UMR 225 IRD - Paris-Dauphine

DOCUMENT DE TRAVAIL DT/2016-02

Co-Authorship And Research Productivity In Economics: Assessing The Assortative Matching Hypothesis

Damien BESANCENOT,
Kim HUYNH,
Francisco SERRANITO

UMR DIAL 225
Place du Maréchal de Lattre de Tassigny 75775 • Paris •Tél. (33) 01 44 05 45 42 • Fax (33) 01 44 05 45 45
• 4, rue d'Enghien • 75010 Paris • Tél. (33) 01 53 24 14 50 • Fax (33) 01 53 24 14 51
E-mail : dial@dial.prd.fr • Site : www.dial.ird.fr

CO-AUTHORSHIP AND RESEARCH PRODUCTIVITY IN ECONOMICS: ASSESSING THE ASSORTATIVE MATCHING HYPOTHESIS

Damien Besancenot, CEPN and University Paris 13, Sorbonne Paris Cité;

Kim Huynh, LEM and UniversityPanthéon-Assas, Paris 2;

Francisco Serranito[1], Univ. Orléans, CNRS, LEO, UMR 7322 and IRD, LEDa, DIAL UMR 225.

Abstract

This paper estimates the relation between the size and quality of scientists' co-author networks and individual characteristics (notably productivity) in the context of institutional changes in French academia in the mid-1980s. The analysis employs the Two-Stage Residual Inclusion (2SRI) framework to handle endogeneity in individual productivity relative to the quality of co-authors. Data is taken from a novel database of French academic economists. The main finding is that the size and quality of authors' networks are positively related to their productivity; this is understood as evidence of assortative matching. Other effects on co-author networks (life-cycles, specialities) are also identified.

Keywords: Co-authorship, Count Data, Zero Inflate Models, Instrumental Variables, h index.
JEL: A14, C25, D83, I23, J24.1

[1]Corresponding author: francisco.serranito@univ-orleans.fr; Université d'Orléans, Collegium DEG, rue de Blois BP 26739, 45067 Orléans Cedex 2, France.

1 Introduction

Beginning at the end of the 1970s, a vast strand of research has highlighted that co-authorship is no longer the exception but constitutes a new scientific norm (Beaver and Rosen 1978; Stefaniak 1982; Zitt et al. 2000, Laband and Tollison 2000, Hamermesh 2013 and 2015). Research policy often makes collaboration a major goal in order to boost the production of knowledge. Policymakers encourage collaborative research, as most studies find that co-authorship improves individual academic productivity (Bidault and Hildegrand, 2014 or Levitt, 2015). Similarly, understanding the drivers of the growth in co-authorship is an important issue for national research organizations whose goal is to draw up research policy recommendations.

While public policies require researchers to increase their scientific production, the rising trend of co-authorship is better explained by the positive effects of collaboration on both the quantity and quality of research output. With respect to quantity, co-authorship is a simple way to increase the number of papers that a researcher is able to publish in a given period of time. For example, Laband and Tolisson (2000) and Ursprung and Zimmer (2007) document that co-authorship increases acceptance rates by refereed journals; and Durden and Perry (1995) and Lee and Bozeman (2005) find that the total number of publications is significantly and positively related to the number of collaborative publications. Moreover, as two-authored papers are generally worth more than a single-author paper, the incentives to collaborate are strong (Barnett et al., 1988). Empirical evidence concerning the link between research quality and co-authorship is less conclusive, particularly if quality is measured by the number of citations (see Levitt 2015 for a review). On the one hand, Laband (1987) and Johnson (1997), and more recently Levitt (2015) report that co-authored papers are cited more often than single-author ones. On the other hand, Barnett et al. (1988) and Hollis (2001) find no effect.

If the theoretical motivation for scientific collaboration, and more specifically the writing of papers, is now well understood there have been few empirical studies on the matching aspects of co-authorship. In a pioneering paper, McDowell and Melvin (1983) linked the rise in co-authorship to increased scientists' specialization. Researchers expand their networks by finding colleagues with complementary skills, which proves to be an efficient way to increase production. According to Piette and Ross (1992) authors who work in areas outside of their specialty tend to participate more in co-authorship than others. More recently, Chan et al. (2015) focused on Nobel Laureates and showed that scientific collaboration is fostered by conceptual complementarities, which erode over time following repeated interactions. However, if complementarities may lead to scientific collaboration, substitutability may also be at its core (Barnett et al. 1988, Medoff 2007).

In these settings, our work follows Fafchamps et al. (2010), who highlight the matching problem of finding a co-author (see also Hamermesh 2015). They argue that when research output depends on ability, collaboration is most likely between authors with similar abilities – the assortative matching hypothesis. Collaboration between authors with different abilities can only happen if the contribution of the lower-ability author relaxes the time-constraints of his/her co-authors. In this case, collaboration enables higher-ability authors to produce more research, while lower-ability researchers produce better-quality output than they would otherwise. Recently, Bidault and Hildebrand (2014) assessed the determinants of asymmetric co-author teams. They show that while co-authorship is less favorable for the "senior" co-author in the short term, in the long term each author benefits from collaborating.

Our paper contributes to this literature by assessing the link between the "individual research quality" of an academic and that of his/her co-authors. More precisely, we estimate the empirical determinants of co-authorship based on a novel database of all academic economists working in a French university in 2004. The database includes both academics who are (or have been) active in publishing their work, and others. Bibliometric studies do not usually include this second category. Therefore our paper, which uses a bespoke econometric framework, is likely to produce more reliable estimates. However, individual research productivity should be considered as an endogenous regressor, as the quality of an individual researcher's output is more-or-less dependent on the quality of his/her co-authors. We draw upon the Two-Stage Residual Inclusion (2SRI) framework to take into account endogeneity in individual productivity relative to that of co-authors. To the best of our knowledge, our paper is the first to apply such a framework to co-authorship.

Our paper is also related to the literature concerning the impact of national research policy changes on individual behavior. As in most industrialized countries, French academia has undergone profound transformation in the past twenty years.[2] Bibliometrics have been introduced to firstly measure individual productivity and secondly to provide incentives (Académie des Sciences, 2011). The French system is progressively adopting Anglo-Saxon standards; fostering competition between academics in order to fund their research, and rewarding highly-ranked publications. We scrutinize the impact of these institutional changes on individual research productivity by modeling different cohorts of economists. While our results are based on a sample of French academics, they are very similar to those obtained in Germany (Rauber and Ursprung, 2008) or Italy (Cainelli et al., 2012). Hence the findings can be interpreted as the response of continental European science systems to structural and institutional changes.

The paper is organized as follows. The next section describes the empirical methodology. Section 3 describes the database; section 4 presents the results; section 5 is a sensitivity analysis; and finally section 6 discusses the policy implications of our results.

2- Instrumental variables with count data: the 2SRI methodology

Measures of individual research productivity, such as the size or the quality of co-authors' networks, are count data; therefore this paper applies count data econometrics. However, count data suffer from two major drawbacks: overdispersion and excess of zeros. In this case, Zero Inflate models (zero inflated Poisson: ZIP or zero-inflated negative binomial: ZINB) should be applied (see Besancenot et al. 2015 for a detailed presentation). In addition, individual productivity is not independent of that of co-authors. This endogeneity must be addressed by Instrumental Variables (IV) methods.

In non-linear models, IV methods correspond to the Two-stage Predictor Substitution (2SPS) approach. 2SPS substitutes endogenous regressors in the estimated equation with their consistent predicted values obtained in a first-stage auxiliary regression. However, when the conditional expectation model is non-linear, the 2SPS approach tends to produce inconsistent estimates. In this case, Wooldrige (2014) advocates applying the Two Stage Residual Inclusion (2SRI) approach, which provides consistent estimates of the parameters in the structural regression. The 2SRI estimator has the same first stage as the 2SPS. However, in the second stage, endogenous regressors are not replaced. Instead, first-stage residuals of the auxiliary regressions are included as additional regressors in the second-stage estimation.

Recently, Geraci et al. (2014) extended the 2SRI framework to count data models. They consider the following general, non-linear model for the conditional mean of the outcome:

$$(2.1) \quad E\left(Q_{i,co-authors} / x_i, x_{ei}, w_i\right) = M(x_i\beta + x_{ei}\beta_e + w_i\lambda) = M(x_i\beta + \textstyle\sum_{s=1}^{S} \gamma_s\, x_{sei} + \sum_{s=1}^{S} \xi_s\, w_{si})$$

[2] Appendix 1 gives an overview of the French academic system and recent changes.

Where $Q_{i,co-authors}$ is a measure of the quality of co-authors, *M(.)* is a known non-linear function and X_{2i} regressors can be split into two components: $X_{2i} = [x_i \quad x_{ei}]$ where x_i is a set of K exogenous regressors, and x_{ei} which is a set of S endogenous regressors (either discrete or continuous) possibly correlated with the set of S unobservable confounder latent (or omitted) variables w_i. The endogeneity of regressors x_{ei} is modelled by the correlation between the unobserved confounder factors with x_{ei} and $Q_{i,co-authors}$ (Terza *et al.*, 2008):

(2.2) $x_{eis} = r_s(v_i \xi_s) + w_{si}$ s=1, ..., S

Where $v_i = [x_i \quad z_i]$: z_i is a set of at least S instrumental variables satisfying all the necessary conditions, and $r_s(.)$ is a set of S non-linear auxiliary equations.

The 2SRI estimator is then obtained by estimating the following regression:

(2.3) $E\left(Q_{i,co-authors}/x_i, x_{ei}, \hat{u}_i\right) = M(x_i\beta + x_{ei}\beta_e + \hat{u}_i\psi)$

Where $\hat{u}_i$ is a set of S estimated first-stage residuals for individual *i*. Consistent standard errors of second-stage parameters can be obtained by bootstrapping (Wooldrige, 2014).

In count data models, there is no consensus on how to define residuals. Geraci et al. (2014) advocate computing two measures: the raw residual ($\hat{u}_{is} = x_{eis} - E[x_{eis}/w_i]$) and the standardized residual ($\hat{u}_{is}^{std} = \frac{x_{eis} - E[x_{eis}/w_i]}{(V[x_{eis}/w_i])^{1/2}}$). If x_{ei} are count data variables, then the first-stage auxiliary regression can be modelled by a zero inflate model. The exogeneity of x_{ei} can be tested via a conventional Wald-type statistic for $H_0: \psi_1 = \psi_2 = \cdots = \psi_S = 0$.

The 2SRI method has good finite sample properties (Geraci et al. 2014). Empirical evidence shows that the power of the exogeneity test is always higher using standardized residuals. Furthermore, applying standardized residuals leads to smaller bias in endogenous regressors.

3 Data and Sample Characteristics

3.1 The Database

Our database contains data from a variety of sources. We used the *Tableau de classement du personnel enseignant titulaire et stagiaire* - Economics Section - published by the French Ministry for Research. This document lists all academics employed by French Universities on December 31 2004 and provides information about the gender, age and academic status of individual researchers.

In early 2012, we used “Publish or Perish” (PoP, Harzing 2010) software to collect researchers’ CVs from Google Scholar[3]. For each paper listed on an individual CV, PoP provided the number of citations, publishing medium and names of co-authors. In a second stage, we used PoP to get the ‘h’ and ‘g’ indexes of co-authors.

The raw Google scholar data had significant shortcomings. In some cases, CVs of well-known researchers did not list their publications. Authors’ names were often incomplete, incorrect or misspelled. Authors with lastnames that are also firstnames raised the difficult problems of disambiguation: “Philippe Martin” could be listed as Philippe M. or Martin P. Authors with popular French lastnames (such as Petit) were credited with papers from homonymous researchers. Publications by married women who have used different lastnames during their career are often underestimated. To overcome these difficulties, we excluded data that could not be disambiguated. From an initial total of 1830 names in the *Tableau de classement*, we retained 1566 researchers[4].

[3] At the time it was possible to select papers according to specific subject areas. Our data therefore includes all papers classified in the “Business, Administration, Finance, Economics” and “Social Sciences, Arts, Humanities” domains.

[4] Consequently, some of the most productive researchers might be excluded.

The dataset was supplemented by Journal of Economic Literature (JEL) codes for the papers included in our database and listed in Econlit. Finally, for doctoral students, we identified the name of their supervisor and the year they defended their thesis through individual searches.

Individual Productivity Indexes

From this information, we computed three different productivity indexes for each researcher. The h and g indexes provided a synthetic measure of both the quantity (number of papers) and quality (number of citations) of a researcher's production. The h index is defined as equal to *x* if *x* of his/her *N* papers have received at least *x* citations each, and the other (*N*−*x*) papers have received no more than *x* citations each (Hirsh, 2005). One drawback of this measure is that different academics may have similar h indexes even if their respective 'best' papers have a very different number of citations. In order to address this limit, Egghe (2006) proposed the g index as the (unique) largest number such that the top g articles received (all together) at least g^2 citations.

Given the controversy over h and g indexes (see for instance Bornmann and Daniel 2007), we developed a third measure, "CL_index", based on the quality of the journal in which papers were published. This variable will be used in robustness checks. It is developed according to the following formula:

$$\mathrm{CL\_index} = \sum_{i=1}^{n} \frac{W_i}{\sqrt{a_i}},$$

where $\frac{W_i}{\sqrt{a_i}}$ is the score of paper *i* defined as the ratio between a weight W_i associated to its medium of publication and the square root of the number a_i of its authors. CL_index only considers papers published in Econlit Journals and W_i was taken from the Combes and Linnemer (2010) ranking[5]. Here, n is the number of papers published by a given author in an Econlit journal. As the CL_index neglects papers published in books, working papers and non-Econlit journals it is an elitist measure of productivity.

Co-author variables

To summarize both the number and productivity of a researcher's co-authors, we computed two Meta indexes (hh and gg) corresponding to h and g indexes. For an academic, the hh index is defined as equal to *n* if *n* of his/her co-authors have an h index at least equal to *n*, and the other co-authors have a h index less than *n*. In the same way, the gg index is equal to *n* if the sum of the g indexes of his/her *n* best co-authors is greater than or equal to n^2 (the square of the rank) and the sum of the g indexes of the n+1 best co-authors is less than $(n+1)^2$. These two indexes aim to provide a simple measure of both the number and the quality of a researcher's co-authors.

The number of co-authors (NB_COAUTHORS) is considered as a proxy of the size of an academic's network. As this variable only measures quantity it contributes to the analysis of the trade-off between quality and quantity when choosing co-authors.

Control variables

For each author the following control variables were computed:

- FEMALE is a dummy variable equal to 1 if the academic is a woman.
- AGE is the age of the individual. If researchers are investment-motivated, then a decline in productivity should be observed over the course of their career. Indeed if the motivation to engage in research depends on the present value of future financial benefits, then as the scientist ages this return will decrease (Levin and Stephan, 2001).

5 Combes and Linnemer define two scores (Clm and Clh) for each of the 1205 journals listed in Econlit. Both scores reflect a journal ranking, but Clh gives a higher weight to top-tier journals and a lower weight to others. Initially we used both weightings to compute productivity indexes. However, the index built on CLh scores was too selective to provide conclusive results, therefore the CL_index used in this text only reflects Clm weights.

- NUMBER_YEARS is the number of years post-PhD defense. This variable is used as a proxy for professional experience.
- "COHORTYY_yy" is a dummy variable equal to 1 if the individual defended his/her PhD thesis between 19YY and 19yy. It controls for cohort or vintage effects. Cohort effects imply that individual productivity is linked to macroeconomic aggregate shocks.
- Academic position: In France, there are three ranks for full Professors (PR): *Classe-Exceptionnelle* (PR_CE), *Première-Classe* (PR_1C) and *Seconde-Classe* (PR_2C); and two for Assistant Professors (MCF): *Hors-Classe* (MCF_HC) and *Classe-Nornale* (MCF_CN). These variables reflect the quality of an academic, as promotion is largely dependent on the number and quality of their publications.
- Publication language: The percentage of papers published in journals by an individual researcher in English (SHARE_GB), French (SHARE_FR) or other languages (SHARE_OTHER).
- WORK_ALONE_ONLY: This dummy variable is equal to 1 if an academic has published at least one paper since the beginning of his/her academic career and has never co-authored a paper.
- NB_PAPERS: The number of papers listed in Google Scholar for an individual researcher. This variable is a quantitative measure of academic production.
- Article Quality: In France the CNRS ranking of economics journals is the key reference for the assessment of economic research (see Appendix 1). Journals are ranked from 1 (top-tier journals) to 4 (less-influential journals). Recently, two more categories were added to take account of multidisciplinary or promising new journals. Based on this classification, seven variables were developed: NB_PAPERS_CNRS1 to NB_PAPERS_CNRS4 indicate the number of papers published in the four main categories of journals. NB_ PAPERS _CNRS_5 represents papers published in multidisciplinary and promising journals. ECONLIT_NO_CNRS records publications in journals listed by Econlit but with no CNRS classification, Finally MISCELLANEOUS_PAPERS counts all other items.
- Topic: Co-authorship varies greatly according to the economic topic. Following a now standard methodology (Fafchamps et al. 2010 or Bosquet and Combes 2013), we identified topics through the letters of their JEL Classification codes. The topic with the most JEL codes was denoted as the principal field of research (Max_kwX variables). We also computed a normalized Herfindahl index (Herfindahl_JEL_code variable) from the different codes used by a researcher in order to measure degree of specialization. By construction this variable ranges from 0 (no specialization) to 1 (full specialization).
- COWRITE_DR is a dummy variable equal to one if the academic has written at least one paper with his/her supervisor.
- Network effects: The PhD_DEFENDED_AT variable divides the set of authors into 11 categories according to where they defended their thesis.[6]

Choice of Instrument

[6] The overall dataset was divided into eleven institutions or group of institutions : University Paris 1, Paris 9, Paris 10, other Paris universities, Aix-Marseille, Strasbourg, Toulouse 1, the *Grandes Ecoles,* other French universities, European universities, and American universities.

It is well-known that the efficiency of IV estimators relies on the quality of the instruments. Unobserved heterogeneity can lead to endogeneity (potentially due to unobserved individual characteristics), which imply that the dependent variable is correlated with one or more regressors. In our case, a reliable instrument requires two assumptions: (i) it is highly correlated with individual research productivity; and (ii) it is uncorrelated with co-author quality. Obtaining such an instrument is challenging as many potential variables may explain both individual and co-author quality.

The instrument we use in this paper is the best-quality paper published alone by an academic, which is assumed to reflect their intrinsic skill level. This measure is based on the CNRS classification of journals (see above). For example, BEST_ALONE_CNRS1 is a dummy variable equal to 1 if the researcher's best article as a sole author was published in a CNRS 1 category journal.

At this stage a caveat is called for. Even if the researcher's network has no direct influence, it is important to note that co-authorship implies a process by which a researcher improves his/her skills and learns how to publish better papers. We thus acknowledge that a certain level of endogeneity cannot be ruled out in our estimations. However, we consider that this learning effect is weak compared to the intrinsic skill level measured by our instrument. Thus the bias, if any, should be low.

3.2 Descriptive Statistics

Table 1 reports descriptive statistics. In 2004, 28% of French academic economists were women; 35% were Full Professors and the average age was 47. In 2012 the average French academic had around 22 years of professional experience and had published an overall total of around 8 papers, rising to 11 in the sub-sample of publishing academics. There is a huge degree of heterogeneity in production as the number of papers ranges from 0 to 157. Mean h and g indexes are respectively 3.25 and 6.02. Again, there is huge heterogeneity in "quality": h and g indexes range from 0 to 39 and 0 to 84 respectively.

In 2012, 22.4% of French academics had no paper referenced by Google Scholar. Among publishing academics, only 1.98% had never published in a journal referenced by the CNRS or Econlit, and 85.2% of papers were published in journals listed by the CNRS. Authors prefer to publish in CNRS-ranked journals as their career progression depends on this classification. Despite this bias, mean paper quality is low: around 79% of the total number of published papers was categorized as CNRS3 and CNRS4 (the least-influential categories), compared to CNRS2 (13%) and CNRS1 (7%). This result is partly explained by the fact that the majority (62.5%) of papers are written in French, and Francophone journals tend to be lower ranked. However, there is a generational effect of English proficiency. Academics that started their career pre-1968 drafted more than 75% (respectively 17%) of papers written in French (English) compared to 52.9% (45.1% respectively) for those that started in 1999 (see table 2).

On average, each French academic has 4.5 co-authors (for the whole sample) and 6.9 for the sub-sample engaged in co-authorship (see table 1). Overdispersion is important in this case as the variance is equal to twelve times the mean. It is worth noting that 34.8% of the individuals in the sample have no record of collaboration, while 17.4% have never written a paper alone. Among publishing academics, 44% have between one and three co-authors; and about 15% of individuals had written at least one paper with their PhD supervisor. The mean of the hh Meta index (which summarizes both the number and quality of co-authors) is 3.2 and the hh index ranges from 0 to 29. Here again, there is overdispersion as the hh index variance is 11.7. A similar result is obtained with the gg index, while variation is even greater: the mean and variance are equal respectively to 7.4 and 74.1 (see table 1).

With respect to specialization, overall it is quite low. Although the mean for Herfindahl_JEL_Code is 0.31, there is a high level of heterogeneity (see table 1). Few academics have only published in one field (8.2%).

The dramatic institutional changes observed in France over this period seem to have created clear life-cycles in research productivity (see table 2). We computed individual productivity (h, g, and CL_index) and co-authorship (hh and gg) indexes per years of experience by entry cohort. All variables highlight a similar structural break in the mid-1980s. For instance, the number of co-authors per year is stable for researchers belonging to the first four cohorts, but steadily increases for individuals who begin their academic career later (the mid-1980s; cohort84_88 and following). Individual productivity indexes follow a similar pattern. Younger cohorts are more productive and seem to participate more in co-authorship.

Descriptive statistics related to the instrument variables are reported at the bottom of table 1. The number of sole authors decreases with the quality of the journal. Around 6% of sole authors published their best paper in a CNRS1-category journal; this compares to 9.5% (CNRS2), 27% (CNRS3) and 11% (CNRS4).

4 Discussion of Empirical Results

This section applies the 2SRI methodology to co-authorship. In a first step, individual productivity is modeled with exogenous regressors and instrumental variables in order to compute standardized residuals. In a second step, the determinants of the size and quality of co-author networks are estimated including standardized residuals as additional explanatory variables in the regression.

First-stage empirical results: exogenous determinants of productivity measures

In the first step, the main drivers of individual productivity indicators are estimated based on exogenous determinants (i.e. funding or academic position are excluded). Therefore, the first-stage regression only controls for exogenous demographic variables (e.g. age, gender), individual talent, institutional determinants and the researcher's main discipline (max_kwX JEL codes). Individual talent is non-observable and is proxied by instrumental variables. Finally, the academic institution where the PhD was defended is included in order to control for the institutional environment. Results are reported in table 3. Overdispersion and excess of zeros are taken into account through a ZINB model applied to both the h and g indexes (as recommended by Vuong and likelihood ratio tests – see the bottom of table 3, columns 1 and 2).

Productivity is the result of two decisions. First, the researcher has to decide whether they will undertake a research activity, and then the extent of their contribution. Regarding the first decision, inflate coefficients of the two productivity measures suggest similar conclusions. Gender has a significant effect, while age does not (see table 3, part inflate logit model, columns 1 and 2).

Following the introduction of dummy variables to model cohort effects (see below for further details), only the last two cohort variables seem to have a significant effect on the decision to publish. Being an academic who defended his/her thesis post-1994 decreases the log odds of an inflated zero.

We turn now to the results for the parent model based on citation scores (the h and g indexes). The number of years of professional experience (NUMBER_YEARS) is the offset variable as it is different for each researcher. Support for the idea that productivity declines over time is mixed: although AGE coefficients are negative (IRR < 1) they are only significant for the h index specification. Working alone (WORK_ALONE_ONLY) decreases productivity: failure to co-author reduces the h index by 58.4%[7]. Finally, GENDER does not have a significant impact on individual productivity.

There is strong evidence of vintage effects. Six out of seven cohort dummies are significant for all productivity measures[8]. Incidence Rate Ratio (IRR) estimates of cohort dummies increase over time, implying that coefficients decrease over time (see table 3, Part ZINB, columns 1 and 2). For example, the h index of an academic who defended his/her thesis between the years 1969–1973 is 30% (1−0.701) lower than the reference group. The gap falls to 13% when the defense took place in 1989–1993[9]. According to Levin and Stephan (2001), such effects may be the consequence of hiring better researchers and thus only reflect the state of the job market at the time the thesis defense took place in the US. However, they may also be interpreted as the consequence of structural changes in Western Europe that incentivized productivity and led to higher output (Rauber and Ursprung 2008, Cainelli et al. 2012).

Most importantly, our instrumental variables are significant and their coefficient have the expected signs. BEST_ALONE publication coefficients increase as a function of journal quality for all productivity measures. For example, in the case of the h index, estimated IRR coefficients for BEST_ALONE_CNRS1 and BEST_ALONE_CNRS4 are respectively equal to 3.094 and 1.315 (see table 3, Part ZINB, columns 1 and 2) meaning that an individual who was the sole author of an article published in a CNRS category 1 journal (best quality) has an average 209.4% increase in expected productivity. This can be compared to a 31.5% increase if it is in a CNRS category 4 journal (lowest quality)[10].

Second-stage empirical results: 2SRI estimates of the determinants of co-authorship

The second stage of the 2SRI methodology considers the empirical determinants of co-authorship. Here we control for exogenous demographic variables (gender and age), individual talent, academic position, life cycles dummies, number and quality of publications, specialities, publication language, and network. Network size is measured by the number of co-authors. Furthermore, as hh and gg indexes provide a measure of both the quality and size of the co-author network, controlling for the number of co-authors in the hh and gg specification makes it possible to explicitly model the quality of co-author networks. Results are reported in table 4.

In every case, standardized residual variables are significant. Furthermore, the four Wald tests always reject at the 1% level the null hypothesis of exogeneity of the individual research productivity variables (see the bottom part of table 4, columns 1 to 4)[11]. Therefore, the 2SRI methodology must be implemented to adress endogeneity.

Regarding the determinants of the decision to collaborate, the inflate coefficient of the individual research productivity variable (either h or g) has a consistently negative and significant effect (see inflate model in table 4 columns 1 to 4). For example, in the hh specification, the inflate coefficient of h suggests that for each unit increase in h, the log odds of an inflated zero decrease by 2.82 when the language variable is included and 2.14 if it is not (table 4, inflate model columns 1 and 2)[12]. We find similar results for NB_COAUTHORS specifications (table 4, inflate model columns 5 and 6): higher productivity is consistent with increased interest in collaboration.

[7] IRR is equal to 0.416 (see table 3, part ZINB, column 1), thus the expected change in the h index is 0.584 (1−0.416).

[8] The reference variable is academics who defended their thesis post-1999.

[9] Rauber and Ursprung (2008) found similar results in Germany. In both cases, publication incentives contribute to the increase in output.

[10] The reference is no published journal papers.

[11] For instance, the Wald statistic for hh is equal to 35.11 without the language dummy variable and 21.19 with it (see the bottom part of table 5, columns 1 and 2). In both cases, the null hypothesis of exogeneity is rejected (significance levels are equal to 0%)

[12] For gg, estimated coefficients are respectively −1.34 and −0.58 (table 5, inflate model columns 3 and 4).

We turn now to results for the parent models of quality and size of co-author networks. Here again, professional experience (NUMBER_YEARS) is the offset variable. In terms of quality, the h (respectively the g) index has a positive and significant effect at the 1% level effect on the hh (respectively the gg) index. The IRR estimate for the h index is around 1.04 (table 4, parent model, columns 1 and 2), meaning that an increase of 1% increases the quality of co-author networks by 4%[13]. In the case of the NB_COAUTHORS specification, both h and g indexes have a negative impact on the size of the co-author network (both IRR estimates are below one), but both variables are only significant at the 10% level: higher-quality scientists tend to work with fewer co-authors. Thus individual productivity only seems to improve the quality of co-author networks. All in all, these results confirm the assortative matching hypothesis, namely that the quality of co-author networks is a function of individual productivity.

AGE has a significant and negative impact on both the quality and size of the co-author network at the 1% level[14]. The mean of IRR coefficient estimates is around 0.99: each year of age is consistent with a decrease of 1% ($= -ln(0.99)$) in both the size and quality of co-author networks. Being a woman has no effect on quality, but it does reduce the size of the network. As expected, academic position has a positive and significant effect on co-authorship: being a Full Professor increases both the quality and size of the network.

Co-authorship with the PhD supervisor (COWRITE_DR) has a consistent, positive and significant effect (at the 1% level) on both the quality and size of the co-author network. Co-writing may be seen as implicit recognition by a supervisor of the student's quality, and this may create a reputation effect. This confirms Bidault and Hildebrandt (2014)'s finding that a junior researcher benefits from joining an asymmetric team.

Like individual productivity, there are significant cohort effects in the composition of the co-author network. In all cases, IRR coefficients of cohort dummies are significant and increase over time. Similar effects are found for both the quality and size of the co-author network. For example, an academic who defended his/her thesis in the period 1964–1968 has an expected average hh index (respectively an expected number of co-authors) that is 68% (resp. 62%) lower than the reference group (thesis defended 1999–2004). These figures are respectively 30% and 22% for the 1994–1998 cohort group (table 4, parent model columns 1 and 5). Younger generations of economists are collaborating with more and better co-authors.

Interesting results are obtained regarding the link between the number of papers, their quality, and the decision to co-author. The IRR coefficient of the number of papers published in top-quality journals (CNRS1) is significant but always below one for all measures of co-author quality (hh or gg indexes). This implies that co-author quality decreases with an increase in the number of CNRS1 papers. However, the IRR coefficient is significant (in three out of four cases) and above one (in all cases) for CNRS category 2 papers, implying an increase in co-author quality (table 4, parent model columns 1 and 4). In general, the number of papers published in other CNRS categories (CNRS3, CNRS4, and CNRS5) has a non-significant effect. For the number of co-authors, variables measuring the number of papers published in CNRS categories 1 and 2 are non-significant (table 4, parent model columns 5 and 6). On the other hand, papers published in other CNRS categories significantly increase the number of co-authors (mainly CNRS4, CNRS5 and Econlit journals).

Therefore, while scientists who have mostly published in CNRS category 1 journals tend to have lower-quality networks (after controlling for their own quality, which has a positive effect on co-author network quality), their networks are not statistically significantly smaller. Economists with more papers published in CNRS category 2 journals tend to have higher-quality networks of similar size. This may be the result of a strategic approach to collaboration. Collaborating with better-quality co-authors could be an efficient strategy for

[13] In the case of the gg index, estimates are lower (around 1%) but they remain significant (columns 3 and 4).

[14] We also introduced the age squared variable, but all results were non-significant.

publication in CNRS2 journals. On the other hand, academics who publish in top-ranked (CNRS1) journals may find it difficult to find similar high-quality co-authors, but may agree to work with lower-quality co-authors in order to publish more papers. These results partly contradict those of Fafchamps et al. (2010) who argue that collaboration is more likely between authors who differ in the number and quality of their publications. It is possible that this is only true for top-tier researchers and not the majority.

Publication language also plays an important role in the co-authorship decision. The IRR coefficient of SHARE_GB (papers written in English) is above one, and highly significant in all cases. For example, a 1% increase in papers written in English increases the expected hh index (respectively the gg index) by 32.5% (resp. 38.4%) and the number of co-authors by around 73% (table 4, parent model columns 2, 4 and 6). Two arguments may explain this relationship. Firstly, French academic economists may be interested in international collaboration in order to produce papers in better English and publish in more influential journals (see Olney 2015 for the influence of English proficiency on research performance). Secondly, English proficiency is a necessary condition to meet high-quality co-authors and collaborate with more, better-quality authors. In both cases, publishing in English increases the number and quality of co-authors.

Co-authorship is also linked to economic specialities. Almost all JEL dummy variables are significant in all cases. A high degree of specialization reduces the number and quality of co-authors: a 1% increase in the JEL Herfindahl index decreases the expected hh index (respectively the gg index) by around 32% (resp. by 25%) and the number of co-authors by 57% (table 4, parent model columns 2, 4 and 6).

Finally, the effect of the dummy variable PhD_DEFENDED_AT (which controls for an institutional network effect) is rather weak. With respect to the quality of the co-author network, there is no difference between French universities[15]. As for network size, the number of co-author is only statistically significantly lower for Paris 10 and the group of "other non-Parisian universities". The great divide in the French higher education system between universities and the *Grandes Ecoles* also has an impact on co-authorship. IRR estimates of the *Grande Ecole* dummy variable are all significant and above one in all cases. Doctoral students who defend their thesis at a *Grande Ecole* can expect future increases in both the quality and number of co-authors. Those who studied in other European countries can expect benefits in terms of both the quality and number of co-authors, while those who studied in the United States can only expect an increase in the size of their network.

5 Sensitivity Analysis

The sensitivity analysis was based on Econlit publication scores (the CL-index) as a measure of individual productivity. Robustness tests focused on the impact of the CL-index on the quality and number of co-authors. Here again, the BEST_ALONE dummies were the instruments. First-stage estimates are reported in the first column of table 5 and 2SRI results are reported in columns 2–4. We applied a Heckman selection model as individual productivity measures are not count data. The selection variable is the number of publications: for academics who list no publications, the CL-index is equal to zero by definition. The Wald test of the null of independent equations is rejected at the 10% level: this supports the use of the Heckman selection model (see table 5, bottom part of the selection model, column 1). Results for the CL_index are very similar to those obtained with h and g indexes. Most importantly, our instruments are highly correlated with productivity measures and increase as a function of CNRS journal classification. Being the sole author of a paper published in a

[15] The reference variable is Paris 1; the exception in Paris 9, where the quality of the PhD network is statistically significantly lower.

CNRS1 journal increases the CL_index by 213.08. This can be compared to 51.24 (CNRS2), 23.17 (CNRS3), 10.82 (CNRS4), and 0 (CNRS5) (see table 5, Heckman Selection, column 1).

We finally consider the effect of the CL_index on co-author networks. Wald tests found that the null hypothesis of exogeneity is always rejected (see table 5, bottom part of inflate logit model, columns 2 to 4). Thus the individual productivity index (measured by the quality of Econlit publications) is endogenous to the number and quality of co-authors. Once this bias is corrected (by applying the 2SRI methodology), the quality of Econlit publications and citation scores produce similar results with respect to the main determinants of co-authorship. Most importantly, the CL_index has a positive and significant effect on all co-authorship variables[16]. This provides robust support for the conclusion that more productive academics tend to have higher-quality co-author networks (the assortative matching hypothesis). However, the impact of productivity on the size of co-author networks depends on how it is measured. Network size increases with EconLit publication scores and decreases with citation scores.

6 Conclusion and Policy Implications

This paper estimates the determinants of co-authorship in economics. More specifically, we test the existence of a relationship between individual research performance and that of co-authors. As it is clear that the quality of an individual researcher's publications is more-or-less linked to the quality of his/her co-authors there is an endogeneity bias that we handle by applying the 2SRI methodology. Tests are run using data from a novel database of French academics working in the field of economics. Although our results are necessarily focused on France, they could be applied to other European countries and other domains.

Our main finding is the existence of a positive and robust relationship between individual performance and the productivity of co-authors. This confirms the assortative matching hypothesis in the French case. Furthermore, the results show that individual productivity measured by citation scores is an important determinant of the quality, but not the quantity, of co-author networks. With the exception of the very best researchers, individual academic performance improves the matching process leading to efficient collaborations.

Our results have several limitations. Our analysis does not take account of the temporal dimension as our calculations are based on articles published over an entire career. A panel data analysis would help to disentangle the effect of variables such as age or life-cycle. However, including the temporal dimension may introduce other econometric problems. Publication time delays should be taken into account in computing annual productivity, as the time lag between the submission and publication of papers varies between journals. Therefore, the inclusion of the time dimension is left to further research as it would require to self-assemble a totally new database. In addition to a greater range of control variables, other factors should be considered for a deeper analysis. For instance, the size of the researcher's institution, the nationality of co-authors, the geographical location of their institutions, or the research interests of co-authors could be also considered to verify the robustness of our findings.

Despite these limitations, our paper yields interesting results. It highlights some important factors that lead to co-authorship of influential papers. For instance, for non-Anglophones, mastery of the English language increases the probability of collaboration with a broader set of international researchers. Similarly, co-authorship with the PhD supervisor increases the quality of future co-authors. This emphasizes the fact that working with senior

[16] For example, a 1% increase in the CL_index is consistent with a rise of 0.11% in hh. The gg index increases by 0.14%, and the number of co-authors by 0.17%.

colleagues is highly desirable for a junior academic. However, the main indirect result of our model lies in the estimation of large cohort effects.

Over the past thirty years, as in many other developed countries, the French academic system has experienced major institutional changes in the management and assessment of research. Incentives for publications in peer-reviewed international journals, the emergence of new publication norms and the progressive introduction of bibliometric tools to assess researchers have deeply changed scientific practice. Younger cohorts of French economists produce more papers and collaborate with more co-authors (who may also be more productive). The overall effect is to improve scientific production in economics. This result has important policy implications as it indirectly demonstrates the effectiveness of career incentives linked to publication. The above-mentioned results led to the recommendation that the training of researchers, and particularly new generations of young academics, should integrate a significant component of internationalization. Increased international mobility, especially to other European countries (for researchers from non-Anglophone countries) appears to be an efficient way to reinforce research networking and foster knowledge creation.

Finally, this research does not address various other aspects of co-authorship. For instance, there is no analysis of a potential gender effect on co-authorship[17]. Furthermore, we did not test for complementarity versus substitutability. Drawing upon detailed information related to co-authors' expertise (from JEL codes), it should be possible to calculate measures of the distance between individual authors and their co-authors, and to test this dimension of team formation. The question is left for further research.

[17] Our data suggests that if women do underperform in terms of production and coauthorship compared to men, the effect seems to be offset once controls are introduced for specialization.

Appendix 1 The French academic system

The French academic system is characterized by two divides. On the one hand, most public higher education is provided by universities and the *Grandes Ecoles*. Unlike universities, which are accessible to all students who have completed their secondary education, the *Grandes Ecoles* recruit their students through competitive examination. Of the 2.38 million French higher education students in 2013, 59.2% were at university, while 11% were either at engineering or business schools. A demanding selection procedure means that the *Grandes Ecoles* only train a few highly-skilled students, while universities select the best students after they have been admitted. Research, on the other hand, is the domain of either specialized universities run by academics who both teach and carry out research, or public institutions such as the *Centre national de la recherche scientifique* (the French National Centre for Scientific Research: CNRS) that employs pure researchers.

In this landscape, the CNRS plays a central role. It supports research laboratories that are affiliated to higher education institutions (90% of them universities) under different types of partnership agreement that offer facilities, staffing or funding. If we restrict the analysis to Economics and Management, 2119 permanent researchers were affiliated to such labs in 2007.

A doctoral candidate who aspires to become an assistant professor in a university must go through a local, competitive recruitment procedure. In order to limit any insider bias, a national body, the National Council of University (CNU), must first approve the candidature. This is based on an evaluation of the applicant's CV and the quality of his/her PhD. The CNU also plays an important role in promotion decisions and its decisions are primarily based on the scientific assessment of the candidate's CV.

Until the mid-1990s, under CNU regulations, it was possible to become an assistant professor in economics without any publication else than the PhD dissertation. This criterion dramatically changed in later years, with a progressive focus on academic performance.

One important milestone was the creation, in 2006, of a quasi non-governmental organization, the *Agence d'Evaluation de la Recherche et de l'Enseignement Supérieur* (AERES). This body was put in charge of evaluating both the public research system and the higher education service. AERES developed the first standards related to the minimal publication activities of academics. In the domain of economics and management, it published a list of journals and decided that a 'teaching' researcher had to publish at least two papers in one of these journals over a four-year evaluation period.[18]

Two years earlier, the CNRS issued its own classification of economic journals. The first draft of this classification had two objectives: firstly, to help evaluators identify the most influential journals in a particular scientific area, and secondly to provide researchers with a benchmark of the most important media for the dissemination of their results. CNRS assigned a grade ranging from 1 (for the most prestigious) to 4 (for the less influential) to the listed journals. From its inception, this classification has been very influential and subsequent drafts have significantly changed both French publication patterns in economics, and procedures for the evaluation of researchers. While officially, CNU experts use the AERES list, unofficially they frequently use the more demanding CNRS list.

A final manifestation of the impact of this list on the French academic landscape in economics is seen in changes to assessment procedures implemented in the competitive examination that leads to the status of full professor – the *Concours d'Agregation*. The *Concours d'agrégation en sciences économiques* is a nationwide competition through which assistant professors of economics are promoted to full professorship (see Combes et al. 2008

[18] Researcher with no teaching load had to publish four articles over the four-year period.

for a description). Applicants' research and teaching abilities are assessed by a jury of seven recognized French professors in economics. Since 2005, in addition to the classical evaluation by two members of the jury, research abilities are assessed through the computation of a productivity index defined as the sum of each paper's score (scores are proportional to the journals' CNRS ranking) (Levy-Garboua 2008). Although the specific formula has changed over the years, the objective reference to the CNRS ranking remains constant through time. This effect is so pervasive that it guides researchers' publication strategies. In 2003, an average of 43% of French teaching researchers in economics published in CNRS journals; by 2009 this frequency had risen to 51%.

Acknowledgements

We would like to thank Nicolas Debarsy, Etienne Farvaque, Christelle Garrouste, Sofoklis Goulas, El Mouhoub Mouhoud, all the participants at the LEO (Université d'Orléans) seminar and the 64th AFSE Meeting in Rennes (France) in 2015 for their comments. We would also like to thank Andrea Geraci for providing the STATA program of his paper. Finally we would like to thank the two anonymous referees for their comments which have helped to improve this paper. All remaining errors are ours. This research has been conducted as part of the Labex MMEDII project (ANR11-LBX-0023-01).

REFERENCES

Académie des Sciences. 2011. Du bon usage de la bibliométrie pour l'évaluation individuelle des chercheurs, Institut de France, Paris, France.

Barnett, A. H., Ault, R. W., Kaserman, D. L., 1988. The rising incidence of Co-authorship in economics: Further Evidence. Review of Economics and Statistics, 70, pp. 539–543.

Beaver, D. B., R. Rosen. 1978. Studies in Scientific Collaboration, Part I. Scientometrics. 1, pp. 65–84.

Besancenot, D., Huynh, K., F. Serranito, 2015. Thou shalt not work alone, Working paper hal-01175758, 49p.

Bidault F., T. Hildebrand, 2014. The distribution of partnership returns: Evidence from co-authorships in economics journals. Research Policy, 43, pp 1002–1013

Bornmann**,** L**.** H-D., Daniel, 2007. What do we know about the h index? Journal of the

Bosquet C., P. P. Combes, 2013. Are academics who publish more also more cited? Individual determinants of publication and citation records. Scientometrics, vol 97, pp 831–857.

Bruno, B. (2014), "Economics of co-authorship", Economic Analysis and Policy, pp.212–220.

Cainelli G., M.A. Maggioni ,T.E. Uberti, A. de Felice, 2012. Co-authorship and productivity among Italian economists. Applied Economics Letters, 19:16, pp 1609–1613

Chan, H. F., Önder, A. O., B. Torgler, 2015. The First Cut is the Deepest: Repeated Interactions of Coauthorship and Academic Productivity in Nobel Laureate Teams. CREMA Working paper 2015-04.

Chung, K.H., Cox, R.A.K., Kim, K.A., 2009. On the relation between intellectual collaboration and intellectual output: evidence from the Finance Academy. Quarterly Review of Economics and Finance, 49, 893–916.

Combes, PP., Linnemer L., M. Visser, 2008. Publish or Peer-rich? The Role of Skills and Networks in Hiring Economics Professors, Labour Economics 15, pp. 423–441 (2008).

Combes PP., Linnemer L., 2010. Inferring Missing Citations: A Quantitative Multi-Criteria Ranking of all Journals in Economics. WP halshs-00520325.

Durden, G. and Perri, T. 1995. Coauthorship and publication efficiency, Atlantic Economic Journal, 23 (1), pp. 69–76.

Egghe, L., 2006. Theory and practise of the g-index. Scientometrics, 69, 1, pp. 131–152.

Fafchamps, M., Van Der Leij M. and R. Goyal, 2010. Matching and Network Effects, Journal of the European Economic Association, 8, 1: 203–31.

Geraci A., D. Fabbri, C. Monfardini, 2014. Testing exogeneity of multinomial regressors in count data models: does two stage residual inclusion work? Department of Economics, Universita di Bologna, Working Paper DSE n° 921.

Hamermesh, D.S., 2015. Age, Cohort and Co-authorship. NBER Working Paper series n°20938, February. National Bureau of Economic Research, Cambridge.

Hamermesh, D.S., 2013. Six decades of Top Economics publishing: Who and How?. Journal of Economic Literature, 51(1), pp. 162–172.

Harzing. A.-W., 2010. The Publish or Perish Book. TarmaSoftwareResearch Pty Ltd, Melbourne, Australia.

Hirsch, J.E., 2005. An index to quantify an individual's scientific research output. PNAS 102(46), pp. 16569–16572.

Hollis, A., 2001, Co-authorship and the output of academic economists, Labour Economics, 8, pp. 503–530.

Johnson, D., 1997. Getting noticed in economics: the determinants of academic citations, The American Economist, 41(1), pp. 43–52.

Laband D., R. Tollison, 2000. Intellectual Collaboration. Journal of Political Economy, 108, pp. 632–662.

Laband, D.N., 1987. A Qualitative Test of Journal Discrimination Against Women, Eastern Economic Journal, 13, pp. 149–154.

Lee S., B. Bozeman, 2005. The Impact of Research Collaboration on Scientific Productivity, Social Studies of Science, 35, pp. 673–702.

Levin S. G., P. E. Stephan, 2001. Research Productivity Over the Life Cycle: Evidence for Academic Scientists, The American Economic Review, Vol. 81, No. 1 (Mar., 1991), pp. 114–132.

Levitt J.M., 2015. What is the optimal number of researchers for social science research? Scientometrics, vol 102 pp 213–225.

Lévy-Garboua, L., 2008. Rapport sur le premier concours d'agrégation de l'enseignement supérieur pour le recrutement de professeurs des Universités en Sciences Economiques (Année 2007-2008), Revue d'économie politique, 5(118), pp. 603–23.

McDowell, J.M., M. Melvin, 1983. The Determinants of Co-Authorship: An Analysis of the Economics Literature. The Review of Economics and Statistics, 65(3), pp. 155–160.

Medoff, H.M., 2007. The Input Relationship Between Co-Authors in Economics. A Production Function Approach. The American Journal of Economics and Sociology, 66(2), pp. 289–308.

Piette M. J., K. L. Ross., 1992, An Analysis of the Determinants of Co-Authorship in Economics, Journal of Economic Education, 23, pp. 277–283.

Rauber M., H.W. Ursprung, 2008. Life cycle and cohort productivity in economic research: the case of Germany. German Economic Review, vol 9, issue 4, pp 431–456.

Stefaniak, B., 1982. Individual and Multiple Authorship of Papers in Chemistry and Physics. Scientometrics, pp. 4: 331–337.

Terza J.V., Basu A., P.J. Rathouz, 2008. Two-Stage residual inclusion estimation: Adressing endogeneity in health econometric modeling. Journal of Health Economics, vol 27, pp 531–543.

Ursprung, H. W., Zimmer, M., 2007. Who is the “Platz-Hirsch" of the German economics profession? A citation analysis. Jahrbücher für Nationalökonomie und Statistik, 187–208.

Wooldridge J.M., 2014. Quasi-Maximum Likelihood Estimation and testing Nonlinear Models with Endogenous Explanatory Variables. Journal of Econometrics, vol 182(1) pp 226–234.

Zitt, M., E. Bassecoulard, Y. Okubo, 2000. Shadows of the Past in International Cooperation: Collaboration Profiles of the Top Five Producers of Science. Scientometrics, pp. 47: 627–657.

Table 1: Descriptive statistics (1566 observations)

Variable	Mean	Std. Dev.	Min	Max
hh	3.21	3.42	0	29
gg	7.43	8.61	0	83
nb_coauthors	4.51	7.44	0	60
h	3.25	3.99	0	39
g	6.02	8.13	0	84
CL_index	33.7	96.9	0	481.3
age	46.6	10.2	28	68
Prof_experience	21.9	9.36	7	46
cohorte64_68	.012	.112	0	1
cohorte69_73	.060	.238	0	1
cohorte74_78	.126	.332	0	1
cohorte79_83	.139	.346	0	1
cohorte84_88	.108	.311	0	1
cohorte89_93	.189	.391	0	1
cohorte94_98	.227	.419	0	1
cohorte99_04	.136	.342	0	1
Female	.282	.450	0	1
Full Prof.	.347	.476	0	1
Ass. Prof.	.652	.476	0	1
never_published	.224	.417	0	1
cowrite_dr	.148	.355	0	1
nb_papers	8.36	14.1	0	157
nb_papers_Review	6.18	9.07	0	94
nb_papers_other	2.17	6.36	0	91
never_published_Rev	.019	.139	0	1
nb_papers_Rev_cnrs	5.37	8.20	0	85
share_Rev_cnrs	.852	.247	0	2
share_cnrs1	.067	.157	0	1
share_cnrs2	.133	.198	0	1
share_cnrs3	.458	.332	0	1
share_cnrs4	.334	.363	0	1
share_cnrs5	.005	.036	0	0,6
share_papers_fr	.625	.332	0	1
share_papers_gb	.347	.325	0	1

Variable	Mean	Std. Dev.	Min	Max
share_papers_other	.027	.114	0	1
working_alone_only	.123	.328	0	1
never_working_alone	.173	.378	0	1
Herfindahl_JEL_code	.307	.263	0	1
max_kwa	.001	.036	0	1
max_kwb	.063	.240	0	1
max_kwc	.012	.107	0	1
max_kwd	.069	.254	0	1
max_kwe	.066	.249	0	1
max_kwf	.073	.261	0	1
max_kwg	.053	.224	0	1
max_kwh	.021	.144	0	1
max_kwi	.017	.128	0	1
max_kwj	.071	.257	0	1
max_kwk	.006	.076	0	1
max_kwl	.071	.257	0	1
max_kwm	.010	.101	0	1
max_kwn	.012	.110	0	1
max_kwo	.091	.288	0	1
max_kwp	.025	.156	0	1
max_kwq	.045	.207	0	1
max_kwr	.050	.219	0	1
max_kwt	0	0	0	0
max_kwy	0	0	0	0
max_kwz	.003	.056	0	1
best_alone_C1	.059	.236	0	1
best_alone_C2	.095	.293	0	1
best_alone_C3	.266	.442	0	1
best_alone_C4	.109	.311	0	1
best_alone_C5	.000	.025	0	1
best_alone_Econlit	.044	.205	0	1
best_alone_misc	.026	.161	0	1

Table 2: Individual research productivity indicators per year and cohort

	mean	sd	min	max	mean	sd	min	max	mean	sd	min	max
	Cohorte64-68				Cohorte69-73				Cohorte69-73			
hh_peryear	.092	.081	0	.260	.092	.111	0	.564	.080	.105	0	.718
gg_peryear	.216	.189	0	.622	.221	.274	0	1.58	.207	.325	0	2.44
h_peryear	.125	.099	0	.391	.105	.122	0	.743	.095	.121	0	.714
g_peryear	.222	.207	0	.826	.196	.249	0	1.64	.174	.255	0	1.84
CL_index_peryear	.834	1.03	0	4.50	1.13	2.57	0	19.6	.967	2.44	0	18.8
nb_papers_peryear	.294	.388	0	1.67	.258	.399	0	2.05	.226	.444	0	2.48
nb_coauthors_peryear	.122	.236	0	.913	.100	.194	0	1.13	.098	.198	0	1.21
share_gb	.171	.187	0	.555	.208	.271	0	1	.234	.297	0	1
share_fr	.756	.195	.333	1	.727	.311	0	1	.731	.314	0	1
never_published	.2	.410	0	1	.263	.442	0	1	.267	.443	0	1
female	.1	.307	0	1	.105	.308	0	1,0	.111	.315	0	1

	cohorte79-83				cohorte84-88				cohorte89-93			
hh_peryear	.087	.120	0	.666	.161	.207	0	1.20	.175	.174	0	.947
gg_peryear	.209	.293	0	1.66	.374	.484	0	2.77	.396	.423	0	2.19
h_peryear	.091	.146	0	.740	.157	.238	0	1.62	.182	.224	0	1.5
g_peryear	.174	.295	0	1.74	.288	.464	0	3.5	.337	.454	0	3.55
CL_index_peryear	.887	2.60	0	22.5	2.57	8.34	0	61.7	2.09	5.73	0	69.0
nb_papers_peryear	.207	.410	0	2.96	.430	.882	0	6.30	.522	.868	0	8.26
nb_coauthors_peryear	.114	.239	0	1.55	.232	.437	0	2.5	.296	.459	0	2.89
share_gb	.238	.293	0	1	.341	.328	0	1	.359	.316	0	1
share_fr	.715	.323	0	1	.633	.333	0	1	.621	.326	0	1
never_published	.389	.488	0	1	.288	.454	0	1	.229	.421	0	1
female	.215	.412	0	1	.205	.405	0	1	.347	.477	0	1

	cohorte93-98				cohorte99-04				Total			
cohorte94-98hh_peryear	.232	.192	0	1.30	.358	.257	0	1.6	.181	.200	0	1.6
gg_peryear	.526	.459	0	2.92	.779	.642	0	3.9	.413	.477	0	3.9
h_peryear	.219	.194	0	1.31	.312	.286	0	2.1	.176	.215	0	2.1
g_peryear	.405	.423	0	3.06	.572	.605	0	5	.326	.440	0	5
CL_index_peryear	1.87	3.23	0	34.1	2.09	2.87	0	18.4	1.71	4.43	0	69.0
nb_papers_peryear	.575	.758	0	5.53	.756	.838	0	6.4	.456	.738	0	8.26
nb_coauthors_peryear	.337	.452	0	3.46	.462	.500	0	2.9	.257	.414	0	3.46
share_gb	.411	.329	0	1	.451	.327	0	1	.347	.325	0	1
share_fr	.570	.334	0	1	.529	.328	0	1	.625	.332	0	1
never_published	.140	.347	0	1	.084	.278	0	1	.224	.417	0	1
female	.396	.489	0	1	.389	.488	0	1				

Table 3: First-stage estimations: Results for individual research productivity

	(1)		(2)	
Dependant variable:	H		G	
Model	ZINB		ZINB	
	IRR	P>|z|	IRR	P>|z|
Age	.979921***	0.001	.988158	0.118
Female	.8807374	0.118	.9387624	0.373
cohort64-68	.8612403	0.540	.7404993	0.305
cohort69-73	.7011104*	0.063	.5951844**	0.026
cohort74-78	.6800065**	0.022	.5823269**	0.011
cohort79-83	.6301373***	0.003	.5830024***	0.008
cohort84-88	.7556746*	0.067	.6614766***	0.009
cohort89-93	.8298919*	0.065	.8116941*	0.088
cohort94-98	.8492619**	0.033	.8324645**	0.037
Working_alone_only	.4163141***	0.000	.4754622***	0.000
Best_alone_CNRS1	3.094647***	0.000	2.948845***	0.000
Best_alone_CNRS2	2.356779***	0.000	2.31174***	0.000
Best_alone_CNRS3	1.593337***	0.000	1.423981***	0.000
Best_alone_CNRS4	1.315086***	0.001	1.197053*	0.059
Best_alone_CNRS5	2.023828***	0.000	2.099423***	0.000
Best_alone_Econlit_no_CNRS	1.341716**	0.039	1.478307***	0.010
Best_alone_Miscellaneous	1.749711***	0.000	1.752828***	0.000
Max_KW_A	5.97215***	0.001	3.410354**	0.020
Max_KW_B	2.421425***	0.000	1.798521***	0.002
Max_KW_C	3.261491***	0.000	2.493829***	0.000
Max_KW_D	3.066759***	0.000	2.242522***	0.000
Max_KW_E	2.351067***	0.000	1.719579***	0.001
Max_KW_F	3.459356***	0.000	2.709904***	0.000
Max_KW_G	2.926475***	0.000	2.093188***	0.000
Max_KW_H	3.204283***	0.000	2.247758***	0.000
Max_KW_I	3.117897***	0.000	2.197333***	0.000
Max_KW_J	3.047401***	0.000	2.390873***	0.000
Max_KW_K	2.156383***	0.001	1.22162	0.495
Max_KW_L	3.078709***	0.000	2.308664***	0.000
Max_KW_M	1.837837**	0.042	1.383792	0.264
Max_KW_N	2.436778***	0.000	1.448994	0.135
Max_KW_O	3.215247***	0.000	2.370315***	0.000
Max_KW_P	3.356206***	0.000	1.993537***	0.000
Max_KW_Q	3.372556***	0.000	2.224564***	0.000
Max_KW_R	3.610631***	0.000	2.71248***	0.000
Max_KW_Z	1.820709**	0.036	1.655251	0.155
PhD defended at :				
Univ. of Toulouse 1	.9417946	0.630	1.012134	0.938
Other French research institution	.8306353***	0.008	.8234902**	0.011
Univ. of Paris 10	.8961199	0.217	.8761202	0.222
Univ. of Aix-Marseille	.8038877**	0.023	.7415389**	0.013
Univ. of Strasbourg	1.243968*	0.058	1.124496	0.362

Univ. of Paris 9	1.104304	0.484	1.301935	0.263
Grande Ecole	1.102744	0.378	1.164485	0.245
Other Univ. In Paris	1.091772	0.296	.9312122	0.515
European country	1.726903*	0.088	2.431724**	0.037
US	1.591633***	0.007	1.601079***	0.030
Constant	.1580547***	0.000	.3016937***	0.000
Inflate : logit model				
	Coef.	P>|z|	Coef.	P>|z|
Age	-.0287674	0.924	.1000169	0.156
female	19.18457***	0.000	2.164583***	0.000
Working_alone_only	-1.463464	0.413	-.7288353	0.267
cohort64-68				
cohort69-73				
cohort74-78				
cohort79-83				
cohort84-88	-1.777777	0.876	-.0984602	0.900
cohort89-93	-3.376407	0.550	-.3113382	0.813
cohort94-98	-42.41814***	0.000	-15.39363***	0.000
cohort99-04	-5.049723***	0.509	-29.57643***	0.000
PhD defended at :				
Univ. of Toulouse 1	1.92255	0.283	.9298872*	0.102
Other French research institution	.0985543	0.967	.2342164	0.568
Univ. of Paris 10	1.122981	0.406	-.0882637	0.900
Univ. of Aix-Marseille	-.8473311	0.930	-.7632229	0.667
Univ. of Strasbourg	3.910572*	0.085	-.6485223	0.614
Univ. of Paris 9	-16.39208***	0.000	.8489505	0.267
Grande Ecole	1.568933	0.387	1.030994	0.217
Other Univ. In Paris	3.770505**	0.026	-.0288551	0.969
European country	-12.93755***	0.000	.7142122	0.562
US	-17.41771***	0.000	-17.41771***	0.000
constant	-18.37793	0.303	-31.27383***	0.000
lnalpha	-1.232484***	0.000	-.4987948***	0.000
athrho				
N	1566		1566	
Log Likelihood	-3147.398		-4100.567	
Vuong Test	3.21 ***	0.007	4.03 ***	0.000
Likelihood-ratio test of alpha=0	487.34 ***	0.000	2456.55 ***	0.000

The IRR value is the Incidence Rate Ratio of variable i and it is calculated as e^{β_i}; so if regressor i is increased by 1% the dependant variable will be increased by (IRR-1)%. P-values are reported in the P>|z| column (robust standard errors are calculated by bootstrap). The null hypothesis is rejected at the 1% level (***), 5 % (**) and 10% (*).lnalpha indicates the overdispersion parameter of the negative binomial distribution. The offset variable is the professional experience variable in each model.

Table 4: Determinants of co-authorship: 2SRI Estimation Results

Dependant variable:	HH				GG				NB_COAUTHORS			
	(1)		(2)		**(3)**		**(4)**		**(5)**		**(6)**	
Model	ZINB		ZINB		ZINB		ZINB		ZINB		ZINB	
	IRR	P>|z|	IRR	P>|z|	IRR	P>|z|	IRR	P>|z|	IRR	P>|z|	IRR	P>|z|
H	1.042084***	0.000	1.039226***	0.000					.9828172*	0.102		
G					1.014676***	0.001	1.010678***	0.009			.9920077*	0.087
Standardized-Residual	1.112301***	0.000	1.097373***	0.000	1.144926***	0.000	1.163374***	0.000	1.249599***	0.000	1.257356***	0.000
Age	.985219***	0.000	.9873376***	0.002	.9861084***	0.000	.9908974**	0.025	.9767026***	0.000	.9754919***	0.000
Female	.932133*	0.079	.9421786	0.153	.9766983	0.563	.9913404	0.825	.8779666***	0.007	.906711**	0.044
PR_CE	1.328025***	0.001	1.272798***	0.005	1.444356***	0.000	1.387759***	0.000	1.466899***	0.001	1.503297***	0.000
PR_2C	1.141204**	0.022	1.13541**	0.032	1.170767**	0.011	1.147701**	0.022	1.320809***	0.000	1.320599***	0.000
PR_1C	1.1038**	0.033	1.085009*	0.082	1.117463**	0.028	1.107455**	0.032	1.191133***	0.002	1.186777***	0.003
MCF_HC	.8794011	0.119	.8827909	0.158	.9388028	0.416	.8972782	0.174	.8880715	0.266	.9074843	0.370
cohort64-68	.317987***	0.000	.3334748***	0.000	.32785***	0.000	.3428459***	0.000	.3777281***	0.000	.4195257***	0.001
cohort69-73	.4034702***	0.000	.4288772***	0.000	.4147216***	0.000	.4135825***	0.000	.3853646***	0.000	.3748184***	0.000
cohort74-78	.3838905***	0.000	.4052276***	0.000	.4244945***	0.000	.4396829***	0.000	.4070941***	0.000	.4008705***	0.000
cohort79-83	.4175757***	0.000	.4453983***	0.000	.4426466***	0.000	.4445929***	0.000	.5395034***	0.000	.5451224***	0.000
cohort84-88	.5634481***	0.000	.5612336***	0.000	.6018537***	0.000	.5650782***	0.000	.6460625***	0.000	.6462379***	0.000
cohort89-93	.5803074***	0.000	.5919551***	0.000	.592278***	0.000	.6107291***	0.000	.7446976***	0.000	.7486914***	0.000
cohort94-98	.7053876***	0.000	.7057467***	0.000	.7335686***	0.000	.7136612***	0.000	.7789276***	0.000	.7829444***	0.000
nb_papers_Misc	.9999004	0.967	1.0003	0.893	.9995612	0.902	1.00005	0.987	1.023649***	0.000	1.024596***	0.000
nb_papers_EconLit_no_CNRS	1.003359	0.670	.9956444	0.590	1.01398	0.144	1.002135	0.808	1.033192***	0.002	1.035533***	0.001
nb_papers_CNRS1	.9783471***	0.000	.9811615***	0.000	.9709158***	0.000	.9766847***	0.000	1.007223	0.286	1.007139	0.302
nb_papers_CNRS2	1.015072**	0.049	1.007846	0.293	1.031534***	0.002	1.018168**	0.044	1.015464	0.150	1.014697	0.171
nb_papers_CNRS3	.995172	0.284	.9988137	0.792	.990886*	0.106	.9967855	0.530	1.05413***	0.000	1.055889***	0.000
nb_papers_CNRS4	.9965078	0.474	1.000438	0.928	.9867248**	0.033	.992407	0.179	1.04393***	0.000	1.045542***	0.000
nb_papers_CNRS5	.954151	0.448	.9899327	0.868	.9771805	0.729	1.003139	0.958	.9914971	0.908	.9900335	0.894

Max_KW_A	2.97886***	0.002	3.243207***	0.003	2.957538***	0.004	3.765687***	0.000	1.47005	0.641	1.496709	0.626
Max_KW_B	1.709175***	0.000	1.924573***	0.001	1.473293***	0.000	2.024066***	0.000	2.258015***	0.005	2.369907***	0.003
Max_KW_C	2.469432***	0.000	2.67134***	0.000	2.55383***	0.000	3.292041***	0.000	4.985464***	0.000	5.268684***	0.000
Max_KW_D	1.860555***	0.000	2.051891***	0.000	1.649102***	0.000	2.275266***	0.000	3.637877***	0.000	3.821164***	0.000
Max_KW_E	1.944903***	0.000	2.194845***	0.000	1.4613***	0.000	2.023645***	0.000	3.329697***	0.000	3.464074***	0.000
Max_KW_F	1.904109***	0.000	2.161185***	0.000	1.558311***	0.000	2.192218***	0.000	3.206283***	0.000	3.320214***	0.000
Max_KW_G	1.801595***	0.000	2.071904***	0.000	1.445131***	0.000	2.06265***	0.000	3.617831***	0.000	3.7429***	0.000
Max_KW_H	1.828563***	0.000	1.979695***	0.001	1.79121***	0.000	2.248511***	0.000	3.740673***	0.000	3.897239***	0.000
Max_KW_I	2.085403***	0.000	2.338452***	0.000	1.77805***	0.000	2.426356***	0.000	3.551919***	0.000	3.750028***	0.000
Max_KW_J	2.195083***	0.000	2.49477***	0.000	1.709877***	0.000	2.419579***	0.000	4.502826***	0.000	4.677448***	0.000
Max_KW_K	1.805412***	0.004	1.997078***	0.008	1.04632	0.832	1.462413	0.112	3.161944***	0.001	3.258965***	0.001
Max_KW_L	1.932462***	0.000	2.134989***	0.000	1.697882***	0.000	2.335484***	0.000	3.43135***	0.000	3.570446***	0.000
Max_KW_M	1.449688*	0.085	1.767825**	0.041	1.243556	0.257	1.761268**	0.013	3.102597***	0.001	3.222172***	0.001
Max_KW_N	1.833915***	0.000	2.260718***	0.000	1.819956***	0.000	2.571753***	0.000	2.988923***	0.001	3.117473***	0.001
Max_KW_O	1.742426***	0.000	2.000841***	0.000	1.439313***	0.000	2.072455***	0.000	3.666729***	0.000	3.808461***	0.000
Max_KW_P	1.904894***	0.000	2.19255***	0.000	1.664302***	0.000	2.343557***	0.000	2.478825***	0.004	2.602105***	0.002
Max_KW_Q	1.955264***	0.000	2.235366***	0.000	1.609147***	0.000	2.305896***	0.000	4.721441***	0.000	4.899548***	0.000
Max_KW_R	1.927516***	0.000	2.196696***	0.000	1.717664***	0.000	2.459632***	0.000	3.832075***	0.000	4.018506***	0.000
Max_KW_Z	1.839646**	0.039	2.04952**	0.031	1.827516**	0.032	2.411554***	0.002	1.931404	0.169	2.089348	0.125
Herfindahl_JEL_CODE	.6614971***	0.000	.6885356***	0.000	.7133196***	0.000	.7498311***	0.000	.4496104***	0.000	.4276585***	0.000
Cowrite_dr	1.195147***	0.000	1.185693***	0.000	1.149332***	0.001	1.140886***	0.001	1.199914***	0.000	1.200328***	0.000
nb_coauthors	1.017842***	0.000	1.017056***	0.000	1.030136***	0.000	1.027915***	0.000				
PhD defended at (Network effect) :												
Univ. of Toulouse 1	1.091919	0.255	1.045402	0.572	1.012737	0.880	.9669692	0.675	1.122751	0.231	1.126704	0.223
Other French research institution	.990712	0.823	.9695812	0.469	.9335992	0.123	.9046955**	0.017	.9442521	0.268	.943085	0.264
Univ. of Paris 10	1.075504	0.268	1.016827	0.807	1.073756	0.304	.9918353	0.902	.8546621*	0.067	.8476518*	0.058
Univ. of Aix-Marseille	1.052381	0.428	1.043807	0.512	1.041037	0.571	1.020515	0.759	.8313562**	0.029	.8196203**	0.021
Univ. of Strasbourg	1.213255**	0.015	1.135338	0.115	1.203975**	0.049	1.09718	0.291	.9087715	0.381	.8903624	0.294
Univ. of Paris 9	.7914497**	0.012	.7809851***	0.007	.8234189**	0.050	.844345*	0.079	.9998631	0.999	1.036374	0.758
Grande Ecole	1.283229***	0.001	1.223373***	0.006	1.332835***	0.001	1.338129***	0.000	1.066246	0.494	1.06514	0.509
Other Univ. In Paris	1.113196*	0.089	1.109559*	0.108	1.155467**	0.035	1.089803	0.194	.9831769	0.839	.9624741	0.651
European country	1.507732***	0.000	1.472532***	0.001	1.533225***	0.006	1.545232***	0.001	1.572692***	0.007	1.597396***	0.006
US	.9512512	0.777	.8866488	0.486	1.080082	0.719	1.025538	0.894	1.886956***	0.003	1.921799***	0.003
Share_gb			1.325931***	0.000			1.384238***	0.000	1.730591***	0.000	1.742366***	0.000
Share_other			1.222239	0.264			1.137919	0.421	.7991543	0.357	.8222495	0.428
constant	.2714861***	0.000	.1978249***	0.000	.6530717	0.008	.355216***	0.000	.2214688***	0.000	.214967***	0.000

Inflate : logit model	(1)		(2)		(3)		(4)		(5)		(6)	
	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|
H	-2.823099***	0.000	-2.139665***	0.000					-2.573567***	0.000		
G					-1.340126***	0.000	-.5825985***	0.000			-1.461653***	0.000
Standardized-Residual	1.37467***	0.001	1.328562***	0.005	1.845008***	0.000	.2586434	0.559	5.451412***	0.000	6.229798***	0.000
Age	.0002698	0.991	.0017599	0.962	.0023032	0.908	.0423658	0.116	-.0552307	0.182	-.0855823*	0.069
Female	-.027674	0.924	-.124582	0.760	-.036914	0.872	.3393366	0.238	-4.858831***	0.000	-4.409307***	0.000
PR_CE	-.0853099	0.953	-.6597885	0.694	-.2231051	0.812	-.4026852	0.673	.5048468	0.617	.4376303	0.683
PR_2C	-1.109853**	0.037	-.7975822	0.164	-1.001837***	0.009	-.8486715**	0.046	.6690542	0.244	.3718004	0.509
PR_1C	-.4569083	0.389	-.6283158	0.298	-.3739703	0.324	-.0140062	0.972	.0630835	0.920	.0884699	0.889
MCF_HC	.0208497	0.958	-.892451	0.165	-.0212284	0.943	-.4229394	0.320	-.6086455	0.331	-.605903	0.314
cohort64-68	5.105937**	0.029	5.093248*	0.059	2.197106	0.125	1.432296	0.378	7.640476***	0.000	9.513126***	0.000
cohort69-73	2.899146***	0.002	3.898105***	0.005	2.481967***	0.001	1.424851	0.132	5.2875***	0.000	6.882627***	0.001
cohort74-78	1.697518**	0.022	1.972589*	0.057	1.489429**	0.011	1.133333	0.123	3.478864***	0.006	4.874718***	0.007
cohort79-83	.7960128	0.235	1.219621	0.233	.640547	0.225	.2926999	0.684	3.42111***	0.004	4.423214**	0.011
cohort84-88	1.610108**	0.011	1.644094*	0.059	1.255303**	0.012	.5365094	0.397	3.391606***	0.001	4.553844***	0.005
cohort89-93	.9393254*	0.097	1.059767	0.195	.8299438*	0.055	.2013273	0.714	.9358644	0.398	-.2366065	0.904
cohort94-98	.4340404	0.416	.7872186	0.274	.3105463	0.435	.0150064	0.975	.7169281	0.396	1.591115	0.221
constant	.8018058	0.487	.1994622	0.903	1.228041	0.240	-2.474895**	0.049	4.843342***	0.007	5.101366**	0.020
lnalpha	-4.914826***	0.000	-16.17411	0.927	-1.743548***	0.000	-2.139864***	0.000	-1.905398***	0.000	-1.844072***	0.000
N	1566		1183		1566		1183		1183		1183	
Log Likelihood	-2688.252		-2258.393		-3859.97		-3218.27		-2605.685		-2631.585	
Vuong Test (unconstraint)	8.70***	0.000	13.05***	0.000	12.50***	0.000	8.45***	0.000	6.48***	0.000	6.33***	0.000
Likelihood-ratio test of alpha=0	0.92	0.168	1.39	0.112	711.71***	0.000	423.64***	0.000	386.90***	0.000	476.10 ***	0.000
Exogeneity test (Wald test)	35.11 $^{\mu\mu\mu}$	0.000	21.19 $^{\mu\mu\mu}$	0.000	42.05 $^{\mu\mu\mu}$	0.000	29.35 $^{\mu\mu\mu}$	0.000	109.31 $^{\mu\mu\mu}$	0.000	110.38 $^{\mu\mu\mu}$	0.000

The IRR value is the Incidence Rate Ratio of variable i ant it is calculated as e^{β_i}; so if regressor i is increased by 1% the dependant variable will be increased by (IRR-1)%. P-values are reported in the P>|z| column (robust standard errors are calculated by bootstrap). The null hypothesis is rejected at the 1% level (***), 5 % (**) and 10% (*).lnalpha indicates the overdispersion parameter of the negative binomial distribution. The null hypothesis of exogeneity is rejected at the 1% level (μμμ), 5% level (μμ) and 10% level(μ). All parent models include a constant parameter which is not reported in the table. The offset variable is the Experience variable.

Table 5: Robustness checks: Determinants of co-authorship

	First Stage Regression		**Second Stage Regression**					
Dependant variable:	**CL_index**		**HH**		**GG**		**NB_COAUTHORS**	
	(1)		(2)		**(3)**		**(4)**	
Model	Heckman Selection		ZINB		ZINB		ZINB	
	Coeff.	P>|z|	IRR	P>|z|	IRR	P>|z|	IRR	P>|z|
CL-index			1.001114***	0.000	1.001378***	0.000	1.001659***	0.000
Standardized-Residual			1.013446**	0.040	1.0164**	0.029	1.044252***	0.000
Age	-.4947327	0.349	.9813628***	0.000	.9829286***	0.000	.9670202***	0.000
Female	-12.29857***	0.001	.929742	0.129	.9650636	0.462	.8203863***	0.001
PR_CE			1.887019***	0.000	1.942773***	0.000	2.605878***	0.000
PR_2C			1.475731***	0.000	1.446747***	0.000	1.921692***	0.000
PR_1C			1.258107***	0.000	1.292335***	0.000	1.39706***	0.000
MCF_HC			.8818905	0.205	.9576312	0.634	.921926	0.534
cohort64-68	14.78529	0.437	.354216***	0.000	.3245448***	0.000	.4622183**	0.012
cohort69-73	23.96988	0.196	.4030283***	0.000	.4088248***	0.000	.3559446***	0.000
cohort74-78	17.39514	0.215	.4047581***	0.000	.4319376***	0.000	.4278694***	0.000
cohort79-83	21.20129	0.123	.4670122***	0.000	.4815984***	0.000	.5853556***	0.001
cohort84-88	53.96536**	0.013	.6255676***	0.000	.6311596***	0.000	.6632762***	0.004
cohort89-93	24.68417**	0.013	.6470144***	0.000	.6431645***	0.000	.8585262	0.133
cohort94-98	14.19882**	0.016	.7150509***	0.000	.7374271***	0.000	.8440864**	0.036
Working_alone_only	-31.08238***	0.000						
Best_alone_CNRS1	213.0755***	0.000						
Best_alone_CNRS2	51.2407***	0.000						
Best_alone_CNRS3	23.17387***	0.000						
Best_alone_CNRS4	10.81977**	0.025						
Best_alone_CNRS5	6.203222	0.657						
Best_alone_Econlit_no_CNRS	15.03348	0.143						
Best_alone_Miscellaneous	4.271392	0.526						

Max_KW_A	-9.724827	0.272	5.100593***	0.000	3.778635***	0.003	38.64384***	0.000
Max_KW_B	-19.94836**	0.025	2.235167***	0.000	1.566358***	0.000	57.50343***	0.000
Max_KW_C	7.520197	0.663	3.798386***	0.000	3.248843***	0.000	119.3732***	0.000
Max_KW_D	25.46488*	0.071	2.535956***	0.000	1.893472***	0.000	91.35538***	0.000
Max_KW_E	9.420111	0.213	2.418056***	0.000	1.587806***	0.000	87.10895***	0.000
Max_KW_F	16.89251**	0.046	2.922626***	0.000	1.916825***	0.000	105.8776***	0.000
Max_KW_G	22.05428	0.130	2.512777***	0.000	1.718109***	0.000	95.24252***	0.000
Max_KW_H	24.23717	0.470	2.716825***	0.000	2.136986***	0.000	111.3967***	0.000
Max_KW_I	-4.102471	0.714	2.8983***	0.000	2.117202***	0.000	82.01653***	0.000
Max_KW_J	16.56281**	0.047	3.173345***	0.000	2.179079***	0.000	115.5669***	0.000
Max_KW_K	-1.81393	0.908	2.573481***	0.000	1.412671	0.148	101.4353***	0.000
Max_KW_L	8.352658	0.423	2.682643***	0.000	1.951436***	0.000	87.93333***	0.000
Max_KW_M	2.919344	0.714	1.800081**	0.016	1.343985	0.188	54.30431***	0.000
Max_KW_N	-12.93358	0.321	2.797522***	0.000	2.091606***	0.000	81.38917***	0.000
Max_KW_O	3.716523	0.618	2.6176***	0.000	1.774657***	0.000	97.00238***	0.000
Max_KW_P	10.76113*	0.107	2.870926***	0.000	1.953114***	0.000	67.7025***	0.000
Max_KW_Q	15.60789	0.166	2.89453***	0.000	1.9893***	0.000	129.5121***	0.000
Max_KW_R	-2.192024	0.764	3.140688***	0.000	2.24666***	0.000	103.1474***	0.000
Max_KW_Z	-3.276266	0.865	2.591006***	0.006	2.288644**	0.014	49.89448***	0.000
Herfindahl_JEL_CODE			.4738775***	0.000	.5367003***	0.000	.2465645***	0.000
Cowrite_dr			1.342601***	0.000	1.279674***	0.000	1.274668***	0.000
PhD defended at (Network effect) :								
Univ. of Toulouse 1	43.84993	0.120	.8795278	0.173	.9090126	0.341	.8362421	0.154
Other French research institution	-.7645302	0.887	.9086915**	0.050	.8819028**	0.014	.8829447**	0.054
Univ. of Paris 10	-4.131278	0.588	.9835408	0.830	.9725755	0.732	.7963284**	0.031
Univ. of Aix-Marseille	-3.425586	0.687	.9530729	0.533	.9577904	0.603	.7959504**	0.031
Univ. of Strasbourg	8.445213	0.346	1.251424***	0.021	1.216292*	0.082	.9924292	0.955
Univ. of Paris 9	47.35364	0.132	.8217255*	0.079	.8160147*	0.085	.9353773	0.648
Grande Ecole	-11.77238	0.364	1.241279**	0.015	1.334728***	0.004	1.045897	0.710
Other Univ. In Paris	-1.796448	0.823	1.087836	0.264	1.115741	0.178	.9946334	0.959
European country	45.76318	0.192	1.739695***	0.000	1.793192***	0.003	1.528453*	0.057
US	17.54735	0.746	1.04392	0.839	1.038366	0.881	1.167051	0.588
constant	-10.20822	0.604	.2929171***	0.000	.8377577	0.343	.0233336***	0.000

	Selection model: nb_papers = 0		Inflate : logit model					
	(1)		(2)		**(3)**		**(4)**	
	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|
CL-index			-.0878245***	0.000	-.0932751***	0.000	-.1685765***	0.000
Standardized-Residual			.238932**	0.011	.2100567***	0.002	1.608292***	0.000
Age	-.0560476***	0.000	.0580008**	0.021	.0580813***	0.000	-.0122577	0.757
Female	-.4404481***	0.000	.4996365*	0.101	.2724172	0.172	-3.057696***	0.000
Working_alone_only	7.10931***	0.000						
PR_CE			-1.166768*	0.089	-1.302699**	0.025	.3562914	0.700
PR_2C			-1.324146***	0.001	-1.369774***	0.000	.9714578*	0.083
PR_1C			-.6021319	0.162	-.4497443	0.160	-.3928021	0.587
MCF_HC			-.6305575*	0.100	-.3426731	0.172	-.2240672	0.734
cohort64-68	.9611347**	0.028	3.211649**	0.017	1.887027*	0.053	8.715849***	0.000
cohort69-73	.4049214	0.155	2.675933**	0.014	1.751709***	0.006	8.374569***	0.000
cohort74-78	.2252025	0.356	2.270412**	0.022	1.42765***	0.008	5.949673***	0.000
cohort79-83	-.2309051	0.298	2.047851**	0.029	1.21485**	0.017	7.416507***	0.000
cohort84-88	-.1697644	0.421	2.709171***	0.003	1.709105***	0.001	8.965193***	0.000
cohort89-93	-.153463	0.370	1.604793*	0.051	.9969648**	0.017	4.708843***	0.000
cohort94-98	-.0467956	0.769	.6716703	0.427	.398521	0.316	3.307283***	0.002
constant	-31.27383***	0.000	-5.236934***	0.000	-4.066527***	0.000	-4.637296***	0.006
lnalpha			-2.261074***	0.000	-1.208168***	0.000	-.9829132***	0.000
PhD defended at (Network effect) :								
Univ. of Toulouse 1	-.1151487	0.251						
Other French research institution	.037979	0.820						
Univ. of Paris 10	-.1557919	0.339						
Univ. of Aix-Marseille	.424296	0.138						
Univ. of Strasbourg	.0027242	0.990						
Univ. of Paris 9	-.0079141	0.972						
Grande Ecole	1.030994	0.217						
Other Univ. In Paris	-.1859341	0.253						
European country	.5156946	0.318						
US	3.539954***	0.000						
athrho	-.02917*	0.076						

N	1566		1566		1566		1566	
Log Likelihood	-7836.496		-3043.567		-4194.584		-2950.102	
Vuong Test (unconstraint)			5.26***	0.000	9.95***	0.000	1318.80***	0.000
Likelihood-ratio test of alpha=0			91.44***	0.000	1518.12***	0.000	6.82***	0.000
Wald test of indep. eqns. (rho = 0)	3.15*	0.076						
Exogeneity test (Wald test)			9.73 µµµ	0.008	14.02 µµµ	0.001	99.79 µµµ	0.000

The IRR value is the Incidence Rate Ratio of variable i ant it is calculated as e^{β_i}; so if regressor i is increased by 1% the dependant variable will be increased by (IRR-1)%. P-values are reported in the P>|z| column (robust standard errors are calculated by bootstrap). The null hypothesis is rejected at the 1% level (***), 5 % (**) and 10% (*).lnalpha indicates the overdispersion parameter of the negative binomial distribution. The null hypothesis of exogeneity is rejected at the 1% level (µµµ), 5% level (µµ) and 10% level(µ). All parent models include a constant parameter which is not reported in the table. The offset variable is the Experience variable. In the Heckman selection model, the athrho variable is the estimate of the inverse hyperbolic tangent of rho (ρ): athrho=0.5*ln((1+ρ)/(1-ρ)) where ρ is the correlation between the residuals of the two equations. The Wald test of independent equations is the likelihood-ratio test of Ho: ρ = 0 and it is computationally the comparison of the joint likelihood of an independent probit model for the selection equation and a regression model on research productivity index data against the Heckman model likelihood.