Document de Recherche du Laboratoire d'Économie d'Orléans

DR LEO 2015-16

Co-Authorship and Individual Research Productivity in Economics: Assessing The Assortative Matching Hypothesis

Damien BESANCEN0T
Kim HUYNH
Francisco SERRANITO

Laboratoire d'Économie d'Orléans
Collegium DEG
Rue de Blois - BP 26739
45067 Orléans Cedex 2
Tél. : (33) (0)2 38 41 70 37
e-mail : leo@univ-orleans.fr
www.leo-univ-orleans.fr/

CO-AUTHORSHIP AND INDIVIDUAL 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 and IRD, LEDa, DIAL UMR 225.

Abstract:

This paper aims at estimating the determinants of co-authorship in economics. More specifically, we test the existence of a potential relationship between the research efficiency of an individual and that of his co-authors (the so called assortative matching hypothesis) using a novel database of French academic scholars. However, individual research productivity should be an endogenous regressor as the quality of an academic's publication will depend somehow on the quality of his co-authors. We have applied the Two Stage Residual Inclusion (2SRI) approach in order to take into account this endogeneity bias. The main empirical result is that the number and the quality of a researcher's co-authors reflect the productivity of that researcher.There is also a significant gender effect: being a woman has no impact on the probability of never collaborating with other economists but it decreases both the quality and the quantity of co-authors. Finally, life-time cycles are also an important determinant of the co-authorship trend as the social imprinting hypothesis would suggest. So institutional changes occurred in French academia in mid-eighties have had a large impact on individual research productivity.

Keywords: Co-authorship, Count Data, Zero Inflate Models, Instrumental Variables, gender productivity gap, h index.

Resumé:

Cet article estime les déterminants de la co-écriture d'article enéconomie. Nous testons l'existence d'une relation entre la productivité dans la recherche d'un individu et celle de ses co-auteurs utilisant une nouvelle base de données de chercheurs universitaires français. Cependant, il y a y problem d'endogéneité avec la variable de productivité dans la recherché: la qualité des publications d'un chercheur depend de la qualité de ses co-auteurs. Nous avons appliqué l'approche 2SRI afin de tenir compte de ce biais d'endogénéité. Notre principal résultat empirique est de montrer que la qualité de co-auteurs d'un chercheur reflètent sa productivité. Il existe aussi un effet lié au genre: être une femme diminue à la fois la qualité et la quantité de co-auteurs. Enfin, les changements institutionnels survenus dans le milieu universitaire français au milieu des années quatre-vingt ont eu un grand impact sur la productivité des chercheurs.

Mots Clés: Coécriture d'article, modèle de poission, modèle à exces de zeros, variable instrumentale.

JEL: A14, C25, D83, I23, J24.

---

[1]Corresponding author: francisco.serranito@univ-orleans.fr; Univ.orléans, CNRS, LEO, UMR 7322, F45067, Orléans, France, Université d'Orléans, Collegium DEG, rue de Blois BP 26739, 45067 Orléans Cedex 2, France.

1- Introduction

Since the end of the seventies, a vast strand of research emphasizes that coauthoring papers is not the exception but constitutes a new scientific norm (see for instance Beaver and Rosen 1978; Stefaniak 1982; Petry 1988; Zitt et al. 2000, Laband and Tollison 2000, Cardoso et al 2010, Card and DellaVigna 2013, Hamermesh 2013 and 2015). Co-authorship can be considered as a particular form of scientific collaborations. Promoting collaboration in general and multi-authorship in particular is often a major goal of research policy in order to boost the production of knowledge. Indeed, policy makers encourage collaborative research because most studiesfind that co-authorshipis a way to improve individual academic productivity in research (Bidault & Hildegrand, 2014 and Levitt, 2015). So understanding the drivers of long term co-authorship growth should be an important issue for national research organizations whose goal is to design research policy recommendations.

Our paper is also related to the economics of science and more particularly on the strand of literature studying how policies designed by national research organizations can impact individual behavior. Indeed, French academia is undergoing profound transformations over the last twenty years. Measuring individual productivity by bibliometrics has been introduced to provide incentives for high research productivtity (Académie des Sciences, 2011). Furthermore, the French system is more an more adopting the Anglo-Saxon global standards in academia: increasing competition between academics in order to fund their research and rewarding top ranked publications. We scrutinize the impact of these institutional changes on individual academic productivity by modeling different cohorts of economists[2]. We acknowledge that our results remain estimated on French academics only; but they are very similar to those obtained in the German case (Rauber & Ursprung, 2007) or the Italian one (Cainelli et al., 2012). Hence our results could be interpreted as the response of continental European science systems to structural and institutional changes.

Generally, the long term growth trend of co-authorship is explained by the positive effects of scientific collaboration on both the quantity and the quality of the research output.Firstly, as regards quantity, co-authorship is a simple way to increase the number of papers that a researcher may publish during a given period of time. Indeed Laband and Tolisson (2000) and Ursprung and Zimmer (2007) document that co-authorship increases acceptance rate by refereed journals. Durden and Perry (1995) finds that the total number of publications is significantly and positively related to the number of collaborative publications. Hollis (2001) shows that the more co-autorship done in the past, the more prolific an author is likely to be today. Lee and Bozeman (2005) stress that collaboration is a strong predictor of the total number of a researcher's publications.So, even if assessment procedures discount the value of coauthored papers according to the number of authors, the value of two bi-authored papers is

[2] Using a panel of top labor economists, Hamermesh (2015) finds that the increasing trend of co-authorship has something to do with the research styles that one learns from what the profession is doing during one's Ph.D. So intutional changes implemented at the level of PhD programs should have a large impact on academic research outputs.

generally worth more than the value of a single-authored paper. This creates astrong incentive to scientific collaboration (Liebowitz and Palmer, 1983, Barnett et al., 1988 and Bruno, 2014).

Secondly, co-authoring is also an efficient way to improve the quality of a scientific production. Quality is in general measured by the number of citations since Hamermesh et al. (1982) have showned that one additional citation had a larger impact on academic wages than one additional published paper on a sample of 148 full professors of seven large US universities[3]. Empirical evidence on the link between coauthorship and the number of citations are rather mixed and it depends on the research field (see Levitt 2015 for a review). On the one hand, some studies find a positive impact of co-authorship on quality. For instance, Laband (1987) or Johnson (1997) report that citations frequency is significantly higher for co-authored papers compared to single-authored ones. Within the economic field, Levitt (2015) finds that amongst the published papers in 2007, those with multi authors received more citations: he estimates that an additional discrete author in the collaboration team increases the number of citations up to three authors. Chung et al. (2009) show that papers co-authored with a prolific author receive more citations, whereas papers written with colleagues from the same institution does not disclose any increase in quality. On the other hand, Barnett et al. (1988), Hollis (2001) and Medoff (2003) find no support for the quality effect of co-authoship. Recently, Ductor (2012) claims that these negative effects could be explained by some methodological bais. He finds that after controlling for endogenous co-authorship formation, unobservable heterogeneity and time varying factors, then the effect of intellectual collaboration on individual performance becomes positive.

If we exclude the specific cases where collaboration is justified by friendship or is considered as a way to escape academic isolation (Medoff, 2003, Acedo et al., 2006, Hamermesh, 2013), team formation is mainly explained by advocating the role of complementarities in researchers' abilities. In a pioneer paper, McDowell and Melvin (1983) linked the raise of co-authorship to the explosion of knowledge in economics. In an academic world where researchers are involved in increasing specialization, co-authorship allows complementarities and appears as an efficient way to improve scientific production. Under alternative presentations, this seminal argument has been developed in a series of contributions. For instance, Piette and Ross (1992) states that authors who work in areas outside of their specialty tend to engage more in co-authorship than authors with close scientific tools. More recently, focusing on Nobel Laureates' pattern of co-authorship, Chan et al. (2015), shows that scientific collaboration may be induced by conceptual complementarities – complementarities that erode through time after repeated interaction.

On the opposite, Barnett et al. (1988) and Medoff (2007) found that substitutability may be at the heart of collaborative combinations of equally skilled researchers. According to Fafchamps et al. (2006) collaboration is most likely between authors of a similar level of ability: the so-called assortative matching hypothesis. However collaboration between authors with different abilities may also arise if the contribution of the lower ability author relaxes the

[3]Diamond (1985) also found that the monetary return from a citation worth more to the authors when the paper is coauthored than when there is a single author.

time-constraint of his/her co-authors. Thus, as a result of collaboration, higher ability authors produce more research while lower ability researchers produce better quality output than would otherwise be the case. Recently, Bidault & Hildegrand (2014) assess the determinants of asymmetric co-authors team. They distinguish between short-term and long-term relative returns within a two author team and they show that if co-authorship is less favorable for the "senior" co-author in the short run, in the long run all the co-authors benefit from collaboring. Finally, international co-authorship seems to pay off in terms of receiving more citations. According to Bruno (2014), a symmetric or an asymmetric team constitution will depend mainly on the reward system designed by national research organizations.

If the theoretical motivations of team formation in scientific activities, and more specifically in the writing of papers, are now well identified, few empirical studies are however available on the determinants of co-authorships. Our paper aims at filling up that gap. More precisely, we will assess the link between the "individual research quality" of an academic and that of his/her co-authors. In a companion paper, we have developed a theoretical matching model in which authors with different level of ability are randomly matched (Besancenot et al, 2015). Each researcher has to decide if he/she accepts to collaborate or if he/she prefers to work alone. Built in a dynamic setting, the model leads to the characterization of an optimal decision rule leading to the choice of collaboration. Three main conclusions arise from that theoretical model: first, the higher the ability of a researcher, the higher will be the skills of his/her co-authors. Second, the number of papers written during a given period of time is increasing with the productivity of the researcher. Finally, a talented author should more frequently meet authors willing to accept collaboration and should have more co-authored papers than authors with low productivity (Besancenot et al, 2015).

In this paper, we turn rather to estimate the empirical determinants of co-authorship by applying a novel database considering all academic economists with a position in a French university in 2004. Our dataset has two main advantages: it is exhaustive and it includes both publishing and non publishing academics. In general, studies applying bibliometrics never include the second category of academics. Our paper employs a specific econometric framework in order to take account of these non publishing academics; so it is likely to produce more trustworthy estimates.

For each French academic in our dataset, and for each of their co-authors, we computed their h and g indexes. In order to measure the number and thequality of the coauthors in a one-dimensional variable, we also computed two new Meta indexes (hereafter the hh and gg indexes) built by reference to the hand the g methodologies[4]. A high hh(or gg) index will reveal that an author collaborates with a high number of influential co-authors. However, individual research productivity should be an endogenous regressor as the quality of a researcher's publication will depend somehow on the quality of his co-authors. So in order to evaluate the effect of the level of its own research productivity on the co-authors quality, this endogeneity bias should be addressed. Endogenity could be addressed by applying

[4] Meta indexes are often used in bibliometrics. For instance, Schubert (2012) computes a Meta index to measure the characteristics of a researcher's network and Tol (2008) proposes a generalized g index to rank groups of researchers.

Instrumental Variable (IV) procedures. In the case of a linear model the IV framework corresponds to the two Stage Least Square (2SLS) frawemork. In non-linear models, the Two Stages Prediction Substitution (2SPS) approach could be considered as the non-linear counterpart of the 2SLS estimation. However, Wooldrige (2014) highlights that, when the conditional expectation model is non-linear, then 2SPS approach produces in general inconsistent estimates. He advocates applying the Two Stage Residual Inclusion (2SRI) approach which allows getting consistent estimates of the parameters in the structural regression. To our knowledge, our paper is the first one to apply such a framework to the co-authorship determinant issue.

The paper is organized as follows. The next section presents the empirical methodology. Section 3 describes the database; section 4 provides the results of the empirical model; and finally section 5 concludes the paper.

2- Econometric Methodology

According to the assortative matching hypothesis, a highly skilled author should have more co-authors of better quality. Thus, our goal is to estimate a relationship as follows:

$$(2.1)\ Q_{i,co-authors} = f(Q_i, X_{ij}) \text{for i=1, .., N and j=1, ..., M.}$$

where Q_i stands for the quality of researcheri, $Q_{i,co\_authors}$ represents the average quality of his co-authors and $X_{ij} = [X_{1ij} \quad X_{2ij}]$ stands for exogenous variables of researcher i.

Two main issues arise in this case. Firstly, we as individual research productivity or the quality of the co-authors are count data (see next section for a presentation of the data), so we should apply count data econometrics to the former specification. Secondly, the productivity level of an academic is not independent of that of his co-authors[5]. Thus, there is an endogeneity issue in the data which could be addressed by the Two Stage Residual Inclusion (2SRI) approach (Terza *et al.*, 2008).

2.1- Count data with overdispersion and Excess zeros: ZIP and ZINB modeling

Poisson regression models provide a standard framework to analyze count data[6]. However, in practice, count data suffer from two major drawbacks: overdispersion and excess of zeros. Overdispersion could stem from unobserved heterogeneity which causes the conditional variance of the sample to be larger than the conditional mean. The most frequently cited approach to address overdisperson is the negative binomial regression model. Another issue in count data modeling is a situation in which the number of zeros in the data exceeds what would typically be predicted by the Poisson distribution. Lambert (1992) has developed the Zero Inflated Poisson (ZIP) model to handle this case. In order to model both unobserved

[5]For example, in the case of a sample of French physicists, Mairesse& Turner (2005) demonstrate that individual productivity is explained by the quality of other researchers belonging to the same research center.
[6]See Ridout, Demétrio& Hinde (1998) for a review.

heterogeneity and excess zeros a Zero Inflated Negative Binomial (ZINB) model could be applied to the data (Greene, 1994).

Zero inflated models suppose that the data generating process is different for the sample values equal to zero and those positive[7]. There could be also a distinction between "*structural* zeros" (which are inevitable) and "*sampling* zeros" (which occur by chance)[8]. For example, we may assume that there are two different types of academicsin the sample: those who would never collaborate (for ideological reasons, for instance) and the others. Among those wishing to work with other economists, some have not collaborated since they have not found a researcher with whom working. Therefore there are two types of zeros among the observed values, but econometricians cannot distinguish between the two types of individuals. Lambert (1992)introduced the ZIP model in which the zeros values are the result of both a Poisson model and a logit decision process.

In the ZIP model there are two different latent variables: C_i the collaboration decision variable of academic i, and $Q^*_{i,co-authors}$ the potential quality level of his/her co-author. The observed quality level of the co-authors ($Q_{i,co-authors}$) is then a function of these two latent variables:

$$(2.2)\ Q_{i,co-authors} = \begin{cases} Q^*_{i,co-authors} & if \quad C_i = 1 \\ 0 & if \quad C_i = 0 \end{cases}$$

The probability function of the quality level of co-authors is then the following:

$$(2.3) \quad f\left(Q_{i,co-authors}\right) = \begin{cases} p_i + (1 - p_i) \times g(0) & if \qquad Q_{i,co-authors} = 0 \\ (1 - p_i) \times g\left(Q_{i,co-authors}\right) & if \qquad Q_{i,co-authors} > 0 \end{cases}$$

where$p_i \in [0,1]$ is the probability that academic i will not collaborate (or the probability of a structural zero), and g(.) is the probability function of the parent count model. Excess zeros occur whenever $p_i > 0$. The collaboration decision C_i will depend on a new latent variable C_i^* and it will be modeled with a logistic model:

$$(2.4) C_i = \begin{cases} 1 & if \quad C_i^* = X'_{1ij}\delta_1 + \varepsilon_i \geq c_i \\ 0 & if \quad C_i^* = X'_{1ij}\delta_1 + \varepsilon_i < c_i \end{cases}$$

Where X_{1ij} are the exogenous variables involved in the decision process, c_i is a threshold value and ε_i is a residual following a logistic density function. Accordingly, the probability of a structural zero is defined as follows:

$$(2.5)\ p_i = \frac{exp\left(X'_{1ij}\delta_1\right)}{1+exp\left(X'_{1ij}\delta_1\right)}$$

[7]See Garay et al. (2011) for a detailed analysis of zero inflate models.
[8]Staub & Winkelmann (2014) do the distinction between *structural* or *strategic* zeros and *incidental* zeros.

A fully parametric zero-inflated model is then obtained once the probability function of the parent count model is specified. If *g(.)* is a Poisson probability function, then we get the ZIP model (Lambert, 1992):

$$(2.6)\begin{cases} g\left(Q_{i,co-authors},\lambda_i\right) = \frac{exp(-\lambda_i)\times\lambda_i^{Q_{i,co-authors}}}{Q_{i,co-authors}!}, \quad \lambda_i > 0 \\ \lambda_i = exp\left(X'_{2ij}\delta_2\right) \end{cases}$$

Where X_{2ij} are the exogenous variables explaining the expected value of the quality level of academic i's co-authors. The mean of the zero-inflated count data model (i.e. the expected value of the quality of the co-authors) and its variance are then:

$$(2.7a)\ \mathrm{E}\left(\mathrm{Q_{i,co\text{-}authors}}/\mathrm{X_{1i}},\mathrm{X_{2i}}\right) = (1\text{-}\mathrm{p_i})\lambda_\mathrm{i} = \frac{\exp\left(\mathrm{X}'_{2ij}\delta_2\right)}{1+\exp\left(\mathrm{X}'_{1ij}\delta_1\right)}$$

$$(2.7b)\ V\left(Q_{i,co-authors}/X_{1i},X_{2i}\right) = (1-p_i)\lambda_i(1+p_i\lambda_i)$$

The ZINB model is obtained if the *g(.)* function is a negative binomial distribution function, the new probability function of the quality level of co-authors is then (Garay*et al*, 2011):

$$(2.8)\ f\left(Q_{i,co-authors}\right) = \begin{cases} p_i + (1-p_i)\times\left(\frac{\phi}{\lambda_i+\phi}\right)^{\phi} & if \quad Q_{i,co-authors} = 0 \\ (1-p_i)\times\frac{\Gamma(\phi+Q_{i,co-authors})}{\Gamma(Q_{i,co-authors}+1)\Gamma(\phi)}\left(\frac{\lambda_i}{\lambda_i+\phi}\right)^{Q_{i,co-authors}}\left(\frac{\phi}{\lambda_i+\phi}\right)^{\phi} & if \quad Q_{i,co-authors} > 0 \end{cases}$$

Where $\alpha \equiv \phi^{-1}$is the dispersion parameter and *Γ(.)* is the gamma function, then the first two conditional moments are defined as follows:

$$(2.9a)\ E\left(Q_{i,co-authors}/X_{1i},X_{2i}\right) = (1-p_i)\lambda_i$$

$$(2.9b)\ V\left(Q_{i,co-authors}/X_{1i},X_{2i}\right) = (1-p_i)\lambda_i(1+\alpha\lambda_i+p_i\lambda_i)$$

Unobserved heterogeneity is linked to the parameter α. Indeed if the coefficient α is different from zero then the conditional variance is no more equal to the conditional expectation. So there is unobserved heterogeneity in the data and the binomial model should be used instead of the Poisson model. Unobserved heterogeneity can be tested by a likelihood ratio test on parameter α with theVuong test.It is worth noting that the X_{2i} variables can be identical to the X_{1i} ones, overlap with X_{1i} or be completely distinct from X_{1i}. The parameters δ_2 and δ_1 can be interpreted respectively as the semi-elasticities of the parent model and the changes in the log-odds of strategic zeros.

The former specification implies that each subject is observed for the same time interval, referred to as the exposure. If different subjects have different exposures (t_i), then the natural logarithm of the exposure must be included as an offset, a covariate with regression coefficient set to 1 in the specification (Rabe-Hesketh & Skrondal, 2005):

$$(2.10)\ \lambda'_\mathrm{i} = \lambda_\mathrm{i}\times\mathrm{t_i} = \exp\left(\mathrm{X}'_{2ij}\delta_2 + \ln(\mathrm{t_i})\right)$$

The parameters of the zero-inflated model will be estimated by Full Maximum Likelihood (ML) framework.

2.2 Addressing the endogeneity issue in count data: the 2SRI approach

Instrumental variables (IV) methods are the most common framework for addressing the endogeneity. In linear models, the IV methodology corresponds to the Two Stage Least Squares (2SLS) which is a two step procedure. In non linear models, the Two Stages Prediction Substitution (2SPS) approach can be considered as the non-linear counterpart of the 2SLS estimation. 2SPS substitutes the endogenous regressors in the estimated equation with their consistent predicted values obtained in a first stage auxiliary regression. However, Wooldrige (2014) highlight that, when the conditional expectation model is non-linear, then 2SPS approach produces in general inconsistent estimates. He advocates applying the Two Stage Residual Inclusion (2SRI) approach which allows getting consistent estimates of the parameters in the structural regression. Terza et al. (2008) provide the formal proof of consistency for the 2SRI approach.

The 2SRI estimator has the same first-stage as the 2SPS. However, in the second stage regression, the endogenous regressors are not replaced. Instead, the first-stage residuals of the auxiliary regressions are included as additional regressors in the second-stage estimation. Recently, Geraci*et al* (2014) extends the 2SRI framework to count data models. They consider the following general non linear model for the conditional mean of the outcome (for instance the variable $Q_{i,co-authors}$):

$$(2.11)\quad E\left(Q_{i,co-authors}/x_i, x_{ei}, w_i\right) = M(x_i\beta + x_{ei}\beta_e + w_i\lambda) = M(x_i\beta + \sum_{s=1}^{S}\gamma_s\, x_{sei} + \sum_{s=1}^{S}\xi_s\, w_{si})$$

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

$$(2.12)\quad x_{eis} = r_s(v_i\xi_s) + w_{si} \qquad \text{s=1, ..., S}$$

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:

$$(2.13)\quad E\left(Q_{i,co-authors}/x_i, x_{ei}, \hat{u}_i\right) = M(x_i\beta + x_{ei}\beta_e + \hat{u}_i\psi)$$

Where $\hat{u}_i$ is a set of Sestimated residuals of the first stagefor individual i. Consistent standard errors of the second-sage parameters can be obtained by bootstrap (Wooldrige, 2014).

In count data models, there is no consensus on how to define the residuals. Geraci et al (2014) advocate to compute two different measures: the raw residual ($\hat{u}_{is} = x_{eis}\text{-}E[x_{eis}/w_i]$) and the standardized residual ($\hat{u}_{is}^{std} = \frac{x_{eis}\text{-}E[x_{eis}/w_i]}{(V[x_{eis}/w_i])^{1/2}}$). If x_{ei} are count data variables, then the first-stage auxiliary regression can be modelled either by a ZIP or by a ZINB model and the two conditional moments can be computed as stated in equations (2.7) to (2.9). The exogeneity of x_{ei} can be tested via a conventional Wald-type statistics forH_0: $\psi_1 = \psi_2 = \cdots = \psi_S = 0$.

Geraci et al (2014) extend the literature by analyzing the small sample properties of the 2SRI estimators in count data models. They use Monte Carlo simulations in order to study the power of the exogeneity test and measure the bias of the structural coefficients. Their results show that the 2SRI method has good finite sample properties. Their empirical evidence show that the power of the test is always higher using standardized residuals. Furthermore, applying standardized residuals lead to smaller bias in the endogenous regressors.

3 Data and Sample Characteristics

3.1- The Database

In order to study the relationship between the scientific performance of a researcher and the one of his/her co-authors, we built an original database mixing various sources of data. We used the "Tableau de classement du personnel enseignanttitulaire et stagiaire", economics section from the Ministry for Research (2004) to identify all the academics employed by French Universities on December 31th 2004. For each researcher, this official file allowed us to get information about gender, age, academic status (Full Professor and Assistant Professor) and the university assignment during the year of 2004. As it is now common in the literature, we will use Google Scholar citation indexes in order to compute individual research productivity (Bosquet & Combes, 2012 and 2013). There are large empirical evidences showing that citations have a more important effect on academic earnings than the number of publication (Hamermesh et al.,1982).

So at the beginning of 2012, we used the software "Publish or Perish" (PoP, Harzing 2010) to collect each academic CV from Google Scholar[9]. With this CV came information about the set of published papers, and, for each paper, the numbers of citations, the language used and the name of the authors. In a second round, we used PoP to compute the h and the g-indexes of each co-author. The raw data extracted from Google scholar present some important shortcomings. Authors with names identical to first names raise difficult problems of disambiguation. A query "Philippe Martin" would thus be credited indifferently with the work of Philippe M or Martin P. Authors with frequent French last name such as Petit are often

[9] In 2012, the software PoP offered the option to select papers according to specific subject areas. Our data record all papers identified within the option “Business, Administration, Finance, Economics” and “Social Sciences, Arts, Humanities”. Results obtained within the option “Business, Administration, Finance, Economics” only lead to similar conclusions (see Besancenot et al. 2015).

credited with papers from homonymous researchers. In the same way, married women who used different author names during their academic life often present underestimated academic resume. In order to avoid these difficulties, we removed from the database the name of any author for which the disambiguation was hazardous. From an initial number of 1830 names in the "Tableau de classement", we kept only 1566 researchers[10].

In order to complete the database, we got information about the research topics of each researcher by collecting the JEL codes of the papers included in our database and listed in Econlit. Finally, we used the dataset "Fichier Central des Thèses"[11] to identify the name of the PhD supervisor and the year of the PhD defense. For foreign PhD or unrecorded thesis in this dataset, information was obtained through individual searches on the net.

With this information, we then computed four different indexes of productivity for each researcher.

- *Individual Productivity Indexes*

We computed first the authors' h and g indexes in order to have a synthetic measure of both the quantity (number of papers) and the quality (number of citations by paper) of the researcher's academic production. By definition, the h index of an researcher is equal to x if x of his N papers have received at least x citations each, and the others (N-x) papers have received no more than x citations each (Hirsh, 2005). One drawback of this measure is that two different academics may exhibit similar h indexes even if their respective best papers get a very different number of citations. In order to address this limit, Egghe (2006) has proposed the g index as the (unique) largest number such that the top g articles received (together) at least g citations. By definition the g index is at least equal or larger than the h index.

There are large empirical evidences showing that citations have a more important effect on academic earnings than the number of publication (Hamermesh et al.,1982). However, given the load of criticisms addressed to the h and the g indexes (see for instance Bornmann and Daniel 2007), we also built two additional productivity indexes grounded on the quality of the medium in which papers were published (journals, books, or working papers series). Both indexes are built following the methodology used since 2005 by the juries of the French "Concours National d'Agrégation pour le recrutement des professeursd'économie" leading to the hiring of new full professors in economics (for a description of this nationwide competition, see Combes et al., 2013c). These indexes are defined as the sum of individual score values given to each recorded paper. The scores are computed as the ratio between the weight of the medium of publication and the square root of the number of the paper's authors.

The first index, denoted LLG index, is computed according to the discrete weight function implemented by the jury of the "Concours 2008" (Levy-Garboua 2008) and grounded on the CNRS (2012) ranking of economic journals. This ranking considered 6 categories of economic journals: the main journals are graded from 1 (the top tier journals) to 4 (the less

[10] This choice implies that some of the most productive researchers could have been removed from the database.
[11] The "Fichier Central des Thèses" is a French database created in 1968 to identify every thesis being prepared in French universities.

influential). Two additional categories: MAD (multidisciplinary) and NR (new journals) were also introduced in order to consider multidisciplinary or promising new journals. Following the framework advocated by the jury of the "Concours 2008", we therefore gave 6 points for each publication in a journal graded 1 by the CNRS, 4 for any journal with a grade of 2, 2 when the journal is credited with a grade of 3 and 1 for a grade of 4. Articles published in journals listed in categories MAD (multidisciplinary) and NR (new journals) by the CNRS were credited with a weight of 1. Finally, any publication listed in Econlit but not in the CNRS ranking received a weight of 0.5. This allowed distinguishing non producing researchers from active researchers publishing only in books or in journals with weak economic impact.

The second index, denoted CL, follows the same methodology but considers only papers listed in Econlit journals. In this index, weights are taken from the Combes and Linnemer (2010) ranking of the Econlit journals. In order to make their classification, Combes and Linnemer defined two scores values (Clm or Clh) from 0 to 100 for each of the 1205 journal listed in Econlit. Both scores reflect the same ranking of the journals but the CLh is more selective giving higher weights to the top tier journals and lower one to less influential journals. In our paper, the CL index is built according to the CLm weight[12]. Note that the CL index neglects papers published in books, working papers or journals ignored by Econlit and therefore gives an elitist measure of productivity. According to this methodology, a researcher with two papers: one paper written with one coauthor and published in the American Economic Review[13] and one paper published alone in a book would have respectively a LLG index equal to 4.74 ($= 6/\sqrt{2} + 0.5$) and a CL index of 69.37 ($= 98.1/\sqrt{2}$). These two additional variables will allow us to perform some robustness checks.

- *Co-authors index*

In order to summarize both the productivity and the number of a researcher's co-authors, we computed two Meta indexes denoted hh and gg by reference to the h and the g indexes. By definition, the hh index of a researcher will be equal to n if n of his/her N co-authors have at least a h index equal to n, and the other (N-h) co-authors have a h index less than n. In a same way, the gg index will be equal to n if the sum of the g indexes of his/her n best coauthors is superior or equal to n (the square of the rank) and the sum of the g indexes of the n+1 best coauthors is inferior to (n+1) .

These two indexes aim at giving in a one-dimensional variable a measure of both the number and the quality of a researcher's coauthors. A high hh index indicates that a researcher presents a high number of productive co-authors (with high h indexes). For instance, an author who has 5 coauthors with the following h index, h1 = 15; h2 = 12; h3 = 4; h4 = 3; h5 = 2 will have an hh index equal to 3. Indeed only three co-authors have a h index higher than

[12] The index built on the CLh weight scheme was too selective to provide conclusive results, specially when it was compared to the h and g indexes.

[13] The AER has a CLm weight equal to 98.1 in the Combes and Linnemer's ranking and it belongs to the category 1 list of reviews of the CNRS classification.

3[14]. The main advantage of these two Meta indexes is that they take into account both a quality and quantity dimension in the co-authorship issue. We will also analyze the total number of different co-authors ("NB_COAUTHORS" variable) of an academic i. This variable may be interpreted as the network size of an academic. Furthermore, as this variable measures only the quantity dimension, we will also analyze if there is a trade-off between quality and quantity when choosing coauthors.

- *Control variables*

For each author the following additional variables have been computed:

- ✓ "FEMALE" " is a dummy variable equal to 1 if the individual academic is a woman. This variable allows taking into account a gender effect on the publishing strategy of an individual, if any (the so-called gender productivity gap).
- ✓ "AGE" stands for the age of the individual so this variable could control for a kind of generation effect. The influence of this variable on the quality and quantity of co-authors may be ambiguous. Indeed, due to the increasing pressure for publication, "young" researcher may want to write more papers and therefore may be looking for an increasing number of co-workers. On the other hand, young researchers may want to signal their quality by avoiding publishing their first papers with co-authors.
- ✓ "NUMBER_YEARS" stands for the number of years since the PhD defense. This variable is a proxy for the academic professional experience. It is worth noting that for each academic, the h, g, hh or gg indexes are computed from the beginning of his/her academic career with a different time exposure for each individual of the sample. In the empirical model we will apply the professional experience variable as an offset one.
- ✓ Academic position: We also control for the rank position of the individual in the academic career: assistant professor ("MCF") or full professor ("PR") with dummy variables[15]. There are three type of rankings for full Professor: -"ClasseExceptionnelle" (PR_CE), "Première Classe" (PR_1C) and "SecondeClasse" (PR_2C) - and two for Assistant Professor- "Hors Classe" (MCF_HC) and "ClasseNornale" (MCF_CN)-. These variables could also reveal the quality of an academic as the promotion from one position to another one (say from MCF to PR_2C or PR_2C to PR_1C) depends on obtaining a national examination based in part on the number and the quality of publications.
- ✓ Language of publication: when papers are published in journals, we identified its language of publication and compute for each researcher the share of their papers published in English (SHARE_GB), in French (SHARE_FR) or in other languages (SHARE_OTHER).

[14] A high gg index also indicates that an author works with authors presenting high g indexes (some of the co-authors have published very influential papers). For instance, if the g index of these co-authors are g1 = 15; g2 = 12; g3 = 4; g4 = 3; g5 = 2; g6=0, then the gg index will be equal to 6: the sum of the g index for the 6 best co-authors is equal to 36 which is equal to 6 , the square of the rank.

[15] In French academia it remains a very old status that does not exist anymore: « *Maître Assistant* ». We merge academics with this status in the group of MCF_CN.

- ✓ "WORK_ALONE_ONLY": this dummy variable is equal to 1 if academic i has published at least one paper during his/her academic career and has always refused to co-author a paper.
- ✓ "NB_PAPERS" stands for the number of papers listed in google scholar for the individual researcher. This variable can be considered as a quantitative measure of the level of production of an individual academic.
- ✓ Papers' Quality: the CNRS classification of economics journals plays a dramatic role in the assessment of economic research in France, both at individual and institutional levels. So we applied the CNRS classification to measure papers' quality by splitting up the researchers papers in 7 different categories: CNRS_1 to CNRS_5, ECONLIT_NO_CNRS and MISCELLANEOUS_PAPERS. Variables CNRS_1 to CNRS_4 indicate the number of papers published by a given researcher in the four main categories of journals. CNRS_5 stands for papers published in multidisciplinary journals. The variableECONLIT_NO_CNRS records the publications in journals that are referenced by the Econlit database but not by the CNRS and, at least, variable MISCELLANEOUS_PAPERS counts all other items (papers published in journals not referenced by either the CNRS or the Econlit database, working papers and books).
- ✓ HERFINDAHL_JEL_CODE: The propensity to publish and to engage in co-authorship relationships varies greatly according to the various economic topics. Economic fields of research might exhibit very different rates of co-authorship because they do not belong to the same economic tradition (Witte & Schulze, 2009). Following a now standard methodology (see for instance Fafchamps et al., 2010, Kelchtermans & Veugelers, 2011 or Bosquet & Combes 2013a), we collected the Journal of Economic Literature (JEL) codes of the papers included in our database and listed in Econlit and we identified the economic topics through the letter of the papers' JEL Classification codes[16].We then computed a Herfindahl index of the different letters' JEL code used by a researcher in order to measure the researchers' degree of specialization: the higher the value of the Herfindahl index, the more the researcher is specialized. A Herfindahl index equal to 1 means that all the publications of the researcher are classified in only one JEL category[17].
- ✓ WORK_ALONE_ONLYis a dummy variable equal to 1 if academic ihas published at least one paper during his/her academic career and he has no co-authors.
- ✓ "COWRITE_DR" is a dummy variable equal to one if the academic has written at least one paper with his/her supervisor. This variable could be interpreted as a signal of quality for a young researcher; its effect is expected to be positive.
- ✓ Finally we also control for network effects with the two different variables. Firstly, the "UNIVERSITY_NAME" variable is the university assignment of the individual in 2004. This is a dummy variable equal to 1 if the individual works in the assigned university. In our dataset there are 90 different institutions (universities, "Grande Ecole"). We assume that belonging to academic institutions with large economic

[16] We also controlled on the researcher's main field of research considering this field as the one in which a researcher has the highest JEL figure. In this paper, this variable didn't prove to be significant.

[17] We compute a normalized Herfindahl index which means that this variable ranges from 0 (no specialization) to 1 (full specialization).

departments which are recognized nationally and internationally can facilitate the matching with complementary co-authors[18].

- ✓ Secondly, we also compute the "PhD DEFENDED AT" variable. This variable divides the set of authors into 11 categories according to the academic institution where the individual researcher has defended his PhD. It comprises nine French academic institutions (University of Toulouse 1, Paris 1, Paris 9, Paris 10, the others universities in Paris, Aix-Marseilles, Strasbourg, a group of 12 different "Grande Ecole" institutions and all the others French universities)[19] and two categories listing foreign institutions (the PhD has been defended in either an European country or in the United States).

- *Choice of Instruments*

It is well-known that IV estimators' efficiency relies on the quality of the instruments. Endogeneity issue arises because of unobserved heterogeneity in the data, possibly stemming from unobserved individuals' characteristics, which implies that the dependent variable is correlated with some regressors in the equation. So in our case a good instrument requires two assumptions: (i) to be highly correlated with the individual research productivity level and (ii) to be uncorrelated with the quality of co-authors. Getting such an instrument is a challenging issue because a lot of potential variables explain both individual and co-authors quality.

In this paper we will use as instrument the best quality paper published alone by an academic. This variable could be interpreted as a signal regarding the researcher's intrinsic skill level. Again we will turn to the CNRS classification in order to measure papers quality and we will rely on the seven categories already discussed. For example, "BEST_ALONE_CNRS1" is a dummy variable equal to 1 if the researcher has published alone at least one paper in the first category of the CNRS classification.

3.2- Some Descriptive Statistics

In Table 1 are reported some descriptive statistics of our database. In 2004, 28% of French academic economists were women and the average academic is 47 years old. There is an over representation of Assistant Professors in the data as only 35% of academics are Full Professors. In 2012 an average French academic has around 22 years of professional experience[20] and 22.5% of French academics have never produced a paper referenced by Google Scholar during their career (see table 1). In order to analyze the impact of different generations of academics, we have split-up the data set into 8 different cohorts. For example, the variable "Cohort64_68" is a dummy variable equal to 1 if the individual has defended his PhD dissertation between 1964 and 1968. Individuals belonging to this cohort have on

---

[18] As recently showed by Bosquet& Combes (2013b), the network effect should be better measured at the level of the economic departments rather than universities. However in our dataset, we were unable to obtain this information.

[19] This classification corresponds broadly to the international departments of French academia defined recently by Bosquet *et a.* (2013c).

[20] The professional experience variable ranges from 7 to 46 years.

average 44 years of post PhD experience against 10 years for individuals belonging to the last cohort ("Cohort99_04"). Around 1.3% of academics are members of the first cohort and they are on average 65 year old. For the last cohort, the figures are respectively 13.6% and 34 years old. Around 30% of French academics in our sample have started their career between 1983 and 1993 and 36% between 1994 and 2004.

Considering researchers' productivity indexes now, the average French academic economist has published around 8 papers in his career in the whole sample. The average number of papers is equal to 11 in the sub-sample of publishing academics[21]. There is a huge heterogeneity between academics as regards their scientific production as the number of papers ranges from 0 to 157. On average about 40% of the publishing academics have produced between 1 and 4 papers during their career[22]. As regards the quality dimension of the research outcome, the estimated mean of the h and g indexes are respectively 3.25 and 6.02. So according to the h index, French academics have published 3.25 papers on average with 3.25 citations each[23]. Again there is a huge heterogeneity in "quality" among French academics as the h and g indexes range from 0 to 39 and 0 to 84 respectively.

Some simple correlations between the different individual research productivity indicators are reported in table 2. All correlation estimates are significantly different from zero. As expected, there is almost a perfect correlation between the h and g indexes (0.95). It is worth noting that the correlation is also positive and high (0.66) between the citations scores (the h or g indexes) and the Econlit publication scores (the CL_index). Academics who have published more papers on top ranked journals receive more citations for each publication[24]. It is also worth noting that the LLG_index is more correlated with the quality measures of Econlit publication (0.90) than the citations scores (0.73).

Regarding the quality of publications, only 1.98% of the publishing academics have never published in a journal referenced either by the CNRS or by Econlit (see table 1). French academic economists who develop a research activity choose to publish in journals belonging to the CNRS classification as their career evolution depends mainly on that particular classification: it represents 85.2% of the total number of papers published in a journal. The quality of the papers as measured by the CNRS classification is quite low. Indeed around 79% of the total number of published papers belong to the two lowest quality categories (CNRS3 and CNRS4), whereas 13% belong to the CNRS2 category and only 7% to the top ranked

[21] The average numbers of papers per year of experience are respectively 0.46 and 0.59. This implies that on average each researcher has published one paper every two year. It is worth noting that this figure corresponds to the new requirement of the HCERES (Haut Conseil de l'Evaluation de la Recherche et de l'EnseignementSupérieur), the national agency for the research evaluation in France, to consider an academic as a publishing one.

[22] 52% of the publishing academics have produce between 1 and 6 papers during their career.

[23] Our results are in line with those obtained by Bosquet& Combes (2013a). Indeed these authors estimated an average g index of 7.25 on a sample of 2,782 French academic economists between 1969 and 2008. One may explain this discrepancy by different kinds of population under study. In our sample, we only take into account academics that have a position in a French university whereas in Bosquet& Combes' sample they add all full-time researchers from the CNRS and the INRA (Institut National de la RechercheAgronomique). These last two kinds of academics do not have teaching loads so they may have higher research productivity.

[24] These results are again in line with those obtained by Bosquet& Combes (2012) on a sample of 2,832 French economist academics for the year 2010.

category CNRS1[25]. Note that 62.5% of the published papers are drafted in French against 34.8% in English (see table 1). Again there are huge discrepancies between the different generations of academics: more than 75% (respectively 17%) of the papers are drafted in French (English) for academics that started their career before 1968 (our first generation) against only 52.9% (45.1% respectively) for those that started in 1999 (our last generation, see table 3).

We turn now to comment co-authorship indicators (the hh and gg indexes and the nb_coauthors variable). On average each French academic has 4.5 different co-authors in the whole sample and 6.9 co-authors in the sub-sample of academics which engage in co-authorship (see table 1)[26]. It is worth noting that 34.8% of the individuals in the sample have never had a co-author whereas 17.4% have never published a paper by their own. 44% of publishing academics of the sample have between 1 and 3 co-authors; and about 15% of the individuals in the sample have written at least one paper with their PhD supervisor. The mean of the hh Meta index which summarizes both the number and the quality of co-authors is 3.2 and the hh index ranges from 0 to 29. There is overdispersion in the data as the hh index variance is 11.7. A similar result is obtained with the gg index with an even larger range of variation: the mean and variance are equal respectively to 7.4 and 74.1 (see table 1).

As regards fields of economics, the specialization of French academics is quite low. The average value of the Herfindahl JEL Code index is 0.31, but there is a high heterogeneity between French academics (see table 1). Few academics have all their publications in one main field (only 8.2% of the sample). The ranking (by decreasing importance) of the most cited fields are the following: D (Microeconomics), O (Economic Development, Technical Change and Growth), F (International Economics), L (Industrial Organization), E (Macroeconomics and Monetary Economics, B (History of economic thought, Methodology and Heterodox Approaches) and J (Labor and Demographic Economics).

Cohort effects seem to play a critical role in France. According to Raubert & Ursprung (2008), countries whose national academic system has been subject to important institutional changes may be characterized by significant cohort effect in research productivity as it was the case in Germany. A similar argument may be raised in France as can be observed from statistics reported in table 3. We compute individual productivity and co-authorship indexes per experience year by entry cohort. All variables have a similar pattern and exhibit a structural break in the mid-eighties. For example, the number of co-authors per year remains quite stable for researchers belonging to the first four cohorts, then there is a huge increase for individuals entering in the academic career in the mid-eighties (cohort84_88). Since then there has been a steadily increase in the number of co-authors per year. A similar evolution is obtained for individual productivity indexes (h, g, CL_index and LLG_index). Younger cohorts are more productive and they engage more in co-authorship activities.

[25] This result might be explained by the fact that the majority of the French economic journals are classified by the CNRS as belonging mainly to the third and fourth category. In the 2008 classification, only one French economic journal is classified as category 2 and none is classified as category 1.

[26] The overdispersion issue is even more important in this case as the variance is equal to twelve times the mean.

We will use the best paper published alone as the instrumental variable in the paper. We created a set of dummy variables representing for each CNRS category the best publication reached by an acamedic by his own. Some descriptive statistics on this instrument are reported in the bottom part of table 1. As expected the number of researchers publishing alone declines when the quality level of journals increases. Indeed there are around 6% of the 1566 academics that have succeeded to publish alone at least one paper in the CNRS1 category, 9.5% in the CNRS2, 27% in the CNRS3 and 11% in the CNRS4.

4 - Discussions of Empirical Results and Sensitivity Analysis

Studying the relationship between the characteristics of a researcher and those of his/her co-authors is tricky as individual research productivity should be an endogenous regressor: the quality of a researcher's publication will depend somehow on the quality of his co-authors. According to Azoulay et al. (2010), co-authors of a superstar suffer a lasting 8 to 18% decline in their quality-adjusted publication output following his death. Furthermore, as regards the quality side measure of research output, co-authored papers are cited more frequently, and in the case of asymmetric partnerships collaborating with a higher quality author seems to pay off (Chung et al, 2009). Recently, Bosquet& Combes (2013a) with a panel of 2,782 French academic economists get the following results. Firstly, in 2008 a published academic who increase on average his/her number of co-authors from two to three will be cited 53.4% more and his/her g index will be 41.8% higher. Secondly, if the number of co-authors is increased by one (say from 5 to 6) then their own measure of the CL index will be increased by 4%.

So in order to evaluate the effect of the level of its own research productivity on the co-authors quality, this endogeneity bias should be addressed. We start by commenting the results of the 2SRI estimations and then we will check the robustness of our results.

4.1 Empirical Results: the 2SRI estimation outputs

2SRI is a two-step approach. In the first step, the endogenous regressors are modelled with the exogenous regressors and the instrumental variables in order to compute the standardized residuals. In the second step, the structural model is estimated and the standardized residuals are included as additional explanatory variables in the regression.

First-Stage empirical results: exogenous determinants of research productivity measures

Research productivity variables (both the h and g indexes) are explained by some exogenous demographic variables such as the age and the gender effect and by the instrumental variables. As there is overdispersion in the data (see table 1), ZINB modeling has been applied to research productivity indicators. Results are reported in table 4[27].The Vuong test compares zero-inflated models with an ordinary Poisson regression model (the null hypothesis). As the Vuong test rejects the null hypothesis whatever the productivity measure applied, a zero-inflated model is better than a Poisson regression (see table 4 column 1)[28]. Furthermore, the likelihood ratio test of Ho: $\alpha=0$ indicates that the null hypothesis of no unobserved heterogeneity is rejected at the 1% level whatever the productivity variable applied (again in the case of the h index, the statistic is equal to 589.77 with a p-value of 0%). So these two tests indicate that individual research productivity measures should indeed be estimated with a ZINB model.

Regarding the determinants of the decision to undertake some research activity, the inflate coefficients of the two productivity measures provides similar conclusions. The age and the gender variables have a significant positive effect on the log odds of an inflated zero, and the dummy variable "Working_alone_only" has a non-significant effect. In the case of the age variable, the estimates range from 0.16 to 0.22. For example, regarding the h index results, one additional year old will increase the log odds of an inflated zero by 0.22 (see the inflate model in table 4 column 1). In other words, the older the researcher, the more likely is the fact that not having published a paper is a deliberate decision. There is a gender effect as being a woman increases the log odds of not publishing by 2.33 in the case of the h index[29]. We have also introduced in the regression some cohort dummy variables in order to model life cycles in research productivity (see below for further details). In the case of the h index specification, five out of the six cohort dummies are significant against only three out of 6 for the g index case.

We turn now to comment the results for the parent model of the research quality measures based on citation scores. The level of these individual research productivity measures may be interpreted as the effort allocated by each academic to research. The number of year of professional experience (the "number_years" variable) is the offset variable as each researcher has different time exposure. The evidence regarding the existence of a gender effect is mixed as the female variable is significant in the h index specification but not in the g index specification (see table 4 columns 1 and 2). Being a woman decrease the expected h index estimate by 14% (1-0.86). The age variable has a negative impact on the research productivity of an individual as the IRR estimates are below 1 in both cases. So younger economists tend to be more productive: for example, the expected change of the h index if an individual has an additional year old is -0.012 (=ln(0.988491))[30]. However, the age effect on the research productivity is quite fragile as the age variable is significant only at the 10% level in the case of the h index. So contrary to Hamermesh(2015)'s results age by itself seems not to be an important determinant of French academics research productivity.

---

[27] According to Staub&Winkelmann (2013) identification of all parameters in a ZINM model is achieved if at least one variable in X_2 is not included in X_1.

[28] For example, in the case of the h index, the statistic is equal to 3.58 with a p-value equal to 0.2%.

[29] The estimate is lower for the other productivity measure, but it remains significant.

[30] We have also introduced in the specification the age squared variable, but this variable was never significant.

Research productivity may also vary across historical times because of institutional changes (Raubert & Ursprung, 2008). We include cohort dummies in the specification in order to capture this effect. There is strong evidence for the presence of life cycles in research productivity in our sample of French academics. The reference dummy cohort is the last one (i.e. academics that have defended their PhD after the year 1999), all the seven cohort dummies are significant whatever the measure of research productivity applied. Most of these dummy variables are significant at the 1% level. The IRR estimates of the cohort dummies are increasing over time implying that the coefficients are decreasing over time as expected (see table 4 columns 1 and 2). For example, an academic having defended his PhD between the years 1969-1973 exhibits a 41% lower h index than that of the reference group. If the PhD has been defended between during the period 1989-1993, the decrease in the h index is only around 22%. The reason of this phenomenon should be very similar to the one put forward by Rauber&Ursprung (2008) in the case of Germany. Over the last forty years, French academics have been increasingly exposed to the Anglo-Saxon research tradition that rewards researchers according to their own research productivity efforts. Furthermore, bibliometrics has become increasingly significant in evaluating individual researchers in French academic system since the nineties (Académie des Sciences, 2011).

Most importantly, our selected instrumental variables are significant and they have the expected signs on research productivity. Indeed the coefficients of the "best alone publication" dummies are increasing according to the quality of journals whatever the measure of research productivity applied. For example, in the case of the h index, the estimated IRR coefficients for "Best_alone_CNRS1" and "Best_alone_CNRS4" are respectively equal to 4.87 and 1.86. These results imply that an individual that has published alone in a journal belonging to the CNRS category 1 (the best quality) has on average a 387% increase in his expected productivity whereas the expected increase in productivity will be solely of 86% if his best publication alone is the CNRS category 4 (the lowest quality)[31].

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

In Table 5 are reported the empirical results of the determinants of co-authorship by the 2SRI methodology. We analyze the co-authorship decision both in terms of quality with the number of citations (the hh and gg indexes) and in terms of quantity (the number of co-authors). According to Geraci et al (2014)'s Monte Carlo results, for applied research the best model is a ZINB or a ZIP model with standardized residuals with non-corrected standard errors of the parameters. In every specification the 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. So the 2SRI methodology must be implemented in order to address the endogeneity issue in the data (see the bottom part of table 5, columns 1 to 4)[32].

[31] The reference here is not having published a paper in a journal.

[32] For instance in the case of the hh specification the Wald statistic is equal to 32.94 with a p-value equal to 0%.

In the case of the hh index, once a network effect is taken into account, then the likelihood ratio test no more rejects the null hypothesis of no unobserved heterogeneity. In this case the better model is the ZIP model. If the dependent variable is the gg index, then the selected model is the ZINB: indeed the likelihood ratio test the null of no heterogeneity is always rejected and the Vuong test also rejects the null hypothesis of a Poisson model (see table 5 columns 3 and 4).

Regarding the determinants of the collaboration decision, the inflate coefficient of the individual research productivity variable has always a negative and significant effect (see inflate model in table 5 columns 1 to 4). For example, if the co-authorship quality is measured by the hh index, then the inflate coefficient of h suggests that for each unit increase in h the log odds of an inflated zero decrease from a minimum value of 3.37 up to a maximum of 3.41 (table 5, inflate model columns 1 and 2)[33]. In other words, the higher the research productivity level of an academic is, and the less likely is the decision of never collaborate. More productive academics tend to engage more in co-authorships. The variable “Age” has a negative but never significant effect on the decision of not co-writing papers whatever the research productivity measure. Estimates of the gender effect on the collaboration decision are never significant. Again, some cohort dummies are also significant in the decision of engaging in co-authorship.

We turn now to comment the results for the parent model of the quality measures of co-authorship based on citation scores. Again, the professional experience variable is used as the offset variable. The h index (respectively the g index) of an individual has a positive and significant at the 1% level effect on the hh index (respectively the gg index). The IRR coefficient for the h index ranges from 1.109 to 1.126 (table 5, columns 1 and 2), implying that if the h index increases by 1%, then the expected value of co-authors quality increase will range from 10.9% to 12.6%. In the case of the gg index, estimates are lower (around 2.1%) but they remain significant (columns 3 and 4). These results confirm the assortative matching hypothesis that the quality in research of the co-authors of an economist will depend on his own research productivity.

As expected the age variable has a significant and negative impact on the quality of the co-authors at the 1% level[34]. The average of the IRR coefficient estimates is around 0.985: thus the expected decrease of quality of co-authors if an individual has an additional year old is 1.5% ($= -ln(0.985)$). Again the existence of a gender effect is quite fragile. The gender variable is only significant with the hh measure of co-authorship quality. In this case, being a woman reduces the expected quality of co-authors by around 15%. Publishing a paper with his/her PhD supervisor has a positive and significant effect (at the 1% level) on the quality of the future co-authors. The estimated IRR for “cowrite_dr” is around to 1.26, implying that an individual that has published at least one paper with his/her PhD supervisor has on average a 26% increase of the expected quality of his co-authors. This result could be explained by a reputation effect. Co-writing with his/her student may signal an implicit recognition by a

[33] As regards the gg index, the estimated coefficients are around -0.75.
[34] We have also introduced in the specification the age squared variable, but this variable was never significant.

supervisor of the quality of the student. For a junior researcher in French academia, publishing with his PhD supervisor may be a signal of efficiency. We thus confirm the Bidault& Hildebrandt (2014)'s result that a junior researcher benefits from joining an asymmetric team. We estimate a new mechanism which goes through an increase in the quality of future co-authors for the junior academic.

The rank position has also a positive and significant effect on the quality of co-authors. Compared to an assistant professor at the standard level (MCF_CN), being a full professor improves the hh index (respectively the gg index) by 20.4% (resp. 27.2%) for a full professor at the highest level (PR_CE) and by 13.25% (12.9%) for a full professor at the first level (PR_2C)[35].

As in the case of the individual research productivity, there are significant life-time cycles in the quality of co-authors. Indeed the IRR coefficients of the cohort dummies are increasing over time either with the hh index as dependent variable or with the gg index. Thus younger cohorts are collaborating more and with better quality co-authors. As most cohorts dummies are non-significant in the inflate model, this implies that the life-time cycles should have a very small impact on the probability of never collaborating with other economists but they have a positive effect over time on the expected co-authors quality level.

Interesting results are obtained regarding the link between the number of papers, their quality and the co-authorship decision. As expected the CNRS classification of journals has a huge impact on the selection process of co-authors, but surprisingly the effect of journals quality goes in the reverse direction of that expected. Indeed, the IRR coefficient of the number of papers published in top quality level journals (CNRS1) is significant but always below one whatever the measure of the quality of co-authors applied (hh index or gg index), implying that the number and the quality of co-authors decrease with the increase in the number of papers published in the CNRS1 category. On the contrary, the IRR coefficient is significant (only in the case of the gg index) an above one (in all cases) as regards papers published in the CNRS2 category, implying an increase in the number and in the quality of co-authors in this case. In general, the number of papers published in the others CNRS categories (CNRS3, CNRS4, and CNRS5) have a non-significant effect[36].These results could be interpreted in the following way: academics may have a strategic behavior when dealing with the collaborating issue. To successfully publish in journals classified as CNRS2, an academic decides to collaborate with a higher number of coauthors or with more productive coauthors[37]. On the contrary, academics who have a sufficient level of research productivity to publish in top

[35] Oddly enough, being an assistant professor at the last level (MCF_HC) reduces by around 12% the hh index and by 11% the gg index but coefficients are not significant. This result might stem from the fact that the transition from MCF_CN to MCF_HC should depend more on the age of an individual than on the quality of his publications in the French academic system.

[36] If the dependent variable is the gg index, then the number of papers published in the CNRS3 category has also a significant effect.

[37] Another interpretation is the following:French academics who are trying to signal their skills ability in research by publishing in English journals ranked mainly in CNRS2 classification decide to collaborate with more coauthors of better quality.

ranked journals (here CNRS1 journals) prefer working with less coauthors in order to reap a higher prestige of the publications[38].

Publication language also plays an important role in the model. This result is not surprising because in general English journals have a higher impact factor than French ones in all rankings (CNRS, Combes & Linnemer). The IRR coefficient of the share_gb variable is larger than one and highly significant: an increase by 1% in the share of paper written in English raises the hh index (respectively the gg index) by 32.4% (48.9%). Two arguments may explain this relationship. Firstly French academic economists may be interested in developing international collaborations in order to produce papers in better English and to publish in more influential journals (see Olney (2015) for the influence of English proficiency on research performance). Secondly proficiency in English is a necessary condition to meet efficient co-authors and engage collaboration with more authors of better quality. In both cases, publishing in English contributes to increase the number and the quality co-authors.

Finally economic fields are also an important determinant in the co-authorship issue. As expected an increase in the level of economic specialization tends to reduce the number and the quality of co-authors: an increase in the JEL Herfindahl index by 1% decreases the hh or gg indexes by around 20%. Being a very specialized academics limits the number of efficient co-authors with whom working.

With the hh or gg indexes both the number and the productivity of a researcher's coauthors are measured. We turn now to a strict quantitative measure, namely the total number of co-authors. In column 5 of table 5, the number of co-authors is explained by the h index whereas in column 6 it is explained by the g index. The Wald tests again reject the null hypothesis of exogeneity of the h and g indexes. Results are relatively similar to those obtained with the quality measure of co-authorship except for the individual research productivity variable.

As regards the number of co-authors, there is no difference between male and female academics. Again, there is a significant age effect: younger economists tend to collaborate with more co-authors. The number of co-authors is also marked by significant cohort effects as all cohort dummies are significant. The IRR estimates are increasing over time: younger cohorts are collaborating with more co-authors. Finally, the most striking results are that the h or g indexes estimates are no more significant (see table 5 columns 5 and 6). Thus individual research productivity is mainly a determinant of the quality of co-authors and has no effect on the quantity measure.

4.2 Sensitivity Analysis

We run a sensitivity analysis by relying on Econlit publication scores as a measure of the individual research productivity, namely the CL and LLG indexes. So we will test for the robustness of our results to different measures of individual research productivity. Our comments will emphasis mainly the impact of these indexes on the quality and the number of

[38] Authors who publish in the top tier journals might also have difficulties to find efficient co-authors.

co-authors only. Again we will use the best alone publication as the instrumental variable. The first stage estimates are reported in table 4 (columns 3 and 4) and the 2SRI results are reported in table 6. In the first stage estimation, we have applied a Heckman selection model as the individual research productivity measures are no more count data. In both cases, the Wald test of the null of independent equations is rejected at least at the 5 % level: this clearly justifies the Heckman selection equation in our data. Results are very similar with these two different measures of productivity. Most importantly, our instruments are highly correlated with the productivity measures (see table 4)[39].

According to the Wald tests, the null hypothesis of exogeneity is always rejected (see table 6). Thus individual research productivity index measured by the quality of Econlit publications is also endogenous to the number and the quality of co-authors. Once this bias has been corrected by applying the 2SRI methodology, quality of Econlit publications and citation scores produce similar results as regards the main determinants of co-authorship. There is a significant gender effect: being a woman reduces by about 10% (respectively 3 or 4%) the quality (respectively the number) of co-authors (see table 6). The age variable has still a negative and significant effect on both the quality and quantity of co-authors. Life-time cycles are also significant: again as the cohort dummies estimates are increasing over time, younger economists are collaborating with a higher number co-authors. The specialization index and the language of the publications have the same impact as already estimated. The most important results is that the CL and LLG indexes have a positive and significant effect on the co-authorship variables[40]. Thus the conclusion that more productive academics will collaborate with higher quality co-authors which is the assortative matching hypothesis is a robust one.

5- Conclusion

This paper aims at estimating the determinants of co-authorship in economics. More specifically, we test the existence of a potential relationship between the research efficiency of an individual and that of his co-authors, the so-called assortative matchning hypothesis, using a novel database of French academics. However, individual research productivity should be an endogenous regressor as the quality of a researcher's publication will depend somehow on the quality of his co-authors. Furthermore, as regards the quality side measure of research output, co-authored papers are cited more frequently and in the case of asymmetric partnerships collaborating with a higher quality author seems to pay off. In the empirical model, we take into account this endogeneity bias by applying the 2SRI methodology. In this paper we have applied as instrument the best quality paper published alone by an academic. This variable could be interpreted as a signal send by an individual researcher to the academia

[39] The main differences between individual research productivity measured either by Google scholar citations or by publication scores rely on the gender effect and to a lesser extent on the cohort effect. Being a woman reduces significantly both the CL and the LLGindexes. Furthermore the dummy cohort variables are less significant with the LLG index. Therefore, a cohort effect seems to be at work mainly with citation scores index.

[40] For example, if the CL index increases by 1% then the quality of co-authors (the hh index) will be increased by 0.16% and the number of co-authors by 0.25%.

regarding his intrinsic skill level. We apply the CNRS classification in order to measure papers quality.

We then have estimated the relationship between the researchers' h index or g index and a meta-index synthesizing the number and the efficiency of his/her co-authors. Our main finding is that there is a positive and robust relationship between the h index of a researcher and his/her hh meta index. So the assortative matching hypothesis is confirmed in French academia. Furthermore, we obtained a new interesting result. Indeed we proved that individual research productivity is mainly the determinant of the quality and not of the quantity of co-authors. These conclusions shed a new light on the collaboration issue in academic activities. While co-writing is generally seen as the way for low skilled researchers to increase the quality and the quantity of their research output, our paper shows on the opposite that the quality of his/her co-authors constitutes a "good" signaling device for the quality of a researcher. In order to improve the information about academic's ability, co-authorship should be encouraged.

Others factors appear important in the decision of co-authoring a paper. There is a significant gender effect: being a woman has no impact on the probability of never collaborating with other economists but it decreases both the quality and the quantity of co-authors. This is a new mechanism that could explain the gender productivity gap in science. As expected the academic position of an individual has strong impact on the expected quality of co-authors: Full professors tend to collaborate more with higher productive co-authors. Life-time cycles are also an important determinant of the co-authorship issue. We have demonstrated that younger cohorts of French academics are collaborating with more co-authors who are also more productive. This result has a very important economic policy implication. Firstly, national public agencies should apply different evaluation criteria according to the time cohort of the academic. Secondly, institutional changes occurring at the aggregate level have a huge impact on individual research productivity. Finally, it is worth noting that publishing with his/her PhD supervisor contributes to increases the quality of future co-authors. This is a new mechanism from which a junior academic can benefit from joining an asymmetric team.

In order to be fully conclusive, additional variables should be considered to analyze a wider dimension of the publication activity. For instance, the size of the institution hiring the researcher, the influence of the academic network (national or international) or the research topics on the academic fellows' resume should be also considered to evaluate the robustness of our results. This is let for future research.

Acknowledgements

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

REFERENCES

Acedo, F.J., Barroso, C., Casanueva, C., Galan, J. L., 2006. Co-Authorship in Management and Organizational Studies: An Empirical andNetworkAnalysis. Journal of Management Studies, 43(5), pp. 957-983.

Azoulay, P., Graff Zivin, J.S., Wang, J., 2010. Superstar extinction. Quarterly Journal of Economics 125, 549–589.

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., 2001. Reflections on Scientific Collaboration (and Its Study): Past, Present and Future. Scientometrics, 52, pp. 365-377.

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

Bornmann, L., H-D. Daniel, 2007. What do we know about the h index?. Journal of the American Society for Information Science and Technology, 58(9), pp. 1381–1385.

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

Butler, R.J., 2007. JRI, JF, and the internet: Coauthors, New authors and the empirical research. Journal of Risk and Insurance, 74, pp. 713-37.

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

Bosquet C., P. P. Combes, 2013b. Do large departments make academics more productive? Agglomeration and peer effects in research. CEPR Discussion Paper n° 9401.

Bosquet C., P. P. Combes, 2012. Un panorama de la recherche française en économie comparant les approches Google Scholar et Econlit. Revue d'économie politique, Vol. 122, p. 477-545.

Bosquet C., P. P. Combes, C. Garcia-Penalosa, 2013. Gender and competition: evidence from academic promotion in France. CEPR Discussion Paper n° 9711, october.

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

Card, D., S. DellaVigna, 2013. Nine Facts about Top Journals in Economics. Journal of Economic Literature, 51(1), pp. 144-161.

Cardoso, A. R., Guimaraes, P., K.F. Zimmermann, 2010. Trends in Economic Research: An International Perspective. Kyklos, 63(4),pp. 479-494.

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 Academe. Quarterly Review of Economics and Finance, 49, 893–916.

Combes P.P., Linnemer L., 2001. La Publication d'Articles de Recherche d'Economie enFrance. Annales d'Economie et de Statistique, n° 62, p.5 – 47.

Courtault J.M., Hayek N. , Rimbaux E., Zhu T., 2010. Research in Economics andManagement in France: A Bibliometric Study Using the h–index. Journal of Behavioral and Experimental Economics, 39(2), p. 329 – 337.

Diamond, A. M., 1985. The money value of citations to single-authored and multiple-authored articles. Scientometrics, 8, pp. 315-320.

Ductor, L., 2012. Does co-authorship lead to higher academic productivity.Working Paper. Massey University, New Zealand.

Egghe, L., 2006. Theory and practise of the g-index. Scientometrics, vol.69, No 1, pp. 131-152.

Fafchamps, M., Van Der Leij M. and R. Goyal, 2006. Scientific networks and co-authorship. Discussion Paper Series n°256, University of Oxford.

Garay A.M, Elizabeth M. Hashimoto, E.M.M. Ortega, V.H. Lachos, 2011. On estimation and influence diagnostics for zero-inflated negativebinomial regression models. Computational Statistics and Data Analysis, vol 55 pp 1304–1318.

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, Università di Bologna, Working Paper DSE n° 921.

Glanzel, W., 2002. Coauthorship Patterns and Trends in the Sciences (1980-1998): A Bibliometric Study with Implications for Database Indexing and Search Strategies. Library Trends 50(3), pp. 461-473.

Greene W., 1994. Accounting for excess zeros and sample selection in Poisson and Negative Binomial regression models. Department of Economics, Stern School of Business, New York University.

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.

Hamermesh,D.S., Johnson G. E., Weisbrod B. A., 1982. Scholarship, Citations and Salaries: Economic Rewards in Economics. Southern Economic Journal, 49, p. 472-81.

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 scientificresearch output. PNAS 102(46), pp. 16569-16572.

Hudson, J., 1996. Trends in Multi-authores Papers in Economics. The Journal of Economic Perspectives, 10 (3), pp. 153-158.

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

Lambert D., 1992Zero-Inflated Poisson regression, with an application to defects in manufacturing. Technometrics, vol 34, pp. 1-14.

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

Liebowitz, S., J. Palmer, 1983. Assessing Assessments of the RelativeQuality of Economics Departments. Quarterly Review of Economics and Business, 28(2), pp 88-113.

Mairesse, J., Turner, L., 2005. Measurement and Explanation of the Intensity of Co-Publication in Scientific Research An Analysis at the Laboratory Level. NBER working paper series no. 11172, National Bureau of Economic Research, Cambridge.

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 BetweenCo-Authors in Economics. A Production Function Approach. The American Journal of Economics and Sociology, 66(2), pp. 289-308.

Petry, G. H., 1988. A Statistical Analysis of Worldwid Co-authorship Relationships in Scholarly Journals of Business. Journal of Economics and Business, 40(2), pp. 169-176.

Rabe-Hesketh S., A. Skrondal, 2005. Multilevel and Longitudinal Modeling using Stata. Published by Stata Press Publication, StataCorp LP, CollegeStation, Texas USA.

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.

Ridout M., Demétrio C. and J. Hinde, 1998. Models for count data withmany zeros. Proceeding of the XIXth International Biometric Conference, CapeTown, December, Invited Paper, pp. 179-192.

Schubert, A., 2012. A Hirsh-type index of co-author partnership ability. Scientometrics, 91(1), pp. 303-308.

Staub E. K., R. Winkelmann, 2013. Consistent estimation of zero-inflated count models. Health Economics, 22(6), pp 673-6886.

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

Stephan, P. E., 1996. The Economics of Science. Journal of EconomicLiterature, vol. 34(3), pp. 1199-1235, September.

Kelchtermans, S., R. Veugelers, 2011. The great divide in scientific productivity: Why the average scientist does not exist. Industrial and corporate Change, 20(1) pp 295-336.

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.

Tol, R.S.J., 2008. A rational, successive g-index applied to economics departments in Ireland. Journal of Infometrics, vol. 2, pp. 149--155.

Witte, S., G.G. Schulze; 2009. Do Birds of a Feather Flock together? Hiring Patterns of academic economists in Germany. Mimeo University of Saarlandes, Germany.

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

Variable	Obs	Mean	Std. Dev.	Min	Max
hh	1566	3.210728	3.428919	0	29
gg	1566	7.43742	8.611137	0	83
nb_coauthors	1566	4.510217	7.447595	0	60
h	1566	3.251596	3.991914	0	39
g	1566	6.02235	8.130808	0	84
CL_index	1566	33.77775	96.90919	0	481.349
LLG_inde	1566	13.72596	29.03717	0	76.6924
age	1566	46.69796	10.20819	28	68
Prof_experience	1566	21.92593	9.360218	7	46
cohorte64_68	1566	.0127714	.1123225	0	1
cohorte69_73	1566	.0606641	.2387894	0	1
cohorte74_78	1566	.1264368	.3324471	0	1
cohorte79_83	1566	.1392082	.3462742	0	1
cohorte84_88	1566	.1085568	.3111818	0	1
cohorte89_93	1566	.1890166	.3916469	0	1
cohorte94_98	1566	.2273308	.4192419	0	1
cohorte99_04	1566	.1360153	.3429143	0	1
Female	1566	.2828863	.4505455	0	1
Male	1566	.7171137	.4505455	0	1
Full Prof.	1566	.3473819	.4762904	0	1
Ass. Prof.	1566	.6526181	.4762904	0	1
never_published	1566	.2247765	.4175684	0	1

cowrite_dr	1566	.1487867	.3559918	0	1
nb_papers	1566	8.360792	14.17236	0	157
nb_papers_Review	1566	6.182631	9.071662	0	94
nb_papers_other	1566	2.178161	6.367514	0	91
never_published_Rev	1566	.0197957	.139342	0	1
nb_papers_Rev_cnrs	1566	5.376756	8.205418	0	85
share_Rev_cnrs	1183	.8520602	.2479001	0	2
share_cnrs1	1133	.0679517	.1574878	0	1
share _cnrs2	1133	.1335283	.1982947	0	1
share _cnrs3	1133	.4583329	.3324206	0	1
share _cnrs4	1133	.3349856	.3631996	0	1
share _cnrs5	1133	.0052015	.0360094	0	.6
share_papers_fr	1183	.625243	.3329255	0	1
share_papers_gb	1183	.3474892	.3250533	0	1
share_papers_other	1183	.0272677	.1147627	0	1
working_alone_only	1566	.1232439	.328822	0	1
never_working_alone	1566	.1736909	.3789645	0	1
Herfindahl_JEL_code	1566	.3073441	.2636275	0	1
max_kwa	1566	0	0	0	0
max_kwb	1566	.0102171	.1005942	0	1
max_kwc	1566	.0006386	.0252699	0	1
max_kwd	1566	.0070243	.0835427	0	1
max_kwe	1566	.0102171	.1005942	0	1
max_kwf	1566	.0102171	.1005942	0	1
max_kwg	1566	.0019157	.0437408	0	1
max_kwh	1566	.0025543	.0504914	0	1
max_kwi	1566	.0025543	.0504914	0	1
max_kwj	1566	.0038314	.0617995	0	1

max_kwk	1566	0	0	0	0
max_kwl	1566	.0031928	.056433	0	1
max_kwm	1566	.0012771	.0357257	0	1
max_kwn	1566	0	0	0	0
max_kwo	1566	.0057471	.0756158	0	1
max_kwp	1566	.0031928	.056433	0	1
max_kwq	1566	.0025543	.0504914	0	1
max_kwr	1566	.0051086	.0713141	0	1
max_kwt	1566	0	0	0	0
max_kwy	1566	0	0	0	0
max_kwz	1566	.00447	.0667297	0	1
best_alone_C1	1566	.059387	.236423	0	1
best_alone_C2	1566	.0951469	.2935114	0	1
best_alone_C3	1566	.2662835	.4421555	0	1
best_alone_C4	1566	.1091954	.3119838	0	1
best_alone_C5	1566	.0006386	.0252699	0	1
best_alone_Econlit	1566	.0440613	.2052969	0	1
best_alone_misc	1566	.0268199	.1616085	0	1

Table 2: Correlations between individual research productivity indicators

	h	g	CL_index	LLG_index
h	1.0000			
g	0.9503*	1.0000		
CL_index	0.6636*	0.6372*	1.0000	
LLG_index	0.7302*	0.6838*	0.8970*	1.0000

* All correlations are significant at the 5% level. N=1566

Table 3: Individual research productivity indicators per year and by cohorts

cohorte variable	mean	sd	min	max
cohorte64-68hh_peryear	.0923135	.0817156	0	.2608696
gg_peryear	.2161584	.1898998	0	.6222222
h_peryear	.1251731	.0997574	0	.3913043
g_peryear	.2227694	.207504	0	.8260869
LLG_index_peryear	.4454247	.5097879	0	2.037797
CL_index_peryear	.8343734	1.033474	0	4.509589
nb_papers_peryear	.2942663	.3880721	0	1.673913
nb_coauthors_peryear	.1229395	.2368148	0	.9130435
share_gb	.1710219	.1875821	0	.5555556
share_fr	.7564798	.1953899	.3333333	1
never_published	.2	.4103913	0	1
female	.1	.3077935	0	1
cohorte69-73hh_peryear	.0929661	.1111759	0	.5641026
gg_peryear	.2210608	.2746659	0	1.589744
h_peryear	.1056686	.1228009	0	.7435898
g_peryear	.196483	.2491615	0	1.641026
LLG_index_peryear	.4949158	.8717781	0	6.357573
CL_index_peryear	1.135906	2.575137	0	19.68135
nb_papers_peryear	.2589028	.399317	0	2.051282
nb_coauthors_peryear	.1008967	.1949082	0	1.131579
share_gb	.2084839	.2715722	0	1
share_fr	.7277525	.3117807	0	1
never_published	.2631579	.4426835	0	1
female	.1052632	.3085203	0	1
cohorte74-78hh_peryear	.0809532	.1051675	0	.71875

```
gg_peryear|  .2075475  .3257318         0  2.441176

h_peryear|  .0951696  .1210431         0  .7142857

g_peryear|  .1746357  .2553125         0   1.84375

LLG_index_peryear|  .4144785  .8375098         0  5.029756

CL_index_peryear|  .9674282  2.448345         0   18.8523

nb_papers_peryear|  .2262017   .444681         0  2.485714

nb_coauthors_peryear|  .0984183  .1981759         0   1.21875

share_gb|  .2348792  .2979328         0         1

share_fr|  .7313249  .3142713         0         1

never_published|  .2676768  .4438704         0         1

female |  .1111111  .3150663         0         1

----------------------------------------------------------------------------

cohorte79-83hh_peryear |  .0871213  .1201428         0  .6666667

gg_peryear|  .2099158  .2938712         0  1.666667

h_peryear|  .0912466  .1464147         0  .7407407

g_peryear |   .174917  .2953089         0  1.740741

LLG_index_peryear |   .434929  .9996627         0  8.562396

CL_index_peryear|  .8879347  2.604961         0  22.54853

nb_papers_peryear|  .2075978  .4104167         0  2.964286

nb_coauthors_peryear|  .1142663  .2397139         0  1.555556

share_gb|  .2380916  .2938016         0         1

share_fr|  .7153819    .32319         0         1

never_published|  .3899083  .4888517         0         1

female |  .2155963  .4121819         0         1

----------------------------------------------------------------------------

cohorte84-88hh_peryear |   .161737  .2078012         0  1.208333

gg_peryear|  .3747277  .4848179         0  2.772727

h_peryear |   .157683  .2389585         0     1.625

g_peryear|  .2884432  .4640665         0       3.5

LLG_index_peryear|  1.020523  2.418725         0  18.24871
```

```
CL_index_peryear|  2.574519  8.349001         0  61.72287
nb_papers_peryear|  .4309983   .882481         0  6.304348
nb_coauthors_peryear|  .2328522  .4375852         0       2.5
share_gb|  .3418605  .3280554         0         1
share_fr|  .6336796  .3337274         0         1
never_published|  .2882353  .4542793         0         1
female |  .2058824  .4055394         0         1
-----------------------------------------------------------------------------
cohorte89-93hh_peryear |  .1753634  .1747176         0  .9473684
gg_peryear|  .3967337    .42326         0  2.190476
h_peryear |   .182196  .2249889         0       1.5
g_peryear|  .3379134  .4548101         0  3.555556
LLG_index_peryear|  1.202878  2.344683         0  20.15555
CL_index_peryear|  2.094658   5.73278         0  69.01778
nb_papers_peryear|  .5226029  .8686731         0  8.263158
nb_coauthors_peryear|  .2969896  .4593647         0  2.894737
share_gb|  .3594232    .31641         0         1
share_fr |     .6215  .3266946         0         1
never_published|  .2297297  .4213714         0         1
female |   .347973   .477134         0         1
-----------------------------------------------------------------------------
cohorte94-98hh_peryear |  .2321876  .1929647         0  1.307692
gg_peryear|  .5267201  .4591398         0  2.923077
h_peryear|  .2192839  .1945358         0    1.3125
g_peryear|  .4058047  .4235059         0  3.066667
LLG_index_peryear|  1.320411  1.960414         0  18.70189
CL_index_peryear|  1.872301  3.235122         0  34.11071
nb_papers_peryear|  .5752364  .7581522         0  5.538462
nb_coauthors_peryear|  .3379599  .4525615         0  3.466667
share_gb|  .4117107   .329633         0         1
```

```
share_fr|  .5704174  .3349058         0         1

never_published|  .1404494  .3479417         0         1

female |  .3960674  .4897671         0         1

----------------------------------------------------------------------------

cohorte99-04hh_peryear |   .358184  .2572319         0       1.6

gg_peryear|  .7796462  .6420484         0       3.9

h_peryear|  .3126832  .2864607         0       2.1

g_peryear|  .5729656   .605065         0         5

LLG_index_peryear|  1.905336  2.112206         0  12.06301

CL_index_peryear|  2.099203  2.878418         0  18.47889

nb_papers_peryear|  .7562915  .8380927         0       6.4

nb_coauthors_peryear|  .4621283  .5000604         0       2.9

share_gb|  .4511526  .3277379         0         1

share_fr|  .5293931  .3284885         0         1

never_published |   .084507   .278802         0         1

female |  .3896714  .4888245         0         1

----------------------------------+-----------------------------------------

Total   hh_peryear|  .1813883  .2000601         0       1.6

gg_peryear|  .4130868   .477887         0       3.9

h_peryear|  .1766795  .2154634         0       2.1

g_peryear|  .3265625  .4407232         0         5

LLG_index_peryear|  1.046137  1.915127         0  20.15555

CL_index_peryear|  1.712053  4.433391         0  69.01778

nb_papers_peryear|  .4561684   .738855         0  8.263158

nb_coauthors_peryear|  .2571403  .4146959         0  3.466667

share_gb|  .3474892  .3250533         0         1

share_fr |   .625243  .3329255         0         1

never_published|  .2247765  .4175684         0         1
```

Table 4: First-Stage Estimations: Results for the individual research productivity variables

	(1)		(2)		(3)		(4)	
Dependant variable:	H		G		CL_index		LLG_index	
Model	ZINB		ZINB		Heckman Selection		Heckman Selection	
	IRR	P>|z|	IRR	P>|z|	Coeff.	P>|z|	Coeff.	P>|z|
Age	.988491*	0.056	.9926215	0.263	-.2711126	0.573	-.0020783	0.988
Female	.8630865**	0.014	.9326821	0.300	-9.133664***	0.005	-2.749063***	0.010
cohort64-68	.639886*	0.079	.5788128*	0.056	11.36906	0.523	-16.63519***	0.002
cohort69-73	.5865155***	0.010	.5446831***	0.006	29.23648*	0.103	-6.493137	0.218
cohort74-78	.5313103***	0.000	.5011169***	0.006	22.62574	0.115	-8.24296**	0.033
cohort79-83	.5673286***	0.001	.6145378***	0.008	29.77376**	0.030	-1.641273	0.676
cohort84-88	.6655879***	0.005	.6295683***	0.002	62.13896***	0.004	8.849404	0.110
cohort89-93	.7785253**	0.015	.7785826**	0.034	31.64009***	0.003	5.458532**	0.039
cohort94-98	.8393262**	0.035	.830293**	0.048	18.24714***	0.001	3.227771*	0.057
Working_alone_only	.3966277***	0.000	.4517704***	0.000	-41.56424***	0.000	-16.34336***	0.000
Cowrite_dr	1.417472***	0.000	1.427765***	0.000	9.533981	0.178	2.903077*	0.094
Best_alone_CNRS1	4.871913***	0.000	4.244295***	0.000	211.1036***	0.000	66.26043***	0.000
Best_alone_CNRS2	3.446526***	0.000	3.04466***	0.000	42.6116***	0.000	22.48153***	0.000
Best_alone_CNRS3	2.274868***	0.000	1.823829***	0.000	15.86695***	0.000	9.011317***	0.000
Best_alone_CNRS4	1.855814***	0.000	1.515237***	0.000	3.246687	0.369	1.191356	0.319
Best_alone_CNRS5	2.071003***	0.000	1.870775***	0.000	8.04704	0.191	3.649929**	0.046
Best_alone_Econlit_no_CNRS	2.027623***	0.000	1.915145***	0.000	9.57467	0.353	1.144091	0.614
Best_alone_Miscellaneous	2.759744***	0.000	2.396928***	0.000	.4210625	0.931	2.107959	0.208
Constant	.2143665***	0.000	.3815675***	0.000	-2.720161	0.869	-9.986753**	0.045

	Inflate : logit model				Selection model: nb_papers = 0			
	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|	Coef.	P>|z|
Age	.2174416***	0.000	.1578476***	0.000	-.0646431***	0.000	-.0638421***	0.000
female	2.330623***	0.000	1.795803***	0.000	-.235634***	0.010	-.23097***	0.010
Working_alone_only	-2.034985	0.165	-.8145498	0.126	9.545215***	0.000	9.748365***	0.000
cohort64-68	.5527885	0.923	-.3973625	0.877	-1.085461**	0.033	-1.269771**	0.021
cohort69-73	14.68225***	0.007	1.028852	0.520	-.7991227**	0.015	-.8572475***	0.009
cohort74-78	15.23061***	0.004	1.682535	0.271	-.6080159**	0.018	-.6350715**	0.013
cohort79-83	17.04227***	0.001	3.065615**	0.035	-.5913037***	0.008	-.6033947***	0.007
cohort84-88	16.48781***	0.001	2.505392*	0.083	-.5503679**	0.012	-.5573013***	0.010
cohort89-93	16.68975***	0.001	2.753779**	0.045	-.2803182	0.106	-.2938074*	0.086
cohort94-98					-.1119017	0.490	-.1161586	0.466
PR_CE					2.724658***	0.000	2.961577***	0.000
PR_2C					1.832338***	0.000	1.873646***	0.000
PR_1C					1.236857***	0.000	1.225056***	0.000
MCF_HC					.5129406***	0.000	.5031609***	0.000
constant	-30.67345***	0.000	-12.73915***	0.000	3.595324***	0.000	3.549177***	0.000
lnalpha	-1.06703***	0.000	-.4579795***	0.000				
athrho					-.327532***	0.000	-.4281149***	0.000
N	1566		1566		1566		1566	
Log Likelihood	-3256.107		-4151.489		-7746.526		-6189.231	
Vuong Test (uncorrected)	3.58***	0.002	5.45***	0.000				
Likelihood-ratio test of alpha=0	589.77***	0.000	2741.61***	0.000				
Wald test of indep. eqns. (rho = 0)					48.44***	0.000	30.13***	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. 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.

Table 5: Determinants of co-authorship: 2SRI Estimation Results

Dependant variable:	HH				GG				NB_COAUTHORS			
	(1)		(2)		**(3)**		**(4)**		**(5)**		**(6)**	
Model	ZIP		ZIP		ZINB		ZINB		ZINB		ZINB	
	IRR	P>|z|	IRR	P>|z|	IRR	P>|z|	IRR	P>|z|	IRR	P>|z|	IRR	P>|z|
H	1.039757***	0.000	1.042237***	0.000					1.004025	0.688		
G					1.020522***	0.000	1.021633***	0.000			.9983641	0,727
Standardized-Residual	1.126451***	0.000	1.109018***	0.000	1.133134***	0.000	1.102424***	0.000	1.167272***	0.000	1.1922***	0.000
Age	.9852736***	0.000	.9854045***	0.000	.9906243**	0.023	.9888942***	0.007	.975415***	0.000	.9729548***	0.000
Female	.9241381**	0.038	.9306992*	0.059	.9840804	0.694	.985529	0.724	.9218153*	0.099	.9233338	0.113
Cowrite_dr	1.261285***	0.000	1.26846***	0.000	1.238887***	0.000	1.255716***	0.000	1.270383***	0.000	1.328744***	0.000
PR_CE	1.319049***	0.001	1.204146**	0.049	1.426929***	0.000	1.271924***	0.010	1.4579***	0.001	1.568339***	0.000
PR_2C	1.123684***	0.001	1.132486***	0.005	1.124639**	0.017	1.129122**	0.014	1.195797***	0.002	1.205116***	0.002
PR_1C	1.201338***	0.005	1.140478**	0.031	1.17343***	0.010	1.126719*	0.052	1.333257***	0.000	1.330581***	0.000
MCF_HC	.8917328	0.197	.8833541	0.188	.890631	0.155	.8942705	0.172	.9197727	0.454	.8825842	0.262
Working_alone_only	.531641***	0.000	.5334222***	0.000								
cohort64-68	.3690495***	0.000	.4155215***	0.000	.3122876***	0.000	.3633593***	0.000	.3394253***	0.000	.4128732***	0.001
cohort69-73	.4246863	0.000	.4468632***	0.000	.3755773***	0.000	.4010743***	0.000	.3581325***	0.000	.3702197***	0.000
cohort74-78	.3903951***	0.000	.4242865***	0.000	.4108896***	0.000	.4422769***	0.000	.3807995***	0.000	.3983065***	0.000
cohort79-83	.4608758***	0.000	.475241***	0.000	.44187***	0.000	.4694847***	0.000	.5035105***	0.000	.5392617***	0.000
cohort84-88	.5589912***	0.000	.5661185***	0.000	.5598416***	0.000	.5795794***	0.000	.6137121***	0.000	.6607299***	0.000
cohort89-93	.5962211***	0.000	.6119466***	0.000	.605443***	0.000	.6270194***	0.000	.7133021***	0.000	.7490796***	0.001
cohort94-98	.7079183***	0.000	.7329668***	0.000	.7258089***	0.000	.7541094***	0.000	.7670176***	0.000	.7864654***	0.000
nb_papers_Misc	1.006274***	0.003	1.005524***	0.006	1.00685**	0.023	1.007065**	0.019	1.016252***	0.000	1.021924***	0.000
nb_papers_EconLit_no_CNRS	1.004665	0.527	1.001992	0.774	1.009589	0.290	1.00799	0.370	1.03505***	0.001	1.03261***	0.000
nb_papers_CNRS1	.9893791**	0.027	.9897854**	0.040	.987247*	0.052	.9847426**	0.018	1.011172	0.118	1.005861	0.440
nb_papers_CNRS2	1.01123	0.217	1.010868	0.264	1.023509**	0.012	1.028064***	0.003	1.01944	0.079	1.024409**	0.034
nb_papers_CNRS3	1.006837	0.136	1.006638	0.146	1.013484***	0.006	1.013384***	0.007	1.05563***	0.000	1.060079***	0.000
nb_papers_CNRS4	1.00393	0.507	1.006825	0.214	1.002371	0.678	1.00501	0.395	1.053119***	0.000	1.050871***	0.000
nb_papers_CNRS5	.9734185	0.607	.9681081	0.565	1.022522	0.718	1.0128	0.836	.9457282	0.459	1.007413	0.924
Herfindahl_JEL_CODE	.8011764***	0.009	.8183969**	0.018	.7890423***	0.005	.7870318***	0.005	.4493046***	0.000	.4381293***	0.000
Share_gb	1.308479***	0.000	1.323982***	0.000	1.531048***	0.000	1.489122***	0.000	1.747907***	0.000	1.881927***	0.000
Share_other	1.09702	0.550	1.106846	0.509	1.108722	0.534	1.072947	0.673	.8145187	0.412	.8954796	0.663
PhD defended at (Network effect I) :												
Univ. of Toulouse 1	1.067021	0.413	1.090307	0.319	1.008725	0.917	1.008725	0.917	1.231663*	0.067	1.212608*	0.062
Other French research institution	.9622981	0.345	1.024923	0.610	.9336406	0.107	.9336406	0.107	.9717338	0.645	.9738609	0.621
Univ. of Paris 10	.9767195	0.670	.978346	0.712	.9826961	0.798	.9826961	0.798	.8433606*	0.064	.8400375**	0.049
Univ. of Aix-Marseille	1.072036	0.222	1.057769	0.464	1.089994	0.203	1.089994	0.203	.891283	0.260	.8785316	0.143
Univ. of Strasbourg	1.105696	0.227	1.125961	0.283	1.076385	0.417	1.076385	0.417	.7756882	0.141	.876032	0.255
Univ. of Paris 9	.7674918**	0.023	.7374286***	0.007	.8468169*	0.095	.8468169*	0.095	.9912081	0.943	1.01117	0.928
Grande Ecole	1.182339**	0.014	1.155048*	0.056	1.309115***	0.001	1.309115***	0.001	1.003747	0.971	1.064606	0.536
Other Univ. In Paris	1.098862*	0.091	1.052457	0.359	1.108811	0.136	1.108811	0.136	.9464552	0.535	1.008636	0.924
European country	1.395722***	0.007	1.439115***	0.002	1.560409***	0.002	1.560409	0.002	1.38816*	0.060	1.268299	0.191
US	.890038	0.523	1.002826	0.985	.9419018	0.760	.9419018	0.760	1.52725*	0.056	1.47037*	0.099
Universities (Network effect II)	NO		Yes		NO		Yes		Yes		NO	

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	-3.414301***	0.001	-3.37412***	0.010					-7.548156***	0.000		
G					-.7498065***	0.000	-.7498065***	0.000			-5.219176***	0.000
Standardized-Residual	1.736803*	0.107	1.605957	0.186	1.465291**	0.011	1.465291**	0.011	14.75788***	0.000	22.05012***	0.000
Age	-.0435219	0.454	-.0350149	0.544	.0178888	0.527	.0178888	0.527	-.1305535*	0.063	-.5140615***	0.003
Female	-.1805496	0.753	-.1502756	0.808	.0666068	0.818	.0666068	0.819	-38.75202	0.952	-37.45614	0.936
PR_CE	.9114115	0.784	.913769	0.805	-.4992077	0.642	-.4992077	0.642	1.603895	0.168	1.160897	0.383
PR_2C	-.4850895	0.506	-.5655367	0.484	.2461645	0.516	.2461645	0.516	.6723288	0.380	.5102187	0.627
PR_1C	-.4845668	0.544	-.5105686	0.553	-.6838853*	0.101	-.6838853*	0.101	1.402785**	0.047	1.875419*	0.088
MCF_HC	-1.272412	0.284	-1.119334	0.326	.3343665	0.421	-.3343665	0.421	.3779083	0.566	-.0398384	0.967
cohort64-68	6.98064**	0.023	6.837184**	0.035	2.612548	0.167	2.612548	0.167	32.54643***	0.000	75.41312	0.973
cohort69-73	4.503021**	0.024	4.380308**	0.033	1.906422*	0.057	1.906422*	0.057	19.76823***	0.001	63.7169	0.978
cohort74-78	2.326547*	0.099	2.382214	0.125	1.48731**	0.050	1.48731**	0.050	15.54651***	0.002	57.10527	0.980
cohort79-83	1.490585	0.258	1.364695	0.322	.5237884	0.472	.5237884	0.472	14.23103***	0.002	55.29417	0.981
cohort84-88	1.874462	0.184	1.70459	0.301	.7831614	0.230	.7831614	0.230	15.30598***	0.001	54.52447	0.981
cohort89-93	1.40844	0.178	1.415473	0.229	.4452705	0.436	.4452705	0.436	12.68754***	0.003	45.61329	0.984
cohort94-98	.6074362	0.475	.6644003	0.519	.2522305	0.607	.2522305	0.607	9.68772***	0.004	19.89594	0.993
constant	2.420922	0.390	1.853168	0.525	-.6780755	0.631	-.6780755	0.631	4.485352	0.215	-10.7505	0.993
lnalpha					-1.952367***	0.000	-1.952367***	0.000	-1.901721***	0.000	-1.568005***	0.000
N	1183		1183		1183		1183		1183		1183	
Log Likelihood	-2262.49		-2213.927		-3272.022		-3836.815		-2601.264		-2680.350	
Vuong Test (unconstraint)	4.73***	0.000	4.38***	0.000	8.28***	0.000	8.10***	0.000	6.56***	0.000	6.55***	0.000
Likelihood-ratio test of alpha=0					575.91 ***	0.000	352.01***	0.000	370.94***	0.000	553.21***	0.000
Exogeneity test (Wald test)	32.94µµµ	0.000	20.93µµµ	0.000	29.61µµµ	0.000	15.94µµµ	0.000	42.85µµµ	0.000	45.05µµµ	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 6: Robustness checks: Determinants of co-authorship

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|
CL-index	1.0016***	0.000			1.001304***	0.000			1.002478***	0.000		
LLG-index			1.005915***	0.000			1.004632***	0.000			1.005347***	0.000
Standardized-Residual	1.003008	0.691	1.007984**	0.022	1.015563**	0.038	1.011484***	0.002	1.057704***	0.000	1.03337***	0.000
Age	.9772159	0.000	.9792032***	0.000	.991479*	0.078	.9928148	0.123	.9592894***	0.000	.9705186***	0.000
Female	.9013064**	0.036	.9091849**	0.044	.9703375	0.542	.9898528	0.827	.779914***	0.000	.8767449***	0.019
Cowrite_dr	1.366051	0.000	1.37816***	0.000	1.275365***	0.000	1.274232***	0.000	1.357346***	0.000	1.32112***	0.000
PR_CE	2.389134***	0.000	1.928204***	0.000	1.874848***	0.000	1.621036***	0.000	2.652294***	0.000	1.990655***	0.000
PR_2C	1.352706***	0.000	1.301035***	0.000	1.23523***	0.000	1.196042***	0.001	1.443428***	0.000	1.257602***	0.001
PR_1C	1.708428***	0.000	1.523754***	0.000	1.427667***	0.000	1.313868***	0.000	1.958431***	0.000	1.58221***	0.000
MCF_HC	.9944974	0.955	.9465338	0.569	.3269134***	0.000	.8734581	0.147	.9103242	0.472	.9019052	0.400
cohort64-68	.3022697***	0.000	.328389***	0.000	.3839133***	0.000	.3354482***	0.000	.5507499*	0.059	.5376615**	0.031
cohort69-73	.337953***	0.000	.3543386***	0.000	.4138367***	0.000	.3902346***	0.000	.4179275***	0.000	.4084654***	0.000
cohort74-78	.3453074***	0.000	.3533337***	0.000	.4932791***	0.000	.4081982***	0.000	.5364283***	0.001	.4928987***	0.000
cohort79-83	.4105089***	0.000	.4148241***	0.000	.5715696***	0.000	.4786778***	0.000	.7452103*	0.088	.6635999***	0.008
cohort84-88	.5303776***	0.000	.5384428***	0.000	.6581819***	0.000	.5646688***	0.000	.7358037**	0.038	.7110265**	0.011
cohort89-93	.6078139***	0.000	.5976638***	0.000	.7476915****	0.000	.6373786***	0.000	.9940561	0.956	.9207603	0.395
cohort94-98	.7065424***	0.000	.7030873***	0.000	.8242373***	0.000	.7241441***	0.000	.9793003	0.811	.9057413	0.198
PhD defended at (Network effect I) :												
Univ. of Toulouse 1	.8588143	0.119	.8906226	0.218	.8242373*	0.052	.8774745	0.175	.8375818	0.173	.8475816	0.161
Other French research institution	.8947545**	0.027	.9215736*	0.095	.8568114***	0.002	.880456***	0.009	.886165*	0.064	.9329208	0.248
Univ. of Paris 10	.9984095	0.984	1.000657	0.993	.9067076	0.223	.9152171	0.256	.7922757**	0.030	.78888**	0.017
Univ. of Aix-Marseille	.965878	0.666	.9764135	0.761	.9841681	0.844	.9910739	0.909	.8162309*	0.059	.8223728**	0.048
Univ. of Strasbourg	1.281718**	0.015	1.273284**	0.013	1.099083	0.388	1.10262	0.355	1.013742	0.921	.9879296	0.925
Univ. of Paris 9	.7912667**	0.040	.7929879**	0.036	.8063962*	0.063	.8264638*	0.088	.8250456	0.199	.8278389	0.166
Grande Ecole	1.292688***	0.006	1.318144***	0.002	1.319735***	0.000	1.358476***	0.001	1.106039	0.421	1.041512	0.727
Other Univ. In Paris	1.091465	0.266	1.100146	0.214	1.056966	0.506	1.069775	0.402	1.034177	0.761	1.010941	0.915
European country	1.889908***	0.000	1.987807***	0.000	1.813658***	0.001	1.876908***	0.000	1.513432*	0.076	1.584631**	0.027
US	1.083623	0.724	1.224009	0.352	.9617424	0.869	1.048616	0.834	1.22257	0.501	1.355687	0.256
Herfindahl_KW	.7764863***	0.006	.7905826***	0.010	.6078884***	0.000	.6587948***	0.000	.3408761***	0.000	.3240986***	0.000
Share_gb					1.903197***	0.000	1.89281***	0.000			1.934445***	0.000
Share_other					1.134149	0.509	1.179378	0.369			.8166748	0.454

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|
CL-index	-.1713406***	0.000			-.0353674***	0.000			-.4324095***	0.000		
LLG-index			-.400623***	0.000			-.0787158***	0.000			-2.6447***	0.000
Standardized-Residual	.095492*	0.088	.0787085*	0.074	.3386137***	0.000	.1747955***	0.000	1.374328***	0.000	4.515943***	0.000
Age	.0508537***	0.010	.0537892***	0.005	.045454*	0.080	.0689766***	0.006	-.0121094	0.579	.0969078	0.245
Female	.3428557	0.140	.3750159*	0.103	.3049845	0.291	.4559854*	0.095	-1.76851***	0.000	-6.621425***	0.000
PR_CE	-1.04646	0.117	-1.134538	0.133	-1.756188*	0.060	-1.644496**	0.045	-.783238	0.339	-1.219885	0.692
PR_2C	-.7530014*	0.073	-.6183909	0.131	.117625	0.756	-.1741986	0.637	-.9368175**	0.025	.2196244	0.869
PR_1C	-1.321026***	0.000	-1.312619***	0.001	-1.066789**	0.012	-1.141776***	0.005	-.1840026	0.666	.4426868	0.693
MCF_HC	-.2544673	0.389	-.3414806	0.265	-.505504	0.240	-.5819776	0.158	-.0793486	0.818	.2285113	0.843
cohort64-68	3.007732**	0.011	2.885046**	0.018	3.174666**	0.025	2.068796	0.121	11.6493***	0.000	35.07727***	0.000
cohort69-73	1.794842**	0.021	1.768232**	0.022	2.707837***	0.006	1.965276**	0.036	8.646351***	0.000	29.04559***	0.000
cohort74-78	1.603819**	0.024	1.508179**	0.029	2.912361***	0.001	2.025999**	0.011	9.815295***	0.000	32.86846***	0.000
cohort79-83	1.430626**	0.031	1.281215**	0.042	2.37005***	0.004	1.523386**	0.049	10.26298***	0.000	33.82659***	0.000
cohort84-88	1.587691***	0.008	1.468928**	0.011	2.471345***	0.001	1.676028**	0.021	8.135476***	0.000	27.14741***	0.000
cohort89-93	1.173924**	0.029	1.118926**	0.037	1.816784***	0.009	1.274693*	0.052	8.024847***	0.000	29.14987***	0.000
cohort94-98	.5039553	0.337	.407361	0.421	1.213887**	0.043	.5977456	0.301	5.777882***	0.000	13.93092***	0.000
constant	-3.612173***	0.000	-3.586909***	0.000	-4.673654***	0.000	-5.488118***	0.000	-.2726039	0.754	-12.0514***	0.001
lnalpha	-1.936493***	0.000	-2.10015***	0.000	-1.357579***	0.000	-1.46246***	0.000	-.8564862***	0.000	-1.112601***	0.000
N	1566		1566		1183		1183		1566		1183	
Log Likelihood	-3128.945		-3094.878		-3469.983		-3457.168		-3130.962		-2739.573	
Vuong Test	8.14***	0.000	8.27***	0.000	6.91***	0.000	7.08***	0.000	14.38***	0.000	10.99***	0.000
Likelihood-ratio test of alpha=0	157.60***	0.000	115.50***	0.000	1204.56***	0.000	1070.03***	0.000	1480.87***	0.000	1162.63	0.000
Exogeneity test (Wald test)	2.92	0.232	$7.58^{\mu\mu}$	0.0226	$38.47^{\mu\mu\mu}$	0.000	$29.47^{\mu\mu\mu}$	0.000	$139.44^{\mu\mu\mu}$	0.000	$75.36^{\mu\mu\mu}$	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.