Spillovers at the Research Unit Level *

Walter Garcia–Fontes and Raul Gonzalez

Universitat Pompeu Fabra

INCOMPLETE DRAFT. FOR DISCUSSION ONLY.

December 2004

*We thank the comments received from Sergi Jimenez, Joel Shapiro and Ernesto Villanueva. Walter Garcia-Fontes thanks project "The Value of European Patents", HPSE-CT-2002-00146, financed by the European Commission.

Abstract

This paper looks at the impact of knowledge spillovers using a survey of inventors. We propose to assess the existence of spillovers at the research unit level, defining as such the network of inventors that arises from the research done to create new patentable knowledge. We propose to value a research unit in a novel way, by means of valuation techniques similar to the ones used in the theory of the firm, and therefore our measure can be interpreted as the present discounted value adjusted by the renewal cost of intellectual capital. With this model, we assess empirically the importance of spillovers at the research unit level, by estimating the effect of spillovers on the present discounted value of the research unit adjusted by the replacement cost of intellectual capital. We construct three different types of spillover pools. After correcting for sample selection and endogeneity, we find that our results do not change with respect to the different spillover pools, which come up positive and significant. The point estimates are nevertheless unplausibly high.

1 Introduction

Quantifying the extent and impact of knowledge spillovers is crucial for the design of an adequate scientific and technological policy. Theoretically, it is clear that knowledge may have a public good nature, due to non-rivalry in consumption and non-appropriability of research returns. This may lead to external benefits of private investments in knowledge creation, and a market failure that justifies government intervention by subsidization of R&D investments or the creation of public research laboratories[1].

There has been extensive research in the identification and quantification of knowledge spillovers, at least since the seminal contribution of Griliches (1979), where he proposed different strategies to measure the contribution of R&D to economic growth. The main lines of research have dealt with the construction of "knowledge pools", by finding "close" firms, using geographical, technological or firm characteristics as distances. But most researchers recognize that spillovers cannot be measured exactly, because there are almost no observable that can be associated with the appropriation of external knowledge, and different sources that increase the knowledge pool of a firm cannot be identified separately[2]. In a recent paper, Kaiser (2002) tried to compare different approaches to the identification of spillover pools, and he found that pools based on uncentered correlations of firm characteristics seem to fit best.

One line of research that has tried to overcome this unobservability of knowledge flows is the one advocated by Jaffe (1986,1988) and his coauthors, who have proposed to use patent citations to identify knowledge flows between firms. The main assumption is that patent citations are sufficiently correlated with knowledge flows to identify spillovers. This approach is subjected to the shortcomings of patent statistics indicated by Griliches(1990), who questioned the quality of patents and related indicators, such as citation on patents, due to the heterogeneity of these indicators as a measure of the intensity of innovation activities.

[1] See Jaffee (1986,1988) for a discussion on options of technological and scientific policies to face the possibility of knowledge spillovers.

[2] See for instance Krugman (1991).

Within this fairly vast literature, there has been fewer examples of attempts to identify spillovers looking at the particular mechanisms of communication that actually permit knowledge to flow. An exception can be found in more recent work by Jaffe et. al. (2000). Through a more detailed survey on inventors, these authors look at modes and mechanisms through which knowledge flows actually take place.

This paper also looks at the possibility of knowledge flows using a survey of inventors. We propose to try to assess the existence of spillovers at the research unit level, defining as such the network of inventors that arises from the research done to create new patentable knowledge. We propose to value a research unit in a novel way, using valuation techniques similar to the ones used in the theory of the firm, and therefore our measure can be interpreted as the renewal cost of intellectual capital. Our approach is first to determine the optimal rule of capital accumulation for the research unit. For this, we construct a stochastic dynamic program for the production of ideas and allow for the interaction of knowledge spillovers and knowledge capital accumulation. Secondly, we are interested in assessing empirically the importance of spillovers at the research unit level, by estimating the effect of spillovers on the present discounted value of the research unit adjusted by the replacement cost of intellectual capital.

The paper starts by setting the basic theoretical framework in section 2. In section 3 we propose the empirical specification of the model. A description of the data and the construction of variables can be found in section 4. Section 5 presents a discussion of the main econometric issues and results, while section 6 concludes.

2 The basic framework

We consider a research unit (RU) created at some given moment in time to produce ideas that can be patentable. The production of ideas relies on two types of intangible capital: one is the stock of R&D, denoted by R and the other is the knowledge capital K. As in Griliches (1979) we assume that both types of knowledge are complementary inputs in the production of ideas. While the first type of capital reflects the investment effort to *acquire* knowledge; the second stands for all the know-how, skills and other knowledge that have been already acquired. The basis for this differentiation is that even though both types of capital can be complements in the production of ideas, not all R&D investments are materialized into new knowledge. A crucial feature of the model is that the stock of R&D investment lasts only while the RU unit arrives to a new patent, and then only a portion of this investment is converted into knowledge capital. Thus, each time a new idea is patented the stock of R&D is set again to zero. With these assumptions we want to capture several features of knowledge in the production of ideas: first, that most of the R&D effort is innovation specific, and only a part can be added to the stock of knowledge capital. Second, the degree of present R&D effort that can be added to the knowledge capital depends on the absorptive capacity of the RU and finally, knowledge spillovers affect the productivity of the current R&D effort, but not directly the stock of knowledge.

The only way the research unit can obtain a patent is discovering an idea that it is technically superior than other existing ideas. More specifically suppose that m currently patented ideas exist in the same technological field where the RU wants to patent. These patented ideas can be ordered according to their technical attributes. Let e_l for all $l = 0, 1, ..., m$ denote the technical attributes associated with the patented idea, where e_0 represents the technological attributes of the first patent in a given field and e_m is the state of the art in terms of technical attributes. In order to be patented, an idea must satisfy $e_{m+1} > \max\{e_0, ..., e_m\}$.

Patentable ideas arrive in a Poisson fashion and the probability of arrival depends on the stock of the two types of capital and the degree of knowledge spillovers. This is the approach proposed by Reinganum (1985), though in this case we assume that there no strategic effects and therefore there is not a patent

race.

Let $h(R(s),K)$ be the idea generation function determining the speed of arrival of patentable ideas. We assume that spillovers affect the efficiency of R&D investments increasing the probability of discovery. The higher the spillovers the less concave the h function. The stock of knowledge capital K affects the production of ideas setting a minimum threshold satisfying $h(0) = K$. The h function is non increasing in R, twice continuously differentiable, with $h^{'}(R) > 0$ and $h^{''}(R) < 0$ for all $R \in [0,\infty)$. These assumptions on the ideas generation function guarantee that more knowledge capital assures a higher probability of discovery but at a decreasing rate (see figure 1). Given that the date of discovery is random the probability that an idea with technological attributes equal to e_{m+1} arrives at some unknown date τ is $\Pr\{\tau(R(s),K) \leq \tau\} = e^{-h(R(s),K)}$, the rest of ideas produced in the interval $[0,\tau]$ have no technological value.

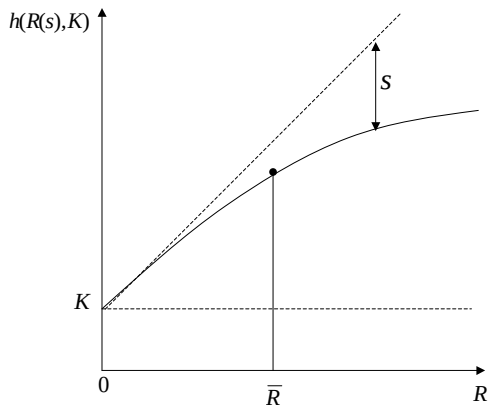

Figure 1: Idea generating function

2.1 The research unit's problem

The problem of the research unit, is to find the optimal rule of capital accumulation that maximizes the value of the research unit today given the initial stock of knowledge K. We assume that patented ideas can be valued given their technological attributes, thus a patent with a technological attributes e_{m+j} will have an

associated value of $V(e_{m+j})$ being $V = \{V_1, V_2, ..., V_T\}$ the sequence of expected patent values created by the research unit. We can interpret all the elements in the value sequence V as the present value expected profits that the RU can obtain selling the patent.

The objective function for the RU can be stated as

$$V(0) = \int_0^\infty [f(R(s),K)V_t - p\Phi(r,K)]\, e^{-\rho t} dt \tag{1}$$

where $f(R,K) = e^{-h(R(s),K)t} h(R(s),K)$ represents the probability density that a patent with value V_t arrives at some t, r represents the investment in R&D, p is the price of the R&D capital, $\Phi(r,K)$ is a linearly homogeneous adjustment cost function and ρ represents the discount rate. The objective function is maximized subject to the capital accumulation equation,

$$\dot{K} = \theta r + (1-\delta) K \tag{2}$$

where δ is the rate of depreciation of the knowledge stock and $\theta \in (0,1)$ represents the degree of absorptive capacity of the RU. The higher the value of θ the more the current investment in R&D is transformed into knowledge capital.

This dynamic optimization problem can be solved using the Pontryagin Maximum principle, introducing a shadow price λ for the constraint. The first-order conditions for optimality are

$$\lambda = p\Phi_r e^{-\rho t} \tag{3}$$

$$-\dot{\lambda} = [\pi_K - p\Phi_K]\, e^{-\rho t} - \lambda\delta \tag{4}$$

where $\pi_K = \partial f(R(s),K)V_t / \partial K$ and the transversality condition

$$\lim_{t\to\infty} \lambda(0) K(0) e^{-\rho t} = 0 \tag{5}$$

Expression (3) characterizes the investment function of the research unit. This expression relates the marginal value of investment to the shadow price of capital

$\Phi_r = \lambda/p$. In other words, this expression shows us how much increase in the value of the research unit is associated to a unit worth of investment. As in the literature of the Tobin's Q, the main problem is that we can not observe the marginal value of the investment. What we can observe indeed is the ratio of the existing capital to its replacement cost (average Q).

Following the approach introduced by Hayashi (1982) it is possible to show that the marginal and the average Q are equivalent under certain circumstances. Using the transversality condition it is possible to show that (see appendix for derivation)

$$\lambda(0)K(0) = V(0) \tag{6}$$

Which means that the shadow value o the initial capital stock is equal to the value of the firm today. Using this property we can rewrite equation (3) to obtain,

$$\frac{V}{K} = p\Phi_r(r,K) \tag{7}$$

There are several different features of this approach compared to the traditional Tobin's Q approach. First, even though there are two types of capital (R&D and Knowledge), the only stock that is considered in the long term is the knowledge capital K. Second, and since the investment in R&D affect both the current stock of R&D and the (long term) stock of knowledge capital the shadow price of capital implicitly embodies both types of capital.

3 Empirical implementation

On the spirit of Griliches(1981), Jaffe (1986), Cockburn and Griliches (1987) and Hall (many years) we assume that the value of the research unit is equivalent to the value of its assets,

$$V(t) = ZH(R(s),K) \tag{8}$$

where $V(t)$ is the value of the research unit today $Z = \exp(\sum_j \omega_j D_j + u)$ is a vector capturing those factors affecting the value of the RU and $H(R(s), K)$ is a function representing a linear relationship between the assets involved in the production of ideas. More specifically we assume $H(R(s), K) = sR + K$. Recall that R is the R&D capital stock, K represents the stock of knowledge. This specification makes that spillovers (s) affect directly the size of the R&D capital. Substituting in equation (8) we obtain

$$V = Z\left[(\gamma_1 + \gamma_2 s)R + K\right] \tag{9}$$

taking logs, letting $p = R/K$, $q = \log V - \log K$ and using the approximations $\log(1+x) \cong x$

$$\begin{aligned} \log V - \log K &= \log Z + (\gamma_1 + \gamma_2 s)\, p \\ q &= \gamma_1 p + \gamma_2 s p + \sum_j \omega_j D_j + u \end{aligned} \tag{10}$$

4 Data

Our data comes from the survey of inventors "The Value of European Patents"[3]. In this paper we use the survey run in Spain in 2003 on the universe of granted patents in the period 1994-1996. These were 623 patents. Inventors were localized through their home and work addresses (registered in the patent application). Considering the time elapsed, there were quite a few of inventors who change work or home addresses. Some of these mobile inventors were localized throught different methods (telephone directories, Internet, later patents, various public and private directories). The final number of inventors interviewed (phone and mail) amounted to 270.

The questions in the survey dealt with the following issues:

[3]This data was collected under the project Patval (INRP-2001-00007) financed by the European Comission.

1. Age and background of inventor
2. Type of institution (big firm, SME, public lab, University, etc.)
3. Commercialization
4. Economic costs to produce the patent
5. Licensing
6. Why was a patent sought
7. Family of patents
8. Sources of knowledge
9. Estimate of the value of the patent

In Table 1 some descriptive statistics can be found. There is a fairly high amount of patents which are produced by research teams, as the number of patents per inventor is 1.83. Both the R&D investment and the value of the patents are clearly right-skewed distributions, with a fairly reduced number of patents with large value and large R&D investment.

From the data in the survey, coupled with data coming from the patent application, we aggregated the data for patents that were produced for more than one inventor, while we kept individual data for patents that were produced by an individual inventor. By this procedure, and keeping only observations for which we had complete data, we ended up with 85 observations (from which 17 are aggregations of various patents through the network of inventors).

4.1 Dependent variable:

Our dependent variable is the market to replacement value of knowledge capital (q). This is the ratio between the value of the research unit today and the current stock of knowledge capital. As a proxy for the latter, we construct an index capturing a pool of factors, such as the reported value of the RU's output, the likelihood of commercialization, and other elements associated to the research team that can

affect the market value (i.e. promotions, rewards, etc). Let n be the productivity of research unit i, measured in terms of patented inventions, then we define $V_i = \sum_{j=1}^{n} \varphi_j b_j$. Where b_j is the individual patent value and φ_j is an adjustment factor that weights the value reported according to several parameters. In measuring the knowledge capital (K), we would ideally capture the cost in current prices of reproducing the knowledge needed to create new patents. The main problem we face for this variable is how to model the stock of knowledge associated to a research unit. With this variable we want to capture the skills, and accumulated knowledge of the members of the research team. Our proxy in this case is the number of patented inventions associated to all the members of the research team at some given point in time. Consider for example a research team with five researchers, then assuming that each one has patented one invention before joining the team the total knowledge stock is five. Finally the expression for our dependent variable is $q = \sum_{j=1}^{n} \varphi_j b_j / K$.[4]

4.2 Explanatory variables

Vector Z: This variable captures other effects that can influence the value of the research unit.

We included in these controls the following variables:

`Litigation`: this is a dummy variable gathering if any of the patents of the research team was ever litigated in a court (litigation meaning court proceedings other than opposition or appeal at the European Patent Office). We assume that a litigation procedure can influence positively the value of the research team. `Rewards:` This is a dummy variable that is equal to 1 if the research taem received any personal monetary compensation expressly offered because of the production of their patents. `Scale:` As a measure of scale of the research unit we use a variable that takes values from 1 to 7 gathering the scale of research time devoted to the research unit patents.

`R&D investment:` as a proxy for the R&D investment we use how much the RU has invested in developing the latest patent.

[4]Note that even for the case when $n = 1$, $q = \sum_{j=1}^{n} \varphi_j b_j / K$ is never equal to b_j.

`Spillovers:` This variable measures the pool of knowledge available to the research team outside their research lab. There are different ways to construct spillover pools that have been considered in the literature.

The identification and estimation of spillovers is a controversial issue, since spillovers cannot be observed directly. Therefore the most common strategy is to consider different ways to construct these pools to check for the robustness of the results.

We constructed three types of pools: `Spillover1:` The first type of measure that we used was a direct measure based on the stock of patents within the same technological class. This type of measure has been criticized because it may simply gather heterogeneity across technological classes. `Spillover2:` Our second measure was based on the methodology proposed by Jaffe (1986), which consists on constructing vectors for each research unit, f, that gather different characteristics of this research units, and constructing weights, $\omega_i j$ for each research unit i with respect to research unit j of the same class, based on the uncentered correlations between vectors f_i and f_j:

$$\omega_{ij} = \frac{f_i f_j'}{((f_i f_i')(f_j f_j'))^{1/2}}$$

The idea is that if the research activity of research unit i and research unit j coincide, this weight takes the value of 1 and we will consider that there will be high spillovers between the research units, while if the research activity of these research units are very different, the weight will have a low value and we will not consider any spillover between the research units.

We fill the f vector with individual characteristics that may make likely that there may be spillovers between the research units, namely the average age, type of academic degree, year of a academic degree and type of employment (firm / organization, selfemployed, student or other). This was originally proposed by Adams (1990), who replaced the patent citation data with the firms' shares of scientists in the different scientific fields.

`Spillover3:` This is the same as `Spillover2` but instead of filling the vector with individual data we fill it with dummy variables that express the likeli-

hood of reseach contacts between the research units: variables that gather if there was any external collaboration for the research achievement, and different variables that gather interactions for the research.

5 Econometric issues and discussion

The main econometric issues in the estimation of the model above come from the posssibility of sample selection and endogeneity.

5.1 Sample selection

Our model implies that we consider only patents for which a positive value of the research unit has been assessed by the inventors composing it, that is successful inventors who granted patents. This implies a potential selection bias, since there are also research units which have not been successful in their research effort. n We have tried to correct for this bias by applying Heckman's two-step estimation procedure that controls for sample selection. We consider that research units for which we got a zero value (considering that they have answered negatively on all the questions in the survey that have asked for a positive impact from their research, such as the direct value of the patent, indirect rewards from it, commercialization of it, and so on) are censored observations for which we do not observe the value of the research unit, and therefore imply a censored dependent variable in our model. The procedure implies a first step that estimates the research unit's propensity to report a positive value, and then use the information from the first stage in the estimation of our model. The specification for the selection equation include a dummy variable expressing if the research unit used external collaboration for its research, as well as a series of dummies asking from the sources of knowledge used in the research (internal, external from academics, other patents, informal contacts, etc.).

In Table 2 we report the results from the selection equation. The dummy for external collaboration does not appear significant, though it may be redundant with the other variables included, which deal with the sources of knowledge used

for the production of the research. From these academic contacts, the patent literature, costumers and users, suppliers and other sources appear as significant, while private labs, technical meetings, the scientific literature and competitors do not. This agrees with the existing literature that places a lot of emphasis on the importance of users and suppliers of technology in the production of new knowledge.

5.2 Endogeneity

The second issue that can arise in the estimation of our model is the potential endogeneity of the right-hand side own R&D stock variable. This endogeneity can come both from measurement error and from simultaneity.

It is also likely that research units will differ in ways that cannot be described by our variables. These unobservables will be included in the error terms from our research valuation equations. With panel data the classical solution would be to assume that these unobservables do not change with time, and therefore the problem could be taken into account with a "within" fixed-effects estimator, for instance. In our case, since we have a cross-section, we will hope that this problem is also solved by instrumenting the potential endogenous variable.

We instrumented this variable using all the exogenous variables in the model plus gender, year of academic degree and a dummy for type of academic degree.

5.3 Valuation equation

The results of the estimation for the valuation equation are reported in table 3.

The different spillover pools do not change the results too much, which is not surprising since the basic spillover variable (stock of patents of the whole technological class) is the same for all specifications, and the only thing that changes across specifications are the weights used in constructing the pools. Other specifications including dummies for the main type of employer of the researchers (big/small firm, university, hospital, etc.) did not come up significant.

On the other hand instrumenting the own R&D stock only reduces the signif-

icance of the coefficient for the model with spillovers based on interactions. The sample selection equation does not change the sign or significance of the coefficients, it only reduces the value of the coefficients.

Own R&D stock and the interaction with the spillover pool appear significant across specifications. It is hard to come up with a marginal effect of the size of spillovers on the valuation of the research unit, since the model is semi–log and the variables have been scaled and modified in various ways. Nevertheless, to provide a rough approximation, we can use the point-estimate corrected for sample selection, which is approximately 1.5, and considering the interaction with own R&D stock the effect on the present discounted value of each patent of the research unit is $e^{0.5}$, that is 1.64. This is the relationt between mean-corrected variables, so it cannot be interpreted as a monetary assessment of the impact of common knowledge stocks on the individual research value of the patent. This is a relatively high estimate, but not inconsistent with similar estimates using patents citantions (see for instance Jaffee (1986).).

6 Concluding remarks

In this paper we try to assess the extent and impact of spillovers at a fairly disaggregated level which has not been treated, the research unit level. After developing a theoretical framework based on the valuation research unit, we propose to estimate a model where the dependent variable can be interpreted as the value of the research unit adjusted by the replacement cost of intellectual capital.

Our main results show that spillover pools, constructed under three alternatives across technological classes, turn out to be positive and significant. Under the current specification it is not easy to come up with a quantitative assessment of the size of spillovers, but a rough approximation shows that the size is on the high end.

Future refinements of the current model should include the inclusion of firm characteristics, to try to account to the unobserved heterogeineity that may be present in the current data. This is possible by crossing the survey of inventors with data from firm databases.

References

Adams, J. (1990). Fundamental stocks of knowledge and productivity growth. *Journal of Political Economy, 98(4)*, 673–02.

Cockburn, I. & Griliches, Z. (1987). Industry effects and appropriability measures in the stock market's valueation of R&D and patents. Working Paper 2465, NBER.

Griliches, Z. (1979). Issues in assesing the contribution of research and development to productivity growth. *Bell Journal of Economics, 10(1)*, 92–116.

Griliches, Z. (1981). Market value, R&D and patents. *Economic Letters, 7*, 183–187.

Griliches, Z. (1990). Patent statistics as economic indicators: a survey. *Journal of Economic Literature, 28(4)*, 1661–1707.

Hayashi, F. (1982). Tobin's marginal q and average q: A neoclassical interpretation. *Econometrica, 50(1)*, 213–224.

Jaffe, A. (1986). Technological opportunity and spillovers of R& D: evidence from firms' patents, profits, and market value. *American Economic Review, 76(5)*, 584–1001.

Jaffe, A. (1988). Demand and supply influences in R& D intensity and productivity growth. *The Review of Economic and Statistics, 70(3)*, 431–437.

Jaffe, A. (1996). Economic analysis of research spillovers: Implications for the advanced technology program. Research memorandum, Advanced Technology Program.

Jaffe, A., Trajtenberg, M. & Fogarty, M. (2000). Knowledge spillovers and patent citations. *American Economic Review, 90(2)*, 215–218.

Kaiser, U. (2002). Measuring knowledge spillovers in manufacturing and services: an empirical assessment of alternative approaches. *Research Policy, 31*, 124–144.

Krugman, P. (1991). *Geography and Trade*. Cambdridge: MIT Press.

Reinganum, J. (1985). Innovation and industry evolution. *The Quarterly Journal of Economics, 100(1)*, 81–99.

Table 1: Descriptive statistics for the survey

	Mean	Min	Max
Number of inventors per patent	1.83	1	11
Number of research units	123	1	245
R&D investment per patent (Eur)	$155,826$	$1,000$	$8,000,000$
Patent value (Eur)	$2.72e+07$	$30,000$	$5.0e+08$

Table 2: Selection equation (First-step estimation) on the probability of observing a positive research unit value

External collaboration	0.007
	(0.008)
Sources of knowledge:	
Academic contacts	-0.013**
	(0.005)
Private labs	0.005
	(0.004)
Technical meetings	0.003
	(0.005)
Scientific literature	-0.001
	0.005
Patent literature	0.021**
	(0.008)
Costumers and users	0.009*
	(0.005)
Suppliers	-0.022**
	(0.009)
Competitors	0.001
	(0.004)
Other sources	0.091**
	(0.201)
Mills ratio	5,796**
	(0.708)

Note:** 5% and * 1% significantly different from 0, respectively.

Table 3: Research unit valuation model: estimation results

	Pool 1			Pool 2			Pool 3		
	OLS	Instrumental Variables	Sample Selection	OLS	Instrumental Variables	Sample Selection	OLS	Instrumental Variables	Sample Selection
R&D Stock	1.795**	3.225**	0.939**	1.759**	4.574**	1.138**	1.832	3.248*	0.956**
	(0.692)	(1.825)	(0.480)	(0.644)	(1.459)	(0.432)	(0.668)	(1.790)	(0.442)
Spillover	3.475**	2.474**	1.636**	3.940**	2.028*	1.619**	3.751	2.768**	1.449**
(0.705)	(1.379)	(0.611)	(0.697)	(1.146)	(0.569)	(0.708)	(1.359)	(0.564)	
Litigation	4.733**	4.452**	3.205**	3.954**	3.686*	2.946**	5.147	4.953**	2.892*
(2.529)	(2.211)	(1.509)	(2.054)	(2.243)	(1.490)	(2.193)	(2.270)	(1.593)	
Rewards	6.469**	6.532**	1.527	5.316**	5.886**	1.322	4.975**	5.461**	0.820
(1.439)	(1.480)	(1.003)	(1.427)	(1.575)	(0.904)	(1.511)	(1.656)	(1.005)	
Scale	0.053	0.046	0.033*	0.029	0.024	0.022	0.025	0.023	0.020
(0.037)	(0.039)	(0.020)	(0.030)	(0.033)	(0.018)	(0.031)	(0.032)	(0.019)	
R^2	0.66			0.70			0.68		

Note:** 5% and * 1% significantly different from 0, respectively.