Knowledge production in European Union: Evidence from a national level panel data

Pinto, Hugo

CIEO – Centro de Investigação sobre o Espaço e as Organizações

7 December 2010

Knowledge Production in European Union: Evidence from a National Level Panel Data

WORKING PAPER

Revised version in December, 2010

Hugo Pinto

hpinto@ualg.pt

Centre for Social Studies, University of Coimbra, and Research Centre for Spatial and Organizational Dynamics, University of the Algarve

Abstract:

The knowledge production function framework is used to understand how territories transform specific inputs into knowledge outputs. This article focuses knowledge production function estimation at European Union with twenty five member-states using a data panel analysis between 1999 and 2003. The importance of different variables in knowledge production is tested. The econometric results give relevant insights for EU decision-makers and the creation of a more integrated European Research Area and innovation cooperation within Europe.

Key-words:

Knowledge Production Function, Panel Data, European Union

Knowledge Production in European Union: Evidence from a National Level Panel Data

Introduction

The creation of a European Research Area requires a strategy and a coherent framework to establish common measures for a territory that should, at least, have some shared features. European policies, in particular since the launching of the Lisbon Agenda in 2000, have been focusing innovation as a central topic for development. One of the crucial debates is the possibility of *one size fits all* innovation policies at European level and the capacity of different territories to accommodate satisfactorily the results of the same innovation instruments.

Knowledge production, the process that a specific territory has to transform knowledge inputs in knowledge outputs, is particularly useful to test econometrically hypothesis regarding the existing differences. The idea of a Knowledge Production Function (KPF) was popularized since the works of Griliches (1979) and adapted for different contexts (for a review of KPF applications, verify Pinto and Rodrigues, 2010). A KPF tries to understand the impacts of input variables, such as R&D expenses, scientific workforce, qualification of human resources or economic structure, in a measure of knowledge and innovation productivity, commonly patent numbers. To estimate a significant KPF each statistical unit should represent the central systemic relation in the innovation process. This regards a central assertion of considering the national level as the main systemic level for knowledge production in EU level. Having, of course evident limitations especially because of the role of geographical proximity in knowledge spill-overs (Paci and Usai, 2009), the nation-states remain a central analytical and political unit mainly because of the relevance of national governments in policy making and institutional building (Hancké, 2009).

In this article, using a panel data approach - for twenty-five member-states from 1999 to 2003 - two main aspects will be explored: i) firstly, the variables with a major impact in knowledge production will be discussed, and secondly, the analysis of nature of the effects for the KPF estimation will permit some findings about the homogeneity of European countries regarding innovation.

Econometric evidences from a European Countries Panel

Presentation of Data

This section intends to comprehend the main drivers of patents by estimating an econometric model that underlines the relations of several science and technology indicators with patents at European national levels. Even if patents are not the perfect knowledge production metric, patent-based indicators assume a huge relevance in innovation studies and research evaluation because they are based on inventions which have an industrial application and cover a broad range of technologies on which there are often few other sources of data (Godin, 2005; WIPO, 2008). The interest in analysing macro-level variables is crucial as a preliminary approach to understand patenting dynamics. The integration of the model facilitates the understanding of what kind of R&D expenses have the central role in patent numbers in Europe in a context characterized by the relevance of patent indicators and its migration from being a means to becoming an end.

This estimation follows from a previous analysis (Pinto and Rodrigues, 2010) where evidences at regional scale in EU were found about the central importance of private R&D to patenting dynamism. In that opportunity the data only permitted a cross-sectional analysis but the interest in taking into account also patterns of relative evolution induced the search for relevant data.

In this way, this new estimation uses RIS - Regional Innovation Scoreboard 2006 database (European Commission, 2006) with twenty five member-states (Belgium, Czech Republic, Denmark, Germany, Estonia, Greece, Spain, France, Ireland, Italy, Cyprus, Latvia, Lithuania, Luxembourg (Grand-Duché), Hungary, Malta, Netherlands, Austria, Poland, Portugal, Slovenia, Slovakia, Finland, Sweden and United Kingdom, between 1999 and 2003. The selected variables (Table 1) are related to knowledge workers (HRSTC), life-long learning (LLL), public R&D (PUBRD), business R&D (BERD) med-high tech manufacturing employment (MHTMAN), high-tech services employment (HTSER), and EPO patents (PATENT).

The data collected was indexed in each year to EU average in order to eliminate problems related with the diversity of units and to homogenize the understanding of the coefficients. In this way it can be detected variations of the relative positions of countries from year to year, understanding the comparative evolution of each member state.

NAME	DEFINITION	RELEVANCE
HRSTC	Human Resources in Science and Technology – Core (% of population)	A rapidly changing economic environment and a growing emphasis on the knowledge-based economy have seen mounting interest in the role and measurement of skills. Meeting the demands of the new economy is a fundamental policy issue and has a strong bearing on the social, environmental and economic well-being of the population. Data on Human Resources in Science and Technology (HRST) can improve our understanding of both the demand for, and supply of, science and technology personnel — an important facet of the new economy.
LLL	Participation in life-long learning per 100 population aged 25-64)	A central characteristic of a knowledge economy is continual technical development and innovation. Individuals need to continually learn new ideas and skills or to participate in life-long learning. All types of learning of valuable, since it prepares people for "learning to learn". The ability to learn can then be applied to new tasks with social and economic benefits.
PUBRD	Public R&D expenditures (%of GDP)	R&D expenditure represents one of the major drivers of economic growth in a knowledge based economy. As such, trends in the R&D expenditure indicator provide key indications of the future competitiveness and wealth of the EU. Research and development spending is essential for making the transition to a knowledge-based economy as well as for improving production technologies and stimulating growth.
BERD	Business R&D expenditures (% of GDP)	The indicator captures the formal creation of new knowledge within firms. It is particularly important in the science-based sector (pharmaceuticals, chemicals and some areas of electronics) where most new knowledge is created in or near R&D laboratories.
MHTMAN	Employment in medium-high and high-tech manufacturing (% of total workforce)	The share of employment in medium-high and high technology manufacturing sectors is an indicator of the manufacturing economy that is based on continual innovation through creative, inventive activity. The use of total employment gives a better indicator than using the share of manufacturing employment alone, since the latter will be affected by the hollowing out of manufacturing in some countries.
HTSER	Employment in high-tech services (% of total workforce)	The high technology services both provide services directly to consumers, such as telecommunications, and provide inputs to the innovative activities of other firms in all sectors of the economy. The latter can increase productivity throughout the economy and support the diffusion of a range of innovations, in particular those based on ICT.
PATENT	EPO patents per million population	The capacity of firms to develop new products will determine their competitive advantage. One indicator of the rate of new product innovation is the number of patents. This indicator measures the number of patent applications at the European Patent Office.

Table 1: The Variables included in the Estimation Process
Source: European Commission (2006: 4-5) adapted

Variables	Mean	Median	Maximum	Minimum	Std. Dev.	Skewness	Kurtosis	Jarque-Bera
PATENT	75,85	27	273	0	83,65	0,91	2,54	18,24
HRSTC	102,67	101	176	51	33,28	0,42	2,20	7,03
LLL	104,51	72	363	13	79,79	1,41	3,91	45,41
MHTMAN	84,91	95	167	9	38,09	-0,23	2,34	3,32
HTSER	96,85	94	167	38	31,11	0,26	2,24	4,45
PUBRD	79,57	80	155	20	33,93	0,14	2,35	2,61
BERD	71,57	60	263	1	62,07	1,04	3,76	25,55

Table 2: Descriptive Statistics
Source: Own Elaboration

A first glance of descriptive statistics (Table 2) underlines some interesting features:

- The high dispersion of PATENT and BERD variables;

- The lowest dispersion of PUBRD when compared with BERD;

- PATENT and BERD assumes a non-normal distribution.

Estimating a National level KPF for an EU Panel

A preliminary general-to-particular approach, inspired in Hendry´s methodology (Hendry, 1979), permitted the simultaneous insertion of all variables in study and eliminate one-by-one the non significant ones based in a t-test. The method used was Pooled Least Squares (PLS) with White heteroskedasticity-consistent standard errors and covariance.

<table>
<tr><td>Variable</td><td>C</td><td>HRSTC</td><td>PUBRD</td><td>BERD</td><td>R-squared</td><td>Adjusted R-squared</td><td>S,E, regression</td><td>F-statistic</td></tr>
<tr><td>Coefficient</td><td>-47,29</td><td>0,28</td><td>0,29</td><td>1,06</td><td>0,86</td><td>0,86</td><td>31,78</td><td>246,07</td></tr>
<tr><td>Std. Error</td><td>10,61</td><td>0,09</td><td>0,12</td><td>0,086</td><td rowspan="2">Mean dep. Var.</td><td rowspan="2">S.D. dep. Var.</td><td rowspan="2">S.S. resid</td><td rowspan="2">Prob(F-statistic)</td></tr>
<tr><td>t-Statistic</td><td>-4,46</td><td>3,06</td><td>1,96</td><td>12,48</td></tr>
<tr><td>Prob.</td><td>0,00</td><td>0,00</td><td>0,05</td><td>0,00</td><td>75,85</td><td>83,65</td><td>122193,50</td><td>0,00</td></tr>
</table>

Table 3: PLS Regression Results
Source: Own Elaboration

The total balanced panel had 125 observations. The final model using homogeneous intercepts and coefficients is synthesized in Table 3[1]. For the specific estimation of the panel data model some preliminary steps must be done to assure the reliability of the analysis. Is relevant to confirm the poolability of the data to understand the heterogeneity of the cases, i.e., if we use common intercepts and coefficients, heterogeneous intercepts but common coefficients or if the analysis must be based in conditional variation of some variables. Commonly homogeneous intercepts and coefficients assumption is an unrealistic approach especially with the preliminary notion about the diversity in national behaviours on patent registration. The use of pool data methods was validated by an F-test as recommended in Baltagi (2001) and Woolridge (2006). Due to the lack of degrees of freedom two different F-tests were conducted[2]. The nature of effects and detect the type of patterns among the intercept and the coefficients of the different cases is central in panel data. Taking into account the observations of our dependent variable y in i=1,…, N cases

[1] The software used was E-Views version 4.1.
[2] F-test 1=A restricted model with homogeneous intercept and coefficients *vs* an unrestricted model with heterogeneous intercept and common coefficients. F-test 2=A restricted model without intercept and homogeneous coefficients: *vs* an unrestricted model without intercept and heterogeneous coefficients. Null hypothesis of homogeneous intercept and coefficients were accepted.

in t=1,..., T periods and k=1, …, K explicative variables defined by a vector K * 1 x, the classic linear regression model assumes the following form:

$$y_{it} = \alpha_i + b_i^{'} x_{it} + \varepsilon_{it} \qquad (1)$$

The error is independent identically distributed, iid (0, σ2ε). If the intercepts (αi) are correlated with the explicative variables coefficients (xit) a fixed effect estimation procedure is adequate. If the αi are not correlated with the xit a random effect model is more suitable.

C	HRSTC?	LLL?	MHTMAN?	PUBRD?	BERD?	HTSER?
195.1250	-0.900800	0.459857	-0.458026	-0.975404	0.777280	-0.951099
-0.900800	0.016686	-0.002924	0.015175	-0.004010	0.003612	-0.017194
0.459857	-0.002924	0.004755	-0.001126	-0.001042	6.46E-06	-0.005027
-0.458026	0.015175	-0.001126	0.024430	-0.008343	0.006587	-0.028347
-0.975404	-0.004010	-0.001042	-0.008343	0.016569	-0.008846	0.015982
0.777280	0.003612	6.46E-06	0.006587	-0.008846	0.012449	-0.018524
-0.951099	-0.017194	-0.005027	-0.028347	0.015982	-0.018524	0.057233

Table 4: Covariance Coefficient Matrix
Source: Own Elaboration

To understand this correlation a first procedure was to use the previously estimated PLS model and analyse the coefficient covariance matrix (table 4). In the first column of table it can be observed a relevant relation between the intercept and the coefficients. This analysis suggests that using fixed effects may be more adequate for our data patterns.

To conclude about the nature of the effects it is important to perform a more robust test. The Hausman test, frequently used in the literature for this outcome, analyses given a model and data in which fixed effects estimation would be appropriate, whether random effects estimation would be as good (Hausman, 1978).

The Hausman test is a test of hypothesis (H0: random effects are consistent and efficient versus H1: random effects are inconsistent when compared to fixed effects). In our case the Hausman statistic supports the rejection of the null hypothesis of the intercept not being correlated with the explicative variable[3]. In this way the individual fixed effects model is the adequate procedure to carry on the estimation. The procedure used was a general-to-specific modelling approach with a Generalized Least Squares Estimator (GLS) and White

[3] Hausman = 1.785.082,00 compared to a Chi-squared distribution critical value of 12,592 (Sig.=0,05 and six degrees of freedom)

Heteroskedasticity-Consistent Standard Errors and Covariance. The final model is synthesized in table 5.

Dependent Variable: PATENT?
Method: GLS (Cross Section Weights)
Sample: 1999 2003
Included observations: 5
Number of cross-sections used: 25
Total panel (balanced) observations: 125
One-step weighting matrix
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
BERD?	0.012103	0.013520	0.895130	0.3729
Fixed Effects				
_BE--C	106.5840			
_CZ--C	7.664167			
_DK--C	158.0969			
_DE--C	224.7395			
_EE--C	6.023303			
_GR--C	6.006360			
_ES--C	19.51348			
_FR--C	107.0518			
_IE--C	58.78429			
_IT--C	61.87717			
_CY--C	10.35159			
_LV--C	3.669293			
_LT--C	1.305600			
_LU--C	146.2509			
_HU--C	12.68049			
_MT--C	10.52738			
_NL--C	179.5326			
_AT--C	125.6518			
_PL--C	1.967632			
_PT--C	3.584575			
_SI--C	26.38671			
_SK--C	3.969150			
_FI--C	250.3562			
_SE--C	247.9962			
_UK--C	93.97523			

Weighted Statistics			
R-squared	0.998045	Mean dependent var	122.1955
Adjusted R-squared	0.997551	S.D. dependent var	155.5501
S.E. of regression	7.697814	Sum squared resid	5866.378
F-statistic	2021.331	Durbin-Watson stat	1.899483
Prob(F-statistic)	0.000000		

Unweighted Statistics			
R-squared	0.990879	Mean dependent var	75.84800
Adjusted R-squared	0.988575	S.D. dependent var	83.65097
S.E. of regression	8.941106	Sum squared resid	7914.395
Durbin-Watson stat	1.456391		

Table 5: Fixed Effects Regression Results
Source: Own Elaboration

Policy Implications and Concluding Remarks

The results of both models, pooled least squares and fixed effects, emphasize the crucial impact of business R&D expenditures (BERD) and the existence of human resources in Science and Technology in the number of patents. Business R&D is the only significant variable in both models. Patent registration and licensing are important mechanisms to bring to market new ideas and transfer new knowledge across institutional borders. Countries where firms demonstrate a minor capacity to invest in R&D has also a smaller absorptive capacity as demonstrated by this interesting field of inquiry introduces by Cohen and Levinthal (1990). In this way the knowledge transfer processes may be ineffective as no linkages between research and economic activities exist. A necessary requirement can be a minimum threshold of human capital operating in private and public bodies, as confirmed by the estimated models, which permits the creation and utilization of new knowledge and its successful share, production and protection for appropriating related benefits. The estimation results follow others underlined by different authors when estimating knowledge production functions using patent numbers as a proxy to innovation and public and private R&D as an inputs (inter alia, Jaffe, 1989). Nevertheless the importance of public expenses in research and development activities they seem to have a secondary role in patenting dynamics when compared with the direct impact of private efforts. Firms remain the central actor in appropriating the value of knowledge through the commercialisation of products and to incorporate the relevant innovations derived from scientific research and academic institutions.

In sum, the panel data macro level models, even if only a rough approximation and suffering from several limitations, confirm in EU member-states the direct impact of the private expenditures in R&D in the dynamics of innovating, measured by patenting numbers. Firms remain central to transform knowledge in inventions with innovative potential. The model underlines a interesting aspect for an effective ERA structure, even if national level variety in terms of departure points exist, proved by the existence in heterogeneous intercepts, a similar capacity to transform innovation inputs in outputs in relative terms, the homogeneous coefficients, subsists.

The results of the current article also increase the interest in the utilization of a KPF to test the importance of different types of proximities in the knowledge production in European Union. Following the ideas that proximity is not limited to geographical distance (Boschma, 2005 or Torre and Rallet, 2005), the utilization of data panel and spatial econometric techniques can be useful to test, in a future analysis, the relevance of physical distance (measured in kilometres between the capital city's distances), geographical contiguity (a dummy that assumes the value 1 if bordering countries, 0 if not), linguistic distance (differences regarding the percentage of population with English proficiency), institutional proximity (belonging of the similar type of capitalism, e.g., Amable and Lung, 2008), technological distance (differences of knowledge-intensive workers share, and finally, the economic distance (measured by differences in GDP level) in knowledge production and spill-over generation.

Acknowledgements

Financial support from *Fundação para a Ciência e a Tecnologia* (grant ref. SFRH/BD/35887/2007) is gratefully acknowledged. A preliminary version of this model was presented in Manchester (June 2009) at the PRIME PhD and Early Career Researcher Conference.

References

Amable, B. and Lung, Y. (2008) *The European Socio-Economic Models of a Knowledge-based society. Main findings and conclusion*. Cahiers du GREThA 2008-26. Groupe de Recherche en Economie Théorique et Appliquée.

Baltagi, H. B. (2001) *Econometric Analysis of Panel Data*, Second Edition, New York, John Wiley & Sons LTD.

Boschma, R. (2005) *Proximity and Innovation: A Critical Assessment*, Regional Studies, 39, pp. 61-74.

Cohen, W. and Levinthal, D. (1990) *Absorptive Capacity: A New Perspective on Learning and Innovation*, Administrative Science Quarterly, 35, pp. 128-152.

European Commission (2006) *European Regional Innovation Scoreboard 2006*, Report prepared by Hugo Hollanders, MERIT – Maastricht Economic and social Research and training centre on Innovation and Technology.

Godin, B. (2005) *Measurement and Statistics on Science and Technology: 1920 to the Present*, London: Routledge.

Griliches, Z. (1979) *Issues in Assessing the Contribution of Research and Development to Productivity Growth*, Bell Journal of Economics, The RAND Corporation, 10, pp. 92-116.

Hancké, B. (2009) *Introducing the Debate*, in Hancké, B. (2009) (Ed.) Debating the Varieties of Capitalism – A Reader, Oxford: Oxford University Press.

Hausman, J.A. (1978) *Specification Tests in Econometrics*, Econometrica, 46 (6), pp. 1251–1271.

Hendry, D.F. (1979) *Predictive failure and econometric modelling in macroeconomics: The transactions demand for money*. *In* P Ormerod, (ed.) Economic Modelling. Heinemann, London.

Jaffe, A. (1989) *Real effects of academic research*, American Economic Review, 79, pp. 957-970.

Paci, R. and Usai, S. (2009) *Knowledge flows across European regions*, The Annals of Regional Science, 43, pp. 669-690.

Pinto, H. and Rodrigues, P. (2010) *Knowledge Production in European Regions: The Impact of Regional Strategies and Regionalization on Innovation*, European Planning Studies, Vol.19, N.10. pp. 1731-1748.

Torre, A. and Rallet, A. (2005) *Proximity and Localization*, Regional Studies, 39, pp. 47-59.

WIPO (2008) *World Patent Report – A Statistical Review*, 2008 Edition, WIPO Publication No. 931(E), World Intellectual Property Organization.

Woolridge, J. M. (2006) *Introdução à Econometria: uma abordagem moderna* [Introduction to Econometrics: A modern approach]. Tradução de Rogério Cezar de Souza, Revisão Técnica de Nelson Carvalheiro, São Paulo, Pioneira Thomson Learning.

Annex 1

Initial Model: before non-significant Variable Elimination

Pooled Least Squares Model

Dependent Variable: PATENT?
Method: Pooled Least Squares
Sample: 1999 2003
Included observations: 5
Number of cross-sections used: 25
Total panel (balanced) observations: 125
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-47.27719	13.96871	-3.384506	0.0010
HRSTC?	0.211062	0.129175	1.633923	0.1049
LLL?	-0.082621	0.068958	-1.198143	0.2333
MHTMAN?	-0.136112	0.156300	-0.870838	0.3856
PUBRD?	0.315984	0.128719	2.454828	0.0156
BERD?	1.079878	0.111576	9.678447	0.0000
HTSER?	0.198460	0.239234	0.829563	0.4085

R-squared	0.862380	Mean dependent var	75.84800
Adjusted R-squared	0.855382	S.D. dependent var	83.65097
S.E. of regression	31.81132	Sum squared resid	119411.3
F-statistic	123.2389	Durbin-Watson stat	0.157108
Prob(F-statistic)	0.000000		

Annex 2

Initial Model: before non-significant Variable Elimination

GLS Method

Dependent Variable: PATENT?
Method: GLS (Cross Section Weights)
Sample: 1999 2003
Included observations: 5
Number of cross-sections used: 25
Total panel (balanced) observations: 125
One-step weighting matrix
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
HRSTC?	0.052737	0.020034	2.632420	0.0099
LLL?	-0.024499	0.020004	-1.224667	0.2238
MHTMAN?	-0.052774	0.012022	-4.389764	0.0000
PUBRD?	-0.040140	0.022942	-1.749648	0.0834
BERD?	0.061184	0.032302	1.894143	0.0613
HTSER?	-0.052664	0.017314	-3.041619	0.0030
Fixed Effects				
_BE--C	110.9567			
_CZ--C	17.53462			
_DK--C	166.4643			
_DE--C	232.3418			
_EE--C	12.15090			
_GR--C	8.286226			
_ES--C	24.12721			
_FR--C	113.4751			
_IE--C	65.21562			
_IT--C	72.05540			
_CY--C	9.911629			
_LV--C	8.126951			
_LT--C	4.320619			
_LU--C	143.2694			
_HU--C	22.94439			
_MT--C	20.69059			
_NL--C	187.4141			
_AT--C	133.0154			
_PL--C	8.340378			
_PT--C	8.162529			
_SI--C	36.32940			
_SK--C	12.81635			
_FI--C	257.5608			
_SE--C	253.9637			
_UK--C	105.2836			

Weighted Statistics			
R-squared	0.997885	Mean dependent var	104.8917
Adjusted R-squared	0.997210	S.D. dependent var	129.8792
S.E. of regression	6.859669	Sum squared resid	4423.175
F-statistic	1478.612	Durbin-Watson stat	1.629756
Prob(F-statistic)	0.000000		

Unweighted Statistics			
R-squared	0.991122	Mean dependent var	75.84800
Adjusted R-squared	0.988289	S.D. dependent var	83.65097
S.E. of regression	9.052489	Sum squared resid	7703.070
Durbin-Watson stat	1.452147		

Annex 3

Residuals of GLS Method: Cross-sectional Units