The Manchester School Vol 81 No. 6 876–903 December 2013
doi: 10.1111/j.1467-9957.2012.02331.x

HUMAN CAPITAL, SOCIAL CAPITAL AND SCIENTIFIC RESEARCH IN EUROPE: AN APPLICATION OF LINEAR HIERARCHICAL MODELS*

by
MATHIEU GOUDARD
Aix-Marseille Université and Amse-Greqam, Centre de la Vieille Charité
and
MICHEL LUBRANO[†]
AMU-Cnrs and Amse-Greqam, Centre de la Vieille Charité

The theory of human capital, even if it reckons the importance of time in science, is too short for explaining the existing diversity of scientific output. The paper introduces social capital as a necessary complement to explain the creation of scientific human capital. It connects these two concepts by means of a hierarchical econometric model. Bibliographical databases contain much information which is exploited to figure out collaboration, mobility, publishing habits and institutional characteristics. The two level hierarchical model is estimated on 14 European countries using bibliometric data in the fields of economics.

1 Introduction

What are the motivations for academics to engage in scientific research and which outside factors influence their publication performance? These are two questions among others addressed in the field of economics of science. In her survey paper, Stephan (1996) suggests a diversity of explanations for the individual production of scientific papers. Above the simple satisfaction of scientific curiosity, one of the main motivations of scientists is the recognition awarded by the scientific community for being the first to publish a main discovery. There is no prize for being second, remember the dispute around the discovery of the HIV. Because of this race, scientific activity becomes such a risky adventure that wages depend only for a fraction on scientific output. The American sociologist Merton (1968) pointed out the *Matthew effect* which shows that mature and recognized scientists are rewarded, both financially and by citations, above their real merit. The combination of these two characteristics (race for being the first and over recognition of already

* Manuscript received 13.1.12; final version received 3.7.12.
[†] We are grateful to Nicolas Carayol and Stephen Bazen for pointing out several references concerning the economic of science and the theory of human capital, to Pascal Lavergne for useful discussions on multilevel modeling and to Russell Davidson for useful comments and suggestions concerning exogeneity. We are grateful to Suzanne de Chevigné for arousing our interest in sociology. We would like also to thank Tony Atkinson for fruitful discussions and suggestions concerning the characterization of social capital in research institutions. Of course remaining errors are solely ours.

matured scientists) might explain the fact that we can observe scientists with an important and continuous production together with scientists who have a more cyclical production. This is illustrated for instance by Lotka (1926) whose famous statistical law models the concentration of publications among very few scientists.

A major ingredient for individual scientific production is human capital which is a combination of basic intelligence and accumulation of efficient knowledge. The life-cycle theory predicts that, due to the finiteness of life, investment declines over time. Combined with a depreciation of human capital, this explains the inverted U shape of scientific output. Several models were developed around this idea, notably by McDowell (1982), Diamond (1984, 1987) and Levin and Stephan (1991). These models recognize the importance of time in scientific discovery.

However, these models, mainly based on time trends and cohorts effects, even if they do find an age–publishing relationship, do lack an explanatory power for irregularities in the flow of output. Using panel data, Levin and Stephan (1991) introduce individual fixed effects to take into account the differences in productivity which are not explained by a life-cycle effect. But apart from displaying individual effects, no rational explanation for diversity is provided.

Life-cycle models are based strictly on individual behavior, ignoring one fundamental aspect of human capital which is increased by sharing. Surroundings, contextual effects, networking are determinant. Following Coleman (1988), the accumulation of human capital needs another ingredient that Bourdieu (1980) was the first to call *social capital*. If the concept of human capital comes from the economics literature and was the object of considerable modeling efforts, the notion of social capital comes mainly from the sociological literature which includes very little modeling. Coleman (1988) provides justifications for showing how the two notions (human capital and social capital) can work together, taking the example of education to build his demonstration. But he provides no formal mathematical model.

The aim of this paper is to combine into a single econometric model the individual publishing behavior explained by the life-cycle model together with individual effects and the social capital 'model' represented by institutional variables. This combination will be operated by means of a two-level hierarchical model. The paper is organized as follows. After reviewing the traditional human capital model of scientific production and proposing an estimable equation in Section 2, we expose in Section 3 how individual publishing strategies can be derived from a bibliometric database and used to model individual fixed effects. In Section 4, we show how the notion of social capital and its two accepted definitions can be transposed to the framework of economics of science. The following Section 5 is devoted to identifying institution variables which characterize scientific cooperation and social capital using bibliometric databases. Section 6 reviews hierarchical linear models as in Raudenbush and Bryk (2002). In Section 7, we estimate our econometric model on a sample of European economists coming from 14 different European countries and covering the period 1991–2007. Section 8 concludes.

2 A Model of Life-cycle Productivity for Scientists

Most of the human capital models explaining the research productivity of scientists are based on the model in continuous time of Ben-Porath (1967). This model describes the accumulation of human capital and explains the life-cycle profile of earnings. Individuals invest in their human capital when they are young, anticipating future earnings. They continue to invest in their human capital after their initial training, but at a lower rate which becomes zero at the end of their career. For instance the model used in Levin and Stephan (1991) is based on this theoretical framework. However, when it comes to estimation, we have to consider discrete time and calendar years. So instead of linearizing a model initially devised in continuous time, we prefer to start directly from an economic model specified in discrete time and will follow, at least partly, Diamond (1984) and adapt ideas taken in Diamond (1987), McDowell (1982) and Heckman *et al.* (2006).

2.1 Intertemporal Optimization

In models for explaining scientific production, the main decision variable is $s_t \in [0, 1]$ which allocates time within a year between using human capital K_t in a proportion $(1 - s_t)$ for earning money and in a proportion s_t for increasing the existing stock of human capital. In academia, $(1 - s_t)$ is the proportion of time devoted to routine academic occupation such as teaching, supervising PhD students, refereeing papers, participating to administrative tasks, while s_t is the proportion of time devoted to the writing of articles or books that will increase the prestige of the scientist, his or her number of citations, the recognition he or she has from his or her peers. The production function for supplementary human capital Q_t takes the simple form

$$Q_t = \beta(s_t K_t)^{\alpha} \tag{1}$$

In Diamond (1987), human capital is seen as the prestige gained by a scientist and measured by the citations that other scientists make to his or her work. Due to the continuous progress of science, citations decrease over time and human capital experiences an obsolescence at rate δ so that the yearly variation of K_t is given by

$$\Delta K_t = -\delta K_{t-1} + Q_{t-1} \tag{2}$$

The objective function of the scientist is assumed to be the maximization of his or her discounted future income. Current income Y_t is provided by the exercise of his or her routine academic work, which consists globally in renting his or her human capital for a unit wage w

$$Y_t = w(1 - s_t)K_t \tag{3}$$

This assumption is coherent with the observation that wages in academia do not depend directly on current scientific production, but mostly on routine academic work. Scientific output is primordial only for promotion or tenure acquisition. In the initial period of training, $Y_t = 0$ because $s_t = 1$. In the second period of scientific activity, $s_t < 1$, because this is the period when human capital is rented for earnings. With an actualization rate of r, the objective function simply writes

$$U = \sum_t^T \frac{1}{(1+r)^t} w(1 - s_t)K_t \tag{4}$$

T corresponding to the age of retirement. Diamond (1984) notices that with a general production function, the maximization of U implies that s_t will decrease as $t \to T$. In order to formalize this life-cycle effect, we shall follow the procedure used in the derivation of a Mincer wage equation as detailed in Heckman *et al.* (2006). We will thus propose for s_t an *ad hoc* expression which fulfils the time diminishing property. For an alternative derivation of a complete model in continuous time, see Diamond (1987).

2.2 An Explicit Solution

Let us first solve by successive substitutions the combination of (1) and (2) for $\alpha = 1$ so as to obtain

$$K_t = \prod_{j=0}^{t-1} (1 + \beta s_j - \delta)K_0 \tag{5}$$

Distinguishing between a first period of training where $s_t = 1$, devoted to the writing of the PhD dissertation and a second period where $s_t < 1$ and using logs, we have

$$\log K_t = \log K_0 + \sum_{j=0}^{p-1} \log(1 + \beta - \delta) + \sum_{j=p}^{t-1} \log(1 + \beta s_j - \delta) \tag{6}$$

The first period lasts p years while the maximum length of the second period is $T - p$ years. Using the approximation $\log(1 + x) \simeq x$, we get

$$\log K_t \simeq \log K_0 + p(\beta - \delta) - (t - p - 1)\delta + \beta \sum_{j=p}^{t-1} s_j \tag{7}$$

or expressed in term of academic experience $e = t - p$:

$$\log K_t \simeq \log K_0 + p(\beta-\delta) - (e-1)\delta + \beta\sum_{j=0}^{e-1} s_{j+p} \tag{8}$$

We now introduce the assumption that s_t is time decreasing with the following linear expression:

$$s_e = \kappa\left(1-\frac{e}{T-p}\right) \tag{9}$$

so that $s_{T-p} = 0$. The optimal capital stock is given by

$$\log K_t = \log K_0 + p(\beta-\delta) + \delta + e\left[\beta\kappa\left(1+\frac{1}{2(T-p)}\right)-\delta\right] - e^2\frac{1}{2(T-p)} \tag{10}$$

as $\Sigma_{j=0}^{e-1}(1-j/(T-p) = e[1+1/2(T-p)] - e^2/2(T-p)$. Its variation in percentage being

$$\Delta\log K_t = -\delta + \beta\kappa\left(1-\frac{t-p-1}{T-p}\right)$$

The stock of capital decreases linearly due to the effect of depreciation and of diminishing investment.

2.3 An Estimable Equation

The log of the human capital stock of a scientist with experience e at time t is a function of his or her initial conditions ($\log K_0$), a term related to the initial period of training ($p(\beta - \delta)$), a trend and a squared trend in experience. In usual Mincer equations, p represents the number of years of schooling. In science, the period of formation is entirely devoted to writing a PhD dissertation and we will suppose that this period is identical for everybody. However, because of the secular progress of science, the date of the PhD can be of importance. It is usually supposed that younger cohorts are more productive than older ones. At least, this has to be tested. Levin and Stephan (1991) introduce individual fixed effects since they propose a panel structure for their model. We have not this possibility here as our equation explains a stock of capital measured at the end of a period. So we have to introduce individual characteristics by mean of exogenous variables x_i to be determined later on, leading to an estimable equation of the following form where i denotes an individual:

$$\log K_i = \beta_0 + \rho p_i + \beta_1 e_i - \beta_2 e_i^2 + x_i'\beta + v_i \tag{11}$$

β_0 is a constant term measuring a global mean score, p_i the date of PhD for individual i, e_i is his or her academic experience and x_i a set of personal characteristics. The dependant variable is a stock of weighted publications.

3 The Database and its Informational Content

Formally, the information we need to estimate this first model would be contained in three different databases: a list of PhD recipients, a list of department members and a bibliographical database, such as that provided for instance by the Web of Science.[1] We claim that most of the information we need is contained in bibliographical databases, provided they include detailed affiliations. We shall use the ECONLIT database, because it is the only one that contains detailed affiliations (the SSCI and SCI do not). We have thus to restrain ourselves to the economic profession. Our data cover the period 1991–2007 which represents a maximum span of 17 years.[2] We have selected in the database 14 European countries, Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Spain, Sweden, Switzerland[3] and the UK. They represent the major EU countries having an important higher education sector.

3.1 Measuring Individual Scores and Experience

The stock of human capital is difficult to define and to measure. Both McDowell (1982) and Levin and Stephan (1991) have chosen to estimate a production equation where the dependent variable is the number of publications in a given year. For measuring the stock of human capital, (2) favors a measure based on the discounted sum of weighted publications. We want to take into account the obsolescence of knowledge. A paper written 20 years ago for an active researcher has less value than an article written last year. McDowell (1982) has estimated the yearly obsolescence of a paper in different fields using the data of the *Journal Citation Reports*. He found $\delta = 13$ per cent for economics. We evaluate the stock of scientific human capital using

$$K_i = \sum_{t=fpy_i}^{lpy_i} \sum_{j=1}^{n_{it}} v_j (1-\delta)^{lpy_i - t + 1} \tag{12}$$

In this formula, n_{it} is the number of articles published by author i in year t; v_j an index measuring the quality of the journal: 1 for low quality till 10 for the narrow list of top journals. fpy_i and lpy_i are respectively the first and last publishing years of author i.

We have decided to use different values of δ according to the journal in which a paper is published. Papers in top journals have on average a longer citation life than papers published in lower ranked journals. The *Journal*

[1]Levin and Stephan (1991) underlined that, for matching those three files, they had to ask the help of the *National Research Council*, because of the confidential nature of some of the lists.

[2]We could not consider a longer period, because affiliations are not reported before 1991 in ECONLIT. Note also that ECONLIT was much more difficult to access for bibliometric studies after 2007. For more details on this database, see Lubrano *et al*. (2003).

[3]Switzerland is not a formal member of the EU, but it shares many educational programs with the EU and has an important higher education sector.

TABLE 1
SAMPLE CHARACTERISTICS

Country	*Academic authors*	*Registered in RePEc*	*Productive authors*
Austria	832	243	282
Belgium	1729	536	575
Denmark	904	230	376
Finland	915	134	243
France	5570	1794	1767
Germany	4897	2275	1898
Greece	1296	272	375
Ireland	438	157	181
Italy	4132	1920	1092
Netherlands	3480	903	1682
Spain	4704	1259	1282
Sweden	1601	400	703
Swiss	1529	435	508
UK	12729	2324	5786
Total	44756	12882	16750

Note: RePEc stands for Research Papers in Economics. It is a collaborative project maintaining a decentralized bibliographic database of working papers, journal articles, books, books chapters and software components. It includes an Author service requiring voluntary registration and maintenance of a profile. The web site is www.repec.org.

Citation Reports publishes a half-life citation indicator for a number of journals. We used it to compute a variable discount rate which covers the range [6.67, 50 per cent], $\delta = 1 - \exp(\log(0.50)/\text{half})$. For journals not in the *Journal Citation Reports*, we took the mean of their quality group.

We do not know when a person started his or her academic career or the date of his or her PhD. We know only his or her first date of publication fpy_i and his or her last date of publication lpy_i. We have decided to measure total experience e_i as

$$e_i = lpy_i - fpy_i + 1$$

As we are considering a specific period of time (1991–2007), e_i measures the time needed to accumulate a given stock of publications. It does not measure exactly academic experience, because the database does not include for instance book chapters and so does not measure the complete output.

3.2 The Need to Trim the Database

When appearing in the ECONLIT database, an author can be a regular academic member, a PhD student, a visitor or even an author with no academic affiliation. This make a much greater number of individuals than the one given by academic affiliation lists. Even if we eliminate the authors that have no academic affiliation, the database still contains too many persons. Table 1 shows that there is a huge difference between the content of ECONLIT and for instance the number of economists that have registered in RePEc. That difference would be even larger when compared with the number of persons

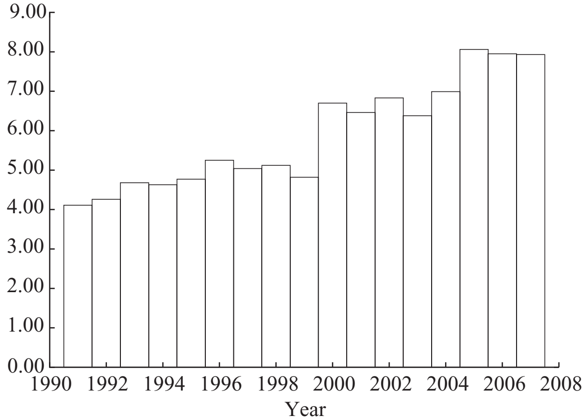

FIG. 1. First Year of Publication
Note: The y-axis indicates the percentage of authors with a first year of publication indicated on the x-axis.

holding an academic tenure. For instance, in their study Henrekson and Waldenström (2011) report a mere total of 90 Swedish economic professors, to be compared with the 1601 initial records of our database.

We have decided to eliminate authors with a too low activity. The chosen criterion is $K_i/e_i < 1.5$. This corresponds to eliminating authors who publish less than three lowest graded papers in two years when $\delta = 0$. With this rule, we get figures that are not too far from those of RePEc. And we have eliminated more than half of the initial sample.

3.3 Cohorts

We identified a vintage to the first year of publication. The empirical distribution of *fpy*, graphed in Fig. 1, illustrates the general slightly increasing number of entries in the profession at the European level.

A vintage effect is usually difficult to identify because we have the linear relation

$$\text{calendar time} = \text{experience} + \text{vintage}$$

We can introduce the first year of publication as an explanatory variable together with experience as suggested in (11). The usual practice (see, for example, Levin and Stephan, 1991 or Rauber and Ursprung, 2008) is to build dummy variables corresponding to intervals of several years, for instance four years for each vintage. The dummy will be one if the first year of publication falls into the corresponding interval, zero otherwise. We have a sample of 17 years, which means three cohorts of four years (the average time for preparing a PhD) and one cohort of five years. We have taken 1991–94, 1995–99, 2000–3 and 2004–7.

3.4 Individual Characteristics

Bibliometric databases can provide much information on the individuals characteristics x_{ij} introduced in (11), in particular concerning their publishing strategies. We can distinguish between the choice of support of publication (national journal, top journal) and their type of collaboration (publishing alone, having international coauthors, choosing coauthors only in the same institution).

Among the more than 1200 journals available for publication in economics, we have pointed out two categories. The first category corresponds to top journals in the field. The six top journals in economics are supposed to be *American Economic Review, Econometrica, Journal of Economic Theory, Journal of Political Economy, Quarterly Journal of Economics, Review of Economic Studies*. They are graded 10 on the scale used in Lubrano *et al.* (2003).[4] An author having achieved to get such publications is supposed to be a potential leader in his or her institution. We constructed a first variable called P_{10_i} which is 1 if an author has managed to publish at least one article in that short list over the whole period and 0 otherwise.

The second category of journals that we pointed out are national journals.[5] Authors publishing mostly in national journals favor national networks and thus avoid international competition. What is the consequence of this practice on their total output? We have computed Pnj_i as the proportion of articles that an author has published in national journals.

As research is a risky activity, there is an increasing tendency to publish papers with a greater number of coauthors. But this choice is not uniform among the disciplines. The number and the characteristics of coauthors is a decision variable reflecting a particular type of collaboration or absence of collaboration.

- We have defined Pal_i as the proportion of papers that an author has written alone, reflecting thus the absence of collaboration.
- Conversely, Psi_i measures the proportion of papers that an author has written with all coauthors belonging to the same institution. It is an indication of the absence of collaboration with the outside world.

[4]We could also have chosen the 10 top journals in term of total citations in 2008, i.e. *American Economic Review, Econometrica, Journal of Political Economy, Quarterly Journal of Economics, Journal of Financial Economics, Journal of Econometrics, Review of Economics and Statistics, Review of Economic Studies, Economic Journal, Journal of Economic Theory*.

[5]A national journal can be easy to define, just by looking at the language it uses. But sometimes a national journal turned for English as *Economic Notes* in Italy or *Journal of Economics* in Germany or *Spanish Economic Review* in Spain. For the UK, the matter is more complex. We have to look deeper into the journal. A journal has a national coverage if it serves as a major means of diffusion for national authors. It is considered as mainly national if in addition it does not serve as a major means of diffusion for other countries. A more precise definition is given in Lubrano *et al.* (2003).

- On the contrary, $Pint_i$ measures the proportion of papers that an author has published with at least one foreign coauthor. This last variable measures international cooperation, but also the belonging to outside social networks.[6]

An empirical question is then to know if these publishing habit variables explain and replace the cohort effects. Before examining this question, we have to introduce a new concept, that of social capital and see its operation in the field of economics of science.

4 Social Capital and Scientific Organization

The life-cycle model gives a rational principle for individual action. But it does not say anything about the institutional framework that facilitates, shapes or limit individual actions. Contrary to physical capital, human capital is expandable and self generating with use. It is transportable and shareable. For instance, Coleman (1988) shows how the initial accumulation of human capital relies on social context with the example of education. In our case, we could ask the question what is a good place for writing a PhD? We have to characterize the influence of the context on the production of scientific knowledge. A department is not a mere collection of individuals, there is something more that can be called the social capital of the department. Following a branch of the sociological literature initiated by Bourdieu (1980), Coleman (1988) and Putnam (1995), we would like to characterize what could be the notion of social capital applied in the domain of the economics of science.

4.1 Two Definitions of Social Capital

Following Coleman (1988), social capital is defined by its functions which are to facilitate individual actions that otherwise would be either more difficult or even impossible to achieve. When we try to go beyond this generality, we discover that social capital can receive two types of contradictory definitions as discussed in Siisiäinen (2003). For Bourdieu (1980, 1986), the social capital is simply the value of an individual social network. This network is used as a resource in social competition and social reproduction. It explains unequal achievements of otherwise equal individuals. On the contrary, for the American tradition represented among others by Coleman (1988) and Putnam (1995), social capital is a collective good made of moral obligations and norms, social values and social networks. A society with a high level of social capital is an integrated society, functioning on trust and collaboration. Coleman (1988) demonstrated the importance of social capital in the process

[6]We also created the variable *Pmul* which indicates multiple affiliations, but it was never significant in the subsequent regressions.

of human capital accumulation using the example of education. He pointed out the influence of a collaborative attitude of parents helping children for their homework to illustrate the influence of social capital at the family level. At the school level, he noted the very small dropout rate of students in catholic schools compared with public schools, explaining this by a common ideology of solidarity. Bourdieu on the contrary is very skeptical with respect to altruistic actions. They cannot be free of any specific interest of the actor. Individuals are engaged in social competition and do not value altruism. His position is thus totally opposed to Putnam's and Coleman's romantic ideas of generalized trust.

If the theory of human capital has been widely used (and criticized) for explaining scientific production, the theory of social capital has rarely been applied in the field of economics of science as underlined by Bozeman *et al.* (2001). This last paper proposes to use some measures of social capital when evaluating research projects.

4.2 Social Capital and Scientific Collaboration

This dual conception of social capital matches perfectly the opposition that exists in science between collaboration and competition. As we recalled in the introduction, scientific research is motivated by peer recognition, collaboration, co-authorship, but at the same time by a race for being the first. In a way or in another, social capital is imbedded in departments and universities characteristics. Researchers have networks, invite visitors, attend seminars thanks to their university. The question is to know whether competition is between departments and collaboration inside departments or whether there is competition everywhere. In hard sciences, we can suppose that competition is limited to competition between departments as the exploitation of large equipments need collaboration. Kim *et al.* (2009) found that the specific university effect on researcher productivity has declined over the last three decades even if top departments still manage to concentrate top researchers. However, Kim *et al.* (2009) consider only the case of economics and finance faculties. This would lead to think that there is more competition than collaboration in these departments.

Coleman (1988) details three forms of social capital: obligations and expectations, information channels and social norms. Coleman (1990) introduces a fourth one with authority relations. We have three quasi identical components in Putnam's (1995) concept of social capital: social values such as trust, social networks (especially voluntary associations) and moral obligations and norms (see Siisiäinen, 2003). We shall now review these general characteristics and see how they can be transcribed in the field of economics of science, showing each time the possible bent induced by scientific competition.

1. *Obligations and expectations, trust and cooperation*: you help somebody once and you expect that in the future he or she will help you in the same

way if necessary. This is mutual reciprocity and trust. At a department level, the best example is given by the relations between a supervisor and a PhD student where altruism plays a major role. The PhD student expect good guidance, the professor expect good work and outside recognition. Relations between colleagues of the same department can be an illustration too. You discuss with a colleague concerning a problem for which you need help. That colleague gives you help and ideas. He or she then might expect either the same kind of service in the future or being the co-author of your work if the discussion goes far enough. But competition between two academics working on the same topic can also exist.

2. *Information channels*: a community facilitates communication and general information. Collecting information is costly while being strategic; the group can provide it. In scientific communities, it is impossible to read all journals in order to maintain up-to-date information: the knowledge of the most recent research. Institutions organize seminars and conferences in order to diffuse this information. This leads to meeting researchers outside the department and possibly to starting new networks.
3. *Social networks*: science is a risky activity, so that co-authorship can be seen as an insurance against risk. A social network is built by gaining new co-authors, either inside the same department, or outside it as a result of conferences, seminars. So the institution, by facilitating inside and outside communication, helps academics to build their social networks. This raises the question of individual mobility and its benefit to the department which may loose members in the process.
4. *Authority relations*: a skillful leadership in a group that is fully accepted and enhance the performance of everybody. At a department level, this is illustrated by the role that a scientific leader can play. Top researchers initiate new lines of research, write papers with other members of the group and have a decisive role in scientific animation and in attracting new people. But also, the presence of a leader can lead to sterile competition and prevent younger researcher to take their full dimension (Matthew's effect).
5. *Social norms*: what is socially accepted and what is socially forbidden. What is imposed to the individual by the community in order to behave according to public interest and not according to personal interest. What is good scientific practice in relation with other scientists? Plagiary, scientific forgery have always existed, but have always been condemned and punished. Beyond saying what is allowed and what is forbidden, the norm can go further by imposing a certain type of publications. The importance of books is for instance declining in economics at the advantage of articles. Finally, the social norm *publish or perish* plays a fundamental role in scientific development. This is an internal norm in a social group, an unwritten norm which became explicit recently in some top departments. This social norm can lead to positive behaviors. By enforcing a given level

of publishing quality, it can favor the building of new networks. But it can also lead to competition between groups or between single authors leading to writing papers alone when one is convinced to have the right idea and does not want to share it with others.

5 Tracking Social Capital in Bibliometric Databases

Bibliographic databases can provide information that goes far beyond counting publications, provided we exploit them on a relatively long period. We have seen how they can provide information on individuals' characteristics. They can also provide information on the main characteristics of the institutions housing the individual researchers. But of course, we will not be able to measure all the five features of social capital described in Section 4.

We are interested in measuring the common characteristics of the human capital hosted by an institution at the end of the period of observation. We have chosen to affect an author according to a rule based on his or her last declared affiliation. When an author arrives in a new institution, he or she comes in with all his or her stock of past publications, so that in our model K_i represents the total stock of human capital of an author at the end of the period of observation, whatever his or his or her past affiliations. This stock will be explained by personal characteristics and by the characteristics (or social capital) of his or her present affiliation.

5.1 Measuring the Social Capital of an Institution

As social capital is a collection of social relations inside a department that facilitates individual scientific production by means of collaboration and of social networks, we can imagine that there is a positive correlation between the production of the different members of the same department and that this correlation could measure the importance of collaborative social capital in a department. The difference of individual performance between departments can be explained by a series of factors characterizing differences in social capital allocation.

- Interactions between authors can be favored by the size of a department. Below a minimal size, the possibilities of cooperation are nearly zero and the amount of time devoted to administrative tasks can be very large. But cooperation can be also impeded by the anonymity created by a too large number of colleagues. We call Nz_j the total number of active authors affiliated to department j.[7]

[7]In fact, Nz_j indicates the total number of active members of a department and whose last affiliation is department j, and not the total number of members of a department because there can be inactive members who are not reported in the database or who have published in different outlets than journals.

- The presence of a leader can have a tremendous effect, both by attracting other top researchers and for supervising PhD students. We can identify a top researcher and perhaps a leader by noting if he or she belongs to the small circle of authors having published at least one paper in a top journal. The variable N_{10_j} indicates the total number of top researchers affiliated to a department.
- The history of a department can be tracked by the number of top researchers that it has managed to attract in the past and that have left it. We measure it with $Nm10_j$. There is apparently a strong correlation between $N10_j$ and $Nm10_j$, meaning that present performance of a department is a function of its past achievements.
- With $Nmul_j$, we measure the proportion of authors having a current multiple affiliation in a department. This might be a sign of a lack of personal investment to the social capital of the department.
- The degree of openness to international cooperation can be measured by the proportion of articles inside the institution which were written with at least one foreign coauthor. We call this variable $Nint_j$.
- The mean seniority in the department is also an important characterization. Is the department composed mainly of seniors or of juniors? Nea_j measures the mean total publishing experience of active members of a department, while Nei_j corresponds to the mean affiliation length in this institution.
- The degree of cooperation inside a department can be measured by two antinomic variables: the proportion of papers that are produced alone Nal_j and the proportion of papers that are written with co-authors belonging all to the same department $Nsia_j$.

5.2 Exploring the Social Characteristics of Departments

Most of the previous variables are supposed to have a positive influence on the accumulation of human capital. It is useful to analyze the correlation matrix between these nine variables in order to statistically figure out positive associations and to check also for the coherency of these.

We have 512 departments. We reproduce the correlation matrix, keeping only the significant correlations. The inspection of Table 2 shows that negative correlation usually results from a mechanical effect and thus are not to be commented.

- *Nal*, the proportion of papers written alone is correlated with no other variable. This seems to be a pure individual characteristics.
- *Nsia*, representing inside collaboration depends positively on the size *Nz* and on *Nei* the mean duration of affiliation in the institution, indicating that it takes time to create fruitful inside collaboration.
- *N*10, the number of current top authors depends positively again on the size, the history of the department (represented by the number of leaders

Table 2
Correlation Matrix for Department Characteristics

	Nz	*N*10	*Nm*10	*Nal*	*Nsia*	*Nint*	*Nei*	*Nea*	*Nmul*
Nz	1.0	0.78	0.72	—	0.17	0.13	0.31	0.22	—
*N*10	0.78	1.0	0.90	—	—	0.20	0.30	0.24	—
*Nm*10	0.72	0.90	1.0	—	—	0.19	0.29	0.24	0.11
Nal	—	—	—	1.0	—	–0.25	—	—	—
Nsia	0.17	—	—	—	1.0	–0.19	0.21	—	–0.14
Nint	0.13	0.19	0.18	–0.25	–0.19	1.0	—	—	0.15
Nei	0.31	0.30	0.29	—	0.21	—	1.0	0.86	—
Nea	0.22	0.24	0.24	—	—	—	0.86	1.0	0.15
Nmul	—	—	0.11	—	–0.14	0.15	—	0.15	1.0

Calculations are based on 512 departments. Significance is determined assuming that $r\sqrt{n-2}/\sqrt{1-r^2} \sim t(n-2)$, using a 5 per cent level of significance. Only significant correlations are reported.

that have left), *Nint* the degree of international collaboration, *Nei* the mean affiliation time in the institution and *Nea* the mean total experience.

- There is a large inertia in the history of institutions as *Nm*10 and *N*10 are strongly correlated (0.90), and they are correlated with the same variables. Top authors know the glorious past of an institution which has a virtuous signaling effect.
- *Nint* represents international collaboration and depends positively again on the size *Nz*, the number of leaders *N*10, the history of the department *Nm*10 and the proportion of multiple affiliations *Nmul*. Former researchers keep relations with their former institution and multiple affiliations can be a by-product of this collaboration.
- *Nei*, the mean affiliation time, may be interpreted as the ability of the institution to keep its members. It depends positively on the size *Nz*, the number of leaders *N*10, *Nm*10 the history of the institution, *Nsia* the ability of people to work together inside the institution, not on international collaboration, and on *Nea* the total mean experience.
- *Nmul*, multiple affiliations seems to be a historical by product of *Nm*10, of international collaboration *Nint* and of total experience *Nea*.

As a summary, we have a bundle of positive association between size, international collaboration, the attraction of new top researchers, institution reputation. We expect these variables to have a positive impact on scientific production. But we have at the same time negative by-products as multiple affiliations, which are associated to the dispersion of social capital, are also positively correlated with past history, international collaboration and total experience.

5.3 Individual Mobility

Individual mobility was identified by Putnam (1995) as a factor explaining the decrease in US social capital. In his case, mobility endamages existing

TABLE 3
INDIVIDUAL MOBILITY WITHIN EUROPE OF PRODUCTIVE AUTHORS

	Authors		*No change*		*National mobility*		*Foreign mobility*	
Country	*All*	*Top*	*All*	*Top*	*All*	*Top*	*All*	*Top*
Austria	270	28	74%	57%	10%	11%	17%	32%
Belgium	551	52	70%	23%	06%	12%	23%	65%
Denmark	365	17	78%	41%	13%	35%	10%	24%
Finland	236	8	78%	25%	11%	25%	11%	50%
France	1538	162	59%	25%	28%	44%	13%	30%
Germany	1835	109	77%	43%	13%	25%	10%	32%
Greece	366	18	68%	33%	11%	11%	21%	56%
Ireland	154	12	77%	42%	01%	08%	22%	50%
Italy	1030	94	62%	33%	17%	30%	20%	37%
Netherlands	1628	94	76%	40%	11%	31%	12%	29%
Spain	1245	126	78%	52%	13%	18%	10%	29%
Sweden	683	40	85%	65%	09%	18%	06%	18%
Swiss	486	42	63%	33%	07%	07%	30%	60%
UK	5573	290	76%	44%	17%	32%	08%	23%

Column 2 gives the number of active authors in the country and column 3 the number of authors having published at least one paper in a top journal. Mobility is indicated as the percentage of these publishing authors that have moved at least once. Pmi_i and Pmo_i correspond to the columns marked All.

networks and volunteer associations. Can we conclude that mobility of researchers between different universities indicates a lack of collective investment of the researcher in his or her institution? Or does mobility correspond to the dissemination of new ideas and thus benefits to the hosting institution which welcomes a new researcher? We have here an example where the specificity of academic science can imply totally different consequences.

Mobility can easily be tracked by means of the declared affiliations. There is mobility whenever the last affiliation is different from the previous or initial affiliation. We can identify two types of individual mobility: Pmi_i indicates past mobility inside the same country; Pmo_i indicates mobility between two different countries. Finally, we have noted with $Pmul_i$ the fact that an author has multiple affiliations. In Table 3, we have reported the percentage of authors that had a national or an international mobility. We made that computation for all researchers and for top researchers separately. We have obtained a contrasted sketch of individual mobility in Europe. Most authors never change, except eventually in countries like France, Italy and Switzerland. In France and in Italy, the centralized procedure for recruiting based on the Napoleonian model of organization may be an explanation. When we turn to top authors, the picture becomes totally different. Most authors do change, except in Austria, Spain and Sweden. In small countries that share a common language with an immediate neighboring large country like Belgium, Finland, Ireland and Switzerland, top authors move most often for a foreign country. This also the case for Greece. In other countries, top authors move in a comparable proportion inside their country and for a foreign country.

6 Hierarchical Multilevel Models

Hierarchical linear models were designed to model observations that are regrouped into clusters. The main domain of application in economics is school achievement where individuals are scholars endowed with a measure of schooling performance and clusters are schools (see the US *High School and Beyond* data set from *The National Center for Education Statistics* and the survey article of Goldstein and Spiegelhalter, 1996). Hierarchical linear models will be the device that we shall use to mix together the log linear individual life-cycle model with the theory of social capital that consider the individual in his or her environment. To quote Coleman (1988) 'The conception of social capital as a resource for action is one way of introducing social structure into the rational action paradigm'. He illustrates his demonstration using an example concerning high school dropouts. He showed that Catholic schools had a lower rate of dropout than public schools, explaining this difference by a common background both at the school and at the parent levels. The same data were analyzed, with similar conclusions in Raudenbush and Bryk (2002) using linear hierarchical models. We shall follow the same route, showing how individual decision parameters of the life-cycle model are modified by contextual variables at the department level and can thus contribute to explain the variability of scientific production.

6.1 A Simple Model with Random Effects

Let us consider log of the total score K_{ij} of an individual i belonging to institution j and measured at the end of the period. Our regression model is

$$\log K_{ij} = \beta_0 + \beta_1 e_{ij} + \beta_2 e_{ij}^2 + x_{ij}'\beta + v_{ij} \qquad v_{ij} \sim N(0, \sigma^2) \tag{13}$$

The log score is explained by a constant term and a set of predictors, all observed at the individual level. These variables are on one side the life-cycle variables (experience and cohort dummies), and on the other side variables representing the individual network characteristics and publishing habits. These variables indicate how the individual has made use of his or her social capital (collaborations and network).

Authors are regrouped into departments (or clusters) where they can share unobserved common individual features, due for instance to a particular recruiting policy but also a place where they share a common social capital. The first way of introducing the possibility of a department effect is to consider a specific constant term per institution called β_{0j}. In a random effect model, the β_{0j} are normally distributed with mean β_0 and variance ω^2 while being independent of the v_{ij}. We have thus a hierarchical structure, here represented in a simplified manner as

$$\log K_{ij} = \beta_{0j} + x'_{ij}\beta + v_{ij} \quad v_{ij} \sim N(0, \sigma^2)$$
$$\beta_{0j} = \beta_0 + u_j \quad u_j \sim N(0, \omega^2) \quad u_j \perp v_{ij}$$

This model introduces a correlation between individuals inside the same department which is given by

$$\rho = \frac{\omega^2}{\omega^2 + \sigma^2} \tag{14}$$

The higher this correlation, the higher will be the unobserved sharing of a common social capital. We must note that this variance decomposition assumes that $u_j \perp v_{ij}$ and $\beta_{0j} \perp v_{ij}$.

The interpretation of constant terms in this model depends heavily on the metric which is used for measuring the predictors. β_0 represents the average score of all the departments while β_{0j} represents the average score of department j. This interpretation is valid when all the predictors are set equal to zero. This particular value of zero might be meaningful for most variables, but certainly not for experience that has to be strictly positive. In order to recover a clear interpretation for the constant terms, the predictors are usually centered around a common value, usually their sample mean. We have the choice between centering around the grand mean (the mean of the whole sample) or around the local mean (the institution or cluster mean). When the predictors are centered around their local mean, $\beta_{0j} = \beta_0 + u_j$ represents the cluster mean of the log individual scores when the predictors are taken equal to their local mean. More precisely:

$$\log K_{ij} = \beta_{0j} + (x'_{ij} - \bar{X}'_{.j})\beta + v_{ij} \tag{15}$$

where $\bar{X}_{.j}$ is a vector of empirical means computed over i for a given j. The $u_j = \beta_{0j} - \beta_0$ represent the deviation of the department means from the grand mean. The u_j can be used to rank departments according to their mean score. This is the usual way to rank schools (see, for example, Goldstein and Thomas, 1996).

How to center the predictors constitutes a large debate in the applied literature, see for instance Raudenbush and Bryk (2002, p. 31). Because it cannot be zero, the experience predictor e_{ij} is centered around a given value L. We have chosen to center it around $L = 1$, so the obtained mean score will be that of authors having one year of experience, which is the modal value in many countries of our sample.

6.2 A More General Model

Cooperative social capital variables are introduced at the department level. They modify the potentialities of the individuals, which means either their mean score or the yield of their experience. We regroup in z_j these surrounding variables.[8] Introducing a second random effect, for instance on experience, the enlarged model is

$$\begin{aligned} \log K_{ij} &= \beta_{0j} + \tilde{e}_{ij}\beta_{1j} + \tilde{e}_{ij}^2\beta_2 + \tilde{x}_{ij}\beta + v_{ij} \\ \beta_{0j} &= \beta_0 + z_j\gamma_0 + u_{0j} \\ \beta_{1j} &= \beta_1 + z_j\gamma_1 + u_{1j} \end{aligned} \tag{16}$$

where $\tilde{e}_{ij}$ is experience centered around L and $\tilde{x}_{ij}$ are the other predictors centered around their local mean. This model says that individual scores vary around a local mean β_{0j}, the mean department score, according to experience and individual characteristics $\tilde{x}_{ij}$. The local mean β_{0j} varies around the grand mean and this variation depends on department characteristics z_j. The yield of individual experience β_{1j} varies around a global mean β_1 and this variation depends also on department characteristics.

Both u_{0j} and u_{1j} are independent of v_{ij}. But u_{0j} and u_{1j} can be correlated. If u_j is the vector formed by the concatenation of u_{0j} and u_{1j}, we have

$$u_j \sim \mathrm{N}(0, \Omega). \tag{17}$$

u_{0j} indicates how much the mean log score of department j deviates from the grand mean β_0, conditionally on z_j. u_{1j} indicates how much the average yield of a supplementary year of experience deviates from its average β_1, conditionally on z_j. A positive correlation would mean that the higher the average score of the department is, the higher the return to one year of experience.

This structural model can be expressed in a reduced form which is convenient for estimation:

$$\log K_{ij} = \beta_0 + \tilde{e}_{ij}\beta_1 + \tilde{e}_{ij}^2\beta_2 + \tilde{x}_{ij}\beta + z_j\gamma_0 + \tilde{e}_{ij}z_j\gamma_1 + u_{0j} + \tilde{e}_{ij}u_{1j} + v_{ij}$$

Note the rather complex structure of the error term $u_{0j} + \tilde{e}_{ij}u_{1j} + v_{ij}$ and the fact that level-two variables appear in a product with the predictors. Due to the particular structure of the error term, this model has to be estimated using either iterated GLS, the EM algorithm or a Gibbs sampler (see Zeger and Karim 1991 for a Bayesian approach).

7 Empirical Results for Europe

We have a pooled sample of 14 European countries covering 16,750 individuals indexed by i distributed over 512 departments indexed by j. For measuring their capital stock, we have used (12) with a δ taken as a function of journal quality.[9]

[8]The level-two variables are usually not centered because we are not interested in the interpretation of the grand mean β_0 or of β_1.

[9]Note that $\delta = 0$ in usual bibliometric studies and rankings.

Table 4
ANOVA

Parameter	*Estimate*	*Std error*	*Z value*
σ_v^2	0.755	0.0084	90.4
ω_u^2	0.0493	0.0055	8.92
ρ	0.061		

7.1 *The Initial Life-cycle Model*

We first estimate our model allowing for a random effect on the constant term, with the clusters being the departments. The first estimation is simply an analysis of variance:

$$\log K_{ij} = \underset{[0.014]}{2.09} + v_{ij} + u_j \qquad -2\log lik = 43\,276 \quad BIC = 43\,288$$

There is definitively a clustering effect because of the large significance of ω_u^2, given in Table 4. However, 93.9 per cent of the variance is located within the departments and only 6.1 per cent between the departments. So that of course the intra-class correlation is small ($\rho = 0.061$), which is not uncommon with education data.

Let us now introduce the life-cycle variables and see how this initial variance is reduced. This will be a test of the validity of the initial life-cycle model:

$$\begin{aligned}\log K_{ij} = &\underset{[0.014]}{1.46} - \underset{[0.013]}{0.016}\, co_{9599} + \underset{[0.014]}{0.085}\, co_{0003} + \underset{[0.014]}{0.10}\, co_{0407}\\ &+ \underset{[0.0029]}{0.27}\, e - \underset{[0.0022]}{0.087}\, e^2 + v_{ij} + u_j\\ &\sigma^2 = 0.220 \qquad \omega^2 = 0.0174\\ &-2\log lik = 22\,745 \qquad BIC = 22\,758\end{aligned} \tag{18}$$

The human capital variables (cohorts and experience) have a very high explanatory power. They manage to explain 71 per cent of the variance at the individual level (1 − 0.220/0.755). The within correlation has increased slightly up to 0.073. Cohort effects are important, the last cohorts being more productive than the initial one by 8 and 10 per cent. This surprisingly high explanatory power is due to the fact that we are explaining a stock of publications at the end of a period and not an annual flux. The experience curve, computed from (19) has an inverted U shape. If we suppose that career begins at 30, decline in productivity begins around 43.[10] Kim *et al.* (2009) estimated a decline in productivity after only four years of career in the USA, mainly explained by tenure acquisition.

[10]Due to the small span of the data (17 years), we cannot reasonably introduce higher powers to have a more precise shape for the life-cycle productivity as was done for instance in Rauber and Ursprung (2008).

TABLE 5
LIFE-CYCLE AND SOCIAL NORMS IN EUROPE

Effect	*Estimate*	*Std error*	*t or Z value*
Intercept	1.5196	0.00785	193.57
Co95–99	–0.0625	0.00808	–7.74
e	0.2412	0.00262	91.89
$e^2/10$	–0.0855	0.00183	–46.59
Pnj	–0.0042	0.00015	–27.09
Pal	–0.0011	0.00012	–8.63
Psi	–0.0005	0.00012	–4.71
Pint	0.0003	0.00013	2.55
*P*10	0.6623	0.01419	46.68
Pmi	0.1120	0.01012	11.07
Pmo	0.1840	0.01165	15.80
σ_v^2	0.1808	0.001992	90.29
ω_u^2	0.0136	0.001510	9.02
ρ	0.070		

The first cohort is taken as the reference. All variables are taken in deviation to their local mean, except for experience which is in deviation to 1. Cohort dummies are treated as the constant term. The constant term (grand mean) measures the average European score (geometric mean). –2 log *lik* = 19,395, *BIC* = 19,407.

Cohort effects have always been difficult to identify in the literature. For instance Levin and Stephan (1991) never found that more recent vintages are more productive than older ones in the USA. Rauber and Ursprung (2008) quote other references for the USA leading to the same conclusion. However, the same Rauber and Ursprung (2008) did find a vintage effect for German economists, which they explain by the enormous and recent changes in the organization of the German academic system. The same type of changes seems to have also occurred in the rest of Europe.

7.2 *The Impact of Individual Characteristics*

Are cohort effects solely a proxy for the change in publishing habits in term of outlet of publication and co-authorship? If in the previous model, we replace the cohort dummies by a set of four publishing habit variables, *Pnj*, *Pal*, *Psi* and *Pint*, we explain 73 per cent of the variance instead of 71 per cent. So publishing habits and co-authorship have a slightly better explanation power than the cohort effect.

Let us introduce the full set of individual variables x_{ij} in the life-cycle model. They regroup the already detailed publication habits, the ability of publishing in top journals and the two mobility variables. They manage to reduce further the individual variance σ^2 by 18 per cent (1 – 0.180/0.220) while the within correlation stays at 0.070. Estimation results for this model are displayed in Table 5.

The cohort effects, which were very significant in the pure life-cycle model, are now reduced to a single dummy, suggesting again that cohort effects were a proxy for the change in publication habits.

Let us now analyze these results in detail:

- Publishing strategies play a very important role in explaining differences of output. In some European countries, the established social norm is or was to publish in national journals. This has a strong negative impact on performance. This should be put in parallel with the findings of Bauwens *et al*. (2011) concerning the use of English as a scientific vehicle. At the other extreme, being able to publish at least once in a top journal has a very strong impact on the performance of an author and has a high signaling power.
- It is profitable to have individual networks as publishing alone has always a negative impact. But not every network is profitable. Being in the same institution as his or her co-authors has a small negative impact while choosing foreign co-authors is profitable. This can be interpreted in two ways. Foreign co-authors can be chosen just because they have a higher publishing score and thus illustrate one of the conclusions in Kim *et al.* (2009). Or the opposition inside-outside networks shows simply the importance of new ideas in scientific development.
- In the social capital literature (see Putnam, 1995), mobility has always be seen as a factor decreasing social capital because it breaks social links. Here the effect is just the opposite. Productive researchers have a tendency to be mobile both inside their country and also outside their country. Mobility is a positive factor for bringing in new ideas and the effect is greater for foreign mobility.

The result on mobility should be taken with care. We have seen in Table 3 that mobility was concentrated mostly among top researchers, which creates the possibility of a bias of endogeneity. We shall check for this now.

7.3 Testing for Exogeneity

We can suppose that mobility inside the same country is motivated by administrative reasons and so cannot be endogenous. International mobility on the contrary can be motivated by scientific reasons and authors with a top score have a tendency to be more mobile as shown in Table 3. This table also shows that there are large country differences. We are going to build a two-level model explaining international mobility, using the same exogenous variables as in model estimated in Table 5 and add instrumental variables. As instruments, we have chosen country dummies (using UK as a reference) plus characteristics of the hosting institution which has managed to attract the new researcher: *Nm*10, *Nal*, *Nsia*, *Nz*, *N*10, *Nei*, *Nea*, *Nint*, *Nmul*. A specification search produced the results displayed in Table 6 with a random coefficient on the grand mean. From this model, we infer that a department manages to attract foreigners when it has already strong international cooperations (*Nint*), researchers with a strong experience (*Nea*), not necessarily

Table 6
Decision to Join a New Department when Leaving a Foreign Country

Effect	*Estimate*	*Std error*	*t or Z value*
Cste	–0.09054	0.01958	–4.62
e	0.02855	0.00168	16.96
$e^2/10$	–0.00902	0.00121	–7.41
*P*10	0.06791	0.00981	6.92
Pal	0.00097	0.00008	11.07
Psia	0.00055	0.00008	6.39
Pint	0.00210	0.00008	23.72
Austria	0.05613	0.02355	2.38
Belgium	0.09170	0.02237	4.10
France	0.03875	0.01130	3.43
Greece	0.07976	0.02012	3.96
Italy	0.1015	0.01276	7.96
Netherlands	0.03836	0.01540	2.49
Spain	0.05444	0.01373	3.96
Swiss	0.1215	0.02045	5.94
Nal	0.1108	0.03722	2.98
Nint	0.3987	0.03170	12.58
Nei	–0.04105	0.00676	–6.07
Nea	0.02766	0.00568	4.87
σ_v^2	0.08730	0.000965	90.47
ω_u^2	0.00176	0.000331	5.32
ρ	0.020		

gained in the hosting institution (*Nei*) which suggests the influence of a past mobility of the members of the hosting institution.

Let us now introduce the predicted value of the endogenous variable in the model first estimated in Table 5. This is a test of exogeneity. This predicted variable appears with a t value of 0.14 which shows that there is no endogeneity bias for the mobility variable *Pmo*.

7.4 Explaining Department Effects

Individuals do have a different mean score according to which department they belong. But differences in mean score are certainly not the sole random effect as departments can have a specific effect on productivity for instance. We have tried to introduce all random effects whenever they were significant. Starting from an initial model with a $BIC = 19{,}407$, we managed to add three extra random effects on the following variables: e, *P*10, *Pmo* to reach a $BIC = 19{,}207$. The yield of experience varies between departments as well as the efficiency of being able to publish in a top journal and the efficiency of international mobility. The importance and significance of these random effects are indicated in Table 7. At level two, the total variance is 0.0531 so that differences in mean score still represent 22 per cent of the variance. Differences in international mobility represents 20 per cent of the variance

TABLE 7
RANDOM EFFECTS WITHOUT *Z* VARIABLES

Parameter	*Estimate*	*Std error*	*Z value*
σ_v^2	0.1738	0.00196	88.46
ω_{cste}^2	0.01148	0.00143	8.02
ω_e^2	0.00016	0.00002	5.73
ω_{P10}^2	0.03082	0.00895	3.44
ω_{Pmo}^2	0.01060	0.00402	2.64

−2 log *lik* = 19,176, *BIC* = 19,207.

while the capacity to publish in top journals represent the greatest part with 58 per cent of the variance. Differences in experience (measured in years) are a negligible part of the variance even if they are very significant.[11] With this simple analysis of variance, we can say that 78 per cent of the differences between departments are due to specific department effects while 22 per cent are due to their capacity in attracting researchers with an important score.

What are the characteristics that explain these differences between departments? We introduce now level-two variables for each of the four level-two equations. Section 5.1 has given us hints on the possible nature of these effects. They are indicators related to the nature of the department social capital.

The final list appearing in Table 8 was obtained as follows. We tried the whole list of variables provided in Section 5.1 and reduced it sequentially, equation by equation. Among the nine variables proposed in Section 5.1, we managed to introduce only six of them: *N*10 the number of top authors in the department, *Nz* the size of the department (in term of productive authors), *Nei* the mean experience within the department, *Nea* the mean total experience, *Nint* the proportion of papers written with foreign co-authors, *Nmul* the number of authors having multiple affiliations. Other variables were not significant (*Nal* the proportion of papers written alone, *Nsi* the proportion of papers written only with members of the same department, *Nm*10 the number of top authors that have left the department). Adding these six variables reduces the level-two variance by 30 per cent.

The mean score of a department β_{0j} is influenced in a positive way by the number of top authors *N*10 who thus play there a leading role. The mean score of a department is also influenced by the proportion of papers written with foreign coauthors *Nint* (with a large coefficient) and by the mean duration in the affiliation *Nei*. Conversely, the mean number of explicitly declared affiliations has a negative influence on the mean total score. Multiple affili-

[11]There is a scaling effect because our model is formulated in terms of number of years of experience. If we scale it differently, the value of ω_e^2 will be different. So that this variance analysis has to be taken with care.

Table 8
Random Effect and Departments Characteristics

Effect	*Estimate*	*Std error*	*t or Z values*
Intercept	1.5318	0.06450	23.75
*Co*00–07	0.0483	0.00800	6.05
e	0.2330	0.00283	82.17
$e^2/10$	–0.0830	0.00179	–46.28
*P*10	0.5473	0.02759	19.84
Pnj	–0.0042	0.00015	–26.84
Pal	–0.0011	0.00012	–8.28
Psi	–0.0006	0.00012	–4.87
Pint	0.00033	0.00012	2.54
Pmi	0.1221	0.01016	12.01
Finland	–0.0989	0.03885	–2.55
France	–0.0830	0.02100	–3.96
Greece	–0.1184	0.03290	–3.60
Ireland	–0.1655	0.05266	–3.14
*N*10	0.0033	0.00087	3.80
Nint	0.2875	0.05826	4.93
Nei	0.0152	0.00573	2.66
Nmul	–0.1418	0.06078	–2.33
*P*10∗*N*10	0.00633	0.00183	3.45
e∗*Nz*	0.000063	0.000015	4.10
Pmo∗*Nea*	0.03724	0.00278	13.38
σ_v^2	0.1742	0.00196	88.47
σ_{cste}^2	0.00665	0.00105	6.34
σ_e^2	0.00013	0.00002	5.58
σ_{P10}^2	0.02249	0.00761	2.95
σ_{Pmo}^2	0.00772	0.00367	2.10

The first cohort is taken as the reference. All variables are taken in deviation to their local mean, except for experience which is in deviation to 1. The constant term (grand mean) measures the average European score (geometric mean because of the logs). –2 log *lik* = 19,153, *BIC* = 19,122.

ations in a department corresponds to a scraping of social capital. We have here for science a correspondence with the diagnostic made by Putnam (1995) for everyday society.

Variations in the yield of experience, in the yield of being able to publish in top journals and in the yield of mobility still represent 82 per cent of the level-two residual variance. The size of a department positively influences the yield of experience, with however a very small coefficient. The yield of being able to publish in top journals is positively influenced by the number of top authors in the department. We have a collaboration enhancing effect, simply due to the presence of other top authors which does not necessarily goes through effective collaboration. As a matter of fact *Nsia* does not enter the model.

The yield of international mobility *Pmo* enters the model solely in interaction with the mean total experience of members of the department, *Nea*. Stated otherwise, mobility brings in new idea; when joining a new department, it is better to have stayed abroad. And this effect is larger when the mean seniority of your new department is higher.

7.5 Bourdieu or Putnam?

Collaboration or competition? We have a mixed view on the internal structure of a department. Social capital and the scientific collaboration it entails have a clear positive influence on the production of science. However, mobility has to be interpreted in a different way as that of Putnam (1995). Mobility brings in new ideas, while multiple affiliations is a waste for collaboration and personal investment in the functioning of the institution.

Formal inside collaboration has a negative effect at the individual level while it does not appear at the institutional level. The yield of being able to publish in a top journal is enhanced by the presence of other top researchers in the department, but this does not result necessarily in formal collaboration (as *Nsia* does not enter the model).

International collaboration, both at the personal level and at the institutional level, has a large positive role. But we have also seen in Section 5.2 that it was correlated with multiple affiliations, which plays a negative role.

Social capital in the fields of economics of science is a complex phenomenon which mixes contradictory aspects. It is a balance between personal and collective strategies. Personal strategies clearly need collective goods in order to be efficient. But if they are pushed too far with multiple affiliations, they lead to a loss of collective efficiency. This fragile balance is illustrated by some of the findings of Kim *et al*. (2009) who show that elite US universities (in economics and finance) are loosing their competitive edge as authors affiliated to second rank universities also manage to publish in top journals.

8 Conclusion

In this paper, we have shown how to derive a theoretical model of scientific production based on the human capital model. We empirically verified the life-cycle assumption which appears to be an important factor explaining scientific production. However, this simple individual decision model is not sufficient to explain the diversity of scientific productivity. We completed this model by introducing individual characteristics, and by situating the individual action in a social context, namely external networks and the hosting institution.

Personal characteristics have a large impact to explain diversity: clearly belonging to the small class of authors being able to publish in top journals is a strong positive marker. Conversely, publishing alone is a negative personal marker.

Personal networks are associated with Bourdieu's view on social capital. We have a kind of selfish use of social networks. International collaboration is profitable for individuals while internal collaboration is not. Personal networks are the occasion of international mobility which brings in new ideas and are profitable for individuals.

At the institutional level, we observe the negative effect of some parts of individual networks. For instance, simultaneous multiple affiliations have no explanatory power at the individual level, but have a negative influence at the collective level. If at the individual level, internal collaboration is negative, at the collective level, there must exist some sort of invisible collaboration because total size, total number of top researchers and mean seniority in the institution have a positive influence. Finally, there exists some common features between individual and collective social capital: international collaboration is profitable at both levels.

The grasp we had about social capital is not perfect. Bibliographical databases reveal a lot of information concerning authors habits, choices and networks. They are however mute on acknowledgments. Acknowledgements refer to conversations, discussions, information release, comments and improvements made on the paper. We are here fully in the characterization of social capital. It is not evident how to collect this type of information and how to process it.

References

Bauwens, L., Mion, G. and Thisse, J. (2011). 'The Resistible Decline of European Science', *Louvain Economic Review*, Vol. 77, No. 4, pp. 5–31.

Ben-Porath, Y. (1967). 'The Production of Human Capital and the Life Cycle of Earnings', *Journal of Political Economy*, Vol. 75, No. 4, pp. 352–365.

Bourdieu, P. (1980). 'Le Capital Social: Notes Provisoires', *Actes de la Recherche en Sciences Sociales*, Vol. 3, pp. 2–3.

Bourdieu, P. (1986). 'The Forms of Social Capital', in J. G. Richardson (ed.), *Handbook of Theory and Research for the Sociology of Education*, New York, Greenwood.

Bozeman, B., Dietz, J. S. and Gaughan, M. (2001). 'Scientific and Technical Human Capital: an Alternative Model for Research Evaluation', *International Journal of Technology Management*, Vol. 22, Nos 7–8, pp. 716–740.

Coleman, J. (1988). 'Social Capital in the Creation of Human Capital', *American Journal of Sociology*, Vol. 94, pp. S95–S120.

Coleman, J. (1990). *Foundation of Social Theory*, Cambridge MA and London, Harvard University Press.

Diamond, A. M. (1984). 'An Economic Model of the Life-cycle Research Productivity of Scientists', *Scientometrics*, Vol. 6, No. 3, pp. 189–196.

Diamond, A. M. (1987). 'An Optimal Control Model of the Life-cycle Research Productivity of Scientists', *Scientometrics*, Vol. 11, Nos 3–4, pp. 251–253.

Goldstein, H. and Spiegelhalter, D. (1996). 'League Tables and their Limitations: Statistical Issues in Comparisons of Institutional Performance', *Journal of the Royal Statistical Society, Series A*, Vol. 159, No. 3, pp. 385–443.

Goldstein, H. and Thomas, S. (1996). 'Using Examination Results as Indicators of School and College Performance'. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, Vol. 159, No. 1, pp. 149–163.

Heckman, J., Lochner, J. L. and Todd, P. (2006). 'Earnings Functions, Rates of Return, and Treatment Effects: the Mincer Equation and beyond', in E. Hanushek and F. Welch (eds), *Handbook of the Economics of Education*, Vol. 1, Amsterdam, Elsevier Science.

Henrekson, M. and Waldenström, D. (2011). 'How Should Research Performance be Measured?', *The Manchester School*, Vol. 79, No. 6, pp. 1139–1156.

Kim, E. H., Morse, A. and Zingales, L. (2009). 'Are Elite Universities Losing their Competitive Edge?', *Journal of Financial Economics*, Vol. 93, No. 3, pp. 353–381.

Levin, S. G. and Stephan, P. E. (1991). 'Research Productivity over the Life Cycle: Evidence for Academic Scientists', *American Economic Review*, Vol. 81, No. 1, pp. 114–132.

Lotka, A. (1926). 'The Frequency Distribution of Scientific Productivity', *Journal of the Washington Academy of Sciences*, Vol. 16, No. 12, pp. 317–323.

Lubrano, M., Bauwens, L., Kirman, A. and Protopopescu, C. (2003). 'Ranking European Economics Departments: a Statistical Approach', *Journal of the European Economic Association*, Vol. 1, No. 6, pp. 1367–1401.

McDowell, J. M. (1982). 'Obsolescence of Knowledge and Career Publication Profiles: Some Evidence of Differences among Fields in Costs of Interrupted Careers', *American Economic Review*, Vol. 72, No. 4, pp. 752–768.

Merton, R. K. (1968). 'The Matthew Effect in Science', *Science*, Vol. 159, No. 3810, pp. 56–63.

Putnam, R. D. (1995). 'Bowling Alone: America's Declining Social Capital', *Journal of Democracy*, Vol. 6, No. 1, pp. 64–78.

Rauber, M. and Ursprung, H. (2008). 'Life Cycle and Cohort Productivity in Economic Research: the Case of Germany', *German Economic Review*, Vol. 9, No. 11, pp. 431–456.

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edn, London, Sage Publications Ltd.

Siisiäinen, M. (2003). 'One Concept, Two Approaches: Bourdieu and Putnam on Social Capital', *International Journal of Contemporary Sociology*, Vol. 40, No. 2, pp. 100–121.

Stephan, P. (1996). 'The Economics of Science', *Journal of Economic Literature*, Vol. 34, No. 3, pp. 1199–1235.

Zeger, S. L. and Karim, M. R. (1991). 'Generalized Linear Models with Random Effects; a Gibbs Sampling Approach', *Journal of the American Statistical Association*, Vol. 86, No. 413, pp. 79–86.