Does Co-authorship Lead to Higher Academic Productivity?*

Lorenzo Ductor†
Universidad de Alicante

December 2011

Abstract

In recent decades, co-authorship and policies aimed at inducing academic collaboration have increased simultaneously. Assuming that intellectual collaboration is exogenously determined, prior studies find a negative relationship between co-authorship and productivity. I examine a panel data on economists publishing from 1970 to 1999 to test the causal effect of intellectual collaboration on intellectual output. As characteristics of the individual and her opportunity set are endogenously related to both collaboration and productivity, I instrument the amount of co-authorship by the common research interest between an author and her potential coauthors. After controlling for endogenous co-authorship formation, unobservable heterogeneity and time varying factors, the effect of intellectual collaboration on individual performance becomes positive. However, this effect varies significantly between high and low productive authors. These findings justify the existence of policies that stimulate intellectual collaboration.

JEL Classification Numbers: A11; J44; O30
Keywords: Co-authorship Formation; Academic Productivity; Scientific Networks

*I am indebted to Maria Dolores Collado and Marco J. van der Leij for their continuous support and advice. I am very grateful to Marcel Fafchamps for providing data on the Journal Quality Impact Factor and Marco J. van der Leij for his data on the co-authorship networks. I appreciate helpful comments from Carlos Aller, Yann Bramoullé, Pierre-Philippe Combes, Habiba Djebbari, Marcel Fafchamps, Sanjeev Goyal, Daryna Grechyna, Gergely Horvath, Silvia Martinez, Francesco Serti, seminar participants at Granada, Cambridge and GREQAM, and conference participants at the ASSET 2010, QED Jamboree 2010, the EEA-ESEM 2011 and the SAEe 2011. Financial support from the Spanish Ministry of Science and Innovation (Programa Formación del Profesorado Universitario) is gratefully acknowledged. All errors are my own. Previous versions of this paper circulated under the title "Co-authorship and Individual Academic Productivity: Evidence from Scientific Networks".

†Address: Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, Campus de San Vicente, 03080 Alicante, Spain. Tel. (+34) 965903400, Fax (+34)965 90 38 98; e-mail: lductor@merlin.fae.ua.es

1 Introduction

In recent decades, governmental policies aimed at inducing collaboration have increased. These policies are based on the assumption that intellectual collaboration results in productivity gains for the researchers. Some examples of these policies are the EU-funded research networks (Commission of European Communities, 2006) and the national Spanish Ingenio 2010 Program (Ministry of Education and Science, 2006). In both programs, researchers are required to collaborate as a condition to obtain research funding.[1] Other examples of policies inducing academics to collaborate are internal departmental policies (such as evaluations or rankings and employment or tenure decisions that require a minimum amount of publications) that not fully discounted articles by the number of authors.

In addition, scientific collaboration between authors has substantially increased over recent decades. Indeed, the data shows that the number of papers written by more than one author stood at 42% during the 1970s, while the proportion of co-authored articles was close to 60% in 1990s. Several authors have provided explanations for this increase, including greater gains from the specialization and division of labor, falling communication costs and a greater pressure to publish among others.[2]

Consequently, scientific collaboration is conditioned by a scientific policy that has progressively stimulated intellectual collaboration (Melin & Persson, 1996). If intellectual collaboration does not lead to gains in terms of research output, a policy change is required.

A need has therefore arisen to study the effect of collaboration on academic productivity and to answer these important questions: Does co-authorship lead to higher academic productivity? Is the effect of co-authorship the same for every individual? What are the channels through which collaboration might affect individual productivity? This paper contributes to these important questions. By using data on economists over a 30 year period, from 1970 to 1999, I find that after taking into account the endogeneity inherent in the co-authorship formation process – through an instrumental variable strategy, co-authorship leads to higher individual

[1] The purpose of the EU-funded network is to facilitate the sharing of research resources and, as a formally organized and coordinated structure, to sustain knowledge sharing among partners. The aim of the funding is to enhance the research potential of participants through the benefits of collaboration (Defazio et al., 2009).

[2] For instance, Laban and Tollison (2000) compare the incidence and extent of formal co-authorship observed in economics against that observed in biology, and discuss the causes and consequences of formal co-authorship in both disciplines. Hudson (1996) analyzes the increase in co-authorship and the reasons for it. Goyal et al. (2006) studies the emergence of an economics small world and the increasing co-authorship as a factor explaining this phenomenon.

academic productivity. However, this effect varies significantly between low and high productive individuals.

On the one hand, co-authorship might positively affect individual productivity through several channels: teamwork may be more productive than working alone as larger teams combine the ideas and talents of individuals (Chung et al., 2009). Another advantage is the reduction of time devoted to a project, as teams enable the researcher to start more projects with other authors at the same time. Hudson (1996) proposes the division of labor as one of the main advantages of coauthoring in economics. Apart from these advantages, coauthoring should lead to higher productivity if knowledge spillovers – i.e. network and/or peer effects – are present.[3]

On the other hand, intellectual collaboration might also decrease individual productivity. According to Hudson (1996), the main disadvantages of a collaboration are that it involves compromises and communication and organization costs. When working in a group, individual authors will have to agree on ideas, texts, approaches or even conclusions proposed by others. These compromises may lead to a reduction in risk taking that might affect the final quality of the article. Hudson (1996) likewise suggests that there is a free-rider problem, which in this context is related to the idea that the higher the number of authors writing a paper, the easier it is for an author to contribute less to the project. Moreover, another possible disadvantage is the negative externality through time devoted by collaborators to other projects with other authors (Jackson & Wolinsky, 1996). This is what I call congestion externality. The main idea is that if our coauthors are busy because they are working on many projects at the same time, they have less time to devote to our project. Therefore, we will have to devote more time to the collaboration, hence we have less time to do other work.

The literature examining the relationship between co-authorship and academic productivity has increased over recent years. However, there is no agreement as to whether this relationship is positive, negative or non-existent. Laban and Tollison (2000) provide evidence that co-authored scientific papers are more likely to be accepted for publication than sole-authored papers.[4] Recently, Chung et al. (2009) test the relation between intellectual collaboration and the quality of the intellectual output using academic papers published in prestigious finance journals. They find that papers with more authors are cited more often, in particular, papers with four authors are cited the most often.

[3] Azoulay (2010) and Waldinger (2010) examine peer effects in science using the unanticipated removal of individuals as a natural experiment.

[4] Presser (1980) and Zuckerman and Merton (1973) also find a positive correlation between co-authored articles and the probability of acceptance.

In contrast to the above studies that suggest coauthoring and productivity are positively related, Medoff (2003) show that collaboration does not result in significant higher quality research in economics.[5] In addition, using individual panel data on 5277 journal publications, Hollis (2001) finds that output is negatively related to co-authorship: the more an individual coauthors, the lower is the research output attributable to that individual. Greater collaboration appears to lead to more frequent, longer, and better publications, but when the publications are discounted by the number of authors, the relationship between research output and teamwork becomes negative. Estimates of the size of this effect are that, on average, adding one more author is associated with a per capita output reduction of between 7% and 20%. Hollis (2001) provides several reasons for the negative relationship: Mismeasurement (i.e. wrong measure of productivity), free rider problem, coordination costs, future but not current benefits from collaboration, etc. Hollis also points out the potential endogeneity of team formation, that is, the amount of co-authorship depends on the opportunity set the individual faces. He defines the opportunity set as the projects conceived by the author, and of projects offered to the individual as a coauthor. This study focuses on the endogeneity of the co-authorship formation as the main econometric problem driving this negative relationship.

Lee and Bozeman (2005) were the first to control for the possibility that co-authorship is formed endogenously, that is, authors choose with whom to work. For example, an author may choose to collaborate because some ideas are hard to be tackled individually or because he or she prefers to work with authors that have similar characteristics or intellectual skills. In particular, a high assortativity in the matching process is observed in the scientific network, which suggest that less able authors mainly collaborate with authors of a similar type (less productive). Ignoring this selection, its effect would be incorrectly attributed to collaboration and biased coefficients obtained. Lee and Bozeman (2005) deal with the endogeneity problem by using the extent to which a scientist's ties are cosmopolitan – i.e. outside the proximate work environment – as an instrument for the number of collaborators. However, there is a potential correlation between the instruments and productivity, as outside ties can directly affect productivity by providing access to new ideas (Singh, 2007). Moreover, they assume that the productivity of an author is only a function of the number of articles published in a period. I consider that the productivity of an author not only depends on the number of article published by the individual but also on the quality and length of each article.

[5]More recently, Acedo et al. (2006) find very weak evidence that co-authored management papers are of higher quality than sole-author papers.

In addition, the previous literature does not control for unobservable heterogeneity with the exception of Hollis (2001) who estimates a fixed effect model. This study is the first to simultaneously control for time invariant unobservable factors, the potential endogeneity inherent in the co-authorship formation and time varying factors, such as changes of research interests, quality of coauthors or the time devoted by coauthors to other projects with other authors.

I attempt to overcome some of these endogeneity difficulties by drawing on methods used for social network analysis. The basic idea is to derive some measures as proxies of time-variant unobservable factors from the observed scientific network and to analyze their effects on academic productivity. Fixed effects are also considered to capture individual unobserved heterogeneity. Moreover, to control for endogenous collaboration formation, an instrumental variables regression is estimated. The instruments are derived from the past network of the author and are based on the common research interests between the author and her potential coauthors (the past coauthors of i's coauthors) and its quadratic term. As Fafchamps et al. (2010) point out, one of the most important factor in determining the likelihood of collaboration is some commonality of research interest between the authors. On the other hand, collaboration is unlikely when there is too much overlap in skills. Hence, the quadratic term of the common research interests is included to allow for potential non-linear effects between collaboration and research overlap. Both instruments capture the idea of homophily, which consists in the tendency of individuals to associate with those who have similar characteristics.

A panel data among economists over a 30 years period, from 1970 to 1999, is used to implement this empirical strategy. This panel includes economists who published in journals which are included in the list of EconLit, a bibliography of economic journals compiled by the editors of the Journal of Economic Literature. I use the proportion of co-authored articles to measure the extent of collaboration, and a productivity measure created using the publication record of each author in the Econlit database. This productivity measure is based on the work of Fafchamps et al. (2010) and combine quantity (number and length of published articles) with quality (journal rank). The main finding is that, once I control for the endogeneity of network formation, co-authorship leads to higher academic productivity. This result is robust and statistically significant.

This paper contributes to the empirical study of social networks, which is a relatively new research area. In a recent work, Fafchamps et al. (2010) examine the formation of coauthor relations among economists over a twenty year period. Their principal finding is that a new collaboration emerges faster among two researchers if they are "closer" in the existing scientific network. In particular, being at a network distance of 2 instead of 3 raises the probability of initiating a collaboration by 27 percent.[6]

More related to this work are the papers that attempt a joint estimation of network formation and network effects. Various recent papers deal with this issue, Weinberg (2006) develops a model of social interactions with endogenous association. He finds that individuals associate with people whose behavior and characteristics are similar to their own. Mihaly (2009) develops an empirical strategy based on a two-step selection model à la Heckman to control for endogenous friendship formation with the aim of measuring the effect of peer interactions on student academic achievement. She found that popularity (measured as the location of the student in their social network) influences academic achievement positively. However, controlling for endogenous friendship formation results in a large drop in the effect of popularity.[7]

The paper is organized as follows. Section 2 presents the estimation model and the variables used in the regression. Section 3 provides a description of how the variables have been created and a descriptive analysis. In Section 4, I examine how co-authorship affects output. Section 5 presents some robustness checks. Section 6 concludes the paper.

2 Estimation framework

This project seeks to estimate the causal effect of co-authorship on individuals' output. The primary equation of interest is the productivity or output function which is given by,

$$\log(q_{i,t}) = \rho C_{i,t} + \alpha t_{i,t} + \omega t_{i,t}^2 + \phi H_{i,t} + \gamma \bar{n}_{i,t} + \beta \bar{q}_{1i,t} + \delta \bar{q}_{2i,t} + \zeta D_{i,t} + \mu_i + \mu_t + \varepsilon_{i,t}, \tag{1}$$

where $q_{i,t}$ is the productivity of an author i, which is obtained taking into account the quantity and quality of the articles published by the author.[8] As the distribution of output is skewed to the right, I apply the log transformation to minimize the effect of highly productive individuals

[6]Mayer and Puller (2008) also study the formation of links. They argue that individual level heterogeneity reflected in differences in age, sex and race plays an important role in the matching process of individuals.

[7]Conti et al. (2009) estimate the effect of popularity (measured as the number of friendship nominations received from schoolmates) on the labor market returns controlling for the selection of friendship. Their empirical strategy is to simultaneously estimate the outcome of interest together with the friendship formation process. They found that early family environment, school composition and school size play a significant role in determining popularity.

[8]The precise definition of all the variables used in the estimation is provided in section 3

on estimates. The variable $C_{i,t}$ is the co-authorship variable obtained as the proportion of co-authored articles.[9] The time-varying control factors are: years since first publication, $t_{i,t}$, and its square, $t^2_{i,t}$, the degree of research specialization of the author, $H_{i,t}$, average number of coauthors' papers, $\bar{n}_{i,t}$, average coauthors' productivity, $\bar{q}_{1i,t}$ and average coauthors' coauthor productivity, $\bar{q}_{2i,t}$.[10] In addition, as the panel data start for each individual with her first publication in the sample and extends till the last observed publication of the author or 1999, an author would have at least one article in her last year of publication. To prevent any bias from this, all regressions also include a dummy variable author's last year of publication, $D_{i,t}$. An individual fixed effects, μ_i, is also included to account for all time-invariant unobserved factors, such as innate ability, nationality, gender, school education, etc. Years dummies are included for each year from 1981 to 1999, these year fixed effects, μ_t, account for any possible time trends in collaboration and individual performance.[11] $\varepsilon_{i,t}$ is the time varying error term. As both the productivity and the co-authorship variables are likely to be correlated over time, I cluster standard errors by authors. The main parameter of interest is ρ, which captures the effect of co-authorship on productivity.

One problem when defining the productivity variable is the lag of economics publication. As pointed out in Fafchamps et al. (2010), there are considerable delays between the moment a paper is submitted to a journal and the moment it is published. Moreover, the time required for publication in economic journals varies greatly both between and within journals. These two facts could lead to a concentration of publications in some periods, which do not correspond to the exact periods where the author was working on the projects. In order to reduce this potential problem, all the variables, except experience, aggregate the information for the last five years, from $t-4$ to t (e.g. the dependent variable is the log-average productivity from $t-4$ to t). Although variables contain information about the last 5 years publications, the frequency used for estimation is annual; I consider a year as a period.

Productivity might be affected by other important factors such as changes in the degree of specialization, the time devoted by coauthors to other projects or the quality of coauthors of author i. Some of these variables might be endogenous to productivity, however, their inclusion

[9]For robustness, I also consider the average number of coauthors as the co-authorship variable. See section 5

[10]Although the average coauthors' productivity, average coauthors' coauthor productivity and the average number of coauthors' papers might be endogenous to productivity, we can observe in table **??** that their inclusion does not affect the co-authorship coefficient. Moreover, their inclusion provides evidence of important network effects.

[11]The first authors' observation in the estimation sample is 1980. See section 3 for more details.

does not affect the co-authorship coefficient since their effect on co-authorship is negligible.[12]

Some of these time-varying factors are controlled by using proxy variables constructed from the scientific network. This network is defined using authors as nodes and co-authorship as networks links. Following Singth (2007), I assume a collaboration tie to last for 5 years.[13] Therefore, the co-authorship network is defined for each year from 1974 to 1999, and each network contains links formed between $t-4$ to t. The proxy variables to control for these time-varying factors are the following:

- The years since first publication, $t_{i,t}$, is included to control for the experience of the author. Experience in any field or job is one of the main factors influencing productivity. Moreover, more experienced authors are more likely to initiate a project with someone else as they have more contacts and therefore more collaboration opportunities.

- The degree of research specialization, $H_{i,t}$, control for the potential effects of specialization on productivity. Specialization allows a scientist to become an authority on a given subject (Hackett 2005).[14] On the other hand, studying a wide range of topics may facilitate the generation of new ideas and enable a researcher to tackle projects that require a broader view – as the researcher has a more diverse knowledge.[15] Moreover, specialization might also affects the amount of collaboration as overly specialized authors may not be able to tackle projects that requires knowledge on different fields. Therefore, they may be more willing to collaborate than authors who have a more diverse source of knowledge.

- The average number of articles published by the coauthors of an author i, $\bar{n}_{i,t}$, is a proxy for the time devoted by i's coauthors to other projects with other authors – i.e. congestion externality. The time allocated by i's coauthors to other projects with other authors might reduce the opportunities of author i to initiate new projects, as he or she will have to devote more time to the collaboration. Thus, authors who work with busier coauthors are likely to have lower productivity.

- The average coauthors' productivity of author i, $\bar{q}_{1i,t}$, captures the coauthors' quality - in terms of productivity.

[12]See Table **??** for the effect of the inclusion of these explanatory variables in the co-authorship parameter, ρ, and Table **??** for the first stage results.

[13]In the robustness section, the sensibility of the results to this assumption is tested.

[14]Leahey (2006) found that specializing improves productivity.

[15]Belmaker et al. (2010) found that both over-specialization and over-generalization are detrimental to academic success.

- The average coauthors' coauthor productivity of author i, $\bar{q}_{2i,t}$ controls for the quality of these nodes at distance 2.

The main important econometric problem is the endogenous co-authorship formation. Authors choose with whom to work and these associations may be influenced by unobservable characteristics. For example, an author may choose to collaborate because some ideas are hard to be tackled individually. Then, the difficulty of the project can lead to greater cooperation and may also lead to higher output. In this case ignoring the selection of co-authorship would lead to spurious results. To correct for this type of bias, an instrumenting strategy is implemented. Equation (1) is estimated by an efficient two-step generalized method of moments (GMM), instrumenting $C_{i,t}$ by: the common research interests between author i and the nodes at distance 2 (the coauthors of i's coauthors) that author i accumulated from $t-10$ to $t-6$, $w_{2i,t-6}$, and its quadratic term.[16] Both variables capture the degree of homophily in terms of research interests between author i and the potential coauthors that author i may have in the future. These instruments control for the endogeneity of co-authorship by capturing the extent of matching on the overlap in research interests of author i with her potential coauthors.

The identification of the co-authorship parameter, ρ, comes from *past* variation in the research overlap between author i and her potential coauthors, $w_{2i,t-6}$. Therefore, to get rid of the individual fixed effects, μ_i, equation (1) is transformed using first differencing instead of within transformation. As applying the latter would create a spurious correlation between the average of the instrumental variables and the productivity.[17] Given the small variability of the instruments from one period to another, I consider the instrumental variables in levels to avoid the problems of weak instruments.[18] The assumption imposed by the efficient two-step GMM estimator is that the variation in the error term is uncorrelated with past research overlap between author i and her potential coauthors. Formally, the two-step efficient GMM estimator

[16]Following Baum et al. (2003): "The efficient GMM estimator minimizes the GMM criterion function J=N*g'*W*g, where N is the sample size, g are the orthogonality or moment conditions and W is a weighting matrix. In two-step efficient GMM, the efficient or optimal weighting matrix is the inverse of an estimate of the covariance matrix of orthogonality conditions. The efficiency gains of this estimator relative to the traditional IV/2SLS estimator derive from the use of the optimal weighting matrix, the overidentifying restrictions of the model, and the relaxation of the i.i.d. assumption. For an exactly-identified model, the efficient GMM and traditional IV/2SLS estimators coincide, and under the assumptions of conditional homoskedasticity and independence, the efficient GMM estimator is the traditional IV/2SLS estimator ".

[17]As the average of the instrument include the period from $t-4$ to t, positive values of the common field overlap in this period will be associated with positive co-authorship and positive productivity, creating an spurious correlation between the average of the instrument and the productivity variable.

[18]This alternative was proposed by Arellano and Bond (1991), who developed a GMM estimator using lagged levels of the endogenous variables –internal instrument– as instrument for the equation in first differences. Instead, I use the lagged level of the common research interest between an author and her potential coauthors –external instrument– as instrument for the difference in the amount of collaboration.

relies upon the following orthogonality conditions:

$$E(\varepsilon_{i,t}w_{2i,t-6}) = E(\varepsilon_{i,t-1}w_{2i,t-6}),$$
$$E(\varepsilon_{i,t}(w_{2i,t-6})^2) = E(\varepsilon_{i,t-1}(w_{2i,t-6})^2).$$

These instruments have a theoretical justification. They capture the idea of homophily, which consists in the tendency of individuals to associate with those who have similar characteristics. Homophily has been documented across several characteristics, such as age, race, gender, religion and occupations, e.g., Fong and Isajiw(2000), Baerveldt et al. (2004), Moody (2001), McPherson et.al. (2001), Conti et al. (2009) and Mihaly (2009). I focus on the homophily between researchers in term of common research interests, measured by the research overlap and based on the idea that authors tend to collaborate with individuals who have some commonality in research interest. The homophily effect between researchers in term of research overlap has been documented by Fafchamps et al. (2010). They show that the emergence of a new collaboration tie is decisively shaped by the commonality of research interest between the authors. They also find that the relationship between the research overlap index and the probability of collaboration follows an inverted-U curve. For example, the likelihood of forming a collaboration is highest when the research overlap index is 0.660. I include the squared of the research overlap variable to capture this potential inverted-U curve relationship. The main idea behind the inverted-U curve relationship is that authors need some common research interest to initiate a collaboration. Hence, research overlap cannot be too small. On the other hand, when research overlap is too strong collaboration is unlikely as there is too much overlap in skills (Fafchamps et al., 2010).

These instruments are valid as long as the above orthogonality conditions are satisfied. These assumptions are plausible as these potential coauthors are not nodes at distance 2 from the current network (obtained from nodes accumulated from $t-4$ to t) but from the past network (obtained from nodes accumulated from $t-10$ to $t-6$). Moreover, I do not expect that the research overlap would affect the future productivity of an individual through other channels, other than co-authorship, as this variable is only related to the matching process.[19]

[19] See section 5 for the robustness of the instrument to a potential internal validity threat.

3 Data

The data used for this paper are the same as the one used by Goyal et al. (2006), Fafchamps et al. (2010), Van der Leij and Goyal (2011) and Ductor et al. (2011). The data come from the EconLit database, a bibliography of journals in economics compiled by the editors of the Journal of Economic Literature.[20] From this database, I use information on all articles published between 1970 and 1999 by less than 4 authors.[21] The panel data start for each individual with her *first publication* in the sample and extends till the last observed publication of the author or 1999.

As already pointed in Ductor et al. (2011) a significant fraction of economists in the EconLit database publish very infrequently and may not publish a single piece over a five year period. To rule out such authors, I restrict attention to individuals who at every point t, have published at least one piece in the previous 5 years – i.e, active authors. Similarly, the model is estimated in first differences and the instruments are based on information from $t-10$ to $t-6$. As a consequence, the first six observations of each author are lost. Moreover, as the co-authorship network combines 5 years of publications, I loose the first 5 years of the sample as starting values.[22] The analysis therefore only considers articles published between 1980 and 1999. Finally, authors who has never collaborated during their academic career are excluded from the sample, since their variables are constant over the time frame and a first differences model is estimated. These authors represent a 13.25% of the active authors population.[23]

3.1 Definition of the variables

In this section, all the variables used to estimate the outcome equation (1) are described.

Co-authorship, $C_{i,t}$. The amount of co-authorship by an author i during a period t is measured as the ratio between the number of co-authored articles and the total number of articles published by the individual during the period $t-4$ to t. Notice that collaboration that does not result in journal publications is not observed. Moreover, neither informal collaboration

[20]See Goyal et al. (2006) for more details.

[21]As Van der Leij and Goyal (2011) pointed out, in the EconLit database 77% of the articles were written by 2 authors, 19% by 3 authors, and 4% by 4 or more authors. Moreover, Van der Leij (2006, pp. 53-56) show that the co-authorship network statistics are practically unaffected when (for a subset of the data) articles with 4 or more authors are included.

[22]Authors whose first article is published before 1974 are not considered since they do not have a defined network from the first year of their career.

[23]On average the productivity of authors without a co-authored article during all their career life is 1.26, whereas the average productivity for authors with at least one collaboration is 5.35. These sole-authors are also younger on average. Given their characteristics, their exclusion might lead to a downward bias in the value of co-authorship.

as valuable comments from colleagues in conferences and/or seminars nor the amount of effort devoted by each author to the project is observed.[24]

Productivity, $q_{i,t}$. The productivity of an author i at period t is measured as follows:

$$q_{i,t} = \sum_{j=1}^{S} \frac{\text{pages}_j * \text{ quality}_j}{\text{Number of authors}_j},$$

where S is the total number of articles published by author i from $t-4$ to t. The variable "quality" is a measure of the quality of the journal proposed by Fafchamps et al. (2010). They construct this measure based on the work of Kodrzycki and Yu (2006) – hereafter KY – who construct an impact index for a large number of economics journals. They complement KY work by 'predicting' the impact index of journals not included in the list compiled by KY. To do this, they regress the KY index on commonly available information such as the number of published articles per year, the impact factor, the immediacy index, the Tinbergen Institute Index, an economics dummy, interaction terms between the economics dummy and the impact factor, and various citation measures. They then use the predicted value obtained from this regression as impact index for journals not included in the KY list.[25] The actual KY impact index is used whenever available.

The "pages" variable measures the length of the article and is given by the number of pages of the article divided by the average number of pages of the articles published in the same journal.[26] Thus, I assume that longer than average papers are more valuable pieces of research.[27]

Observe that research is discounted by the number of authors, n, to account for the individual's contribution to the sum of output, giving $\frac{1}{n}$ credit to any single author.

Field concentration. To measure the degree of specialization of an author I use the Herfindahl index. Formally, this index is defined as

[24]One important effect of co-authorship may be to raise likelihood of publication for a given quality of the research rather than to improve the quality per se. In that case the marginal published single authored paper may be a better paper than the marginal published co-authored paper. This may lead to a downward bias in the value of co-authorship.

[25]Since most of the journals that KY omitted are not highly ranked, their predicted quality index is quite small.

[26]As pointed out by Sauer (1988) if journal editors act as value maximizers to allocate space, longer articles are more valuable than those of lesser length (on average).

[27]The number of pages for each article has been truncated from above at fifty pages. The main idea is not to give so much extra value to literature review articles as in general a literature review paper is much longer than the average article. For robustness, I also use as productivity only the journal quality index divided by the number of authors working in the article. See Section 5

$$H_{it} = \sum_{f=1}^{F} (x^i_{t,f})^2,$$

where $x^i_{t,f}$ is the total fraction of articles published by author i in the field f from $t-4$ to t, and F is the number of fields. To construct this variable, articles are categorized into 121 subfields according to the first two digits of the JEL codes. Articles with multiple JEL codes are divided and assigned proportionally to each of the corresponding fields. This measure takes value from $1/F$, reflecting the maximum degree of diversity, to 1 if the author write all her articles in the same field. Higher values of this index indicates a higher degree of specialization of the author.

Average number of articles of the coauthors. This variable is computed as the average number of papers published by the coauthors of author i from $t-4$ to t, excluding the papers published with author i. Although, there is a strong correlation between this variable and the average coauthor productivity, peer effects or network effects are expected to affect mainly the quality of the article. This may be because the effects come through the flow of information from one node to another in the co-authorship network. Thus, while the average number of papers of the coauthors of author i captures how busy the coauthors of an author are, the average productivity of the coauthors (which take into account the quality of the article) is more related to the presence of knowledge spillovers.

Average Coauthors' productivity. To control for the quality of coauthors, the average productivity of coauthors from $t-4$ to t is included in equation (1). To compute this variable, the productivity of each coauthor is calculated, excluding the papers published with author i.

Average Coauthors' coauthor productivity. This variable is computed in the same way as the average coauthors' productivity but looking at the productivity of the coauthors of coauthors, and excluding the papers published with the coauthors of author i.

Instrumental variables. The first instrument measures the common research interests between author i and the nodes at distance 2 that this author has had from $t-10$ to $t-6$, $w_{2i,t-6}$, – i.e. the research overlap between author i and her potential coauthors.

To obtain the proxy for research overlap between author i and her potential coauthors, I use a measure that is very close to the one proposed by Fafchamps et al. (2010). They construct an index of research overlap between any two researchers.[28] I extend their definition to construct an index of research overlap between an author and all her potential coauthors.

[28]See Fafchamps et al. (2010) for more details.

To do this, Fafchamps et al. (2010) categorize articles into 121 subfields according to the first two digits of the JEL codes. Articles with multiple JEL codes are divided and assigned proportionally to each of the corresponding fields. Then, the cosine similarity measure is considered to be a measure of field overlap between i and all her nodes at distance 2 accumulated from $t-10$ to $t-6$, $N^2_{t-6}(i)$.[29] This measure is computed as follows: Suppose that $x^i_{t-6,f}$ is the fraction of articles written by i in field f in the period from $t-10$ to $t-6$ (such that $\sum_f x^i_{t-6,f} = 1$) and $x^{N^2_{t-6}(i)}_{t-6,f}$ is the fraction of articles written by the potential coauthors of author i in field f from $t-10$ to $t-6$. Then, the research overlap index is,

$$w_{2i,t-6} = \frac{\sum_f x^i_{t-6,f} x^{N^2_{t-6}(i)}_{t-6,f}}{\sqrt{\sum_f \left(x^i_{t-6,f}\right)^2 \sum_f \left(x^{N^2_{t-6}(i)}_{t-6,f}\right)^2}}.$$

This variable takes values from 0, if i and her potential coauthors did not write any paper in the same field, to 1 if i and her potential coauthors wrote in exactly the same fields and in exactly the same proportion.

The second instrument is the squared of this variable and it is introduced to account for the potential inverted-U relationship between field overlap and collaboration documented by Fafchamps et al. (2010).

The instruments are measured using publications from $t-10$ to $t-6$ to avoid spurious correlation with the productivity variable and its lag, which are measured from $t-4$ to t and from $t-5$ to $t-1$, respectively.

3.2 Descriptive analysis

Figure 1 plots the relationship between the total average output measure and career time using all individuals in the data set.

As it can be observed, there is a rapid increase in total average output between the first publication and the ninth year after the first publication and then a steady decline starts and continues to the sixteenth year. The negative trend in the output after 9 years is due to fewer articles and lower quality. The increase on productivity after the fifteenth year of experience is

[29] "The research overlap measure capture not just having worked in similar research areas but also overlap in research topics. For instance, if a researcher has worked on, say, development economics and microeconomic theory (2 separate categories in JEL codes), she may be more likely to work with another researcher who has also focused on development and micro" (Fafchamps et al, 2010).

a consequence of the definition of the panel in which an author has at least one publication in the last year of her career. Moreover, the proportion of authors who have more than 15 years of experience is high – 40% of the total authors population, thus, these authors will have at least one publication during the 16 and 21 years of their career leading to the positive trend observed in figure 1 at the end of their academic life.[30]

Figure 2 shows the relationship between the total average output and career time using only the sample of active authors. Notice that active authors tend to publish more regularly as there is a rapid increase in the total average productivity until the thirteenth year after the first publication, and then it practically remains constant. The steady decline in the total average productivity in this sub-sample is not observed. The non-linear relationship between experience and productivity is evidence to include the square of experience in the regressions. This quadratic term captures the decreasing return to experience or academics life-cycle effects.

Figure 3 graphs co-authorship against experience for the sample of active authors. Observe that as the authors become more experienced, the average number of coauthors per article increases. This informally suggests that individuals with more contacts are more likely to collaborate. On average, economists in the first years after obtaining their Ph.D. have not established many contacts, thus the opportunities to publish a paper with other colleagues are scarce. Then, as the author becomes more experienced and known by others, the likelihood of collaborating increases.

Table 1 provides a summary statistics of the different variables used in the estimation. Column 1 provides summary statistics for authors with average lifetime co-authorship between 0 and 0.5 (low-average co-authorship). Column 2 provides statistics for authors with average lifetime co-authorship between 0.5 and 1 (high-average co-authorship).

Note that for authors with a low average lifetime co-authorship the mean productivity is 4.35. While for those authors with a high lifetime average co-authorship, the mean productivity is 5.76. Moreover, these authors have higher average coauthors' productivity and higher overlap in research interests with nodes at distance 2. Also note the high variability of the productivity and network variables, whose standard deviation is generally much higher than the mean.

[30]For robustness, the main regressions are also performed extending the panel until the four subsequent years after the last publication. The main results remain under this panel duration and are available upon request from the author. I also consider a different specification including a cubic term of the experience, the results are quantitatively the same.

4 Results

This section presents the results from estimations of the outcome equation (1).

4.1 Does co-authorship lead to higher academic productivity?

The main question of interest is whether co-authorship affects productivity once it is discounted by the number of authors. In this section, I estimate equation (1); I provide results of the estimated model without controlling for the selection of co-authorship, then estimates controlling for the potential endogeneity of co-authorship formation are provided.

Column 1 of Table **??** shows results of the first-difference specification in which the independent variables include co-authorship, years since first publication, a dummy variable author's last year of publication and year dummies. Column 2 provides estimates from a first-difference regression of equation (1) controlling for unobservable individual heterogeneity and time varying factors but not for the endogeneity of co-authorship. The implication of these regressions is that co-authorship lead to lower academic productivity. This result is consistent with Hollis (2001) who finds a negative effect of co-authorship on productivity.

To correct for the possible bias of the co-authorship measure, the instrumenting strategy described in section 2 is implemented. Column 3 of Table **??** presents the results controlling for the co-authorship formation process but not for the time varying factors. Column 4 shows the results from estimating equation (1) controlling for the endogeneity of co-authorship and controlling for the degree of specialization of the author, average number of coauthors' paper, quality of coauthors and quality of the coauthors of coauthors.[31] Observe that the coefficient of co-authorship becomes significantly positive after instrumenting, which shows that the individual productivity increases as authors substitute sole-authorship by teamwork. In other words, for example, the total productivity of two authors collaborating on two published papers is greater than the total productivity expected by each of the two authors writing a sole-authored paper. One possible explanation for the change of the sign of the co-authorship variable after instrumenting could be that authors have some periods where they have better ideas than in other periods. On the one hand, collaboration is more likely in periods where the author has a lack of ideas, since he or she is more willing to accept any co-authored project of any quality. On the other hand, in periods of good ideas, the author will only share ideas that require skills of other researchers. Then, good ideas not requiring specialization will be sole-

[31] First stage estimates are presented in Table **??**.

authored, resulting in a positive correlation between sole-authorship and productivity (Hollis, 2001). Once the instrumental variable strategy is implemented, only exogenous variations of the co-authorship variable through variations on the common research interest between author i and her potential coauthors are considered. As the distribution and fields of specialization of these *potential coauthors* affects the range and diversity of dispersed knowledge that an author can access (Bonacich, 1987), having more diverse contacts might help an individual to create new knowledge combinations, driving the benefits from exogenous co-authorship.

We can observe some evidence of the presence of knowledge spillover: the higher the productivity of coauthors, the higher the productivity of the author is. On the other hand, the quality of authors at a higher distance in the co-authorship network does not affect individual performance. The negative sign of the average number of coauthors papers reflects the congestion externality idea. For example, the busier the coauthors of author i are, the less time they devote to research projects with this author and the lower the output of author i is. The above network variables might be endogenous to productivity. However, their inclusion does not affect significantly the effect of co-authorship on academic productivity.[32] Career time has a negative impact on productivity for authors with more than 6 years of experience, consistent with the academics life-cycle effects.[33] Specialization have a negative effect on productivity consistent with the findings of Belmaker et al. (2010).[34]

Regarding the empirical validity of the instruments. I reject the null hypothesis of weak instrumental variables. Thus, the instrumental variables are sufficiently correlated with the troublesome variable, co-authorship. The Hansen's-J test is used for testing the null that the overidentifying moment conditions are true. From Table **??**, we cannot reject the null that the instruments are valid. Moreover, it is clear from the endogeneity test that the variable co-authorship cannot be treated as exogenous.

4.2 Co-authorship and productivity across individual 'types'

I am also interested in the relationship between co-authorship and productivity across different types of individuals. As already pointed out by Ductor et al. (2011), it is expected that the

[32]From the first stage estimates presented in Table **??**, we can observe that the effect of the average coauthors' productivity and the average coauthors' coauthor productivity on co-authorship is negligible

[33]As a consequence of the empirical strategy, the first observation of an author corresponds to the seventh year after the publication of her first article.

[34]I also consider other specification including field fixed effects using the 121 JEL codes. The results not presented for the sake of brevity are available upon request from the author. All the results are qualitatively the same under this specification

benefits from a collaboration differs across individuals. Access to ideas is an opportunity and it takes ability and effort to publish a high quality article. Thus, it is reasonable to suppose that the potential benefits from a collaboration vary with the abilities and efforts of a person (Ductor et al., 2011). The main hypothesis is that more able researchers can exploit the benefit from co-authorship to greater extent.

In order to analyze the potential difference of co-authorship between the different types of individuals, I divide the sample into two different groups depending on the productivity of the first publication. Summary statistics reported in Table **??** suggest that the first publication of an author is a very good predictor of the future potential performance of an author. Then, the estimation strategy is performed in each sample. Table **??** presents the effect of co-authorship on productivity across high and low productive individuals.[35] Column 1 shows the estimation results for authors whose first publication productivity is above the median – i.e. greater than the 50% distribution of the first publication productivity. Column 2 presents the results for those authors whose first publication productivity is equal or below the median.

In Table **??**, we can observe that the effect of co-authorship on economists' productivity varies significantly between the different types of individuals. As expected, more able authors obtain more benefits from co-authorship. Authors whose first publication productivity is equal or below the median do not obtain benefit from co-authorship – e.g. the coefficient is only statistically significant at the 10.8% level of significance. From the endogeneity test, the null that co-authorship is exogenous is rejected for both individual types. In the summary statistics across authors, see Table **??**, we can see a clear assortativity in the matching process of the authors. More able authors tend to have high-productive coauthors while less able authors collaborate with low-productive researchers. It is more likely that the benefit from collaboration arises when there is a matching between two different types of authors, for example, "mentoring" collaboration, as learning effects are entailed to this type of collaboration. An interesting extension is to study the mentoring effects and the role of the mentors for initiating the co-authorship network of the junior researcher.

5 Robustness

The main result is the positive relationship between scientific collaboration and individual output after controlling for the endogeneity of the co-authorship. However, I need to be concerned

[35]Unfortunately, weakness of the instrument under small sample does not allow to divide the sample into more different groups.

with some potential problems.

5.1 Not appropriate proxy variables.

It is possible that I am using an inappropriate measure of productivity. For example, it might be that the length of an article is not an important factor in measuring the productivity of a journal article.

It is also possible that I am using an incorrect measure of co-authorship and that the ratio of the number of co-authored articles with respect to the total number of articles is not a good proxy for the amount of teamwork. This is as I am assuming that the only factor determining the amount of intellectual collaboration is the proportion of co-authored papers. However, another important factor on measuring the amount of teamwork is the number of authors working on each article.

In this subsection I want to check if the results change with the specification of the productivity and co-authorship variables. Firstly, I redefine the productivity of an article as the journal quality index divided by the number of authors working on the article, that is, the new productivity variable is defined as

$$q_{i,t} = \sum_{j=1}^{S} \frac{\text{Quality}_j}{\text{Number of authors}_j},$$

where S is the number of articles published by author i from $t-4$ to t.

Results presented in Table **??** show that the positive relationship between co-authorship and productivity is not caused by the introduction of length as a factor in measuring productivity.[36]

Then, I redefine the co-authorship variable as the average number of authors for all the articles published by an author from $t-4$ to t. Table **??** presents the results using the new proxy variable for co-authorship. We can observe that the positive relationship between intellectual collaboration and intellectual output is not caused by the proxy variable used for co-authorship.

Therefore, the relation of intellectual collaboration and intellectual output is positive and robust to the specification of the productivity and co-authorship variables.

[36] The results for the FD estimator are also very similar to the main regressions but are not presented for the sake of brevity.

5.2 Co-authorship tie duration.

In the main analysis, I have considered that a collaboration tie lasts for 5 years – i.e. a co-authorship network at period t contains links formed between $t-4$ to t. In this subsection, I check if the results are sensible to this assumption by considering that the effect of a collaboration tie persists for a shorter period, 3 years. Firstly, I compute all the variables combining 3 years of publications – e.g. now, the average coauthors' productivity is obtained using publications and coauthors accumulated from $t-2$ to t, the dependent variable is the average productivity from $t-2$ to t, etc. As in the main analysis, only authors who publish at least one piece of research every five year are considered. Consequently, an author may not publish a piece of research from $t-2$ to t. Therefore, the log transformation to the dependent variable cannot be used, instead the $log(q_{i,t}+1)$ transformation is applied. Finally, the effect of co-authorship on output is estimated using the same empirical strategy described in section 2. Table **??** suggests that the main conclusions remain, however, the numerical magnitude of the coefficient of co-authorship, ρ, is smaller than that of the analogous coefficient in Table **??**. The externalities accrued from the network might takes a long period to affect the productivity of an author, this could explain the smaller effect of co-authorship under this shorter tie duration.

5.3 Research overlap and coauthors' coauthor productivity.

The main identification strategy relies on the assumption that the past common research overlap between an author and her potential coauthors does not affect future changes in productivity through other channels rather than co-authorship. However, it is possible that an author might change her degree of field specialization to meet productive potential coauthors and obtain benefits from them that are not passed to productivity through co-authorship – e.g. "favoritism"in the review process. I do not think this is an internal threat to the validity of the instruments as changes in fields of specialization require an important investment of time and effort, which are probably not compensated by the potential "benefits"that an author might obtain by meeting productive potential coauthors – assuming that such "favoritism"exists. Nevertheless, the aim of this subsection is to test the potential existence of this internal threat. In an attempt to evaluate the validity of the instruments to this threat, I estimate how changes in the research overlap between an author and her coauthors are affected by the average productivity of her potential coauthors. The main specification of interest is,

$$w_{i,t} = \tau t_{i,t} + \varphi t_{i,t}^2 + \lambda w_{2i,t} + \eta \bar{q}_{2i,t} + \theta \bar{q}_{2i,t-6} + \psi D_{i,t} + \mu_i + \mu_t + u_{i,t} \tag{2}$$

where $w_{i,t}$ is the common research overlap between an author and her coauthors from $t-4$ to t, $w_{2i,t}$ is the common research overlap between an author and all her coauthors of coauthors from $t-4$ to t, $t_{i,t}$ is the years since first publication, $D_{i,t}$ is a dummy authors' last year publication and $q_{2i,t-6}$ is the main variable of interest, the average productivity of the coauthors of coauthors from $t-10$ to $t-6$. An individual fixed effect is included to account for the individual unobserved heterogeneity, μ_i. Year fixed effects, μ_t, capture any possible time trends in research overlap and potential coauthors' productivity. Table **??** shows the effect of estimating equation (2) by first difference.

The results suggest that the average potential coauthors' productivity does not affect the common research overlap between an author and her coauthors at period t. Therefore, authors do not change their fields of specialization according to the productivity of their potential coauthors.[37]

6 Conclusions

The aim of this paper is to analyze the effect of intellectual collaboration on individual academic productivity. The approach proposed allows to control for unobservable heterogeneity, time varying factors and for the potential endogeneity of teamwork formation. None of the previous studies have controlled for all these potential source of endogeneity simultaneously. This analysis based on a panel of 129,003 economists over a 30 year period, from 1970 to 1999 reveals the following:

First, greater collaboration leads to higher academic productivity even after discounting by the number of authors working on an article. The positive relationship between intellectual collaboration and intellectual output is in contrast with Medoff (2003) and Hollis(2001), who find a negative relationship between co-authorship and academic output.

Second, co-authorship selection is endogenous – i.e. authors choose with whom to work depending on the quality and difficulty of their projects, which shows that previous results might be spurious. Specifically, the results turn from a significant negative effect of co-authorship on individual academic productivity in the baseline model to a significant positive effect in the specification after controlling for the endogenous team formation.

[37] A different specification using the degree of specialization of the authors as the dependent variable instead of the common research overlap leads to the same conclusion.

Third, having high-productive coauthors enhances individual productivity while the quality of authors at a higher distance in the co-authorship network does not affect individual performance. Moreover, having coauthors working in many projects at the same time affects negatively the productivity of the author, this evidence is consistent with the congestion externality hypothesis.

Finally, the effect of co-authorship on economists' productivity varies significantly between the different types of individuals. More able authors obtain more benefits from teamwork. For example, authors whose first publication productivity is below the median do not obtain statistically significant benefit from co-authorship. This might be a consequence of the high assortativity in the matching process, which suggests that less able authors mainly collaborate with authors of similar type – low productive.

The results are important for a number of reasons. They justify the existence of governmental policies and funding to promote collaboration. However, these policies should not be addressed to all individuals in the same manner, as the benefit from collaboration are not exploded by low-productive individuals. Policies aimed to induce a mixed matching process – e.g. "mentoring" collaboration – could probably facilitate the learning process of the low-productive authors and increase their current and future research output. On the other hand, the result are important for economists, as collaboration between them might enhance their performance, and therefore facilitate the access to research funding, higher salaries and prestige.

Future studies could analyze the effect of the different types of collaboration – e.g. mentoring, external collaboration – on academic productivity; as they might have different policy implications. Moreover, understanding the channel through which collaboration increases productivity is of great importance.

References

[1] Acedo, F. J., Barroso, C., Casanueva, C., & Galán, J. L. (2006). Co-authorshiop in management and organizational studies: An empirical and network analysis. *Journal of Management Studies*, 43, 957–983.

[2] Arellano, M., Bond, S.R., (1991). Some specification tests for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58, 277–298.

[3] Azoulay, P., J. Graff Zivin, J. Wang (2010), Superstar Extinction, *Quarterly Journal of Economics*, 125, 2, 549–589.

[4] Baum, Christopher F., Schaffer, Mark E. and Stillman, S. (2003). Instrumental variables and GMM: Estimation and testing. *Stata Journal*, 3(1), 1–31.

[5] Baerveldt, C. M.A.J. Van Duijn, L. Vermeij, and D.A. Van Hemert (2004). Ethnic boundaries and personal choice: Assessing the influence of individual inclinations to choose intra-ethnic relationships on pupils' networks. *Social Networks*, 26: 55–74.

[6] Belmaker, J., Cooper, N., Lee, T.M. and Wilman, H. (2010). Specialization and the road to academic success. *Frontiers in Ecology and the Environment*, Volume 8 (10).

[7] Bonacich, P. (1987). Power and centrality: a family of measures. *American Journal of Sociology* 92(5): 1170–1182.

[8] Brown, L. D. (2005). The importance of circulating and presenting manuscripts: Evidence from the accounting literature. *Accounting Review*, 80, 55–83.

[9] Chung, K. H., Cox, R. A. K. and Kim K.A. (2009). On the relation between intellectual collaboration and intellectual output: Evidence from the finance academe. *The Quarterly Review of Economics and Finance*, 49, 893–916.

[10] Conti, G., Galeotti, A., Mueller, G.,and Pudney S. (2009). Popularity. Unpublished manuscript

[11] Defazio, D., Lockett, A., & Wright, M. (2009). Funding incentives, collaborative dynamics and scientic productivity: Evidence from the european framework program. *Research Policy*, 38 (2), 293–305.

[12] Ductor, L. , Fafchamps, M. , Goyal, S. and Van der Leij, M. J. (2011). Social networks and research output. Unpublished manuscript

[13] Fafchamps, M. , Goyal, S. and Van der Leij, M. J. (2010). Matching and Network Effects. *Journal of the European Economic Association*, 8(1), 203–231.

[14] Fong, E., and W.W. Isajiw (2000). Determinants of friendships choices in multiethnic society. *Sociological Forum*, 15(2): 249–272.

[15] Goyal, S., Van der Leij, Marco.J., and Moraga, J.L. (2006). Economics: An emerging small world. *Journal of Political Economy*, 114, 2, 403–412.

[16] Hackett EJ. (2005). Essential tensions: identity, control, and risk in research. *Social Studies of Science*, 35: 787-826.

[17] Hollis A. (2001). Co-authorship and the output of academic economists. *Labour Economics*, 8, 503–530.

[18] Hudson, J. (1996). Trends in multi-authored papers in economics. *Journal of Economic Perspectives*, 10, 153–158.

[19] Jackson, Matthew O. and Asher Wolinsky (1996). A Strategic Model of Economic and Social Networks. *Journal of Economic Theory*, 71, 44–74.

[20] Kodrzycki, Y., and Yu, P. D. (2006). New Approaches to Ranking Economics Journals. *Contributions to Economic Analysis & Policy*, Vol. 5: Iss. 1, Article 24.

[21] Laban, D. N., & Tollison, R. D. (2000). Intellectual collaboration. *Journal of Political Economy*, 108, 632–662.

[22] Lee, S. and Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science* 35(5), 673–702.

[23] Leahey, Erin. (2006). Conceptual Metaphors and Empirical Leaps: Measuring Scientific Innovation and its Impact. Unpublished manuscript

[24] Mayer, Adalbert and Steven L. Puller (2008). The Old Boy (and Girl) Network: Social Network Formation on University Campuses. *Journal of Public Economics*, 92, 329–347.

[25] McPherson, M., L. Smith-Loving, and J.M. Cook (2001). Birds of Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27: 415–44.

[26] Medoff, M. H. (2003). Collaboration and the quality of economics research. *Labour Economics*, 10, 597–608.

[27] Melin, G., & Persson, O. (1996). Studying research collaboration using co-authorships. *Scientometrics*, 36(3), 363–377.

[28] Mihaly, K. (2009). Do More Friends Mean Better Grades? Student Popularity and Academic Achievement. RAND Working Paper Series WR-678.

[29] Moody, J. (2001). Race, school integration, and friendship segregation in America. *American Journal of Sociology*, 107(3): 679–716.

[30] Presser, S., (1980). Collaboration and the quality of research. *Social Studies of Science* 10, 95–101.

[31] Sauer, Raymond D., (1988) Estimates of the Returns to Quality and co-authorship in Economic Academia. *Journal of Political Economy*, 96(4), pp. 855–66.

[32] Singh, J., (2007). External collaboration, social networks and knowledge creation: Evidence from scientific publications. In Danish Research Unit of Industrial Dynamics Summer Conference 2007, Denmark, 2007.

[33] Van der Leij, M. J. (2006). The Economics of Networks: Theory and Empiric. Tinbergen Institute Research Series No. 384. Amsterdam: Thela Thesis.

[34] Van der Leij, M. J. and Goyal S. (2011). Strong ties in a small world. Review of Network Economics, 10(2), 1.

[35] Waldinger, F. (2010), Quality Matters: the Expulsion of Professors and Ph.D. Student Outcomes in Nazi Germany. *Journal of Political Economy*, 118(4), 787–831.

[36] Weinberg, B. (2006), Social interactions and endogenous associations. NBER Working Paper No.13038.

[37] Zuckerman, Harriet, and Merton, Robert K. (1973). Age, Aging, and Age Structure in Science. In The Sociology of Science: Theoretical and Empirical Investigations, by Robert K. Merton. Chicago: Univ. Chicago Press.

A First stage estimates

In Table **??**, I provide estimates of the first stage regressions. Column 1 shows the first stage results of the main regressions. Column 2 presents the results of the first stage using the average number of coauthors per article as the proxy for co-authorship. Column 3 shows the first stage results using 3 years window variables.

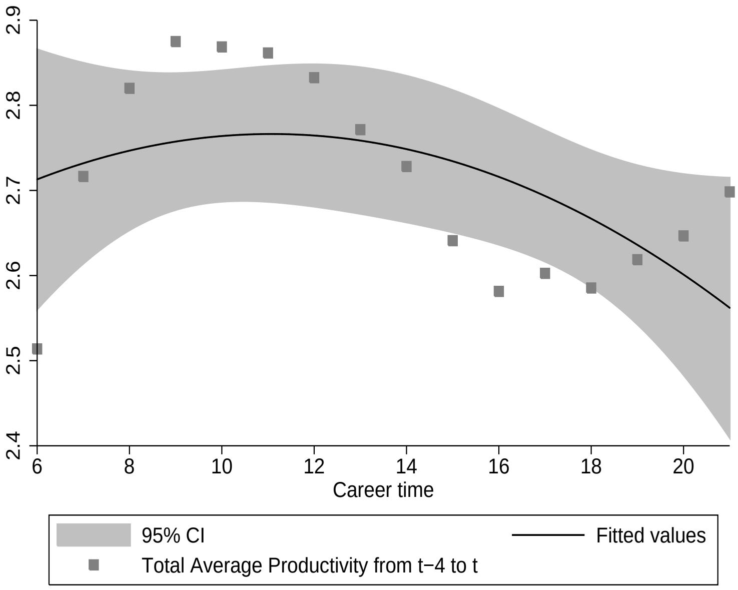
The sings of the instruments suggest that a high common research interest between author i and her potential coauthors - nodes at distance 2 in the past - is needed for these authors to initiate a collaboration in the future. The opposite signs of the research overlap variables is a consequence of the higher correlation of these variables with the lag of co-authorship. The rest of signs are like expected. Note that the average productivity of coauthors and the average productivity of coauthors of coauthors have a negligible effect on co-authorship, -.0003 and .0005, respectively. Thus, their inclusion does not affect significantly the magnitude of the co-authorship effect on academic productivity.

B Summary statistics

Table **??** shows statistics - average and standard deviation - across the different types of individuals. Column 1 shows the mean and standard deviation of the different variables for authors in the top 50% of the distribution of the first publication productivity. Column 2 shows statistics for those authors who are below the 50% of the distribution of the first publication productivity.

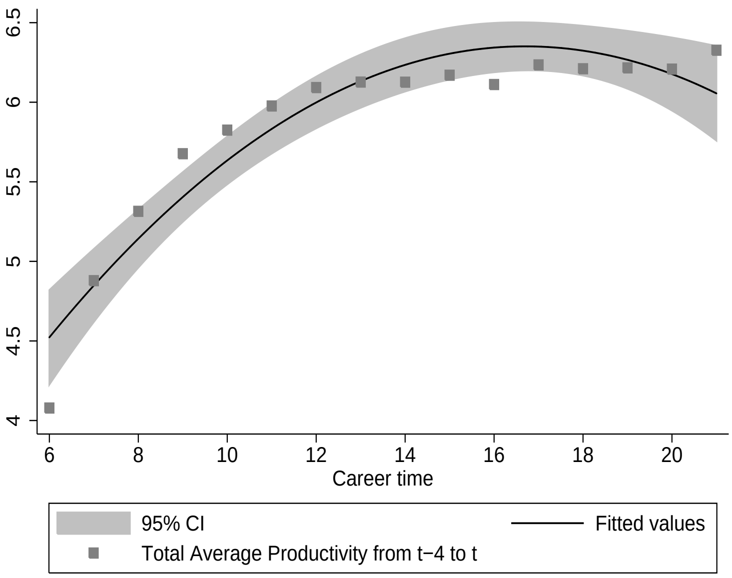
Note that the average coauthors' productivity is very related to the productivity of the author, reflecting the assortativity inherent in the matching process.

Figure 1: Total Average Productivity across Career time. Full sample



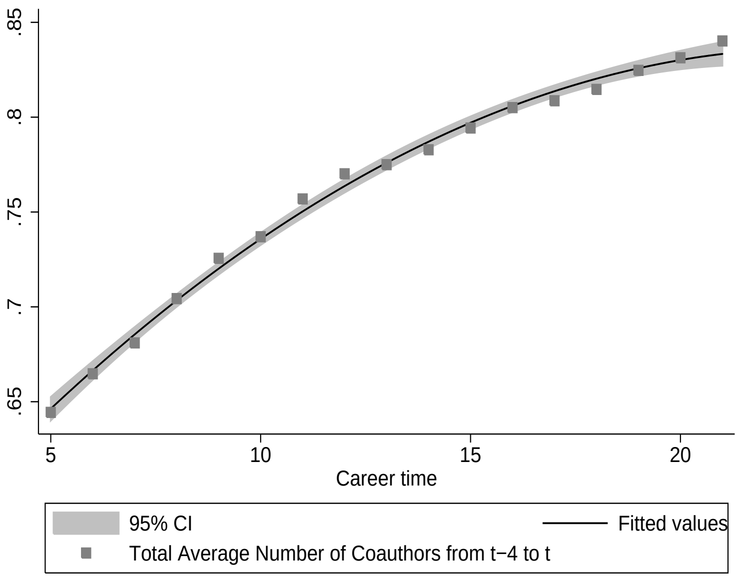
The sample consists of articles published from 1974-1999. All authors are considered except those whose first article was published before 1974.

Figure 2: Total Average Productivity across Career time. Active Sample



The sample consists of articles published from 1974-1999. Only "active "authors are considered.

Figure 3: Total Average number of coauthors on Career time. Active Sample



The sample consists of articles published from 1974-1999. Only “active ”authors are considered.

Table 1: Summary statistics of the data

Variables	Low-co-authorship		High-co-authorship	
	mean	st.d.	mean	st.d.
Avg. Productivity from t-4 to t	4.35	10.12	5.76	11.16
Avg. Pages from t-4 to t	8.62	5.25	9.25	5.32
Avg. Quality from t-4 to t	4.19	8.79	5.18	9.24
Experience	9.60	4.45	9.85	4.53
Co-authorship	.23	.27	.77	.27
Avg. Coauthors' Productivity	11.16	35.87	21.62	41.10
Avg. Coauthors' Coauthor Prod.	9.11	27.25	19.25	33.10
Avg. Coauthors' papers	1.81	3.39	3.28	3.41
Research overlap with Coauthors' Coauthor	.14	.25	.37	.33
Degree of Specialization	.36	.24	.35	.23
Number of Observations	39585	39585	68501	68501
Number of Authors	6726	6726	10652	10652

a. These statistics correspond to the active author sample and publications from 1974 to 1999. The first four years of each author are not considered as the productivity and co-authorship variable are not defined.

Table 2: The effect of co-authorship on academic productivity

	(1) FD	(2) FD	(3) FD-IV	(4) FD-IV
Co-authorship	−.3991***	−.5099***	2.6906***	2.6021***
	(.0288)	(.0281)	(.5270)	(.6570)
Experience	.1395***	.0790**	−.0027	−.0473
	(.0421)	(.0385)	(.0607)	(.0590)
$(\text{Experience})^2$	−.0042***	−.0034***	−.0022***	−.0016***
	(.0003)	(.0003)	(.0005)	(.0005)
Degree of Specialization		−1.5103***		−1.4294***
		(.0325)		(.0466)
Avg. Number of Coauthors Papers		.0162***		−.0193***
		(.0023)		(.0041)
Avg. Coauthors' Productivity		.0009***		.0019***
		(.0002)		(.0003)
Avg. Coauthors' Coauthor Prod.		.0001***		−.0002
		(.0002)		(.0004)
Authors' Last Year of Publication	.2911***	.2010***	.2692***	.1831***
	(.0126)	(.0125)	(.0178)	(.0172)
Year Dummies	*YES*	*YES*	*YES*	*YES*
Number of Authors	10835	10835	10835	10835
Number of Observations	70794	70794	70794	70794
Cragg-Donald Wald F statistic	–	.–	24.30	38.40
Hansen J-test (p-value)	–	–	3.642(.0563)	0.199(.6556)
Endogeneity Test (p-value)	–	–	5.465(.0000)	36.062(.0000)

b. The sample consists of "active "authors and publications from 1980-1999. Year fixed effects were included in the analysis, but are not reported here to conserve space. Standard errors in parenthesis adjusted for clusters. ***Significant at 1% level, **Significant at 5% level.

Table 3: The effect of co-authorship on academic productivity across individual types

	(1) >50%	(2) ≤50%
Co-authorship	3.0472**	1.0904(*)
	(.9815)	(.6778)
Experience	−.1409	.1033
	(.0860)	(.0005)
$(\text{Experience})^2$	.0006	−.0043***
	(.0009)	(.0008)
Degree of Specialization	−1.4679***	−1.3670***
	(.0742)	(.0508)
Avg. Number of Coauthors Papers	−.0503**	−.0105
	(.0179)	(.0131)
Avg. Coauthors' Productivity	.0019***	.0024***
	(.0004)	(.0006)
Avg. Coauthors' Coauthor Prod.	−.0006	.0013**
	(.0005)	(.0006)
Authors' Last Year of Publication	.1469	.2188
	(.0262)	(.0199)
Year Dummies	*YES*	*YES*
Number of Authors	6072	4763
Number of Observations	42490	28304
Cragg-Donald Wald F statistic	12.47	16.00
Hansen J-test (p-value)	.225(.6354)	.408(.5229)
Endogeneity Test (p-value)	23.900(.0000)	6.180(.0129)

c. Column 1 shows the estimation results for authors whose first publication productivity is above the median. Column 2 presents the results for those authors whose first publication productivity is equal or below the median. Standard errors in parenthesis adjusted for clusters.*** Significant at 1% level, ** Significant at 5%,(*) Significant at 10.8%.

Table 4: The effect of co-authorship on academic productivity using other proxy for productivity

	(1) FD-IV
Co-authorship	2.8198***
	(.6792)
Experience	$-.0749$
	(.0613)
$(\text{Experience})^2$	$-.0016$***
	(.0005)
Degree of Specialization	-1.3459***
	(.0484)
Avg. Number of Coauthors Papers	$-.0444$***
	(.0127)
Avg. Coauthors' Productivity	.0018***
	(.0003)
Avg. Coauthors' Coauthor Prod.	$-.0005$
	(.0004)
Authors' Last Year of Publication	.1683***
	(.0125)
Year Dummies	*YES*
Number of Authors	10835
Number of Observations	70794
Cragg-Donald Wald F statistic	24.295
Hansen J-test (p-value)	0.746(.3877)
Endogeneity Test (p-value)	47.939(.0000)

d. In this analysis the productivity of an article only depends on the quality of the article. Standard errors in parenthesis adjusted for clusters. *** Significant at 1% level.

Table 5: The effect of co-authorship on academic productivity using other proxy for co-authorship

	(1) FD-IV
Co-authorship	2.6466^{***}
	(.7918)
Experience	$-.0449$
	(.0652)
$(\text{Experience})^2$	$-.0011$
	(.0007)
Degree of Specialization	-1.3799^{***}
	(.0484)
Avg. Number of Coauthors Papers	$-.0455^{***}$
	(.0162)
Avg. Coauthors' Productivity	$.0023^{***}$
	(.0005)
Avg. Coauthors' Coauthor Prod.	$-.0009$
	(.0007)
Authors' Last Year of Publication	$.1503^{***}$
	(.0252)
Year Dummies	YES
Number of Authors	10835
Number of Observations	70794
Cragg-Donald Wald F statistic	11.517
Hansen J-test (p-value)	0.063(.8017)
Endogeneity Test (p-value)	33.150(.0000)

e. The proxy variable for co-authorship is the average number of authors per article from $t-4$ to t. Standard errors in parenthesis adjusted for clusters.*** Significant at 1% level.

Table 6: The effect of co-authorship on academic productivity assuming a different co-authorship tie duration

	(1) FD-IV (3-years)
Co-authorship	1.0662**
	(.5325)
Experience	−.0121
	(.0450)
$(\text{Experience})^2$	−.0015
	(.0004)
Degree of Specialization	−1.0188***
	(.0192)
Avg. Number of Coauthors Papers	−.0258
	(.0170)
Avg. Coauthors' Productivity	.0018***
	(.0002)
Avg. Coauthors' Coauthor Prod.	.0006
	(.0004)
Authors' Last Year of Publication	.1130***
	(.0119)
Year Dummies	*YES*
Number of Authors	15233
Number of Observations	91917
Cragg-Donald Wald F statistic	13.776
Hansen J-test (p-value)	.730(.3928)
Endogeneity Test (p-value)	8.312(.0039)

f. In this analysis, I assume that collaboration tie lasts for 3 years. The sample of articles analyzed here is from 1978-1999. As in the main analysis, I consider authors who publish at least a piece of research every five year. Year fixed effects were included in the analysis, but are not reported here to conserve space. Standard errors in parenthesis adjusted for clusters. *** Significant at 1% level, ** Significant at 5%.

Table 7: Common research overlap and potential coauthors' productivity

	(1) FD
Average Potential Coauthors' Productivity	.00003
	(.00002)
Current Average Coauthors' Coauthor Productivity	−.00001
	(.00004)
Experience	.00144
	(.00127)
$(\text{Experience})^2$	−.00005
	(.00005)
Current Research Overlap with the Coauthors of Coauthors	.61451***
	(.00635)
Authors' Last Year of Publication	.01189***
	(.00124)
Year Dummies	YES
Number of Authors	14832
Number of Observations	76469

g. In this analysis, I examine how potential coauthors' productivity, measured by the past average coauthors' coauthor productivity from $t-10$ to $t-6$, affects the common research overlap between an author and her coauthors from $t-4$ to t. The sample of articles analyzed here is from 1981-1999. I consider authors who publish at least a piece of research every five year. Year fixed effects were included in the analysis, but are not reported here to conserve space. Standard errors in parenthesis adjusted for clusters. *** Significant at 1% level, ** Significant at 5%.

Table 8: First Stage Regressions

Variables/ Models:	(1) Prop. of Co-authored Papers	(2) Avg. N. of Coauthors	(3) Other Tie Duration
Research Overlap with Potential Coauthors	$-.0255^{***}$	$-.0328^{***}$	$-.02471^{***}$
	(.0065)	(.0097)	(.0083)
(Research Overlap with Potential Coauthors)^2	$.0138^{*}$	$.0242^{**}$	.0134
	(.0077)	(.0118)	(.0102)
Experience	$.0405^{***}$	$.0390^{***}$	$.0483^{***}$
	(.0103)	(.0135)	(.0132)
(Experience)^2	$-.0005^{***}$	$-.0007^{***}$	$-.0005^{***}$
	(.0001)	(.0001)	(.0001)
Degree of Specialization	$-.0261^{***}$	$-.0435^{***}$	$-.0012$
	(.0097)	(.0139)	(.0069)
Avg. Number of Coauthors Papers	$.0180^{***}$	$.0198^{***}$	$.0314^{***}$
	(.0007)	(.0010)	(.0009)
Avg. Coauthors' Productivity	$-.0003^{***}$	$-.0005^{***}$	$-.0003^{***}$
	(.0001)	(.0001)	(.0001)
Avg. Coauthors' Coauthor Prod.	$.0005^{***}$	$.0008^{***}$	$.0007^{***}$
	(.0000)	(.0000)	(.0000)
Authors' Last Year Publication	.0052	$.0175^{***}$	$.0087^{**}$
	(.0036)	(.0052)	(.0041)
Year Dummies	*YES*	*YES*	*YES*
Number of authors	10835	10835	15233
Number of observations	70794	70794	91917

h. Column 1 presents the first stage results using the proportion of co-authored paper variable as a measure of co-authorship. Column 2 shows the first stage results using the average number of coauthors per article as the co-authorship variable. In column 3 the first stage results using a 3 year co-authorship tie duration is considered. Standard errors in parenthesis adjusted for clusters.*** Significant at 1% level, ** Significant at 5%,* Significant at 10%

Table 9: Summary statistics across individual types

Variables/ Percentiles	>50%		$\leq$50%	
	mean	st.d.	mean	st.d.
Avg. Productivity from t-4 to t	7.76	13.19	2.27	6.21
Co-authorship	.56	.53	.55	.39
Experience	12.76	6.12	9.51	4.29
Avg. Coauthors' Prod.	23.95	45.59	8.73	24.51
Avg. Coauthors' Coauthor Prod.	20.42	35.89	8.67	22.31
Avg. Number of Coauthors Papers	2.97	3.52	2.27	3.28
Research overlap with the Coauthors of Coauthors	.30	.32	.24	.32
Degree of Specialization	.33	.22	.37	.24
Number of Observations	80106	80106	50170	50170
Number of Authors	8902	8902	7902	7902

i. Column 1 shows the summary statistics for authors whose first publication productivity is above the median. Column 2 presents the summary statistics for those authors whose first publication productivity is equal or below the median. The sample of articles analyzed here is from 1980-1999 and only consider "active "authors.