Peer effects: influence and knowledge in scientific circles. (DRAFT)

Lukas Kuld *

June 6, 2016

Using a new data set based on the curricula vitae and publication lists of around 1000 eminent economists, this paper analyses the effect of thematic proximity between researchers within institutions and collaboration networks on the research output. Research articles that refer to colleagues and former co- authors are cited significantly more often than other research by the same authors. Applying different specifications and testing for alternative explanations, this can be attributed to peer effects in the researchers' network. It contrasts with recent studies that fail to find peer effects in economic departments but partially matches the empirical evidence in other scientific fields. This analysis contributes to the understanding of the collaborative and social aspect of creative production in science and other knowledge intensive industries.

*Department of Economics, Trinity College Dublin; Electronic address: kuldl@tcd.ie
I thank Greg Mangan, Andrew Morrow, Sarah Mortell, Conor Parle and Michael Stone for valuable research assistance and John O'Hagan, Tara Bedi and Christiane Hellmanzik for help and comments. The financial support by the Irish Research Council is acknowledged.

1. Introduction

In the growing literature on peer effects, several recent articles focus on scientific research as a knowledge-based industry. The empirical evidence for peer effects in science is mixed. Recent articles fail to find localised peer effects in economic research after the 1980s (viz., Bolli and Schläpfer [2015], Kim et al. [2009]). Others observe productivity gains due to a scientific star as a colleague, doctoral advisor or co-author (the "social space"or "geographical and collaborational space"), in particular, when the star's research focus is similar (the "idea space"), e.g. Azoulay et al. [2010] for life scientists, Waldinger [2012] for physics and chemistry in Nazi Germany, Borjas and Doran [2015] for Soviet mathematicians and Agrawal et al. [2014] in biology,.

To understand the discrepancies in the said studies, it is important to notice that the studied spaces, times, and fields differ. Generally, peer effects are rather identified in star to periphery and close work relations, i.e. within co-author networks.[1]

The standard approach to measure peer effects is to examine the effect of the appearance or disappearance (move, promotion or death) of a scientist in the respective space (e.g. department) on the annual productivity of other researchers in the respective space (e.g. colleagues) or the scientist. The annual productivity is commonly measured as an interaction of published journal pages while accounting for the respective journal's impact factor.

I deviate in two key aspects from the literature. First, I measure the determinants for the success of a single article, not the annual productivity of an economist. Second, I use received citations to measure an article's

[1]Besides science, peer effects are frequently sudied in the economics of education. Studying peer effecs in the arts, a field that is traditionally closely related in reflections on the production in science, several recent studies found peer effects using historical data on writers, visual artists and composers (viz., Mitchell [2015], Hellmanzik [2010] and O'Hagan and Borowiecki [2010]).

success instead of number of pages and journal standing.

The novel focus on articles and citations provides a different perspective on the black box "peer effects". It allows to place the researcher within their thematic and social network. Using references, one can directly trace the influence of a colleague on a research article. Crucially, I observe that an economic article is more influential if the research of an eminent colleague is linked, i.e. cited.

The set of research articles is so split depending on whether it is directly linked to the research of a colleague or may only benefit indirectly by the presence of peers. In this setting, the empirical observation allows to asses the benefit of the overlap of social and idea space.

This paper contributes to the literature with this novel apporach which is made feasible by a new CV based data set. These data allow to trace affiliations and phsyical locations more precisely than mere publication data.

The reminder of the paper is organised as follows. Section 2 describes the theoretical framework for peer effects in knowledge production. For the empirical analysis, the data are described in section 3 and the identification strategy in section 4. Section 5 presents the empirical findings and section 6 concludes.

2. Theoretical framework

Weitzman [1998] describes knowledge production as the testing of new combinations of existing knowledge.[2] A key implication of this concept for research is the importance of access and proximity to other influential knowledge.[3] To

[2]How new knowledge depends on the stock of existing knowledge has been a debate and is expressed in different functional forms in the endogenous growth literature, e.g. Romer [1990], Jones [1995], Jones and Romer [2010]. This is taken up by Agrawal et al. [2014] and Borjas and Doran [2015] to analyse peer effects.

[3]Influence indicates high quality and a large interest in the topic. The validity of this concept is less

test the role of access to knowledge in scientific research, one can trace the references in research articles to previous research, i.e. the codified knowledge upon which they build.

Clearly, previous research articles neither absolutely define nor contain all necessary knowledge to conduct a new research project. Much of the used knowledge is tacit (see Polanyi [1958]);[4] Other knowledge is not cited as it is seen as trivial, or secondary, etc.

Importantly, the set of references informs us about the research topic and if we have additional information on the authors, it reveals personal links to related research[5]. In turn, personal links imply privileged access to non-codified knowledge.

We can express this in a functional form as it will be used in the empirical identification. Let Ω_t denote all relevant knowledge at time t: the knowledge relevant for production and dissemination of research. Then, $\Omega_{j,t} \subset \Omega_t$ is the knowledge held by scientist j at time t.

The interest is on $\Omega_{N_i} = \bigcup\limits_{j,t \in N_i} \Omega_{j,t}$, the knowledge held by other scientists in the neighbourhood of article i (N_i) e.g. the last year and departments of the authors. Denoting further characteristics as θ_i , we can express a simple function for the expected influence of article i, Y_i, as

$$Y_i = \theta_i \Omega_{N_i}^{\beta} \tag{1}$$

trivial than it may seem. The lonely genius is a popular image in the conception of science (cf. Schopenhauer's Republic of geniuses, Nietzsche [1954, 1E 1872], p.269: "One giant calls to the other across the waste spaces of time, and the high spirit-talk goes on, undisturbed by the wanton noisy dwarfs who creep among them.").

[4] While all knowledge is somewhat tacit or rooted in tacit knowledge, tacit knowledge is defined here as knowledge that is not easily codifiable and made accessible to other scientists. For example, the appreciation of a mathematical theory by scientists before experimental observation and the theory's relation to the observation thereafter (Polanyi [1958], p.60) or as expressed in Linus Pauling's bonmot: "If you want to have good ideas you must have many ideas. Most of them will be wrong, and what you have to learn is which ones to throw away."

[5] Additionally, it contains much more information, e.g. is the related research mostly conducted in a different language/country, a different time or field?

3. Data

The data used in this study are based on economist who have published in academic journals between 1996 and 2013 and are well cited. This means that tens of thousands of other economists involved in policy work, national and international government organisations, journalism, teaching and report/book writing are omitted.

The construction of this data set started with all economic journals listed by Kalaitzidakis et al. [2011]. These two hundred plus journals were supplemented with a number of other journals, especially other highly ranked titles in Ideas RePEc. From this process we ended up with 257 of the highest ranked journals in economics.

A search was then carried out using the online database Scopus (`scopus.com`) in relation to the chosen journals. There were over 174,266 research articles published in these journals between 1996 and 2013 and all of these were included in the data analysis.

Next, the authors of these articles were ranked by the number of received citations. From this exercise, a total of 967 economists were chosen based on work published in the period 1996 to 2013 and most highly cited in this period. All of these authors were ranked among the top 1,300 economists.

The reason that we did not look at the whole 1,300 economists is that CVs were not available and/or that their name details were not individual enough to be confidently attributed to a single economist.

For the selected economists, complete information on the research career from the undergraduate studies onwards was collected using online CVs and Blaug and Vane [2003]. This was complemented by additional publication data on Scopus using name searches.

4. Identification strategy

The empirical estimation is carried out using the set of published research articles with at least one of the authors in the list of 967 eminent economists as retrievable on Scopus for the years 1996 to 2014. This implies that each observation corresponds to a research article.

The goal of the empricals estimation is it to identify the value/elasticity of linked scientists, i.e. β in Equation 2. It is straightforward to estimate this relation in a log transformation as

$$log(Y_i) = log(\theta_i) + \beta log(\Omega_{N_i}). \quad (2)$$

The influence of article i, Y_i, is empirically estimated as the sum of received citations. The empirical distribution is shown in Figure 1. To account for the distribution of this variable, a Poisson regression ($\mathbb{E}(Y|X) = e^{X\beta}$, therefore the logarithm as link function) or negative binomial regression to account for overdispersion is used. The distribution of the log transformed dependant variable is shown in Figure 4 in the appendix.

To estimate the knowledge in the neighbourhood of article i, Ω_{N_i} and its value, i.e. β in equation 2, the set of all eminent colleagues around the time of the work on the article[6] is identified for each article. Then, the number of colleagues who are referenced is counted, using only references to articles that are not co-authored by one of the authors. The Ω_{N_i} variable for co-authors is constructed analogously.

All shown regressions estimate a variation of the following reduced form:

$$log(Y_i) = \alpha_j + \delta_t + \Omega_{N_i}\beta + X_i\gamma + \epsilon_{ijt} \quad (3)$$

[6] Economists who are included in our data set and share an affiliation one to five years prior to publication.

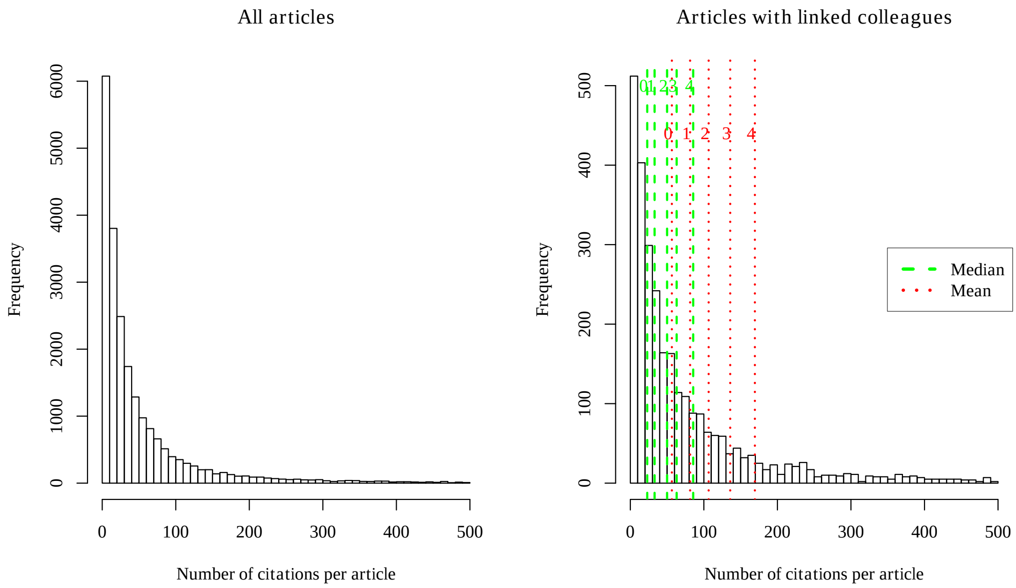

Figure 1: The distribution of received citations for articles published between 1996 and 2010 by the economists in the sample. The right plot shows additionally the mean and median of received citations grouped by number of linked authors. Both plots are truncated at 500 citations. 1.4 % of all in this period published articles and 3.3 % of articles with a linked colleague received more than 500 citations.

Y_i is the sum of citing articles, α_j and δ_t are individual and time fixed effects. Ω_{N_i} is the number of linked colleagues in a given neighbourhood and X_i is a vector of control variables, such as the number of authors, the journal or the number of self citations.

	Dependent variable:			
	Sum of citing articles			
	All articles	All eminent	1 Author	1 Author
Linked co-author	0.103***	0.092***	0.168***	0.082
	(0.018)	(0.032)	(0.048)	(0.053)
Linked colleague	0.112***	0.103***	0.122***	0.154***
	(0.011)	(0.016)	(0.024)	(0.028)
Reference	0.016***	0.015***	0.015***	0.016***
	(0.000)	(0.000)	(0.001)	(0.000)
Self reference	−0.001	−0.010***	−0.006	−0.007
	(0.002)	(0.003)	(0.005)	(0.005)
Number of authors	0.052***	0.708***		
	(0.007)	(0.028)		
Number eminent authors	0.574***			
	(0.018)			
Individual fixed effects	No	No	No	Yes
Time fixed effects	Yes	Yes	Yes	Yes
N (Df)	22098 (22078)	6995 (6976)	4989 (4971)	4989 (4104)
Null deviance	34742	11326.7	7364.6	10957.3
Residual deviance	25904	8313.6	5943.0	5710.1

Table 1: Results of negative binomial regressions using publications between 1996 and 2010. The regressions are using different subsets. Column one shows shows regression results using all articles from this period. The results in column two were obtained using only articles in which all authors are in the set. Columns 3 and 4 present the results of regressions using only single authored papers.

5. Empirical findings

The main interest is on the first two variables, the Ω_{N_i} variables, in Table 1. For example, in the first model (column 1), two linked co-authors and three linked colleagues lead to an expected $(e^{0.103*2+0.112*3} - 1) * 100 = 72\%$ increase in received citations.

The columns show estimates from negative binomial regressions using different subsets of research articles. The data range is truncated after the year 2010 to remove noise. The estimates for all years are reported in the appendix in Table 3.

The parameters in rows and 4 show that the number of references alone is not the cause of the increase in received citations. To test for the effect of citing famous scientists/research, the number of cited dead Nobel prize winners were counted but the observed influence was negative or very small and insignificant.

Interestingly, the effect of linked older co-authors who collaborated 6 to 10 years prior to publication is much smaller than the in Table 1 reported effect of more recent co-authors while the time effect is less pronounced in the case of deapartment colleagues.

The effects for linked colleagues is stable across the different subsets. However, the introduction of individual fixed effects in column 4 halfs the estimated size of the effects of co-authors. On average higher cited economists seem to work more in areas within their co-author network.

6. Conclusion

The here presented empirical evidence underlines the collaborational character of research in economics. The measured influence of thematically and

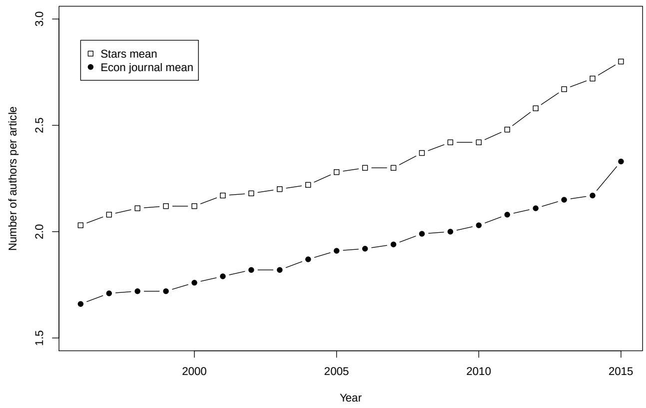

Figure 2: Authors per economic research article: overall and in the sample of eminent economists.

socially linked peers and the significantly higher influence of co-authored articles support the expectation of a further increase in collaborations, a trend that has been observed across scientific fields and is shown in Figure 2 for economics.

A possible explanation is the increased "knowledge burden" Jones [2009], the higher costs to innovate, necessitates a higher specialisation. This would indicate a higher benefit of shared work with other researchers. It is, however not clear to which extend this is geographically bound. Lower costs of communication could account for the increased collaboration and a decline in localised peer effects. Figure 3 shows the internationalisation of collaborations by American economists.

But, even if some aspects of peer effects may disappear, it has been shown here that peers still matter. For a better understanding of localised peer effects more research on the nature of these effects is necessary, e.g. in the creation of collaboration networks[7].

[7] See Freeman and Huang [2014] for a study on homophily in the decision to collaborate

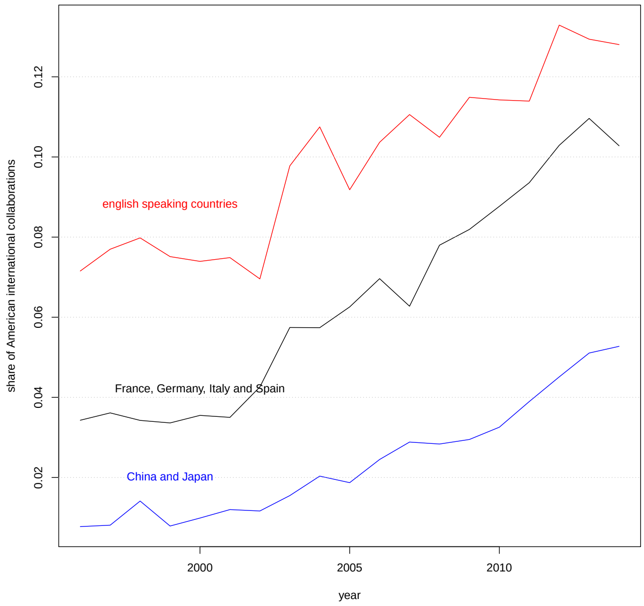

Figure 3: Share of countries of origin in international collaborations involving USA based economists.

References

Agrawal, Ajay K., John McHale, and Alex Oettl, "Why Stars Matter," Working Paper 20012, National Bureau of Economic Research March 2014.

Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang, "Superstar Extinction," *The Quarterly Journal of Economics*, May 2010, *125* (2), 549–589.

Blaug, M. and H.R. Vane, *Who's Who in Economics*, Edward Elgar Pub., 2003.

Bolli, Thomas and Jörg Schläpfer, "Job Mobility, Peer Effects, and Research Productivity in Economics," *Scientometrics*, September 2015, *104* (3), 629–650.

Borjas, George J and Kirk B Doran, "Which peers matter? The relative impacts of collaborators, colleagues, and competitors," *Review of Economics and Statistics*, 2015, *97* (5), 1104–1117.

Freeman, Richard B. and Wei Huang, "Collaborating With People Like Me: Ethnic co-authorship within the US," Working Paper 19905, National Bureau of Economic Research February 2014.

Hellmanzik, Christiane, "Location matters: Estimating cluster premiums for prominent modern artists," *European Economic Review*, February 2010, *54* (2), 199–218.

Jones, Benjamin F., "The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?," *Review of Economic Studies*, 2009, *76* (1), 283–317.

Jones, Charles I, "R&D-Based Models of Economic Growth," *Journal of Political Economy*, August 1995, *103* (4), 759–84.

Jones, Charles I. and Paul M. Romer, "The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital," *American Economic Journal: Macroeconomics*, January 2010, *2* (1), 224–45.

Kalaitzidakis, Pantelis, Theofanis P. Mamuneas, and Thanasis Stengos, "An updated ranking of academic journals in economics," *Canadian Journal of Economics*, November 2011, *44* (4), 1525–1538.

Kim, E. Han, Adair Morse, and Luigi Zingales, "Are elite universities losing their competitive edge?," *Journal of Financial Economics*, 2009, *93* (3), 353–381.

Mitchell, Sara, "London calling? Agglomeration Economies in Literature since 1700," Working Paper, Trinity College Dublin 2015.

Nietzsche, F., *Werke in drei Bänden. Band 1*, Hanser, 1954, 1E 1872.

O'Hagan, John and Karol Jan Borowiecki, "Birth Location, Migration, and Clustering of Important Composers," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2010, *43* (2), 81–90.

Polanyi, Michael, *Personal Knowledge: Towards a Post-critical Philosophy*, The University of Chicago Press, 1958.

Romer, Paul M, "Endogenous Technological Change," *Journal of Political Economy*, October 1990, *98* (5), 71–102.

Waldinger, Fabian, "Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany," *Review of Economic Studies*, 2012, *79* (2), 838–861.

Weitzman, Martin L., "Recombinant Growth," *The Quarterly Journal of Economics*, 1998, *113* (2), 331–360.

A. Figures

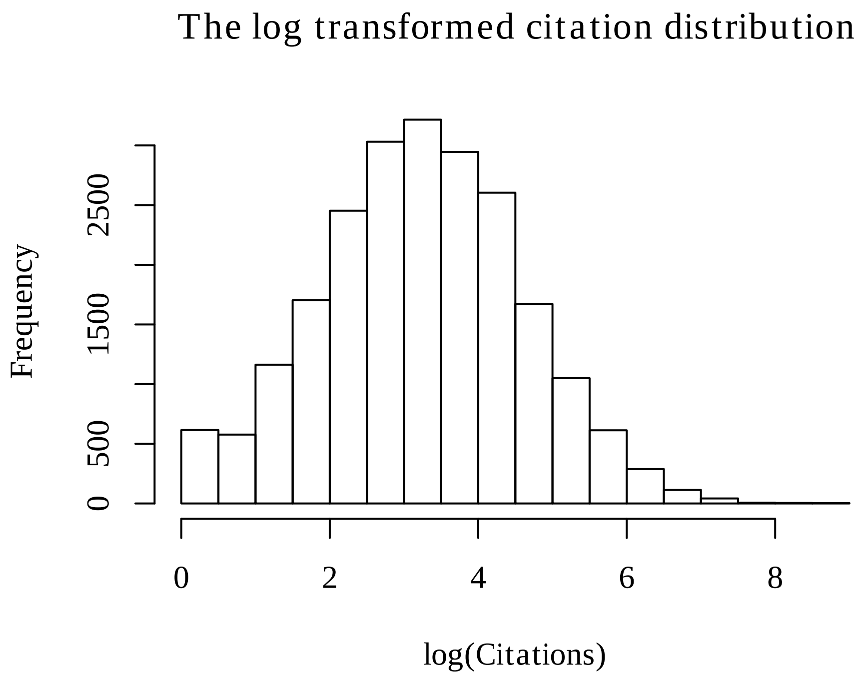

Figure 4: The log transformed distribution of received citations per article published between 1996 and 2010.

B. Tables

C. Robustness checks

Table 2: Results

	Dependent variable:			
	Sum of citing articles			
	All articles	All eminent	1 Author	WRONG FIT!
Linked co-author	0.080^{***}	0.071^{**}	0.143^{***}	0.082
	(0.016)	(0.031)	(0.046)	(0.053)
Linked colleague	0.109^{***}	0.091^{***}	0.109^{***}	0.154^{***}
	(0.010)	(0.016)	(0.024)	(0.028)
Reference	0.015^{***}	0.015^{***}	0.015^{***}	0.016^{***}
	(0.0003)	(0.0005)	(0.001)	(0.001)
Self reference	0.002	-0.014^{***}	-0.013^{***}	-0.007
	(0.001)	(0.003)	(0.004)	(0.005)
Number of authors	0.067^{***}	0.725^{***}		
	(0.006)	(0.026)		
Number eminent authors	0.528^{***}			
	(0.017)			
Individual fixed effects	No	No	No	Yes
Time fixed effects	Yes	Yes	Yes	Yes

Table 3: Results of negative binomial regressions using publications after 1996 . The regressions are using different subsets. Column one shows shows regression results using all articles from this period. The results in column two were obtained using only articles in which all authors are in the set. Columns 3 and 4 present the results of regressions using only single authored papers.