



GREQAM

Groupement de Recherche en Economie
Quantitative d'Aix-Marseille - UMR-CNRS 6579
Ecole des Hautes études en Sciences Sociales
Universités d'Aix-Marseille II et III

Document de Travail
n°2011-29

Human capital, social capital and scientific research in Europe: an application of linear hierarchical models

Mathieu Goudard
Michel Lubrano

June 2011

DT-GREQAM

Human capital, social capital and scientific research in Europe : an application of linear hierarchical models *

Mathieu Goudard Michel Lubrano

June 2011

Abstract

The theory of human capital is one way to explain individual decisions to produce scientific research. However, this theory, even if it reckons the importance of time in science, is too short for explaining the existing diversity of scientific output. The present paper introduces the social capital of Bourdieu (1980), Coleman (1988) and Putnam (1995) as a necessary complement to explain the creation of scientific human capital. This paper connects these two concepts by means of a hierarchical econometric model which makes the distinction between the individual level (human capital) and the cluster level of departments (social capital). The paper shows how a collection of variables can be built from a bibliographic data base indicating both individual behaviour including mobility and collective characteristics of the department housing individual researchers. The two level hierarchical model is estimated on fourteen European countries using bibliometric data in the fields of economics.

Keywords: Economics of science, human capital, social capital, hierarchical models, European science.

JEL codes: I21, D91, C49

*A first version of this paper was presented at the Conference *Measurement and Evaluation of Academic Research Performance: Policy Implications* organised in Braga, Portugal in June 2010. It was also presented in a seminar at Core, university of Louvain in December 2010. Comments of the participants were appreciated. 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 modelling. We are grateful to Suzanne de Chevigné for arousing our interest in sociology. We would like also to thank Tony Atkinson and Russell Davidson for fruitful discussions. Of course remaining errors are solely ours.

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 their survey papers, Stephan (1996) and Diamond (1996) suggest 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 depends only for a fraction on scientific output. The American sociologist Merton (1968) pointed out the *Matthew effect* which shows that mature and recognised 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 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 recognise 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 behaviour, 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 notion 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 behaviour explained by the life-cycle model together with individual effects and the social capital “model” represented by institutional characteristics variables. This combination will be operated by means of a two level hierarchical model. The paper is organised 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 data base 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 next section is devoted to identifying institution variables which characterise scientific cooperation and social capital using bibliometric data bases. Section 6 reviews hierarchical linear models of Lindley and Smith (1972) and Raudenbush and Bryk (2002). In section 7, we estimate our econometric model on a sample of European economists coming from fourteen different European countries and covering the period 1991-2007. In section 8, we show how these models can be used to rank economic departments. Section 9 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 formation, 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 linearising 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. (2003).

2.1 Intertemporal optimisation

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 augmenting 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, her number of citations, the recognition she has from her peers. The production function for supplementary human capital Q_t takes the simple form

$$Q_t = \beta(s_t K_t)^\alpha. \quad (1)$$

In Diamond (1987), human capital is seen as the prestige gained by a scientist and measured by the citations that other scientists make to 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

$$\Delta K_t = -\delta K_{t-1} + Q_{t-1}. \quad (2)$$

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

$$Y_t = w(1 - s_t)K_t. \quad (3)$$

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 formation, $Y_t = 0$ because $s_t = 1$. In the second period of scientific activity, $s_t < 1$, because this is the period when money is earned. With an actualisation rate of r , the objective function simply writes:

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

T corresponding to the age of retirement. Diamond (1984) notices that with a general production function, the maximisation of U implies that s_t will decrease as $t \rightarrow T$. In order to formalise this life-cycle effect, we shall follow the procedure used in the derivation of a Mincer wage equation as detailed in Heckman et al. (2003). We will thus propose for s_t an *had hoc* expression which fulfils the time diminishing property. The derivation of a complete model in continuous time is proposed in the appendix.

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:

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

Distinguishing between a first period of formation 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

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

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

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

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

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

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

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

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

$$\log K_t = \log K_0 + p(\beta - \delta) + \delta + e(\beta\kappa(1 + \frac{1}{2(T-p)}) - \delta) - e^2 \frac{1}{2(T-p)}, \quad (10)$$

as $\sum_{j=0}^{e-1} (1 - j/(T-p)) = e(1 + 1/2(T-p)) - e^2/2(T-p)$. 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 her initial conditions ($\log K_0$), a term related to the initial period of formation ($p(\beta - \delta)$), 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 can 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 in one of their equation, which is possible in their case because of the panel structure of 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. We have finally an estimable equation of the following form where i denotes an individual:

$$\log K_i = \beta_0 + \rho p_i + \beta_1 e_i - \beta_2 e_i^2 + x_i' \beta + v_i. \quad (11)$$

β_0 is a constant term measuring a global mean score, p_i the date of PhD for individual i , e_i is her academic experience and x_i a set of personal characteristics. The dependant variable is a stock of weighted publications. In order to give flesh to these variables, we have now to discuss the content of our data base.

3 The data base and its informational content

Formally, the information we need to estimate this first model would be contained in three different data bases: a list of PhD recipients, a list of department members and a bibliographical data base, such as that provided for instance by the Web of Science. 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. In this paper, we claim that most of the information we need is contained in bibliographical data bases, provided the latter includes affiliations. We shall use the ECONLIT data base, because it is the only one to contain 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.¹ We have selected in the data base fourteen European countries, Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Spain, Sweden, Switzerland and the UK. They represent the major EEC 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

¹We 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 data base, see Lubrano et al. (2003).

production equation where the dependent variable is the number of publications in a given year. For measuring the stock of human capital, (2) favours 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 $\delta = 13\%$ for economics. With this assumption, we have

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

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 used 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 JCR publishes for each journal a half life citation indicator. We used it to compute a variable discount rate which covers the range [6.67, 50%], $\delta = 1 - \exp(\log(0.50)/\text{half-life})$. For journals not in the JCR, we took the mean of their category.

Levin and Stephan (1991) have a complete view of the experience of each author by having access: first to a list of PhD recipients with the date of their PhD, second to a member list for five top US universities and third to publication lists of the Web of Science. But they have introduced a selection bias because they selected only a very small portion of the PhD receivers; they had to introduce a correction for this selection bias. We have no selection bias because we consider all authors that have declared an academic affiliation in the ECONLIT data base. However, we do not know when a person started her academic career or the date of her PhD. We know only her first date of publication FPY_i and her last date of publication LPY_i . We assume that FPY_i is the starting year of scientific career. For the observed ending year of career, we have two possibilities. Either we take the date of the end of the sample or we take LPY_i , the date of the last publication. For the while, 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 data base does not include for instance book chapters and so does not measure the complete output.

3.2 The need to trim the data base

When appearing in the ECONLIT data base, 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 data base 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 to the number of persons holding an academic tenure. For instance, in their study Henrekson and Waldenström (2008) report a mere total of 90 Swedish economic professors, to be compared to the 1 601 initial records of our data base.

Table 1: Sample characteristics

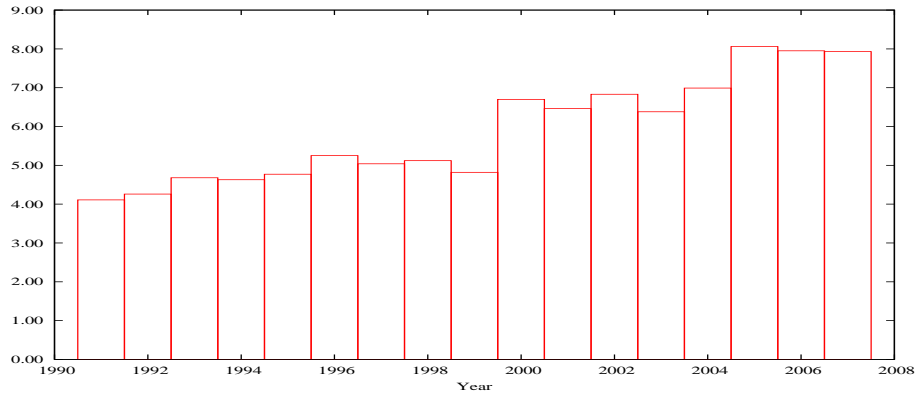
Country	Academic authors	Registered in REPEC	Productive authors
Austria	832	243	282
Belgium	1 729	536	575
Denmark	904	230	376
Finland	915	134	243
France	5 570	1 794	1 767
Germany	4 897	2 275	1 898
Greece	1 296	272	375
Ireland	438	157	181
Italy	4 132	1 920	1 092
Netherds	3 480	903	1 682
Spain	4 704	1 259	1 282
Sweden	1 601	400	703
Swiss	1 529	435	508
UK	12 729	2 324	5 786
Total	44 756	12 882	16 750

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 Figure 1, illustrates the general slightly increasing number of entries in the profession at the European level.

Figure 1: First year of publication



The y axis indicates the percentage of authors with a first year of publication indicated on the x axis.

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 e.g. Levin and Stephan 1991 or Rauber and Ursprung 2008) is to build dummy variables corresponding 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 1991-1994, 1995-1999, 2000-2003 and 2004-2007.

3.4 Individual characteristics

Bibliometric data bases 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).² An author having achieved to get such publications is supposed to be a potential leader in her institution. We constructed a first variable called P_{10i} , 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.³ Authors publishing mostly in national journals favour national networks and thus avoid international competition. What is the consequence of this practice on their total output? We have computed P_{nj_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.
- 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.⁴

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.

²We could also have chosen the ten 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*.

³A 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).

⁴We also created the variable $Pmul$ which indicates multiple affiliations, but it was never significant in the subsequent regressions.

4 Social capital and scientific organisation

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 characterise 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 characterise what could be the notion of social capital applied in the domain of the economics of science. We shall propose in a next section how to measure it.

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 (2000). 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 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 public level, he noted the very small dropout rate of students in catholic schools compared to 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 generalised trust.

If the theory of human capital has been widely used (and criticised) 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. We can have doubts on extending this interpretation to other 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 (1995)'s concept of social capital: social values such as trust, social networks (especially voluntary associations) and moral obligations and norms (see Siisiäinen 2000). 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 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. 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 organise seminars and conferences in order to diffuse this information. This lead to meet researchers outside the department and possibly to start 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 facilitation 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. This is a top researcher who initiates new lines of research, write papers with other members of the group and has 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? Plagiarism, 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 behaviours. By forcing to a given level of publishing quality, it can favour 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.

We now detail how we can build indicators from bibliographical data bases that give an account of these notions.

5 Tracking social capital in bibliometric data bases

Bibliographic data bases 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.

5.1 Allocating authors to an institution

We are interested in measuring the common characteristics of the human capital hosted by an institution at the end of the period of observation. If we had affiliations lists, we could know exactly the composition of a department at a given date. However, we have no such lists and lists available on the web are sometime hard to interpret because they include permanent members, visitors and associate members. We think that an author knows her affiliation when she declares it. We have chosen to affect an author according to a rule based on her last declared affiliation.

When an author arrives in a new institution, she comes in with all 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 past affiliations. This stock will be explained by personal characteristics and by the characteristics (or social capital) of her present affiliation.

The rule of the last declared affiliation for affecting an author to a department has to be applied with caution. First, there are errors in the data base so that the last declared affiliation might not be the real affiliation, coherent with the past declared affiliations. Second, we are going to explain a stock of publications. If an author has just moved to a new affiliation, her stock might not be so related with the characteristics of her new affiliation. Considering these two reasons, we have designed the following procedure:

- An affiliation is a valid affiliation if the author declared that affiliation at least for two different years, not necessarily consecutive. This definition applies only to authors with at least four years of experience. For those with three years of experience or less, we just consider all of their affiliations as valid. This protocol should eliminate some of the errors contained in the data base.
- An author is affected to a given valid affiliation if the last publication

year in this affiliation is either the author's last publication year or the preceding year.

5.2 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 characterising differences in social capital allocation.

- Interactions between authors can be favoured by the size of a department. Below a minimal size, the possibilities of cooperation are nearly zero. 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 .⁵
- 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 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 Nm_{10_j} . There is apparently a strong correlation between N_{10_j} and Nm_{10_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 $Nint_j$.

⁵In 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 data base or who have published in different outlets than journals.

Table 2: Correlation matrix for department characteristics

	<i>Nz</i>	<i>N10</i>	<i>Nm10</i>	<i>Nal</i>	<i>Nsia</i>	<i>Nint</i>	<i>Nei</i>	<i>Nea</i>	<i>Nmul</i>
<i>Nz</i>	1.0	0.78	0.72	-	0.17	0.13	0.31	0.22	-
<i>N10</i>	0.78	1.0	0.90	-	-	0.20	0.30	0.24	-
<i>Nm10</i>	0.72	0.90	1.0	-	-	0.19	0.29	0.24	0.11
<i>Nal</i>	-	-	-	1.0	-	-0.25	-	-	-
<i>Nsia</i>	0.17	-	-	-	1.0	-0.19	0.21	-	-0.14
<i>Nint</i>	0.13	0.19	0.18	-0.25	-0.19	1.0	-	-	0.15
<i>Nei</i>	0.31	0.30	0.29	-	0.21	-	1.0	0.86	-
<i>Nea</i>	0.22	0.24	0.24	-	-	-	0.86	1.0	0.15
<i>Nmul</i>	-	-	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% level of significance.

- The mean seniority in the department is also an important characterisation. 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*.

5.3 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 analyse 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 502 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.
- $N10$, the number of current top authors depends positively again on the size, the history of the department (represented by the number of leaders 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 $Nm10$ and $N10$ 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 $N10$, the history of the department $Nm10$ 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 $N10$, $Nm10$ 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 an historical by product of $Nm10$, 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.

5.4 Individual mobility

Individual mobility was identified by Putnam (1995) as a factor explaining the decrease in US social capital. In his case, mobility endangers existing networks and volunteer associations. Can we conclude that mobility of researchers between different universities indicates a lack of collective investment of the researcher in 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

Table 3: Individual mobility within Europe of productive authors

Country	Authors		No change		National mobility		Foreign mobility	
	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%
Netherds	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.

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 centralised procedure for recruiting based on the Napoleonic model of organisation 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 neighbouring 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 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 drop out than public schools, explaining this difference by a common background both at the school and the parent levels. The same data were analysed, 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

$$\log K_{ij} = \beta_0 + \beta_1 e_{ij} + \beta_2 e_{ij}^2 + x'_{ij} \beta + v_{ij} \quad v_{ij} \sim N(0, \sigma^2). \quad (13)$$

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 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:

$$\begin{aligned}\log K_{ij} &= \beta_{0j} + x'_{ij}\beta + v_{ij} & v_{ij} &\sim N(0, \sigma^2) \\ \beta_{0j} &= \beta_0 + u_j & u_j &\sim N(0, \omega^2), \quad u_j \perp v_{ij},\end{aligned}\tag{14}$$

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

$$\rho = \frac{\omega^2}{\omega^2 + \sigma^2}.\tag{15}$$

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:

$$\log K_{ij} = \beta_{0j} + (x'_{ij} - \bar{X}'_j)\beta + v_{ij}\tag{16}$$

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 e.g. Goldstein and Thomas 1996).

How to center the predictors constitutes a large debate in the applied literature, see for instance Raudenbush and Bryk (2002, page 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.⁶ Introducing a second random effect, for instance on experience, the enlarged model is

$$\begin{aligned}\log K_{ij} &= \beta_{0j} + \tilde{e}_{ij}\beta_{1j} + \tilde{e}_{ij}^2\beta_2 + \tilde{x}_{ij}\beta + v_{ij} \\ \beta_{0j} &= \beta_0 + z_j\gamma_0 + u_{0j} \\ \beta_{1j} &= \beta_1 + z_j\gamma_1 + u_{1j}.\end{aligned}\tag{17}$$

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

$$u_j \sim N(0, \Omega).\tag{18}$$

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.

Model (17) can be expressed in a reduced form which is convenient for estimation:

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

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

⁶The 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 .

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.⁷

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:

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

There is definitively a clustering effect because of the large significance of

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		

ω_u^2 , given in Table 4. However, 93.9% of the variance is located within the departments and only 6.1% 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:

$$\begin{aligned} \log K_{ij} = & \underset{[0.014]}{1.46} - \underset{[0.013]}{0.016} co_{9599} + \underset{[0.014]}{0.085} co_{0003} + \underset{[0.014]}{0.10} co_{0407} \\ & + \underset{[0.0029]}{0.27} e - \underset{[0.0022]}{0.087} e^2 + v_{ij} + u_j \end{aligned} \quad (19)$$

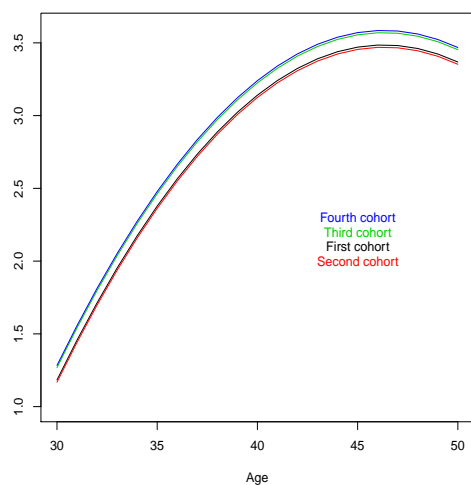
$$\sigma^2 = 0.220 \quad \omega^2 = 0.0174$$

$$-2 \log lik = 22\,745, \quad BIC = 22\,758,$$

⁷Note that in usual bibliometric studies and rankings, $\delta = 0$.

The human capital variables (cohorts and experience) have a very high explanatory power. They manage to explain 71% 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 percent. 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. In Figure 2, we illustrate the

Figure 2: Estimated life-cycle and cohorts for all countries



experience curve and its inverted U shape. The period of estimation covers 17 years and we present these curves for 20 years. If we suppose that career begins at 30, decline in productivity begins around 43.⁸ Kim et al. (2009) estimated a decline in productivity after only four years of career in the US, mainly explained by tenure acquisition.

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 US. Rauber and Ursprung (2008) quote other references for the US 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 organisation of the German academic system. The same type of changes seems to have also occurred in the rest of Europe.

⁸Due 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).

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% of the variance instead of 71%. 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% ($1-0.180/0.220$) while the within correlation stays at 0.070. Estimation results for this model are displayed in Table 5.

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
$P10$	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). $-2 * \log lik = 19\ 395$, $BIC = 19\ 407$.

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 analyse 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. (2007) 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 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 been seen as a factor decreasing social capital because it breaks social links. Here the effect seems to be 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.⁹ 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 (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: $Nm10$, Nal , $Nsia$, Nz , $N10$, 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

⁹This question was raised by Russell Davidson.

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
$P10$	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
Aus	0.05613	0.02355	2.38
Bel	0.09170	0.02237	4.10
Fra	0.03875	0.01130	3.43
Gre	0.07976	0.02012	3.96
Ita	0.1015	0.01276	7.96
Nld	0.03836	0.01540	2.49
Spa	0.05444	0.01373	3.96
Swi	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 suggest 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 model (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 $BIC = 19\ 408$, we managed to add three extra random effects on the following variables: e , $P10$, Pmo to reach a $BIC = 19\ 208$. 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

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

$-2 \log lik = 19\ 176$, $BIC = 19\ 207$

0.0531 so that differences in mean score still represents 22% of the variance. Differences in international mobility represents 20% of the variance while the capacity to publish in top journals represent the greatest part with 58% of the variance. Differences in experience (measured in years) are a negligible part of the variance even if they are very significant.¹⁰ With this simple analysis of variance, we can say that 78% of the differences between departments are due to specific department effects while 22% 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 5.2 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 5.2 and reduced it sequentially, equation by equation. Among the nine variables proposed in section 5.2, we managed to introduce only six of them: $N10$ 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, $Nm10$ the number of top authors that have left the department). Adding these six variables reduces the level-two variance by 30%.

The mean score of a department β_{0j} is influenced in a positive way by the number of top authors $N10$ who thus play there a leading role. The mean score of a department is also influenced by the proportion of

¹⁰There 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.

Table 8: Random effect and departments characteristics

Effect	Estimate	Std. Error	t or Z Values
Intercept	1.5318	0.06450	23.75
$Co00 - 07$	0.0483	0.00800	6.05
e	0.2330	0.00283	82.17
$e^2/10$	-0.0830	0.00179	-46.28
$P10$	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
Fin	-0.0989	0.03885	-2.55
Fra	-0.0830	0.02100	-3.96
Gre	-0.1184	0.03290	-3.60
Irl	-0.1655	0.05266	-3.14
$N10$	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
$P10 * N10$	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). $-2 \log lik = 19153$, $BIC = 19122$.

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 affiliations in a department corresponds to a scraping of social capital. We have here for science a correspondance 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% 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 5.3 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 losing their competitive edge as authors affiliated to second rank universities also manage to publish in top journals.

8 Ranking European economic departments

Mixed linear models are used for combining data collected at different levels and take advantage of particular correlation structures. In the above sections, we have detailed how fixed factors, at the individual level and at the department level, had a significant impact on individual research output. A random effect on the mean indicates how institutions deviate from

an average European score. Random effects have been used in the literature to rank institutions. Goldstein and Spiegelhalter (1996), Goldstein and Thomas (1996) are good examples of what can be done in this respect for secondary schools and other public institutions.

8.1 Rankings using multilevel models

The best way to understand how these models can be used for ranking is to start from the pure variance component model which includes only an overall constant term (the grand mean) and a random effect which considers random deviations from that grand mean at the institution level. When one wants to rank schools using student results at standardised tests such as those collected in PISA,¹¹ the best school is the one for which the average rate of success or the average grade is the highest. In the pure variance component model of individual scores y_{ij} ,

$$\begin{aligned} y_{ij} &= \beta_{0j} + v_{ij} \\ \beta_{0j} &= \beta_0 + u_j, \end{aligned} \tag{20}$$

u_j represents the gap between the grand mean β_0 and the local mean of school j , so that $\beta_0 + u_j$ can thus be directly used for ranking. Standard deviations are of course necessary to appraise the uncertainty attached to the ranking. In this model, the u_j are normally distributed with:

$$u_j \sim N \left(\frac{\sum_i (y_{ij} - \bar{y}_{.j})}{n_j + \sigma^2/\omega^2}, (n_j + \sigma^2/\omega^2)^{-1} \right) \tag{21}$$

where n_j is the number of scholars in school j .

In our case, the grand mean represents the author mean log score computed over the entire population and u_j represents the the gap between the grand mean and the mean log score of department j . In order to get the total log score of department j , we must consider an expression like

$$sct_j = (\beta_0 + u_j) \times n_j \tag{22}$$

where n_j represents a department size to be discussed later on. Departments are usually ranked according to their total score. Eventually according to the total number of authors. It was shown in Lubrano and Protopopescu (2004) that this corresponds to a notion related to stochastic dominance at the order one or two when a minimum level of activity is defined. Ranking departments according to the mean score of their members makes no sense.

¹¹OECD Programme for International Student Assessment.

The initial model was enriched by adding individual exogenous variables, leading to

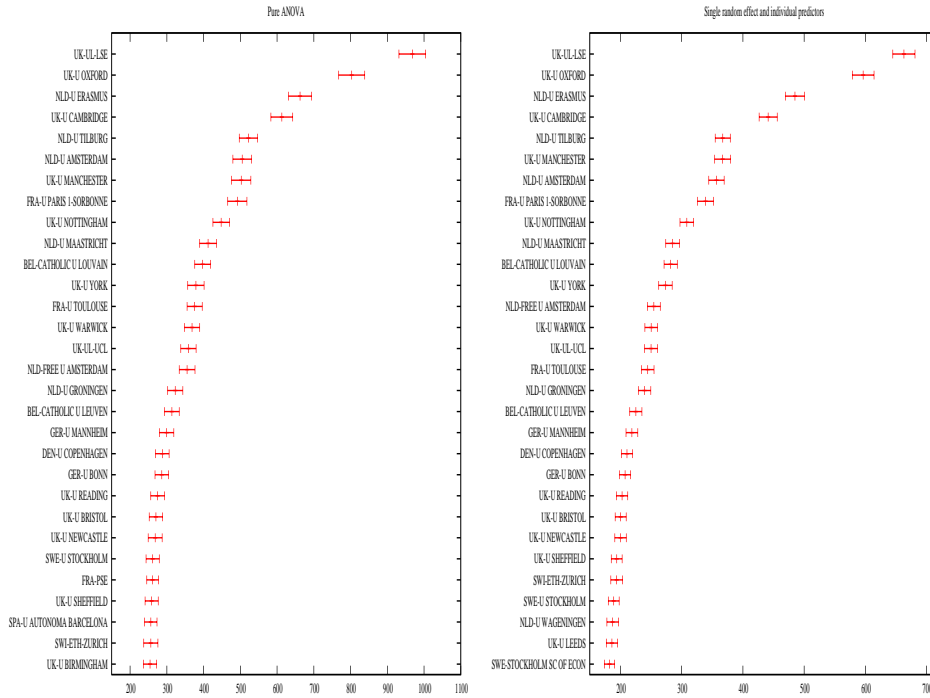
$$\begin{aligned} y_{ij} &= \beta_{0j} + x_{ij}\beta_1 + v_{ij} \\ \beta_{0j} &= \beta_0 + u_j. \end{aligned} \tag{23}$$

The ranking can be done here according to the same principle, except that this time, the u_j are distributed according to:

$$u_j \sim N\left(\frac{\sum_i(y_{ij} - x_{ij}\beta_1)}{n_j + \sigma^2/\omega^2}, (n_j + \sigma^2/\omega^2)^{-1}\right), \tag{24}$$

leading to smaller standard deviations than in the pure ANOVA model.

Figure 3: European ranking using a random intercept



In Figure 3, we give the ranking of the first thirty European departments, using these two methods. On the left panel, the ranking is obtained using the pure variance component model, while in the right panel we have added individual exogenous variables. There are some marginal variations between the two rankings, especially around the bottom. The ranking which makes use of exogenous variables provides smaller standard errors.

Confidence intervals show that the significant differences are at the top of the ranking, while most of the departments are equivalent when reaching

the bottom. With the pure ANOVA model, UK has 14 departments in this ranking, the Netherlands 5, France 3, Belgium 2, Germany 2, Denmark 1, Sweden 1, Spain 1 and Switzerland 1. We have given countries in the rank of their first appearance. Note the performance of the Netherlands that manage to have 40% of its departments in the top 30.

8.2 Size effects

In the two rankings given in Figure 3, there is a huge size effect which explains incidently why smaller departments with a high reputation like the European Institute of Florence do not appear favourably, while larger departments like Leeds, Manchester, Nottingham, Groningen or Wageningen have a surprising good position. The size effect appears because we have chosen $n_j = Nz_j$ in (22). We could perfectly use another size indicator based on for instance:

$$n_j = Nz_j^\alpha \times N10_j^{1-\alpha}. \quad (25)$$

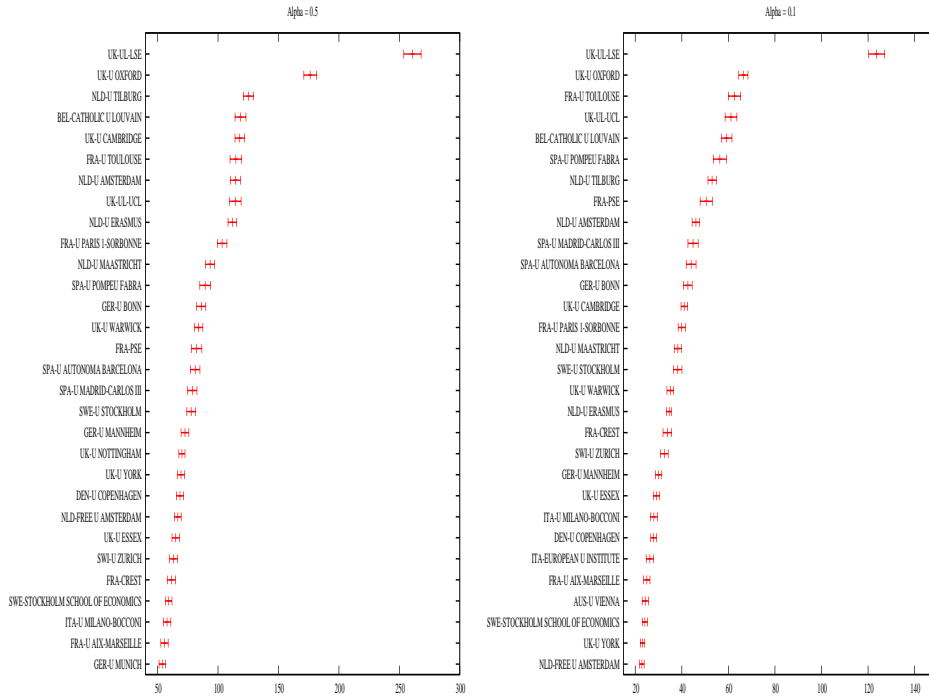
Nz_j represents the total number of productive authors in department j while $N10_j$ represents the number of top authors present in department j at the end of the period of observation. (25) entails a weight of zero if there is no top researcher in a department and then lowers the size effect as a function of α . With $\alpha = 1$, we recover the previous ranking based solely on Nz_j . α monitors the trade-off between size Nz and quality $N10$. We present in Figure 4 two alternative rankings: one with $\alpha = 0.50$ in the left panel, one with $\alpha = 0.10$ in the right panel.

Already with $\alpha = 0.50$, the resulting ranking is totally modified. Most of the large departments like Leeds, Manchester, Reading and Sheffield in the UK or Groningen and Wageningen in the Netherlands have left the top thirty group. However, we need $\alpha = 0.1$ for allowing the small and elitist European Institute to enter the list. With this last value for α , the UK loses some ground with 7, but keeps its first position, France comes next and has 5 ranked departments, Belgian 1, Spain climbs up in the ranking and has 3, the position of the Netherlands drops in the ranking and has 5, Germany 3, Sweden 2, Switzerland 1, Italy 2, Denmark 1 and Austria 1 enters the list.

8.3 Ranking and social capital

Ranking is a very crude ordinal operation that usually piles up weighted articles, independently of any type of explanations. We have tried in section 7 to find some of the determinants to scientific production and we have used one of our models to propose rankings with associated confidence intervals. We try now to answer the question concerning the link between rankings and social capital. We have measured the social capital of an institution with nine department variables. How are these variables related to our rankings

Figure 4: European ranking using a composite size



and when we vary α , what does it mean in terms of weights given to the different components of social capital. Table 9 reports the significant correlations between our different rankings and the nine social capital variables. All rankings seem to be independent of internal collaboration or multiple af-

Table 9: Ranking and social capital

	Nz	$N10$	$Nm10$	Nal	$Nsia$	$Nint$	Nei	Nea	$Nmul$
$\alpha = 1.0$	-0.99	-0.68	-0.70	-	-0.20	-0.20	-0.19	-	-
$\alpha = 0.5$	-0.91	-0.90	-0.80	-	-	-0.32	-0.25	-0.14	-
$\alpha = 0.1$	-0.77	-0.98	-0.78	-	-	-0.37	-0.27	-0.20	-
$\alpha = 0.0$	-0.68	-0.98	-0.77	-	-	-0.41	-0.27	-0.22	-

A negative correlation is associated to a better ranking. Correlations were computed for the 212 departments that had at least one top author, as the other departments could not be ranked whenever $\alpha < 1$. The non-parametric Spearman correlation is used to cope with any type of non-linearity.

filiations.¹² All rankings are highly correlated with past reputation ($Nm10$). A greater weight put on quality (a smaller α) corresponds to a greater influence given to international collaboration, a greater weight given to the mean length of affiliation (the ability of an institution to keep its members) and to a smaller extend a greater weight given to total experience.

9 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 data bases reveal a lot of information concerning authors habits, choices and networks. They are however mute on a tiny part of papers, we mean acknowledgments. Acknowledgements refer to conversations, discussions, information release, comments and improvements made on the paper. They relate to a participation to social capital, not necessarily the social capital of the authors' institution, but sometime yes. This fact was pointed out to

¹²However with a simple linear correlation coefficient, multiple affiliations seems to be inversely related to a better rank.

us by Tony Atkinson who alluded to some department members who were not productive in the usual meaning of this word, i.e. they did not write many papers, but who contributed very much to the work of others by their remarks and implications. These people are outside any type of statistical evaluation, but clearly belong to the social capital of the institution.

Appendix: A life-cycle model in continuous time

The proposed model in continuous time draws extensively on the papers by Diamond (1987) and McDowell (1982). The production function for supplementary human capital Q_t has the form of a Cobb-Douglas

$$Q_t = \beta(s_t K_t)^\alpha. \quad (26)$$

The stock of human capital has a rate of obsolescence δ so that

$$\dot{K}_t = Q_t - \delta K_t. \quad (27)$$

The objective function of the scientist is to maximise his discounted future income. Current income Y_t is given by

$$Y_t = w(1 - s_t)K_t, \quad (28)$$

and the present value at age t of disposable future income:

$$U = \int_t^T e^{-r(\tau-t)} Y(\tau) d\tau = \int_t^T e^{-r(\tau-t)} w(1 - s(\tau)) K(\tau) d\tau, \quad (29)$$

A solution to this problem is found by writing the Hamiltonian

$$H = e^{-rt} Y_t + \lambda(Q_t - \delta K_t). \quad (30)$$

Its solution expresses Q_t as a function of the parameters of the model and of the remaining time to retirement, $T - t$. Production or investment in human capital Q_t is a non-linear decreasing function of time with:

$$\log Q_t = \frac{1}{1-\alpha} \log \beta + \frac{\alpha}{1-\alpha} \log \frac{\alpha}{\delta+r} + \frac{\alpha}{1-\alpha} \log(1 - e^{-(\delta+r)(T-t)}). \quad (31)$$

The solution in Q_t is zero when $t = T$ (the age of retirement). The existence of the solution requires that $\alpha < 1$ (contrary to our discrete time model where $\alpha = 1$). When combining this equation with the capital variation equation, K_t has an inverted U shape life-cycle profile.

References

- Bauwens, L., Mion, G., and Thisse, J. (2007). The resistible decline of European science. CORE-DP 2007-92, CORE, Université catholique de Louvain.
- Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *The Journal of Political Economy*, 75(4):352–365.
- Bourdieu, P. (1980). Le capital social: Notes provisoires. *Actes de la Recherche en Sciences Sociales*, 3(2-3).
- Bourdieu, P. (1986). The forms of social capital. In Richardson, J. G., editor, *Handbook of theory and research for the sociology of education*. Greenwood, New York.
- 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*, 22(7-8):716–740.
- Coleman, J. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94:S95–S120.
- Coleman, J. (1990). *Foundation of Social Theory*. Harvard University Press, Cambridge MA and London.
- Diamond, A. M. (1984). An economic model of the life-cycle research productivity of scientists. *Scientometrics*, 6(3):189–196.
- Diamond, A. M. (1987). An optimal control model of the life-cycle research productivity of scientists. *Scientometrics*, 11(3-4):251–253.
- Diamond, A. M. (1996). The economics of science. *Knowledge, Technology and Policy*, 9(2-3):6–49.
- 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*, 159(3):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)*, 159(1):149–163.
- Heckman, J. J., Lochner, L. J., and Todd, P. E. (2003). Fifty years of Mincer earnings regressions. Working paper, University of Chicago.
- Henrekson, M. and Waldenström, D. (2008). How should research performance be measured? *The Manchester School*. Forthcoming.

- Kim, E. H., Morse, A., and Zingales, L. (2009). Are elite universities losing their competitive edge? *Journal of Financial Economics*, 93(3):353–381.
- Levin, S. G. and Stephan, P. E. (1991). Research productivity over the life cycle: Evidence for academic scientists. *The American Economic Review*, 81(1):114–132.
- Lindley, D. V. and Smith, A. F. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B*, 34(1):1–41.
- Lotka, A. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16:317–323.
- Lubrano, M., Bauwens, L., Kirman, A., and Protopopescu, C. (2003). Ranking European economics departments: a statistical approach. *Journal of the European Economic Association*, 1(6):1367–1401.
- Lubrano, M. and Protopopescu, C. (2004). Density inference for ranking European research systems in the field of economics. *Journal of Econometrics*, 123(2):345–369.
- McDowell, J. M. (1982). Obsolescence of knowledge and career publication profiles: Some evidence of differences among fields in costs of interrupted careers. *The American Economic Review*, 72(4):752–768.
- Merton, R. K. (1968). The Matthew effect in science. *Science*, 159(3810):56–63.
- Putnam, R. D. (1995). Bowling alone: America’s declining social capital. *Journal of Democracy*, 6(1):64–78.
- Rauber, M. and Ursprung, H. (2008). Life cycle and cohort productivity in economic research: The case of Germany. *German Economic Review*, 9(11):431–456.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications Ltd, London, 2nd edition.
- Siisiäinen, M. (2000). Two concepts of social capital: Bourdieu vs. Putnam. Technical report, University of Jyväskylä, Department of Social Sciences and Philosophy. Paper presented at ISTR Fourth International Conference: The Third Sector: For What and for Whom? Trinity College, Dublin, Ireland, July 5-8, 2000.
- Stephan, P. (1996). The economics of science. *Journal of Economic Literature*, 34(3):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*, 86(413):79–86.