estimation	1.01	1.00	1.01	1.00	1.00	1.01
	[0.93, 1.08]	[0.93, 1.07]	[0.94, 1.07]	[0.77, 1.23]	[0.93, 1.07]	[0.93, 1.09]
	Germany					
	ln(R&D)	ln(A)	ln(Ã)	Δln(R&D)	Δln(A)	Δln(Ã)
CD test	14.75	32.97	33.86	9.34	17.90	32.98
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\mathring{\alpha}$ estimation	0.96	1.01	1.01	0.74	0.96	1.01
	[0.83, 1.09]	[0.92, 1.10]	[0.92, 1.10]	[0.64, 0.85]	[0.85, 1.06]	[0.90, 1.11]
	Italy					
	ln(R&D)	ln(A)	ln(Ã)	Δln(R&D)	Δln(A)	Δln(Ã)
CD test	43.17	51.89	53.83	8.28	33.04	52.18
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\mathring{\alpha}$ estimation	1.00	1.00	1.01	0.77	1.00	1.01
	[0.92, 1.08]	[0.90, 1.11]	[0.92, 1.09]	[0.67, 0.88]	[0.90, 1.10]	[0.92, 1.10]
	Spain					
	ln(R&D)	ln(A)	ln(Ã)	Δln(R&D)	Δln(A)	Δln(Ã)
CD test	52.84	41.34	55.83	8.91	8.81	53.54
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\mathring{\alpha}$ estimation	1.00	0.99	1.01	0.89	1.29	1.01
	[0.93, 1.08]	[0.88, 1.11]	[0.93, 1.08]	[0.80, 0.98]	[1.15, 1.44]	[0.91, 1.10]

Note: CD is the weak cross-sectional dependence test statistic developed by Pesaran (2015), p-values reported in parentheses. $\mathring{\alpha}$ is the exponent of cross-sectional dependence estimated using the bias-corrected method developed by Bailey et al. (2015). 90% level confidence bands in brackets.

Table 3: Unit root testing for R&D employment and knowledge stocks.

lags	ln(R&D)	ln(A)	ln(Ã)	cv 5%	cv 10%	Δln(R&D)	Δln(A)	Δln(Ã)	cv 5%	cv 10%
					France					
0	−3.06	−1.32	−1.30	−2.66	−2.58	−5.06	−2.52	−2.43	−2.15	−2.07
	[0.07]	[0.58]	[0.28]			[0.10]	[0.83]	[0.00]		
1	−2.34	−2.19	−1.99	−2.66	−2.58	−3.34	−2.17	−2.08	−2.15	−2.07
	[0.09]	[0.50]	[0.48]			[0.14]	[0.92]	[0.00]		
2	−2.17	−1.82	−1.78	−2.66	−2.58	−2.60	−1.71	−1.66	−2.15	−2.07
	[0.30]	[0.35]	[0.84]			[0.28]	[0.79]	[0.51]		
					Germany					
lags	ln(R&D)	ln(A)	ln(Ã)	cv 5%	cv 10%	Δln(R&D)	Δln(A)	Δln(Ã)	cv 5%	cv 10%
0	−1.36	−0.83	−0.52	−2.76	−2.66	−2.09	−2.10	−2.28	−2.17	−2.07
	[0.48]	[0.12]	[0.03]			[0.46]	[0.12]	[0.02]		
1	−3.56	−1.59	−1.84	−2.76	−2.66	−3.54	−1.46	−1.70	−2.17	−2.07
	[0.00]	[0.11]	[0.02]			[0.04]	[0.08]	[0.02]		
2	−1.87	−1.81	−1.18	−2.76	−2.66	−1.90	−1.40	−1.39	−2.17	−2.07
	[0.41]	[0.03]	[0.03]			[0.13]	[0.07]	[0.02]		
					Italy					
lags	ln(R&D)	ln(A)	ln(Ã)	cv 5%	cv 10%	Δln(R&D)	Δln(A)	Δln(Ã)	cv 5%	cv 10%
0	−1.80	−1.43	−1.92	−2.73	−2.63	−2.87	−3.09	−3.09	−2.21	−2.10
	[0.68]	[0.34]	[0.77]			[0.82]	[0.70]	[0.59]		
1	−2.42	−2.02	−2.29	−2.73	−2.63	−2.75	−2.14	−2.34	−2.21	−2.10
	[0.04]	[0.48]	[0.94]			[0.13]	[0.86]	[0.80]		
2	−1.82	−2.10	−3.05	−2.73	−2.63	−1.82	−1.89	−2.22	−2.21	−2.10
	[0.04]	[0.01]	[0.11]			[0.07]	[0.59]	[0.61]		
					Spain					
lags	ln(R&D)	ln(A)	ln(Ã)	cv 5%	cv 10%	Δln(R&D)	Δln(A)	Δln(Ã)	cv 5%	cv 10%
0	−2.83	−2.22	−1.96	−2.72	−2.63	−5.15	−4.38	−4.00	−2.20	−2.11
	[0.03]	[0.78]	[0.00]			[0.01]	[0.80]	[0.01]		
1	−2.44	−2.53	−2.08	−2.72	−2.63	−3.90	−3.37	−2.89	−2.20	−2.11
	[0.04]	[0.14]	[0.02]			[0.01]	[0.38]	[0.01]		
2	−1.99	−1.38	−2.17	−2.72	−2.63	−2.65	−1.99	−2.60	−2.20	−2.11
	[0.04]	[0.10]	[0.06]			[0.01]	[0.91]	[0.04]		

Note: These figures correspond to the CIPS panel unit root test proposed by Pesaran (2007). The null hypothesis of non-stationarity is tested against the alternatives of trend stationarity (variables in levels) and of stationarity (first differences). Figures in brackets correspond to the p-value of the application of the CD test statistic (Pesaran, 2015)to the residuals of the auxiliary regressions from which the individual unit root test statistics are calculated.

Table 4: Long-run cointegration relationship testing.

	Model 1: Constant					Model 2: Constant and trend				
					France					
lags	$\tilde{A}$	$\tilde{A}_{5nn}$	$\tilde{A}_{dist}$	cv 5%	cv 10%	$\tilde{A}$	$\tilde{A}_{5nn}$	$\tilde{A}_{dist}$	cv 5%	cv 10%
0	1.70	0.85	0.56	−2.34	−2.24	−6.54	−0.08	0.43	−2.93	−2.84
1	−1.78	−1.79	−2.24	−2.36	−2.26	−1.78	−2.75	−2.68	−2.97	−2.87
2	−0.36	−0.24	−0.61	−2.31	−2.20	−0.36	−0.47	−0.19	−2.90	−2.79
					Germany					
lags	$\tilde{A}$	$\tilde{A}_{5nn}$	$\tilde{A}_{dist}$	cv 5%	cv 10%	$\tilde{A}$	$\tilde{A}_{5nn}$	$\tilde{A}_{dist}$	cv 5%	cv 10%
0	2.27	2.29	2.68	−2.34	−2.24	−2.98	−2.20	−3.13	−2.93	−2.84
1	−0.83	−0.04	1.34	−2.36	−2.26	−5.00	−3.50	−3.68	−2.97	−2.87
2	−2.01	−2.08	−0.27	−2.31	−2.20	−6.64	−4.84	−4.72	−2.90	−2.79
					Italy					
lags	$\tilde{A}$	$\tilde{A}_{5nn}$	$\tilde{A}_{dist}$	cv 5%	cv 10%	$\tilde{A}$	$\tilde{A}_{5nn}$	$\tilde{A}_{dist}$	cv 5%	cv 10%
0	4.37	2.27	−4.05	−2.34	−2.24	−3.05	−2.04	−3.13	−2.93	−2.84
1	2.95	−1.30	−4.17	−2.36	−2.26	−3.17	−5.26	−0.45	−2.97	−2.87
2	1.56	−2.21	−4.04	−2.31	−2.20	−3.27	−5.39	−0.52	−2.90	−2.79
					Spain					
lags	$\tilde{A}$	$\tilde{A}_{5nn}$	$\tilde{A}_{dist}$	cv 5%	cv 10%	$\tilde{A}$	$\tilde{A}_{5nn}$	$\tilde{A}_{dist}$	cv 5%	cv 10%
0	−2.45	−5.84	−2.45	−2.34	−2.24	0.24	−0.99	0.24	−2.93	−2.84
1	−2.77	−7.89	−2.77	−2.36	−2.26	−1.35	−6.29	−1.35	−2.97	−2.87
2	−4.01	−6.62	−4.01	−2.31	−2.20	−0.48	−7.91	−0.48	−2.90	−2.79

Note: Reported values correspond to the $CADFC_p$ residual-based test statistic developed by Banerjee and Carrion-i-Silvestre (2017). The null hypothesis is that of no cointegration between the levels of R&D employment and of the regional and external stocks of knowledge.

Table 5: Long-run cointegration relationship estimation.

Dependent variable: ln(A)	France			Germany			Italy			Spain		
	$\tilde{A}$	$\tilde{A}_{5nn}$	$\tilde{A}_{dist}$	$\tilde{A}$	$\tilde{A}_{5nn}$	$\tilde{A}_{dist}$	$\tilde{A}$	$\tilde{A}_{5nn}$	$\tilde{A}_{dist}$	$\tilde{A}$	$\tilde{A}_{5nn}$	$\tilde{A}_{dist}$
ln(R&D)	0.17**	0.01	0.17**	0.07	0.14	0.08	0.06	0.02	0.06	0.11	0.40*	0.19
	(0.08)	(0.04)	(0.08)	(0.11)	(0.14)	(0.11)	(0.07)	(0.05)	(0.07)	(0.22)	(0.24)	(0.21)
$\ln(\tilde{A})$	0.72***	0.80***	0.73***	0.95***	1.00***	0.99***	1.11***	1.11***	1.08***	1.08***	1.13***	1.12***
	(0.18)	(0.21)	(0.20)	(0.19)	(0.24)	(0.19)	(0.16)	(0.22)	(0.15)	(0.24)	(0.26)	(0.23)
Trend	0.01**	0.00	0.01***	0.00	0.00	0.00	−0.01	−0.00	−0.01	–	–	–
	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	–	–	–
RMSE	0.04	0.04	0.04	0.04	0.04	0.04	0.13	0.10	0.13	0.69	0.67	0.69
Observations	484	484	484	253	253	253	361	361	361	425	425	425
Regions	22	22	22	11	11	11	19	19	19	17	17	17

Note: These long-run parameters have been obtained using the mean group estimator proposed by Pesaran and Smith (1995). Standard errors in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10% levels, respectively.

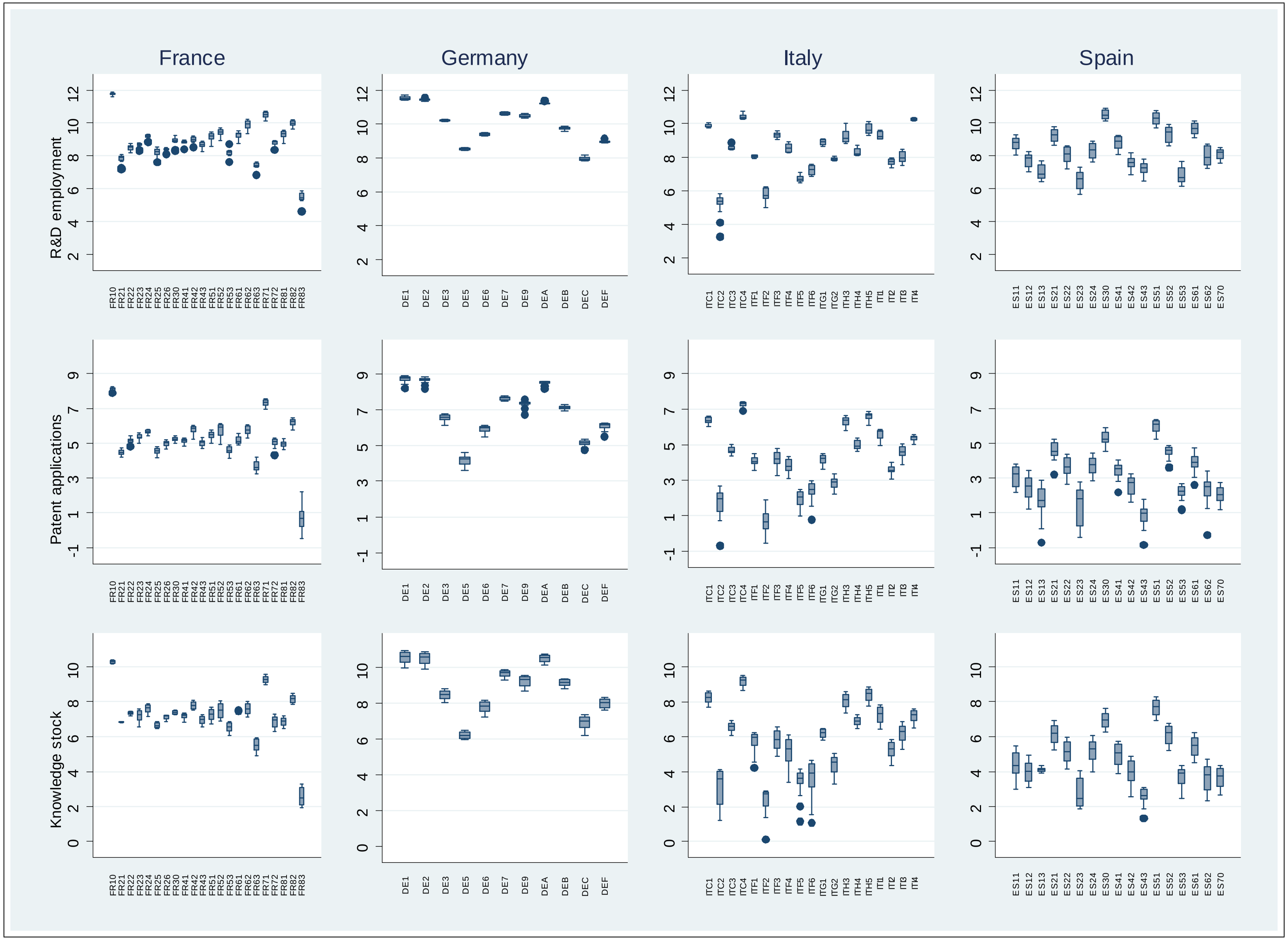

Figure 1: Sample description. Box-plots of R&D employment, patent applications and knowledge stocks across countries and regions (in natural logarithms, 1995-2010).

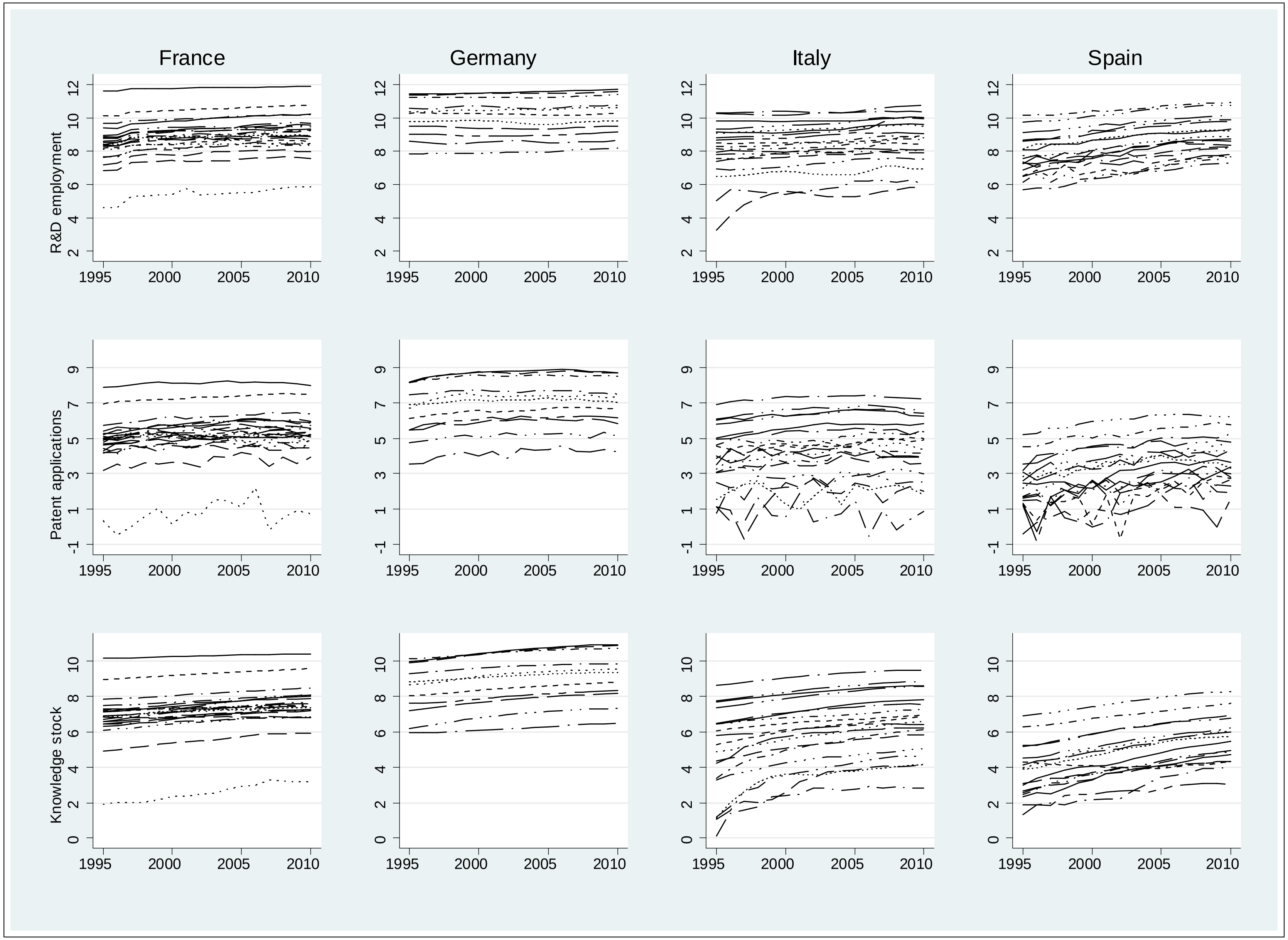

Figure 2: Evolution of regional R&D employment, patent applications and knowledge stocks (in natural logarithms, 1995-2010).