Does Co-authorship Lead to Higher Academic Productivity?*

LORENZO DUCTOR†

†*School of Economics and Finance, Massey University, Private Bag 102904, Auckland New Zealand (e-mail: l.ductor@massey.ac.nz)*

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 found a negative relationship between co-authorship and productivity. I examine a panel data on economists publishing from 1970 to 2011 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 co-authors. After controlling for endogenous co-authorship formation, unobservable heterogeneity and time varying factors, the effect of intellectual collaboration on individual performance becomes positive.

I. Introduction

Scientific collaboration between authors has substantially increased in recent decades. In journals listed in EconLit, the proportion of papers written by more than one author rose from 24.7% during the 1970s to nearly 52% in the 2000s to 62.7% in 2011. Several authors have provided explanations for this increase, including greater gains from specialization and division of labour (McDowell and Melvin, 1983), falling communication costs (Hudson, 1996), a greater pressure to publish, increasing opportunity cost of time of reviewing papers (Barnett, Ault and Kaserman, 1988), increasing uncertainty in the editorial review

process (Barnett *et al*., 1988) and the possible increase in productivity through collaboration (Laband and Tollison, 2000), among others. Governmental policies aimed at encouraging collaboration have also increased in recent decades. 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. Other examples of policies encouraging 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 do not fully discount articles by the number of authors. Consequently, scientific collaboration is affected by scientific policies that have progressively stimulated intellectual collaboration (Melin and Persson, 1996), and if intellectual collaboration did not increase the sum of research produced, a policy change would be required.[1]

This paper studies the effect of research collaboration on research output and contributes to answering these questions: Does co-authorship lead to a 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? Examining data on economists over a 42-year period, from 1970 to 2011, 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 academic productivity. However, this effect varies significantly between high- and low-productivity individuals.

II.   Related literature

The empirical literature examining the relationship between co-authorship and academic productivity has increased in recent years. However, there is no agreement, as to whether this relationship is positive, negative, or non-existent. Laband and Tollison (2000) provide evidence that co-authored scientific papers are more likely to be accepted for publication than sole-authored papers. Recently, Wuchty, Jones and Uzzi (2007) and Chung, Cox and Kim (2009) find that papers with more authors are cited more often. In contrast, Medoff (2003) finds that collaboration does not affect significantly research output in economics and Hollis (2001) finds that co-authorship leads to lower academic productivity.[2] In a model of teamwork formation, an author would decide to collaborate if the expected utility from collaboration is larger than the expected utility from sole-authorship. Therefore, if the utility only depends on research output we should expect a positive effect of co-authorship on productivity, as two authors would engage in a collaboration if the expected productivity from this collaboration is higher than the expected productivity from working alone. This raises an interesting question: why is the effect of co-authorship not always unambiguously

---

[1]Ubfal and Maffioli (2011), Bozeman and Corley (2004) and Lee and Bozeman (2005), Defazio, Lockett and Wright (2009) find a positive relationship between research grants and intellectual collaboration.

[2]Acedo *et al.* (2006) find very weak evidence that co-authored management papers are of higher quality than sole-authored papers.

positive?[3] The negative effect found in previous studies can be explained by the following reasons:

(i) As Hollis (2001) points out, other factors like the increase in salary through collaboration (when individual departments do not discount for promotion articles by the number of authors), risk diversification and friendship may enter into the utility function. From a social point of view, co-authorship would be desirable if it increased total research output, i.e. society would not care about the private increase in salaries and the private enjoyment of friendship through collaboration. Therefore, salaries and friendships may induce authors to collaborate on projects that would be more efficiently done through sole-authorship. For example, authors may choose to collaborate with a mismatched co-author (same-skill co-author) or distant co-authors only to enjoy the consumption benefit from social activities (Hamermesh and Oster, 2002), even if the coordination costs and duplication of effort are more likely in such cases.

(ii) There are also *negative externalities* from collaboration that may lead to a negative effect on productivity, like the free-rider problem (Hudson, 1996) and the externality through time devoted by collaborators to projects with other authors (Jackson and Wolinsky, 1996), which I call *congestion externality*. If your co-authors are busy because they are working on many projects at the same time, they have less time to devote to your project, therefore, you will have to devote more time to the collaboration and hence have less time to do other work. However, busier co-authors may also have a positive effect on productivity by stimulating authors to work harder, so the net effect of the congestion externality is ambiguous.

(iii) The negative relationship found in previous studies can also be explained by the *potential endogeneity* of teamwork formation, i.e. the amount of co-authorship depends on the opportunity set the individual faces. Hollis (2001) defines the opportunity set as the projects conceived by the author and projects offered to the individual as a co-author. The creativity of researchers may be cyclical, i.e. authors may have periods with better ideas than in other periods. In periods with a lack of ideas, authors will be more willing to collaborate, while in periods of inspiration, authors are likely to share only the ideas that require the skills of others. Then good ideas not requiring the specialization of others will be sole-authored, resulting in a negative correlation between co-authorship and productivity (Hollis, 2001). I focus 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, i.e. that authors choose whom to work with. They deal with the endogeneity problem by instrumenting co-authorship using a 'cosmopolitan scale' that ranges from 0 to 5 depending on the location of the co-author. For example, the instrument is 0 if the author collaborates with a co-author from the same institution and in the same work group, and 5 if the author collaborates with a researcher from a foreign institution. However, there is a potential correlation between the instruments and productivity, as links with international colleagues will provide access to new ideas and resources. Moreover,

---

[3] I am grateful to an anonymous referee for pointing out this question.

they assume that the productivity of an author is only a function of the number of articles published in a given period. Instead, I consider that the productivity of an author not only depends on the number of articles published by the individual, but also on the quality of each article proxied by journal quality impact factor.

While previous studies have improved our knowledge on the role of intellectual collaboration in explaining changes in research output, data limitations and especially difficulty in finding exogenous variation in teamwork formation has resulted in treating co-authorship as exogenous. I attempt to overcome some of the endogeneity difficulties by drawing on methods developed for social network analysis. Recently, Bramoullé, Djebbari and Fortin (2009) and Lee, Liu and Lin (2010) provide the necessary and sufficient conditions under which peer effects can be identified using network data. Bramoullé *et al.* (2009) propose the use of peers' peers (and peers' peers' peers) characteristics as instrumental variables for the peers' behaviour. Calvó-Armengol, Patacchini and Zenou (2009) and Lin (2010) use friendship networks of adolescents derived from the Add-Health data to identify peer and social network effects in education.

A few recent studies attempt a joint estimation of network formation and network effects to deal with the selection effects in the formation of links. For example, Mihaly (2009) developed 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. Conti *et al.* (2013) estimate the effect of popularity (measured as the number of friendship nominations received from schoolmates) on the labour market returns controlling for the selection of friendship. Liu *et al.* (2012) used the Add-Health data to develop a network formation model to identify key players in criminal activity and to estimate peer effects in delinquent adolescent networks, extending the identification strategy proposed by Bramoullé *et al.* (2009) and Lee *et al.* (2010).

In this paper, I estimate the causal effect of co-authorship on individual productivity and provide evidence of the existence of peer effects (positive knowledge spillovers) and congestion externalities in academia.[4] I use network data among economists to exploit exogenous variation of co-authorship through variation in the commonality of research interests between the author and the co-authors' co-authors accumulated in the past. This identification strategy is related to the one proposed by Bramoullé *et al.* (2009) to estimate peer effects with network data, as I use the structure of the network in the past to infer exogenous variation in future co-authorship. My main finding is that once I simultaneously control for time invariant unobservable factors and for the potential endogeneity inherent in the co-authorship formation, co-authorship leads to a higher academic productivity. This result is robust and statistically significant.

III. Estimation framework

This section defines the co-authorship networks, and describes the empirical model, the main factors related to research output and the identification strategy to estimate the causal effect of co-authorship on academic productivity.

---

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

Co-authorship networks

Let $G_{t,s}$ denote a co-authorship network obtained using all joint publications from year $t - (s - 1)$ to t. Two authors have a link in this co-authorship network, $G_{t,s}$, if they have published a co-authored article sometime in the period $t - (s - 1)$ to t. The duration of links is defined by s.

For example, in a 5-year co-authorship network ($s = 5$) if an author i publishes an article with author j at year 2001, the link between these two authors is assumed to last until year 2005. This is consistent with the recent empirical network literature, where 3-year co-authorship networks (Gonzalez-Brambila, Veloso and Krackhardt, 2013), 5-year co-authorship networks (Ductor *et al.*, forthcoming), or even 10-year co-authorship networks (Fafchamps, Goyal and Van der Leij, 2010) are considered. As in Ductor *et al.* (forthcoming), I define $N_i(G_{t,s})$ as the co-authors of author i in the co-authorship network $G_{t,s}$ and $N_i^2(G_{t,s})$ as the co-authors' co-authors accumulated from $t - (s - 1)$ to t.

The research output and co-authorship are measured in the same time window as the co-authorship network, $G_{t,s}$. The long time horizon of the research output, co-authorship and network variables is standard in the literature (Hollis, 2001; He, Geng and Campbell-Hunt, 2009; Ubfal and Maffioli, 2011; Carillo, Papagni and Sapio, 2013; Ductor *et al.*, forthcoming; Gonzalez-Brambila *et al.*, 2013) and takes into account the time that it takes for the externalities obtained from intellectual collaboration to be transformed into research output. This long horizon also mitigates the problem caused by the delays in the publication process in economics. Ellison (2002) finds that the average time lag between the submission of a paper to a journal and its publication varies greatly both between and within journals. He finds that the average lag varies from 9 to 29 months depending on the journal. This delay in economic publication could lead to a concentration of publications in some periods, which do not correspond to the exact periods where the authors were working on the projects. As a robustness check, I present results without aggregating research output and considering 1-year network variables.

Empirical model

The primary equation of interest is the productivity or output function which is given by,

$$log(q_{i;t,s} + 1) = \rho C_{i;t,s} + D'_{i,t}\omega + \beta_1 H_{i;t,s} + \beta_2 \bar{n}_{i;t,s} + \beta_3 \bar{q}_{1i;t,s} + \beta_4 \bar{q}_{2i;t,s} + \mu_i + \mu_t + \varepsilon_{i;t,s}, \quad (1)$$

where $q_{i;t,s}$ is productivity over the period $t - (s - 1)$ to t. As $q_{i;t,s}$ could be zero for some periods, I follow Ductor *et al.* (forthcoming) and use the $log(q_{i;t,s} + 1)$ transformation to reduce the impact of high-productivity individuals on the estimates. The variable $C_{i;t,s}$ is the co-authorship variable obtained as the proportion of author i's articles that were co-authored from $t - (s - 1)$ to t. The time-varying factors are: career time dummies, $D_{i,t}$; the degree of research specialization of the author from period $t - (s - 1)$ to t, $H_{i;t,s}$; the average number of co-authors' papers accumulated in the co-authorship network $G_{t,s}$, $\bar{n}_{i;t,s}$; the average co-authors' productivity accumulated in the network $G_{t,s}$, $\bar{q}_{1i;t,s}$; and the average co-authors' co-authors productivity in network $G_{t,s}$, $\bar{q}_{2i,t,s}$. An individual fixed effect, μ_i, is also included to account for all time-invariant unobserved factors, such as innate ability, nationality, gender, school education and others. Year dummies, μ_t, are included to account for any possible time trends in collaboration and individual performance. The time varying error

is $\varepsilon_{i,t}$, and the main parameter of interest is ρ, which captures the effect of co-authorship on productivity.

Defining co-authorship and productivity

The amount of *co-authorship* by author i from period $t - (s - 1)$ to t, $C_{i;t,s}$, 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 $t - (s - 1)$ to t. Formally, this variable is defined as:

$$C_{i;t,s} = \frac{nc_{i;t,s}}{n_{i;t,s}}$$

where $nc_{i;t,s}$ is the number of co-authored articles published from $t - (s - 1)$ to t and $n_{i;t,s}$ is the total number of articles published over period $t - (s - 1)$ to t.

The *productivity* of author i accumulated from $t - (s - 1)$ to t is measured as follows:

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

where j denotes an article and S is the total number of articles published by author i between $t - (s - 1)$ and t, i.e. the productivity variable, $q_{i;t,s}$, is the sum of the productivity of each article, $q_j = \frac{\text{pages}_j \times \text{quality}_j}{\text{Number of authors}_j}$, over period $t - (s - 1)$ and t.

The 'pages' variable measures the relative 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. For robustness, I also use as productivity only the journal quality index divided by the number of authors who worked on the article. Thus, I assume that longer than average papers are more valuable pieces of research. This is consistent with Laband (1986), who finds a positive relationship between article length and number of citations.[5]

Observe that productivity is discounted by the number of authors. Since society would benefit from maximizing total research output (the sum of the output produced by all researchers), we should only attribute a share of a co-authored paper to an individual when evaluating their performance (Hollis, 2001). Note that if we do not discount co-authored articles by the number of authors, the effect of co-authorship on total research output is unambiguously positive, as a co-authored publication would account for more than the same publication written by a single author. Another reason to discount co-authored articles by the number of authors is to evaluate the externalities from collaboration related to quality and not to the opportunities to publish more articles within a given period.

The variable 'quality'$_j$ is a measure of the quality of the journal in which the article, j, was published. This measure is used in Ductor *et al.* (forthcoming) and is based on the work of Kodrzycki and Yu (2006) – hereafter KY. KY apply an iterative process and weight citations depending on the influence of the citing journal. In computing the index,

---

[5]The number of pages for each article has been truncated from above 50 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. The results are robust to the truncation and exclusion of length as a determinant factor of productivity.

KY exclude self-citations and use 8-year period citations to control for journal age. The initial step to obtain the KY index is to calculate the number of citations received by an economic journal i (adjusted by the number of articles published in journal i) from other economics journals:

$$Q_{i,0} = \frac{\sum_{j=1}^{n} C_{ij}/a_i}{\sum_{k=1}^{n} C_{kj}/a_j},$$

$$quality_{i,0} = [Q_{i,0}/Max\{Q_{i,0}\}] \times 100,$$

where $Q_{i,0}$ denotes the initial citations index for journal i, a_i and a_j are the number of articles in journal i and j respectively, and n is the number of economics journals. By dividing citations by the number of articles published in that journal, KY obtain a journal impact factor adjusted by its influence per article. The subsequent iterations are given by:

$$Q_{i,t} = \sum_{j=1}^{n} \left[\frac{C_{ij}/a_i}{\sum_{k=1}^{n} C_{kj}/a_j} \right] quality_{j,t-1},$$

$$quality_{i,t} = [Q_{i,t}/Max\{Q_{i,t}\}] \times 100.$$

The iterative process continues until it converges. As the KY index is not available for all the journals listed in EconLit, I use the predicted impact factor obtained by Ductor *et al.* (forthcoming) for those journals not listed in KY.[6] The impact factor, 'quality'$_j$, relies on the assumption that the journal quality is a good predictor of the quality of the article published by the author.[7] As a robustness check, I use the journal citation reports in social sciences (JCR), developed by Reuters (2013). I do not use citations because of the potential delay between the time of publication and the moment the article accumulates citations. This delay would imply that juniors researchers had a lower productivity *per se*.[8]

In most of the tables, the period of analysis is a 5-year over rolling window ($s = 5$). This aggregation accounts for the lags in economic publication and for the time needed to transform the externalities obtained from co-authorship into research output. However, as the 5-year period is arbitrary – since there is no evidence about the time needed to transform the externalities accrued from intellectual collaboration into research output, I also consider 2- and 3-year over rolling windows and results without aggregating at all,

[6] Ductor *et al.* (forthcoming) predict the impact index of journals not included in the KY list by regressing the KY index on 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 an impact index for journals not included in the KY list.

[7] Recently, Laband (2013) shows that 'journal quality is a useful predictor of citations at the time of publication'.

[8] There are also issues related to citations of working paper versions of subsequently published articles, and pertaining to the relative ranking of citing journals and authors. Moreover, as one anonymous referee suggested, using citations as a measure of quality may inflate productivity gains from co-authorship, since well-connected authors have more opportunities to disseminate their research. Another reason for why I do not use citations is that gathering information on citations for more than 550,000 articles and for every period from 1974 to 2011 is prohibitively costly. Every measure of academic productivity is subject to criticism. Even if the journal quality is not the best proxy of research output, the quality of the journal is important for economists, since they are often recruited, promoted, tenured and recognized according to the quality of the journals they have been published in (Laband, 2013). Previous literature examining the effect of research collaboration on research output finds similar results when they consider the quality impact factor and article citations as proxies of research output (He *et al.*, 2009).

i.e. for 1 year. The productivity variable in the 5-year period analysis is given by:

$$q_{i;t,5} = \frac{(q_{i,t-4} + q_{i,t-3} + q_{i,t-2} + q_{i,t-1} + q_{i,t})}{5}.$$

Thus, the dependent variable in the 5-year over rolling window is $log(q_{i;t,5} + 1)$.

Time-varying factors

Productivity might be affected by important time-varying factors such as changes in the degree of specialization, the time devoted by co-authors to other projects, the quality of co-authors, or the quality of the co-authors' co-authors. The proxy variables for these time-varying factors are as follows: *Career time dummies $D_{i,t}$*, are included to control for the effect of each year of experience on academic productivity. 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.

Let $t_{i,0}$ denote the first year publication of author i and define the career time as $c_{i,t} = t_i - t_{i,0}$. Then, as in Ductor *et al.* (forthcoming), I create dummy variables for each value of the career time variable, $c_{i,t}$.

The degree of research specialization, $H_{i;t,s}$, controls for the potential effects of specialization on productivity. Specialization allows a scientist to become an authority on a given subject (Hackett, 2005). 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 more diverse knowledge. Moreover, specialization might also affect the amount of collaboration, as overly specialized authors may not be able to tackle projects that require knowledge of different fields. Therefore, they may be more willing to collaborate than authors who have more diverse knowledge.

To measure the degree of specialization of an author I use the Herfindahl index. This index is obtained using the fraction of articles published by author i in the field f from period $t - (s - 1)$ to t. Formally, this index is defined as:

$$H_{i;t,s} = \sum_{f=1}^{F} \left(\frac{n_{i;t,s}^{f}}{n_{i;t,s}} \right)^2,$$

where F is the number of fields, $n_{i;t,s}^{f}$ is the number of articles published between $t - (s - 1)$ and t in field f and $n_{i;t,s} = \sum_{f=1}^{F} n_{i;t,s}^{f}$ is the total number of articles published over the period $t - (s - 1)$ to t. To construct this variable, I consider, as in Fafchamps *et al.* (2010), the first two digits of the JEL codes to categorize articles into 121 subfields, f. This measure takes value from $1/F$, reflecting the maximum degree of diversity, to 1 if the author writes all their articles in the same field. Higher values of this index indicate a higher degree of specialization of the author.

Average number of articles of the co-authors, $\bar{n}_{i;t,s}$, is a proxy for the time devoted by i's co-authors to other projects with other authors – i.e. congestion externality. The time allocated by i's co-authors 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. On the other hand, working with busier co-authors may increase the flow of ideas transferred from the busier co-authors to author i through co-authors' co-authors. Thus, the effect of congestion externality on productivity is ambiguous. This variable is computed as:

$$\bar{n}_{1i;t,s} = \frac{\sum_{j \in N_i(G_{t,s})} n_{jt}^{-i}}{|N_i(G_{t,s})|},$$

where $\sum_{j \in N_i(G_{t,s})} n_{jt}^{-i}$ is the number of articles published by all the co-authors of author i over the period $t - (s - 1)$ to t, excluding the articles co-authored with author i. The average number of articles of the co-authors variable is set to zero when the author has no co-authors in the network $G_{t,s}$.

Average co-authors' productivity, $\bar{q}_{1i;t,s}$, controls for the co-authors' quality – in terms of productivity. This variable is defined as,

$$\bar{q}_{1i;t,s} = \frac{\sum_{j \in N_i(G_{t,s})} q_{jt}^{-i}}{|N_i(G_{t,s})|},$$

where $\sum_{j \in N_i(G_{t,s})} q_{jt}^{-i}$ is the research output produced by each co-author j from period $t - (s - 1)$ to t, excluding the co-authored articles published with author i. $|N_i(G_{t,s})|$ is the number of co-authors of author i accumulated from $t - (s - 1)$ to t. The average co-authors' productivity is also set to zero when author i has no co-authors in the network $G_{t,s}$.

Average co-authors' co-author productivity, $\bar{q}_{2i;t,s}$, captures the co-authors' co-authors' quality, as an author may benefit indirectly from the diffusion of ideas between co-authors' co-authors and their co-authors. This variable is obtained as:

$$\bar{q}_{2i;t,s} = \frac{\sum_{j \in N_i^2(G_{t,s})} q_{jt}^{-N_i}}{|N_i^2(G_{t,s})|},$$

where $\sum_{j \in N_i^2(G_{t,s})} q_{jt}^{-N_i}$ is the research output produced over period $t - (s - 1)$ to t by all the co-authors' co-authors accumulated from $t - (s - 1)$ to t, excluding the output produced between these co-authors' co-authors and the co-authors of author i, as the latter output is included in $\bar{q}_{1i;t,s}$. $|N_i^2(G_{t,s})|$ is the number of co-authors' co-authors in the network $G_{t,s}$. The average co-authors' co-author productivity is set to zero when the author has no co-authors' co-authors.

Econometric strategy

The identification of the co-authorship parameter, ρ, comes from *past* variation in the research overlap between author i and their potential co-authors. Therefore, to eliminate 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. The first-difference equation is given by:

$$(log(q_{i;t,s}+1) - log(q_{i;t-1,s}+1)) = \rho(C_{i,t,s} - C_{i,t-1,s}) + (D'_{i,t} - D'_{i,t-1})\omega$$
$$+ \beta_1(H_{i;t,s} - H_{i;t-1,s}) + \beta_2(\bar{n}_{i;t,s} - \bar{n}_{i;t-1,s})$$
$$+ \beta_3(\bar{q}_{1i;t,s} - \bar{q}_{1i;t-1,s}) + \beta_4(\bar{q}_{2i;t,s} - \bar{q}_{2i;t-1,s}) \quad (2)$$
$$+ (\mu_t - \mu_{t-1}) + (\varepsilon_{i;t,s} - \varepsilon_{i;t-1,s}).$$

The main econometric problem is the endogenous co-authorship formation. Authors choose whom to work with, and these associations may be influenced by unobservable characteristics. For example, an author may choose to collaborate because some ideas are hard to tackle individually or because they prefer 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 suggests that less able authors mainly collaborate with authors of a similar type. If this selection was ignored, its effect would be incorrectly attributed to collaboration and biased coefficients would be obtained.

To correct for this type of bias, an instrumenting strategy is implemented. Equation (2) is estimated by an efficient two-step generalized method of moments (GMM), instrumenting $C_{i;t,s}$ by: the commonality research interests between author i and the co-authors' co-authors that author i accumulated in the past network, $G_{t-s-1,s}$, and its quadratic term.

As Fafchamps *et al.* (2010) point out, one of the main determinants of collaboration is some commonality of research interest between the authors. On the other hand, too much overlap in skills may discourage collaboration. Hence, the quadratic term of the common research interests is included to allow for potential nonlinear effects between collaboration and research overlap.

I obtain the proxy for research overlap between author i and their potential co-authors (co-authors' co-authors in the past network), $N_i^2(G_{t-s-1,s})$, using the cosine similarity measure as in Fafchamps *et al.* (2010). This measure is computed as the cosine of two different vectors. One of the vectors, $x_{f;t-s-1,s}^i$, captures the degree of specialization of author i and includes the fraction of articles published by author i in each field f over the period $t - s - s$ to $t - s - 1$, where fields are identified using the first two digits of the JEL codes. The other vector, $x_{f;t-s-1,s}^{N_i^2(G_{t-s-1,s})}$, contains the fraction of articles published by all their co-authors' co-authors accumulated over the period $t - s - s$ to $t - s - 1$ in each field.[9] Using these two vectors, the cosine similarity measure is computed as:

$$w_{2i;t-s-1,s} = \frac{\sum_f x_{f;t-s-1,s}^i x_{f;t-s-1,s}^{N_i^2(G_{t-s-1,s})}}{\sqrt{\sum_f \left(x_{f;t-s-1,s}\right)^2 \sum_f \left(x_{f;t-s-1,s}^{N_i^2(G_{t-s-1,s})}\right)^2}}.$$

[9]To illustrate the definition of the instrument I consider the following example: suppose author i published over the period $t - 10$ to $t - 6$ two different articles, a single-authored article with JEL codes A11, J44 and O30 and a co-authored article with JEL codes A14, D85 and Z13, then the element of vector $x_{f;t-6,5}^i$ corresponding to field A1 would take value 2/6, the elements corresponding to fields J4, O3, D8 and Z1 would be assigned 1/6 and the rest of the elements would be equal to 0. Suppose that the co-authored article was written with two co-authors, h and k. Author h has two co-authors different from i, l and m, over the period $t - 10$ to $t - 6$, while k only collaborated with author i during this period. Therefore, the co-authors' co-authors over the period $t - 10$ to $t - 6$, $N_i^2(G_{t-6,5})$, are authors l and m. Finally, the vector $x_{f;t-6,5}^{N_i^2(G_{t-6,5})}$ is obtained as the vector $x_{f;t-6,5}^i$ but looking at the first two digits of the JEL codes of all the articles published by author l and m over the period $t - 10$ to $t - 6$.

This commonality research index takes a value from 0 to 1.[10] The higher the index, the stronger the research overlap between author i and their potential co-authors. The second instrument is the square of the cosine similarity measure, $w_{2i;t-s-1,s}^2$, and it is introduced to account for the potential inverted-U relationship documented by Fafchamps *et al.* (2010). 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 their potential co-authors.

Since the co-authorship networks $G_{t,s}$ and $G_{t-1,s}$ contain many similar links and the fields of specialization of each authors do not change regularly, the instruments $w_{2i;t-s-1,s}$ and $w_{2i;t-s-1,s}^2$ are quite persistent over time, so $\Delta w_{2i;t-s-1,s}$ would not contain relevant information to infer exogenous variation in $C_{i,t,s}$. To avoid the problems of weak instruments, I consider the instrumental variables in levels. This alternative was proposed by Arellano and Bond (1991), who developed a GMM estimator using lagged levels of the endogenous variables – internal instruments – as instruments for the equation in first differences. Instead, I use the lagged level of the commonality research interest between an author and their potential co-authors and its square – external instruments – as instruments for the difference in the amount of collaboration. Thus, the first-stage regression is given by:

$$
\begin{aligned}
(C_{i,t,s} - C_{i,t-1,s}) = {} & (D_{i,t}' - D_{i,t-1}')\gamma + \theta_1(H_{i;t,s} - H_{i;t-1,s}) + \theta_2(\bar{n}_{i;t,s} - \bar{n}_{i;t-1,s}) \\
& + \theta_3(\bar{q}_{1i;t,s} - \bar{q}_{1i;t-1,s}) + \theta_4(\bar{q}_{2i;t,s} - \bar{q}_{2i;t-1,s}) + (\mu_t - \mu_{t-1}) \quad (3) \\
& + \phi_1 w_{2i;t-s-1,s} + \phi_2 w_{2i;t-s-1,s}^2 + (u_{i;t,s} - u_{i;t-1,s}).
\end{aligned}
$$

As both the productivity and the co-authorship variables are likely to be correlated over time, I cluster standard errors by authors. 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 their potential co-authors.[11] Formally, the two-step efficient GMM estimator relies upon the following moment conditions:

$$
E((\varepsilon_{i;t,s} - \varepsilon_{i;t-1,s})w_{2i;t-s-1,s}) = 0 \Longrightarrow \sum_i w_{2i;t-s-1,s}(\varepsilon_{i;t,s} - \varepsilon_{i;t-1,s}) = 0, \quad (4)
$$

$$
E((\varepsilon_{i;t,s} - \varepsilon_{i;t-1,s})w_{2i;t-s-1,s}^2) = 0 \Longrightarrow \sum_i w_{2i;t-s-1,s}^2(\varepsilon_{i;t,s} - \varepsilon_{i;t-1,s}) = 0, \quad (5)
$$

$$
E((\varepsilon_{i;t,s} - \varepsilon_{i;t-1,s})(X_{i;t,s} - X_{i;t-1,s})) = 0 \Longrightarrow \sum_i (X_{i;t,s} - X_{i;t-1,s})(\varepsilon_{i;t,s} - \varepsilon_{i;t-1,s}) = 0, \quad (6)
$$

where $X_{i;t,s}$ includes all the regressors of equation (3) except the instruments. Although, the average co-authors' productivity, average co-authors' co-authors' productivity and the average number of co-authors' papers might be endogenous to productivity, we can observe in the Bootstrap Hausman test provided in Table 3 that their inclusion does not affect the co-authorship coefficient significantly. The instruments, $w_{2i;t-s-1,s}$ and $w_{2i;t-s-1,s}^2$ are

---

[10] To avoid missing information, the commonality research index is set to zero when the authors have no co-authors' co-authors in the past network.

[11] I used the two-step GMM estimator because it is asymptotically more efficient under heteroscedasticity than the instrumental variable (IV) two-stage least squares estimator (Baum, Schaffer and Stillman, 2003). The same conclusions are obtained if the model is estimated by the IV two-stage least squares.

valid as long as the orthogonality conditions (4) and (5) are satisfied. These assumptions are plausible as these potential co-authors are not current co-authors' co-authors but past co-authors' co-authors. Moreover, it is unlikely that the research overlap would affect the future productivity of an individual through channels other than co-authorship, as this variable is only related to the matching process.

IV.   Data

The database contains information on over 550,000 articles published between 1970 and 2011 in journals listed in EconLit.[12] EconLit does not provide authors' names for articles published by more than three authors before 1999. As a result, these articles are excluded from the analysis for the period 1970–99.[13] EconLit starts to provide author's names for all articles from 2000; hence, I consider all the articles, including those published by more than three authors from 2000 to 2011. The panel data start for each individual with their *first publication* and extend to the last observed publication of the author, or 2011.[14]

As already pointed out by Van der Leij and Goyal (2011), a significant fraction of economists in the EconLit database publish very infrequently and may not publish a single article over a 10-year period. Note that whenever $q_{i;t,s} = 0$, i is by definition not linked to anyone in $G_{t,s}$ and co-authorship, $C_{i;t,s}$, is missing. The past network used to create the commonality research interest, $G_{t-s-1,s}$, is also likely to be missing for these non-research-active authors. Moreover, I am interested in explaining the effect of co-authorship on productivity for individuals who are research-active, i.e. authors who devote an important fraction of their time to research. For these reasons, I restrict attention to individuals who at every point t have published at least one piece in the previous 10 years, i.e. *active authors*.

I also exclude observations relative to authors in the earliest stage of their career, since the model is estimated in first-difference and the instruments, which are based on past information, from $t - s - s$ to $t - s - 1$, are not defined. For example, in the 5-year over rolling window analysis, the instruments are obtained using potential co-authors accumulated from $t - 10$ to $t - 6$; hence, I lose the first six observations of each author or authors who have not been in academia for at least 6 years.[15]

Descriptive analysis

Table 1 shows the percentage of co-authored articles across time and journal quality. I consider all articles published from 1970 to 2011 in an economic journal listed in EconLit. The proportion of co-authored articles published in top-tier journals over the period 1970–2011 with a quality index above 50 is higher than the lowest-tier journals with a quality index

TABLE 1

Percentage of co-authored articles across time and journal quality

Decade/quality	$\geqslant 50$	$20-50$	$10-20$	$1-10$	$\leqslant 1$
1970s	28	31	34	26	21
1980s	46	38	45	37	27
1990s	62	49	54	48	38
2000s	50	64	59	59	48
2010 and 2011	59	70	71	68	59
1970–2011	52	47	53	50	40
Number of Journals	6	12	26	136	750
Number of Articles	12,560	22,562	39,685	148,174	338,841

Notes: Columns (1), (2), (3), (4) and (5) show the percentage of co-authored articles for journals with a quality index above 50, between 20 and 50, between 10 and 20, between 1 and 10, and below 1 respectively.

TABLE 2

Summary statistics of the data

Variables	Low co-authorship		High co-authorship	
	Mean	*SD*	*Mean*	*SD*
Productivity	1.92	7.84	2.33	8.79
Quality	2.01	7.55	2.14	7.38
Experience	6.78	6.77	6.77	6.48
Co-authorship	0.13	0.32	0.86	0.32
Avg. co-authors' productivity	2.62	9.82	2.13	8.44
Avg. co-authors' coauthor productivity	3.58	10.23	2.89	9.02
Avg. co-authors' papers	0.59	1.04	0.53	0.96
Research overlap with co-authors' coauthor	0.37	0.32	0.31	0.30
Degree of specialization	0.31	0.34	0.29	0.33
Number of observations	235,136	235,136	684,403	684,403
Number of authors	66,673	66,673	167,903	167,903

Notes: Column (1) presents summary statistics of authors whose average lifetime co-authorship is below or equal to 0.5. Column (2) shows summary statistics of authors with an average lifetime co-authorship >0.5. These statistics correspond to the 'active' author sample and publications from 1974 to 2011. Co-authorship and network variables statistics are not defined for authors without co-authors in a given year. All the variables are obtained using 1-year period.

below 1. In particular, 52% of the articles published in top-tier journals are co-authored, while 40% of the articles published in the lowest-tier journals are co-authored. The rising pattern in co-authorship varies across journals: the increase in co-authored articles from 1970 to 1999 is higher in top-tier journals than in the lowest-tier journals. This pattern is reversed from 2000 to 2011, where the highest increase in the proportion of co-authored articles is in the lowest-tier journals. Table 2 provides summary statistics of the different variables used in the estimation. Column (1) presents summary statistics for authors with average lifetime co-authorship below 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 annual mean productivity is 1.92, while for those authors with a high lifetime average co-authorship the mean productivity is 2.33. Authors who tend to collaborate more during their career have a lower average co-authors' productivity and co-authors' co-authors' productivity. Also note the high variability of the productivity and network variables, whose standard deviation is generally much higher than the mean.

V. Results

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 provide the results of estimating equation (1) without controlling for the selection of co-authorship. Then I present estimates controlling for the potential endogeneity of co-authorship formation.

   The results presented in columns (1), (2) and (3) of Table 3 are obtained using the first-difference estimator (FD) to control for unobservable individual heterogeneity. Column (1) of Table 3 shows the results of the first-difference specification in which the independent variables include co-authorship, career-time dummies and year dummies. Column (2) adds the degree of specialization as a control variable, and column (3) provides estimates from a first-difference regression of equation (1), controlling for the rest of the time-varying factors, but not for the endogeneity of co-authorship. The implication of these regressions is that co-authorship leads to a lower academic productivity. This result is consistent with Hollis (2001), who finds a negative effect of co-authorship on productivity when research output is discounted by the number of authors and individual unobserved heterogeneity is considered.

   To correct for the possible bias of the co-authorship measure, the instrumenting strategy described in section III is implemented. Columns (4), (5) and (6) of Table 3 show the results of estimating equation (2) using the two-step efficient GMM estimator. Column (4) of Table 3 presents the results controlling for the co-authorship formation process, but not for the time-varying factors. Column (5) shows the results controlling for the endogeneity of co-authorship and controlling for the degree of specialization of the author. Column (6) introduces the rest of the control variables: average number of co-authors papers, co-authors quality and co-authors co-authors quality.[16] Observe that the coefficient of co-authorship becomes significantly positive after instrumenting, which shows that individual productivity increases as authors substitute sole-authorship for teamwork, even when co-authored output is discounted by the number of authors. The latter suggests that the positive externalities from collaboration, not related to quantity since co-authored articles are discounted by the number of authors, exceed the coordination costs and negative externalities of collaboration. 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 they are more willing to accept

---

[16]First-stage estimates are presented in the online appendix.

TABLE 3

The effect of co-authorship on academic productivity: first-difference estimator and instrumental variable first-difference GMM estimation using a 5-year over rolling window

	(1) FD	(2) FD	(3) FD	(4) IV-FD	(5) IV-FD	(6) IV-FD
Co-authorship	-0.143***	-0.1539***	-0.1786***	1.791***	1.4736***	1.7165***
	(0.006)	(0.0057)	(0.0057)	(0.194)	(0.1787)	(0.2448)
Degree of specialization	—	-0.5204***	-0.5114***	—	-0.4733***	-0.4797***
		(0.0064)	(0.0063)		(0.0111)	(0.0113)
Avg. co-authors' productivity	—	—	0.0012***	—	—	0.0016***
			(0.0001)			(0.0001)
Avg. co-authors' coauthor productivity	—	—	0.0011***	—	—	-0.0001
			(0.0001)			(0.0002)
Avg. number of co-authors papers	—	—	0.0038***	—	—	-0.0285***
			(0.0005)			(0.0043)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Career-time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of authors	37,783	37,783	37,783	37,783	37,783	37,783
Number of observations	288,341	288,341	288,341	288,341	288,341	288,341
Cragg–Donald Wald F statistic	—	—	—	83.464	80.249	50.238
Hansen J-test (P-value)	—	—	—	5.728 (0.0167)	1.586 (0.2079)	1.771 (0.1833)
Endogeneity test (P-value)	—	—	—	206.110 (0.0000)	147.775 (0.0000)	123.427 (0.0000)
Bootstrap Hausman test (P-value)	—	—	—	—	—	0.15 (0.7016)

Notes: The sample consists of 'active' authors and publications from 1980 to 2011. Year and career time fixed effects were included in the analysis, but are not reported here to conserve space. The Bootstrap Hausman test evaluates if the difference of the co-authorship parameter estimated by model (6) and model (4) are significantly different from zero. Standard errors in parenthesis adjusted for clusters. ***Significant at 1% level.

any co-authored project regardless of quality. On the other hand, in periods of good ideas, the author will only share ideas that require the skills of other researchers. Then, good ideas not requiring specialization will be sole-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 in the common research interest between author i and their potential co-authors are considered. As the distribution and fields of specialization of these *potential coauthors* affect 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, increasing the benefits from exogenous co-authorship.

We can observe some evidence of the presence of knowledge spillover: the higher the co-authors productivity, the higher the productivity of the author. On the other hand, the quality of authors at a higher distance in the co-authorship network does not affect individual performance. I also find that the congestion externality proxied by the average number of co-authors papers has a negative effect on individual academic productivity. For example, the busier the co-authors 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 significantly affect the effect of co-authorship on academic productivity. Using the Bootstrap Hausman test, I cannot reject the null hypothesis that both estimators of the co-authorship parameter are equivalent.[17] Career time has a negative impact on productivity for authors with >8 years of experience, consistent with the academics' life-cycle effects documented in Levin and Stephan (1991). Specialization has a negative effect on productivity consistent with the findings of Belmaker *et al.* (2010) and Bosquet and Combes (2013).[18]

I reject the null hypothesis of weak instrumental variables. Thus, the instrumental variables are sufficiently correlated with the endogenous variable, co-authorship. Hansen's *J*-test is used for testing the null that the overidentifying moment conditions are true. From Table 3, I cannot reject the null that the instruments are valid when the degree of specialization is included in the model. Moreover, it is clear from the endogeneity test that the co-authorship variable cannot be treated as exogenous.

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.* (forthcoming), it is expected that the benefits from a collaboration differ across individuals. Access to ideas provides an opportunity, but it takes ability and effort to publish a high-quality article. Thus, it is reasonable to suppose that the potential benefits from collaboration vary based on the abilities and efforts of the people involved (Ductor *et al.*, forthcoming). The main hypothesis is that more-able researchers can exploit the benefits from co-authorship to a greater extent.

---

[17] See the online appendix for a description of the test.

[18] 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.

TABLE 4

The effect of co-authorship on academic productivity across individual types: instrumental variable fist-difference GMM estimation using a 5-year over rolling window

	(1) $\geqslant 75\%$	(2) $50-75\%$	(3) $25-50\%$	(4) $\leqslant 25\%$
Co-authorship	1.7181***	1.4570***	0.9932***	0.7191**
	(0.4591)	(0.4051)	(0.3582)	(0.3135)
Degree of specialization	−0.7307***	−0.4351***	−0.3363***	−0.3059***
	(0.0225)	(0.0187)	(0.0164)	(0.0138)
Avg. co-authors' productivity	0.0013***	0.0022***	0.0020***	0.0019***
	(0.0002)	(0.0003)	(0.0003)	(0.0003)
Avg. co-authors' co-author productivity	−0.0002	−0.0003	0.0004	0.0010***
	(0.0003)	(0.0004)	(0.0003)	(0.0003)
Avg. number of co-authors papers	−0.0284***	−0.0246***	−0.0170***	−0.0126**
	(0.0079)	(0.0074)	(0.0065)	(0.0052)
Year FE	Yes	Yes	Yes	Yes
Career-time FE	Yes	Yes	Yes	Yes
Number of authors	10,805	9,768	9,137	8,073
Number of observations	100,568	72,596	62,356	52,821
Cragg–Donald Wald F statistic	18.902	13.880	12.365	12.625
Hansen J-test (P-value)	0.005 (0.9441)	0.290 (0.5902)	0.328 (0.5666)	1.654 (0.1984)
Endogeneity test (P-value)	31.719 (0.0000)	31.855 (0.0000)	14.961 (0.0001)	9.575 (0.0020)

Notes: Column (1) shows the estimation results for authors whose first publication's productivity is above the 75th percentile. Columns (2), (3) and (4) present the estimation results for authors whose first publication's productivity is between the 50th and 75th percentile, between the 25th to 50th percentile and below the 25th percentile respectively. Standard errors in parenthesis adjusted for clusters. ***Significant at 1% level, **significant at 5%.

In order to analyze the potential differences in co-authorship between different types of individuals, I divide the sample into four different groups depending on the productivity of the first publication (as defined in section III).[19] Then, equation (2) is estimated for each sub-sample using the two-step efficient GMM estimator and the moment conditions (4), (5) and (6). Table 4 presents the effect of co-authorship on productivity across high- and low-productivity individuals. Column (1) shows the estimation results for authors whose first publication's productivity is above the 75th percentile. Column (2) presents the results for authors whose first publication's productivity is between the 50th and 75th percentile. Column (3) shows those in the 25th to 50th percentile, and column (4) shows authors below the 25th percentile.

In Table 4 we can observe that the effect of co-authorship on economists productivity varies significantly between different types of individuals. As expected, more able authors obtain more benefits from co-authorship. Authors whose first publication's productivity is below the 25th percentile cannot exploit the benefits from collaboration to the same extent as highly talented individuals – those whose first publication's productivity is above the 75th

[19] Summary statistics reported in the online appendix suggest that the first publication of an author is a good predictor of the future potential performance of an author.

percentile. In the summary statistics across authors (see the online appendix) we can see a clear assortativity in the matching process of the authors. More able authors tend to have highly productive co-authors, while less able authors collaborate with low-productivity researchers.

VI.   Robustness

The main result is the positive relationship between scientific collaboration and individual output after controlling for the endogeneity of the co-authorship. In this section, I test the robustness of the results to the assumptions made in the estimation framework.

Different periods of analysis

In the main analysis, I considered as the period of analysis 5-year over rolling window i.e. productivity is given by $\log([(q_{i,t-4} + q_{i,t-3} + q_{i,t-2} + q_{i,t-1} + q_{i,t})/5] + 1)$ and the co-authorship network at period t contains links formed between $t-4$ to t. In this subsection, I check if the results are sensitive to this assumption by considering shorter periods of analysis: 3 years, and 2 years. First, I compute all the variables for each different period, e.g. the productivity in the 3- and 2-year over rolling window is $\log([(q_{i,t-2} + q_{i,t-1} + q_{i,t})/3] + 1)$ and $\log([(q_{i,t-1} + q_{i,t})/2] + 1)$ respectively. As in the main analysis, only authors who publish at least one piece of research every 10 years are considered. Consequently, the number of missing observations increases as the period of analysis is shorted, since co-authorship is not defined when $q_{i;t,s}$ is zero. Finally, the effect of co-authorship on output is estimated using the empirical strategy described in section III.

Table 5 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 3. The externalities accrued from the network might take a long period to affect the productivity of an author, which could explain the smaller effect of co-authorship under this shorter period of analysis – the 3- and 2-year over rolling windows.[20]

Further robustness

In this subsection, I present the robustness of the results to the assumptions made in constructing the productivity and co-authorship variables, the sample selection and the first-difference transformation. I also check the validity of the instruments to a potential threat. The details and corresponding tables are not reported here to conserve space but are available in the online appendix.

First, it is possible that I am using an inappropriate measure of productivity and co-authorship. I examine if the results differ when productivity is defined as the journal quality index divided by the number of authors working on the article. I also consider the JCR index developed by Reuters (2013) as a different proxy for the journal quality impact factor. Then, I redefine the co-authorship variable as the average number of authors for all articles published by an author from $t-4$ to t. The results show that the positive relationship

---

[20]The results based on a 1-year period are presented in the online appendix.

TABLE 5

The effect of co-authorship on academic productivity assuming a different period of analysis: instrumental variable first-differences GMM estimation using a 3- and 2-year over rolling window

	(1) Three-year	(2) Two-year
Co-authorship	1.1818***	0.9409**
	(0.3334)	(0.4187)
Avg. co-authors' productivity	0.0018***	0.0025***
	(0.0001)	(0.0002)
Avg. co-authors' co-authors productivity	0.0006*	0.0011*
	(0.0003)	(0.0006)
Degree of specialization	−0.6458***	−0.6459***
	(0.0088)	(0.0094)
Avg. number of co-authors papers	−0.0288***	−0.0382**
	(0.0092)	(0.0188)
Number of observations	286,751	278,711
Year dummies	Yes	Yes
Career-time dummies	Yes	Yes
Cragg–Donald Wald F statistic	22.975	10.050
Hansen J-test (P-value)	2.810 (0.0937)	0.373 (0.5412)
Endogeneity test (P-value)	24.914 (0.0000)	11.164 (0.0008)

Notes: Column (1) shows the results of estimating the causal effect of co-authorship on productivity using a 3-year over rolling window. The sample of articles analyzed in column (1) is from 1978 to 2011. The results using 2-year over rolling window are presented in column (2) and include all the articles published from 1977 to 2011. I consider authors who publish at least a piece of research every 10 years, i.e. active sample. Year and career time fixed effects are 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%.

between intellectual collaboration and intellectual output is robust to the specification of the productivity and co-authorship variables.

Second, the results presented so far correspond to authors who publish at least one piece of research every 10 year. To check whether the results are driven by the selection, I estimate equation (2) using the full sample of authors by replacing missing observations of co-authorship and network variables with zero. The 10-year selection and the full sample analysis may lead to an excessive number of zeros for those authors who do not publish regularly. To check if the excessive number of zeros may lead to a spurious positive correlation between co-authorship and productivity, I re-estimate equation (2) using authors who publish at least one article every 5-year. The results presented in the online appendix show that the main conclusions remain under these different samples.

Third, I investigate whether results are sensitive to the transformation used to eliminate the individual fixed effects of equation (1). I consider for robustness forward orthogonal deviations instead of first difference. The results are qualitatively the same.

Finally, I evaluate the validity of the instruments to a potential internal threat. The main identification strategy relies on the assumption that the past common research overlap between an author and her potential co-authors does not affect future changes in

productivity through channels other than co-authorship. However, it is possible that an author might change her degree of field specialization to meet productive potential co-authors and obtain benefits from them that are not passed to productivity through co-authorship – e.g. 'favoritism' in the review process. 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 co-authors are affected by the average productivity of her potential co-authors. The results show that authors do not change their degree of specialization according to the productivity of their past co-authors' co-authors.

VII.   Conclusions and discussion

The aim of this paper is to analyze the effect of intellectual collaboration on individual academic productivity. The approach proposed allows one to control for unobservable heterogeneity, time-varying factors and for the potential endogeneity of teamwork formation. No previous studies have controlled for all of these potential sources of endogeneity simultaneously. The analysis reveals the following:

First, greater collaboration leads to a higher academic productivity even after discounting by the number of authors who worked 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.

Third, over-specialization is detrimental to an author's productivity. I also find evidence for the presence of peer effects and congestion externalities in academic research.

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. That is, authors whose first publication's productivity is above the 75th percentile can exploit the externalities obtained from collaboration to a greater extent. This might be a consequence of the high assortativity in the matching process, which suggests that more able authors mainly collaborate with authors of a similar type – high productivity.

The results raise an important question: if co-authorship leads to higher academic productivity, why may individuals not choose the optimal level of co-authorship? Constraints from collaboration can lead authors to exert a level of collaboration below the social optimum: collaboration involves compromises when working in a group, individual authors have to agree on ideas, texts, approaches, or even conclusions proposed by others (Hudson, 1996). There are also costs associated with finding and assessing co-authors, and organization and coordination costs (He *et al.*, 2009). There are also ex-post transaction costs incurred by the need for coordinating, monitoring and enforcing the contractual promises of inputs from co-authors (Landry and Amara, 1998). The impossibility of anticipating all these costs may discourage private collaboration. Other factors that may lead private agents to choose inefficient levels of collaboration are the tendency of authors to attribute their success to their own

characteristics and their failures to external factors (self-serving biases), and uncertainty about the ability, motivation and effort levels of their potential co-authors. Corgnet (2010) finds that incomplete information will lead to an inefficient level of teamwork as authors are not able to identify the abilities of every potential co-author. He also finds that authors with self-serving biases tend to over-value their contribution to the co-authored project and will decide to write sole-authored projects that could be more efficiently done under collaboration. As a result, policies inducing individuals to collaborate may lead to the socially optimal level of collaboration in cases where organizational constraints are important, when there is uncertainty about co-authors abilities (e.g. collaboration involving juniors without an established publication record), and when self-serving biases are present.

The results are also important for economists, as collaboration between them might enhance their performance, and therefore facilitate access to research funding, higher salaries and prestige.

Future studies could analyze the effect of the different types of collaboration – e.g. mentoring and inter-disciplinary collaboration – on academic productivity, as the benefits are likely to be greater. Mentoring collaboration could probably facilitate the learning process of juniors authors and increase their current and future research output, while inter-disciplinary collaboration combines different types of knowledge and ideas that could lead to greater benefits from collaboration.

Final Manuscript Received: August 2012.

References

Acedo, F. J., Barroso, C., Casanueva, C. and Galán, J. L. (2006). 'Co-authorship in management and organizational studies: an empirical and network analysis', *Journal of Management Studies*, Vol. 43, pp. 957–983.

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

Azoulay, P., Graff Zivin, J. and Wang, J. (2010). 'Superstar extinction', *Quarterly Journal of Economics*, Vol. 125, pp. 549–589.

Barnett, A. H., Ault, R. W. and Kaserman, D. L. (1988). 'The rising incidence of co-authorship in economics: further evidence', *The Review of Economics and Statistics*, Vol. 70, pp. 539–543.

Baum, C. F., Schaffer, M. E. and Stillman, S. (2003). 'Instrumental variables and GMM: estimation and testing', *Stata Journal*, Vol. 3, pp. 1–31.

Belmaker, J., Cooper, N., Lee, T. M. and Wilman, H. (2010). 'Specialization and the road to academic success', *Frontiers in Ecology and the Environment*, Vol. 8, pp. 514–515.

Bonacich, P. (1987). 'Power and centrality: a family of measures', *American Journal of Sociology*, Vol. 92, pp. 1170–1182.

Bosquet, C. and Combes, P. P. (2013). Do Large Departments Make Academics More Productive? Agglomeration and Peer Effects in Research, CEPR Discussion Paper No. 9401.

Bozeman, B. and Corley, E. (2004). 'Scientists? Collaboration strategies: implications for scientific and technical human capital', *Research Policy*, Vol. 33, pp. 599–616.

Bramoullé, Y., Djebbari, H. and Fortin, B. (2009), 'Identification of peer effects through social networks', *Journal of Econometrics*, Vol. 150, pp. 41–55.

Calvó-Armengol, A., Patacchini, E. and Zenou, Y. (2009), 'Peer effects and social networks in education', *Review of Economic Studies*, Vol. 76, pp. 1239–1267.

Carillo M. R., Papagni E. and Sapio, A. (2013), 'Do collaborations enhance the high-quality output of scientific institutions? Evidence from the Italian Research Assessment Exercise', *The Journal of Socio-Economics*, Vol. 47, pp. 25–36.

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*, Vol. 49, pp. 893–916.

Commission of European Communities (2006). 'Amended proposal for Council Decision concerning the Specific Programme "Cooperation" implementing the Seventh framework programme (2007-2013) of the European Community for research, technological development and demonstration activities'.

Conti, G., Galeotti, A., Mueller, G. and Pudney, S. (2013). 'Popularity', *Journal of Human Resources*, Vol. 48, pp. 1072–1094.

Corgnet, B. (2010). 'Team formation and self-serving biases', *Journal of Economics and Management Strategy*, Vol. 19, pp. 117–135.

Defazio, D., Lockett, A. and Wright, M. (2009). 'Funding incentives, collaborative dynamics and scientic productivity: evidence from the European framework program', *Research Policy*, Vol. 38, pp. 293–305.

Ductor, L., Fafchamps, M., Goyal, S. and van der Leij, M. J. (forthcoming). 'Social networks and research output', *The Review of Economic and Statistics*, (`10.1162/REST_a_00430`).

Ellison, G. (2002). 'The slowdown of the economics publishing process', *Journal of Political Economy*, Vol. 110, pp. 947–993.

Fafchamps, M., Goyal, S. and Van der Leij, M. J. (2010). 'Matching and network effects', *Journal of the European Economic Association*, Vol. 8, pp. 203–231.

Gonzalez-Brambila, C. N., Veloso, F. M. and Krackhardt, D. (2013). 'The impact of network embeddedness on research output', *Research Policy*, Vol. 42, pp. 1555–1567.

Goyal, S., Van der Leij, M. J. and Moraga, J. L. (2006). 'Economics: An emerging small world', *Journal of Political Economy*, Vol. 114, pp. 403–412.

Hackett, E. J. (2005). 'Essential tensions: identity, control, and risk in research', *Social Studies of Science*, Vol. 35, pp. 787–826.

Hamermesh, D. S. and Oster, S. M. (2002). 'Tools or toys? The impact of high technology on scholarly productivity', *Economic Inquiry*, Vol. 40, pp. 539–555.

He, Z., Geng, X. and Campbell-Hunt, C. (2009). 'Research collaboration and research output: a longitudinal study of 65 biomedical scientists in a New Zealand University', *Research Policy*, Vol. 38, pp. 306–317.

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

Hudson, J. (1996). 'Trends in multi-authored papers in economics', *Journal of Economic Perspectives*, Vol. 10, pp. 153–158.

Jackson, M. O. and Wolinsky, A. (1996). 'A strategic model of economic and social networks', *Journal of Economic Theory*, Vol. 71, pp. 44–74.

Kodrzycki, Y. and Yu, P. D. (2006). 'New approaches to ranking economics journals', *Contributions to Economic Analysis & Policy*, Vol. 5, art. 24.

Laband, D. N. (1986). 'Article popularity', *Economic Inquiry*, Vol. 24, pp. 173–180.

Laband, D. N. (2013). 'On the use and abuse of economics journal rankings', *The Economic Journal*, Vol. 123, pp. 223–254.

Laband, D. N. and Tollison, R. D. (2000). 'Intellectual collaboration', *Journal of Political Economy*, Vol. 108, pp. 632–662.

Landry, R. and Amara, N. (1998). 'The impact of transaction costs on the institutional structuration of collaborative academic research', *Research Policy*, Vol. 27, pp. 901–913.

Lee, S. and Bozeman, B. (2005). 'The impact of research collaboration on scientific productivity', *Social Studies of Science*, Vol. 35, pp. 673–702.

Lee, L. F., Liu, X. and Lin, X. (2010). 'Specification and estimation of social interaction models with network structures', *Econometrics Journal*, Vol. 13, pp. 145–176.

Levin, S. G., and Stephan, P. E. (1991). 'Research productivity over the life cycle: evidence for academic scientists', *The American Economic Review*, Vol. 81, pp. 114–132.

Lin, X. (2010). 'Identifying peer effects in student academic achievement by a spatial autoregressive model with group unobservables', *Journal of Labor Economics*, Vol. 28, pp. 825–860.

Liu, X., Patacchini, E., Zenou, Y. and Lee, L. F. (2012). *Criminal Networks: Who is the Key Player?* CEPR Discussion Papers No. 8772.

McDowell, J. M. and Melvin, M. (1983). 'The determinants of co-authorship: an analysis of the economics literature', *The Review of Economics and Statistics*, Vol. 65, pp. 155–160.

Medoff, M. H. (2003). 'Collaboration and the quality of economics research', *Labour Economics*, Vol. 10, pp. 597–608.

Melin, G. and Persson, O. (1996). 'Studying research collaboration using co-authorships', *Scientometrics*, Vol. 36, pp. 363–377.

Mihaly, K. (2009). Do More Friends Mean Better Grades? Student Popularity and Academic Achievement, RAND Working Paper Series No. 678.

Ministry of Education and Science (2006). 'Spanish National Plan of Research and Development 2008–2011'. Available at: http://www.ingenio2010.es/.

Reuters, T. (2013). 'Journal Citation Reports'. Web of Science. Available at: http://wokinfo.com/products_tools/analytical/jcr/.

Ubfal, D. and Maffioli, A. (2011). 'The impact of funding on research collaboration: evidence from a developing country', *Research Policy*, Vol. 40, pp. 1269–1279.

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

Van der Leij, M. J. and Goyal, S. (2011). 'Strong ties in a small world', *Review of Network Economics*, Vol. 10, pp. 1–20.

Waldinger, F. (2010). 'Quality matters: the expulsion of professors and the consequences for PhD student outcomes in Nazi Germany', *Journal of Political Economy*, Vol. 118, pp. 787–831.

Wuchty, S., Jones, B. and Uzzi, B. (2007). 'The increasing dominance of teams in production of knowledge', *Science*, Vol. 316, pp. 1036–1039.

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix S1. Further robustness analysis.
Appendix S2. First stage results.
Appendix S3. Summary statistics across individual types.
Appendix S4. Bootstrap Hausman test.