ISSN 2532-5604

CSCC Centro di ricerca sulle Scienze Cognitive e la Comunicazione
Cognitive Science and Communication Research Centre

CSCC Working Papers 01/11

Does co-authorship matter for scientific productivity? Evidence from geography's top journals

Lara Togni*

*Universitá Cattolica del Sacro Cuore, Milano and University of Nottingham

CSCC
Centro di ricerca in Scienze Cognitive e della Comunicazione
Università Cattolica del Sacro Cuore
Via Necchi, 5 - 20123 - Milano ITALY

La pubblicazione nella Collana CSCC Working Papers è soggetta a valutazione da parte di due referees indipendenti attraverso una procedura di valutazione *single blind*.

To be published in the CSCC Working Papers series the article must pass a *single blind* referee procedure by two independent referees.

CSCC Working Papers 01/11

CSCC Centro di ricerca in Scienze Cognitive e della Comunicazione
Università Cattolica del Sacro Cuore
Via Necchi, 5 - 20123 - Milano ITALY

ISSN 2532-5604
CSCC Working Papers
[Online]

DOES CO-AUTHORSHIP MATTER FOR SCIENTIFIC PRODUCTIVITY? EVIDENCE FROM GEOGRAPHY'S TOP JOURNALS

*Lara Togni**

* *Università Cattolica del Sacro Cuore, Milano*
University of Nottingham

Abstract

The aim of this paper is to investigate the relationship between scientific productivity and collaborative behaviours (formal and informal). Despite the different approaches available in the literature, we will focus on what we call the "relational trend": our goal is to detect some of the factors which might affect researchers' productivity, considering "relational variables". In particular, the tendency to write papers in co-authorship will be used as a proxy of formal scientific collaboration amongst scientists, while the number of acknowledgements will be assumed as a proxy of the scientist's ability to build informal collaboration networks. Both co-authorships and acknowledgments indices are interpreted as two of the main forces which could affect and drive scientific production, apart from individual talent. Using the dataset developed by Togni (2009) which collects data about geographers' publications on the Top Journals in the years 2000-2007, an econometric analysis using two-stage least squares has been performed, in order to regress *productivity* on a series of other indices, including (amongst the others) a typical SNA index of centrality (*betweenness centrality*). Three results clearly emerged from the analysis: co-authorships networks affect productivity in a negative way, but a variety of co-authors may increase geographers' productivity; on the contrary, informal influence (acknowledgments network) on productivity seems not to have any effect on productivity. Finally, a trade-off between the transaction costs from the collaboration and the need to mutual exchange of skills and knowledge complementarities which boost scientists to vary their co-authors.

Keywords: SCIENTIFIC NETWORKS, CO-AUTHORSHIPS, ACKNOWLEDGMENTS, SCIENTIFIC PRODUCTIVITY, TOP JOURNALS, GEOGRAPHERS, NETWORK ANALYSIS, TWO-STAGE LEAST SQUARES

JEL code: J16, C01

I. INTRODUCTION: A CONCEPTUAL FRAMEWORK

Over the past sixty years the attention devoted to the world of science has increased significantly, due to the recognition of its main role as one of the crucial determinants of the development of each country (Wagner, 2008). This is also one of the reasons why governments are continuously attempting to foster scientific research through the implementation of new incentive schemes (Shrum *et al*, 2007; Boyle, 2008; Frey 2003, 2009).

Following the pioneer works by Crane (1972) and Price (1986), social scientists began to investigate "The Republic of Science" (David, 2008:2) using appropriate tools of analysis which can help to explain its own patters of knowledge creation and go beyond mere bibliometric indicators.

In particular, one may divide this huge research area into two-stream: the first (which we may label "bibliometric trend") is focused on measuring the effectiveness and efficiency of research activity by taking into account inputs and, above all, outputs of the process (Moed, 2005; Weingart, 2005); the second (which we may label "relational trend") analyses the causes and the scientific behaviours which may lead to different patterns of production, such as the scientists' tendency to write in co-authorship with other scientists or the number of acknowledgments they include in their published articles (Hollis, 2001; Goyal, 2005; Goyal *et al*, 2006; Fachamps *et al*, 2006). This second stream of research has also introduced into this field many innovative tools of analysis, such as Social Network Analysis (Wasserman & Faust, 1994; Scott, 2000) and Textual Analysis (Roth and Cointet, 2010) alongside the most traditional ones.

This paper is assuming the relational trend as the starting point of analysis. The aim is to detect some of the factors which may affect researchers' scientific productivity. In particular, the tendency to write papers in co-authorship will be used as a proxy of formal scientific collaboration amongst scientists, while the number of acknowledgements will be assumed as a proxy of the scientist's ability to build informal collaboration networks. Both co-authorships and acknowledgments indices are interpreted as two of the main forces which could affect and drive scientific production, apart from individual talent.

One of the most common reasons which boosted economists to investigate the relationship between scientific productivity and collaborative behaviours (formal and informal) relies on the belief that collaborative behaviours might enhance both the quality and the quantity of publishable scientific articles. This represents also one of the most controversial assumptions researchers could assume, because of the antithetical results provided by applied works on the subjects (Hollis, 2001).

Owing to the fact that each scientific discipline has its own features which contradistinguish itself from any other, the scientific area of Geography has been selected even if (or just because of) the majority of the contributions already published focus on economics or hard sciences (physics and biology).

After having described the data collected and the criteria behind the construction of the dataset (§1), a two-stage least squares regression model is provided (§2) together with the interpretation of the results we obtained (§3).

II. DATA DESCRIPTION

Before proceeding with the estimation of the econometric model, it is necessary to specify both the rationale behind the dataset we built and the variables we may include in the model.

The original data come from *ISI Web of Science*[1], one of the largest bibliographic dataset provided by Thomson Reuters. The information we are interested in concerns the scientific discipline of Geography (which is also characterised by a strong interdisciplinarity with other social sciences, included Economics) and its pattern of publications during eight years (2000-2007). Therefore, all the bibliometric information regarding the first eight "Top Journals" have been downloaded. A scientific journal is classified as "Top" if its Impact Factor index (henceforth, IF) is ranked in the first 5 per each year. IF is calculated as follows:

$$IF = \frac{two - year\_tot\_citations\_t_3}{tot\_art\_t_1 + t_2\_A\_journal} \quad \textbf{(1)}$$

Where *two-year_tot_citations*$-t_3$ is the amount of citations that the papers published in a generic "A journal" in a two-year period receive in the year following the publication; and $tot\_art\_t_1 + t_2\_A\_journal$ is the total number of articles published in such "A journal" in those two years.

Hence, IF rankings per each single year (2000-2007) were accessed in order to build the dataset. **Table 1** shows the annual IF ranking and the eight top journals which were selected.

[1] http://www.thomsonreuters.com [1 December 2010]

Table 1: Top Journals and Impact Factor

		2000		2001		2002		2003		2004		2005		2006		2007	
ID	**JOURNAL**	IF	rank	IF	rank	IF	rank	IF	rank	IF	rank	IF	rank	IF	rank	IF	rank
TIBG	TRANSACTIONS OF THE INSTITUTE OF BRITISH GEOGRAPHERS	4.067	**1**	3.500	**1**	2.218	**3**	2.388	**3**	2.438	**2**	2.574	**3**	3.093	**1**	2.698	**1**
GEC	GLOBAL ENVIRONMENTAL CHANGE - HUMAN AND POLICY DIMENSIONS	3.915	**2**	2.600	**3**	1.952	**4**	/	/	/	/	/	/	/	/	/	/
PHG	PROGRESS IN HUMAN GEOGRAPHY	3.762	**3**	3.440	**2**	2.616	**2**	2.943	**2**	3.653	**1**	2.762	**1**	2.288	**2**	2.386	**2**
AAAG	ANNALS OF THE ASSOCIATION OF AMERICAN GEOGRAPHERS	2.962	**4**	2.141	**5**	/	/	2.115	**5**	1.972	**5**	2.586	**2**	1.855	**3**	2.109	**4**
JEG	JOURNAL OF ECONOMIC GEOGRAPHY	2.679	**5**	2.519	**4**	3.222	**1**	3.139	**1**	/	/	/	/	/	/	/	/
EG	ECONOMIC GEOGRAPHY	/	/	/	/	1.757	**5**	2.325	**4**	/	/	2.455	**4**	/	/	1.909	**5**
PG	POLITICAL GEOGRAPHY	/	/	/	/	/	/	/	/	2.250	**4**	/	/	1.519	**5**	/	/
EPD	ENVIRONMENT AND PLANNING D-SOCIETY & SPACE	/	/	/	/	/	/	/	/	2.269	**3**	2.377	**5**	1.583	**4**	2.152	**3**

The sampling process originated a list of eight “top journals” which published a total of 2,474 articles (in the period 2000-2007) by 2,436 geographers (authors). To be reminded here that the dataset has been painstakingly cleaned in order to avoid misspelling problems and other trivial mistakes[2].

Additionally, we retrieved information about authors’ affiliation (university, department a/o research centre), city, country and number of acknowledgments per paper and per author. Moreover, in order to compute the productivity index, we looked for the total number of articles every single author has published from 1975 (which is the first year covered by ISI Web of Science) to 2007 (regardless IF).

The data concerning acknowledgments are also crucial in the analysis: they may be considered as a proxy of an unobservable variable concerning the authors’ ability to build informal collaboration networks. The total number of acknowledgments is 7,730, while the average number per paper is 3.12. **Table 2** summarises the information described above.

Table 2: Dataset Description

Years	8
Top Journals	8
Published articles	2,379
Total number of acknowledgments	7,730
Total number of acknowledged authors	636
Average number of acknowledgments per article	3.124
Average number of acknowledgments per author	3.110

Finally, we need to specify that 29 authors were dropped from the dataset due to the impossibility to assign a value to the variable gender[3], while other 19 geographers were dropped due to nationality mismatches[4]. Therefore, the total number of observations became 2,379 instead of 2,436.

After having explained the dataset, a description of the variables is needed.

Amongst the data collected, eleven variables were taken into account in order to estimate their impact on the dependent variable, that is the logarithm of scientific productivity. Furthermore, it must be specified that, unless the data are collected per year, we are dealing with them as aggregate data. **Table 3** reports the variables description.

[2] Deriving, for example, from the inclusion or not of middle initials, “Mortimore, M.” appeared in ISI sometimes as “Mortimore M. L.” or as “Mortimore M. J. L.”. We opted for the latter option.

[3] The information regarding gender was collected by online searching, because ISI Web of Science does not retrieve such type of data.

[4] As gender, the information regarding the authors’ country was collected by online searching, because ISI Web of Science does not retrieve such type of data.

Table 3: Variables Description

VARIABLE	DESCRIPTION
productivity	index of productivity.
coauthorship	average number of co-authors per article.
coauthinstab	co-authors' instability index.
moncoauth	= 1 if the author writes *only* alone and = 0 otherwise.
monpubl	= 1 if the author writes only one article in the period 2000-2007 and = 0 otherwise.
artop0007	number of articles published in one or more of the eight "top journals" of the field (ISI Web of Science impact factor ranking) for the period 2000-2007.
acknavg	average number of acknowledgments per author.
acknnumb	total number of acknowledgments per author.
sciage	author's scientific age.
betweenness	Social Network Analysis index of centrality in the authors' network.
female	= 1 if the author is female and = 0 otherwise.
ukusal	= 1 if the author's affiliation is either in England, Wales, North Ireland, Canada, USA, Australia or New Zealand; and = 0 otherwise.
chijap	= 1 if the author's affiliation is either in China or in Japan; and = 0 otherwise.

Firstly, the variable *productivity* is computed as the ratio between the total number of articles published in the eight top journals in the field of Geography for the period 2000-2007 and the author's scientific age. Instead, the author's scientific age is measured as the difference between the last year we are taking into account (2007) and the author's year of entry in ISI Web of Science database. These two adjustments allow us to control for the researchers' experience, without penalising young geographers.

Secondly, the variable *coauthorship* represents the average number of co-authors per article published by each scientist. More importantly, the variable *coauthinstab* is an index of the instability of the authors who a geographer writes with over time. This index is computed as the ratio between the total number of co-authors who always vary over the total number of co-authors per scientist. Furthermore, the variable *acknavg* represents the average number of acknowledgments per author, while *acknnumb* is a measure of the total number of acknowledgments per scientist.

Thirdly, the variable *betweenness* is a Social Network Analysis (SNA)[5] measure chosen amongst the available indices (**table 4**) of centrality in the authors' network[6] (Freeman, 1979).

Table 4: SNA Indices of Network Centrality

	INDEX	AIM	COMMENT
Degree centrality	$C'_G = \frac{C_G}{(n-1)} = \frac{\sum_{i=1}^{n} l(a_i, a_k)}{(n-1)}$ l = direct link between two adjacent nodes (*ai*, *ak*) C_G = un-normalised degree centrality	It measures the number of links per each single node in the network.	The index is normalised according to the network magnitude.
Closeness centrality	$C'_P(a_k) = \frac{n-1}{\sum_{i=1}^{n} d(a_i, a_k)}$ d = geodesic distance (the smallest one) between two nodes (a_i, a_k)	It measures the node's strategic role in a network: it is related to its ability to reach each other node efficiently and autonomously.	The index is normalised according to the network magnitude.
Betweenness centrality	$C'_I(a_k) = \frac{2C_I(a_k)}{(n-2)(n-1)}$ $\frac{(n-2)(n-1)}{2}$ = maximum value of $C_I(a_k)$	It measures the potential strategic role which a node could play in the scientific collaboration network, through connecting sub networks which otherwise would be disconnected from each others.	The index is normalised according to the network magnitude. The un-normalised index is computed as follows: $b_{ij}(a_k) = \frac{g_{ij}(a_k)}{g_{ij}}$ $b_{ij}(a_k)$ = proportion of geodesic distances (the smallest ones) between every two nodes in the network (i, j and i<z>j $g_{ij}(a_k)$ = geodesic distance between i and j (where we also find z) g_{ij} = the shortest path between i and j

Finally, the other independent variables are dummy variables which are computed as described in **table 3**.

[5] For further information concerning this techniques of analysis, see Wasserman and Faust (1994) and Scott (2000).
[6] This index has been computed using two different software for SNA: *Pajek 1.24* and *Ucinet 6*. Please refer to the bibliographical information for further details.

III. ECONOMETRIC ANALYSIS: TWO-STAGE LEAST SQUARES

The relationship we are interested in is the possible influence that co-authorship may have on productivity. Therefore, the crucial question to pose is whether or not writing with other authors may pay off. **Table 5** shows some descriptive statistics concerning the variables which we are going to include into the model.

Table 5: Descriptive Statistics

VARIABLE	MEAN	STANDARD DEVIATION	MINIMUM	MAXIMUM
productivity	.2307216	.2414975	0	2.1
coauthorship	2.563647	2.964842	0	28
coauthinstab	-.7805637	2.699442	0	1
moncoauth	2.143338	3.023049	0	28
monpubl	.7356032	.4411042	0	1
acknavg	3.365031	4.516346	0	34
acknnumb	5.119798	8.164078	0	131
sciage	10.60614	8.66569	0	32
betweenness	.0012535	.0093117	0	.134
female	.2740647	.4461355	0	1
ukusal	.7545187	.4304627	0	1
chijap	.0281631	.1654734	0	1
Number of observations = 2,379				

The dependent variable is the logarithmic form of the productivity index, while we add all the remaining variables as independent, except for *artop0007*.

Additionally, the scientific age is included in its logarithmic form, in order to make the distribution closer to the normal:

$$ln(productivity_i) = \beta_0 + \beta_1\, coauthorship_i + \beta_2\, coauthinstab_i + \beta_3\, moncoauth_i + \beta_4\, monpubl_i + \\ + \beta_5\, acknavg_i + \beta_6\, acknnumb_i + \beta_7\, ln(sciage_i) + \beta_8\, betweenness_i + \beta_9\, female_i + \\ + \beta_{10}\, ukusal_i + \beta_{11}\, fem\_ukusal_i + \beta_{12}\, chijap_i + \beta_{13}\, fem\_chijap_i + u_{it} \quad \textbf{(2)}$$

As we can see from equation **(2)**, two interaction terms have been added: one in order to control for the interaction between female and Anglo-Saxon countries, and the other in order to control for the interaction between female and China-Japan geographical area.

The main problem we are dealing with in this analysis is the presence of reverse causality: we may assume that cooperative patterns of publication could positively affect scientific productivity; but it could also be argued that productivity might affect co-authorship[7].

Therefore, we are facing a problem of endogeneity of one of the explanatory variables (*coauthorship*), which could be correlated with the error term and therefore generate biased and inconsistent OLS estimators. The tool we are using to overcome this limit is the definition of an instrumental variable (henceforth, IV) and then estimating the model through a two-stage least squares (2SLS) regression.

Hence, we need to find a variable which satisfies two requirements: on one hand, it must not be correlated with *u*; on the other hand, it must be correlated with *coauthorship*, which is in turn the instrumented dependent variable. One of the potential candidates is the variable *artop0007* (the total number of articles published on the top journals for the period 2000-2007). As suggested by Wooldridge (2009:512), in order to test our intuition it is allowed to regress the instrumented variable (*coauthorship*) on the IV (*artop0007*). The results of the regression gave evidence of a statistically positive correlation (1%) between the two variables.

Firstly, both a simple OLS[8] and a 2SLS regressions have been run, in order to be able to test which of the two models could better explain the relationships we are testing.

In particular, we performed a Hausman test[9]: we tested against the null hypothesis that the OLS estimator was consistent and we obtained the result that the null hypothesis can be rejected at the 1% of significance. Hence, the two-stage method is preferred over the OLS regression. **Table 6** illustrates the results of both regressions.

---

[7] For example, scientists could be more attracted by writing papers with authors who have an high productivity rather than with others with a low number of publications. It is well known that having a high number of publications improves authors' reputation inside academia.

[8] Both the White test and the Breusch – Pagan test gave evidence for heteroscedasticity. We had to reject the null hypothesis of homoscedasticity with the 1% of confidence. Therefore, we corrected using robust standard errors. The full results are provided in the appendix.
White test: $\chi^2(81) = 687.68$, Prob > $\chi^2 = 0.000$
Breusch-Pagan test: $\chi^2(13) = 2,797.87$, Prob > $\chi^2 = 0.000$

[9] We included the option "constant", in order to include the constant term in the comparison of both the OLS and the two-stages estimates. Moreover, the option "sigma more" was added in order to use the same estimate of the variance of the error term for the two models. We are allowed to do so because the interpretation of the error term is exactly the same in both models we performed.

Table 6: OLS and Two-Stage Least Squares results (robust standard errors)

Dependent variable: log productivity

INDEPENDENT VARIABLES	OLS	2SLS
coauthorship	-.0375531*** (.0068257)	-1.072764*** (.2181598)
coauthinstab	.0080097*** (.0014219)	.0658599*** (.0127584)
moncoauth	.0353329*** (.0067771)	1.046485*** (.2136866)
monpubl	-.9503638*** (.0235148)	-2.604971*** (.3196528)
acknavg	-.0248672*** (.0029571)	.0062774 (.0054912)
acknnumb	.0188264*** (.0027414)	-.0071947 (.0052597)
lnsciage	-.8236862*** (.0047105)	-.8417227*** (.0169712)
betweenness	8.400692*** (1.1945)	27.35173*** (6.086655)
female	.0032183 (.0158871)	-.1204052 (.0932334)
ukusal	-.0127381 (.010743)	-.1619438** (.0775536)
fem_ukusal	.0000665 (.0191505)	.157393 (.1050781)
chijap	-.0188564 (.0272223)	.1106071 (.2892411)
fem_chijap	-.0643664 (.0475665)	.0934453 (.3709824)
constant	.4693892*** (.0265605)	2.371321*** (.3657218)
Observations	2,231	2,231
R-squared	0.9342	-
Instrumented variable	-	coauthorship
Instrument	-	artop0007

*** *1% level of significance*

** *5% level of significance*

<u>Note</u>*: The quantities in parenthesis below the estimated coefficients refer to standard errors.*

IV. RESULTS

As we can easily note, **table 6** does not report the R^2 for the 2SLS. This is an issue which is well documented in the literature (see amongst the others, Wooldridge, 2009:516-517)[10].

Surprisingly, the variables related to the possible influence of the authors' ability to build informal networks of scientific collaborations (*acknavg* and *acknnumb*) are not significant at any of the usual confidence intervals. It can be assumed that having a rich group of informal collaborators does not affect the amount of published articles an author is able to write. Additionally, the variable *female* (which we add in order to test for the presence of gender discrimination) is not significant, even if we know from the data that only the 27.4% of geographers is composed by women. Consequently, also the coefficients of the interaction terms of female with the two geographical variables (*fem_ukusal* and *fem_chijap*) are not significant.

All the other variables are statistically significant at the 1% confidence level, except for *ukusal* which is significant at the 5%. From the regression emerges that the authors' propensity to write papers with other co-authors negatively affects their productivity. In particular, an increase of one unit of *coauthorship* produces a 10.7% decrease in authors' productivity. Moreover, the coefficient of coauthors' instability is in line with the result above. An increase of one point in the stability index raises productivity of the 6.5%. As a result, we might deduce that at least an author could write with co-authors, but he/she would better change them periodically, instead of keep on writing with them constantly. We can confirm our intuition by taking into account the variable *moncoauth*: authors who published *always* alone are the 10.4% more productive than the geographers who wrote at least one article in co-authorship.

Furthermore, being an "old" author (that is, an author with a high scientific age) negatively affects productivity (-.841). The geographical variables we added to test for the presence of different production patterns produced different results: on one hand, Anglo-Saxon countries present a significant coefficient; on the other hand China and Japan does not. We could also infer that *ukusal* may be a proxy of the presence of a possible language-bias in favour of the native English-speakers, but the issue needs more investigation.

Finally, the relational variable *betweenness* confirms the results related to the co-authorship coefficients. Since *betweenness* positively affects productivity, it should be the case that having few but strategic connections pays off; instead, a solid network of formal collaboration does not. It is also true that this SNA index is a measure of the authors' interposition amongst the different scientific networks; on one hand, only few but strategic collaborative connections could be enough in order to have a high index. On the other hand, being a "bridge" between

[10] It is also possible to find useful comments on the issue in Stata website: http://www.stata.com/support/faqs/stat/2sls.html [10 January 2011].

authors with either different specialised knowledge or different research interests may be crucial (Piette & Ross, 1992).

V. CONCLUSION

The main purpose of this paper was to understand the relationship between scientific productivity and the scientists' tendency to collaborative behaviours. It also took into account some relational factors which could affect productivity related to formal and informal scientific networks. The data on the field of geography suggested that the informal influence (acknowledgments network) on productivity seems not to have any effect on productivity on one hand; co-authorship networks affect productivity in a negative way, but a variety of co-authors may increase geographers' productivity, on the other hand.

Finally, it emerged clearly a trade-off between the "transaction costs" from the collaboration and the need to mutual exchange of skills and knowledge complementarities which boost scientists to vary their co-authors.

References

Boyle, G. (2008) 'Pay Peanuts and Get Monkeys? Evidence from Academia'. *The B. E. Journal of Economic Analysis & Policy* 8 (1): art. 21

Crane, D. (1972) *Invisible Colleges. Diffusion of Knowledge in Scientific Communities* The University of Chicago Press, Chicago & London

David, P. A. (2008) 'The Historical Origins of 'Open Science': An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution'. *Capitalism and Society* 3 (2): art. 5

De Solla Price, D. J. (1986) *Little Science, Big Science... and beyond*. Columbia University Press

Fachamps, M., van der Leij, M. & S. Goyal (2006) 'Scientific Networks and Co-authorship'. *Economics Series Working Papers*, n. 256, Oxford, University of Oxford

Freeman, L. C. (1979) 'Centrality in social networks: conceptual clarification'. *Social Networks*1 (3):215-239

Frey, B. S. (2003) 'Publishing as Prostitution? Choosing Between One's Own Ideas and Academic Success'. *Public Choice* 116 (1-2): 205-223

Frey, B. S. (2009) 'Economists in the PITS?'. *International Review of Economics* 56 (4): 335-364

Goyal, S. (2005) 'Strong and Weak Links'. *Journal of European Economic Association* 3:608-616

Goyal, S., van der Leij, M. & J.L. Moraga-Gonzales (2006) 'Economics: An Emerging Small World'. *Journal of Political Economy* 114: 403-432

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

Moed, H. F. (2005) *Citation Analysis in Research Evaluation* Springer, The Netherlands

Piette, M. J. & K. L. Ross (1992) 'An Analysis of the Determinants of Co-authorship in Economics'. *The Journal of Economic Education* 23 (3):277-283

Roth, C. and J.-P. Cointet (2010) 'Social and Semantic Coevolution in Knowledge Networks'. *Social Networks* 32 (1):16-29

Shrum, W., J. Genuth & I. Chompalov (2007) *Structures of Scientific collaboration* MIT Press (MA)

Scott, J. (2000) *Social network analysis. A Handbook* Sage Publications, London

Togni, L. (2009) 'La Geografia dei Geografi. Il ruolo di collaborazioni e acknowledgments in una comunità scientifica', *CSCC Working Paper Series n. 0209*, December 2009, Centro di Ricerca sulle Scienze Cognitive e sulla Comunicazione (Cognitive Science and Communication Research Centre - CSCC), available on-line at http://www3svil.unicatt.it/pls/unicatt/consultazione.mostra_pagina?id_pagina=16644

Wagner, C. S. (2008) *The New Invisible Colleges: Science for Development*. Brookings Institution Press

Wasserman, S. and K. Faust (1994) *Social network analysis: methods and applications* Cambridge University Press, Cambridge

Weingart, P. (2005) 'Impact of bibliometrics upon the science system: Inadverted consequences?'. *Scientometrics* 62 (1): 117-131

Software

Batagelj, V. & A. Mrvar, *Pajek – Program for Large Network Analysis* (version: *Pajek 1.24*) http://vlado.fmf.uni-lj.si/pub/networks/pajek/

Borgatti, S. P., Everett, M. G. & L. C. Freeman (1999) *Ucinet for windows: Software for Social Network Analysis,* Harvard, MA: Analytic Technologies (version: *Ucinet 6*)

StataCorp. (2009) *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP

Websites

http://www.stata.com/support/faqs/stat/2sls.html [10 January 2011]

http://www.thomsonreuters.com [1 December 2010]

Appendices

a. OLS regression and tests for heteroskedasticity

1) OLS regression

```
regress lnproductivity coauthorship coauthinstab moncoauth monpubl acknavg
acknnumb lnsciage betweenness female ukusal fem_ukusal chijap fem_chijap

      Source |       SS       df       MS              Number of obs =    2231
-------------+------------------------------           F( 13,  2217) = 2420.60
       Model |  1317.22572    13  101.325056           Prob > F      =  0.0000
    Residual |  92.8025012  2217  .041859495           R-squared     =  0.9342
-------------+------------------------------           Adj R-squared =  0.9338
       Total |  1410.02823  2230  .632299653           Root MSE      =   .2046

------------------------------------------------------------------------------
lnproducti~y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
coauthorship |  -.0375531   .0044279    -8.48   0.000    -.0462363   -.0288698
coauthinstab |   .0080097   .0017611     4.55   0.000     .0045561    .0114634
   moncoauth |   .0353329   .0046667     7.57   0.000     .0261814    .0444844
     monpubl |  -.9503638   .0144456   -65.79   0.000     -.978692   -.9220355
     acknavg |  -.0248672   .0014906   -16.68   0.000    -.0277903   -.0219441
    acknnumb |   .0188264   .0008643    21.78   0.000     .0171313    .0205214
    lnsciage |  -.8236862   .0051767  -159.11   0.000    -.8338379   -.8135346
 betweenness |   8.400692   .4897773    17.15   0.000     7.440222    9.361162
      female |   .0032183   .0217239     0.15   0.882     -.039383    .0458196
      ukusal |  -.0127381   .0126187    -1.01   0.313    -.0374839    .0120077
  fem_ukusal |   .0000665   .0243759     0.00   0.998    -.0477355    .0478684
      chijap |  -.0188564   .0299824    -0.63   0.529    -.0776529    .0399402
  fem_chijap |  -.0643664   .0712033    -0.90   0.366    -.2039985    .0752657
       _cons |   .4693892   .0203067    23.11   0.000      .429567    .5092114
------------------------------------------------------------------------------
```

2) White test for heteroskedasticity

```
imtest, white

White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(81)     =    687.68
         Prob > chi2  =    0.0000

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |     687.68     81    0.0000
            Skewness |      60.65     13    0.0000
            Kurtosis |       4.65      1    0.0310
---------------------+-----------------------------
               Total |     752.99     95    0.0000
---------------------------------------------------
```

3) Breusch-Pagan test for heteroskedasticity

```
estat hettest, rhs

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: coauthorship coauthinstab moncoauth monpubl acknavg acknnumb
                    lnsciage betweenness female ukusal fem_ukusal chijap
                    fem_chijap

         chi2(13)      =  2797.87
         Prob > chi2  =   0.0000
```

4) OLS regression (robust standard errors)

```
regress lnproductivity coauthorship coauthinstab moncoauth monpubl acknavg
acknnumb lnsciage betweenness female ukusal fem_ukusal chijap fem_chijap,
vce(robust)

Linear regression                                      Number of obs =    2231
                                                       F( 13,  2217) = 2969.83
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9342
                                                       Root MSE      =   .2046

------------------------------------------------------------------------------
             |               Robust
lnproducti~y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
coauthorship |  -.0375531   .0068257    -5.50   0.000    -.0509386   -.0241676
coauthinstab |   .0080097   .0014219     5.63   0.000     .0052213    .0107982
   moncoauth |   .0353329   .0067771     5.21   0.000     .0220427    .0486231
     monpubl |  -.9503638   .0235148   -40.42   0.000    -.9964772   -.9042503
     acknavg |  -.0248672   .0029571    -8.41   0.000    -.0306662   -.0190682
    acknnumb |   .0188264   .0027414     6.87   0.000     .0134503    .0242024
    lnsciage |  -.8236862   .0047105  -174.86   0.000    -.8329238   -.8144487
 betweenness |   8.400692     1.1945     7.03   0.000     6.058237    10.74315
      female |   .0032183   .0158871     0.20   0.839    -.0279368    .0343734
      ukusal |  -.0127381    .010743    -1.19   0.236    -.0338055    .0083292
  fem_ukusal |   .0000665   .0191505     0.00   0.997    -.0374884    .0376213
      chijap |  -.0188564   .0272223    -0.69   0.489    -.0722403    .0345276
  fem_chijap |  -.0643664   .0475665    -1.35   0.176     -.157646    .0289132
       _cons |   .4693892   .0265605    17.67   0.000     .4173031    .5214752
------------------------------------------------------------------------------
```

b. Two-stage least squares regression and Hausman test

1) 2SLS regression

```
ivregress 2sls lnproductivity coauthinstab moncoauth monpubl acknavg
acknnumb lnsciage betweenness female ukusal fem_ukusal chijap fem_chijap
(coauthorship=artop0007)

Instrumental variables (2SLS) regression              Number of obs =     2231
                                                      Wald chi2(13) = 1295.00
                                                      Prob > chi2   =  0.0000
                                                      R-squared     =       .
                                                      Root MSE      =   1.033

------------------------------------------------------------------------------
lnproducti~y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
coauthorship |  -1.072764   .1346329    -7.97   0.000    -1.336639    -.808888
coauthinstab |   .0658599   .0115808     5.69   0.000     .0431619    .0885579
   moncoauth |   1.046485   .1318015     7.94   0.000     .7881584    1.304811
     monpubl |  -2.604971   .2243854   -11.61   0.000    -3.044759   -2.165184
     acknavg |   .0062774   .0085204     0.74   0.461    -.0104223    .0229772
    acknnumb |  -.0071947   .0054939    -1.31   0.190    -.0179625    .0035731
    lnsciage |  -.8417227   .0262399   -32.08   0.000    -.8931521   -.7902934
 betweenness |   27.35173   3.467347     7.89   0.000     20.55586    34.14761
      female |  -.1204052   .1108268    -1.09   0.277    -.3376217    .0968113
      ukusal |  -.1619438   .0665252    -2.43   0.015    -.2923308   -.0315568
  fem_ukusal |    .157393     .12472     1.26   0.207    -.0870538    .4018398
      chijap |   .1106071   .1522932     0.73   0.468    -.1878821    .4090962
  fem_chijap |   .0934453   .3600844     0.26   0.795    -.6123072    .7991977
       _cons |   2.371321   .2645924     8.96   0.000      1.85273    2.889913
------------------------------------------------------------------------------
Instrumented:  coauthorship
Instruments:  coauthinstab moncoauth monpubl acknavg acknnumb lnsciage
betweenness female ukusal fem_ukusal chijap fem_chijap artop0007
```

2) Hausman test

```
predict ivresid, residuals (148 missing values generated)
est store ivreg
hausman ivreg ., constant sigmamore df(1)
Note: the rank of the differenced variance matrix (1) does not equal the number
        of coefficients being tested (14); be sure this is what you expect, or
        there may be problems computing the test.  Examine the output of your
        estimators for anything unexpected and possibly consider scaling your
        variables so that the coefficients are on a similar scale.

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     ivreg          .           Difference          S.E.
-------------+----------------------------------------------------------------
coauthorship |   -1.072764    -.0375531       -1.035211        .0262943
coauthinstab |    .0658599     .0080097        .0578502        .0014694
   moncoauth |    1.046485     .0353329        1.011152        .0256832
     monpubl |   -2.604971    -.9503638       -1.654608         .042027
     acknavg |    .0062774    -.0248672        .0311446        .0007911
    acknnumb |   -.0071947     .0188264       -.0260211        .0006609
    lnsciage |   -.8417227    -.8236862       -.0180365        .0004581
 betweenness |    27.35173     8.400692        18.95104        .4813561
      female |   -.1204052     .0032183       -.1236235          .00314
      ukusal |   -.1619438    -.0127381       -.1492057        .0037898
  fem_ukusal |     .157393     .0000665        .1573265        .0039961
      chijap |    .1106071    -.0188564        .1294634        .0032884
  fem_chijap |    .0934453    -.0643664        .1578117        .0040084
       _cons |    2.371321     .4693892        1.901932         .048309
```

```
------------------------------------------------------------------------------
b = consistent under Ho and Ha; obtained from ivregress
B = inconsistent under Ha, efficient under Ho; obtained from regress

    Test:  Ho:  difference in coefficients not systematic

                  chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =     1550.01
                Prob>chi2 =      0.0000
                (V_b-V_B is not positive definite)
```

3) 2SLS regression (robust standard errors)

```
ivregress 2sls lnproductivity coauthinstab moncoauth monpubl acknavg
acknnumb lnsciage betweenness female ukusal fem_ukusal chijap fem_chijap
(coauthorship=artop0007), vce(robust)

Instrumental variables (2SLS) regression               Number of obs =    2231
                                                       Wald chi2(13) = 3173.21
                                                       Prob > chi2   =  0.0000
                                                       R-squared     =       .
                                                       Root MSE      =   1.033

------------------------------------------------------------------------------
             |               Robust
lnproducti~y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
coauthorship |  -1.072764   .2181598    -4.92   0.000    -1.500349   -.6451783
coauthinstab |   .0658599   .0127584     5.16   0.000     .0408539    .0908659
   moncoauth |   1.046485   .2136866     4.90   0.000     .6276664    1.465303
     monpubl |  -2.604971   .3196528    -8.15   0.000     -3.23148   -1.978463
     acknavg |   .0062774   .0054912     1.14   0.253    -.0044852      .01704
    acknnumb |  -.0071947   .0052597    -1.37   0.171    -.0175036    .0031142
    lnsciage |  -.8417227   .0169712   -49.60   0.000    -.8749857   -.8084598
 betweenness |   27.35173   6.086655     4.49   0.000     15.42211    39.28136
      female |  -.1204052   .0932334    -1.29   0.197    -.3031393    .0623289
      ukusal |  -.1619438   .0775536    -2.09   0.037     -.313946   -.0099416
  fem_ukusal |    .157393   .1050781     1.50   0.134    -.0485563    .3633423
      chijap |   .1106071   .2892411     0.38   0.702    -.4562951    .6775092
  fem_chijap |   .0934453   .3709824     0.25   0.801    -.6336669    .8205575
       _cons |   2.371321   .3657218     6.48   0.000      1.65452    3.088123
------------------------------------------------------------------------------
Instrumented:  coauthorship
Instruments:   coauthinstab moncoauth monpubl acknavg acknnumb lnsciage betweenness
female ukusal fem_ukusal chijap fem_chijap artop0007
```