FEP WORKING PAPERS

The importance in the papers' impact of the number of pages and of co-authors - an empirical estimation with data from top ranking economic journals

Pedro Cosme Costa Vieira

U.PORTO

FEP FACULDADE DE ECONOMIA
UNIVERSIDADE DO PORTO

THE IMPORTANCE IN THE PAPERS' IMPACT OF THE NUMBER OF PAGES AND OF COAUTHORS - AN EMPIRICAL ESTIMATION WITH DATA FROM TOP RANKING ECONOMIC JOURNALS

Pedro Cosme Costa Vieira

Faculdade de Economia do Porto

R. Dr. Roberto Frias, s/n; 4200-464 Porto; Portugal

pcosme@fep.up.pt

ABSTRACT: On a regular basis the scientific output of academic people has to be evaluated, e.g. to decide tenure. A very important item in this evaluation is the published papers' quality that tends to be approximated by its impact in the literature. As the measure of this impact requires a long-term analysis, thus it is used as its estimator the journal average impact where each paper is published. But some papers have a single author while others have several and some papers have one or two pages while others have more than fifty. In this work I validate the conjecture that these two variables have a significant and positive effect in papers' future impact, i.e. in papers quality. Nonetheless, I quantify that this two variables jointly considered merely explain 2.8% of papers' impact variability.

KEYWORDS: Scientific Skill, Co-authorship, Papers' impact

JEL: J24, J31

1. INTRODUCTION

In order to allocate resources to investigation and to compare the scientific performance of academic people it is necessary to calculate the impact of the papers they publish. But this calculation requires a long-term analysis (see, Vieira 2004b), which turns it not operational to perform. To overcome this difficulty, it is usual to compute with past data the average impact of papers published in each journal (see, Liebowitz and Palmer, 1984 and Laband and Piette, 1994), and to use the result as an estimation of the future impact of papers published today.

But some papers have a single author while others have several and some papers have one or two pages while others have more than fifty. Evaluators (e.g, Kalaitzidakis, Mamuneas and

Stengos, 2003) are not unanimous concerning the influence on papers' quality of the number of pages, and the number of authors. While some conjecture that a published paper with more pages contains more arguments that enlarge the applicability of its results others, by the contrary, argue that this it is just a question of style. Similarly, some evaluators conjecture that each author introduces a different point of view, interdisciplinary, that enlarges the applicability of the paper's results while others state that it is a risk diversification policy.

It is important to stress that these conjectures have never been evaluated with data. Here I intend to perform its evaluation using panel data downloaded from ISI Web of knowledge site at isi4.newisiknowledge.com.

Data was downloaded during the first half of December 2004 for the papers published in the 10 years period between 1987 and 1996 in the 21 top ranking economics journals whose average impact per papers is higher than 10 (see, Vieira, 2004a, p. 13). The selected journals were (average number of citations per paper is named AvCit, standard deviation of the number of citations is named StDevCit, the average number of co-authores per paper is named Coaut and the average number of pages per paper is named Pp):

Journal's name	AvCit	StDevCit	Coaut	Pp
Econometrica	65,17	160,98	1,69	24,17
Journal of Economic Literature	63,67	81,15	1,25	27,88
Journal of Political Economy	53,83	72,34	1,63	24,01
Quarterly Journal of Economics	51,45	88,06	1,67	24,59
Journal of Financial Economics	47,07	57,16	1,91	26,29
Journal of Economic Perspectives	29,86	37,90	1,31	15,13
Journal of Monetary Economics	29,67	72,16	1,56	21,18
American Economic Review	28,43	43,67	1,60	10,33
Review of Economic Studies	27,56	39,14	1,60	17,98
Rand Journal of Economics	23,60	24,82	1,60	16,38
Journal of Econometrics	21,60	43,19	1,73	21,87
Journal of Law Economics & Organization	19,69	32,50	1,59	23,13
Journal of Human Resources	18,44	19,26	1,76	24,34
Journal of Labor Economics	18,23	18,55	1,56	24,94
Economic Journal	15,84	30,64	1,64	13,71
Review of Economics and Statistics	15,77	19,60	1,79	9,73
Journal of Economic Theory	15,01	18,98	1,55	21,39
Games and Economic Behavior	12,69	21,97	1,68	19,45
Brookings Papers on Economic Activity	12,08	21,32	1,96	53,91
Journal of Public Economics	11,61	15,06	1,60	19,29
International Economic Review	8,50	10,79	1,58	16,90

Table1 – Selected journals and average results

It resulted in 10253 papers that, on average, have 19.0 pages, 1.63 co-authors and are cited 26.8 times.

I selected the time span 1987-1996 because "approximately 2/3 of all citations occur 13 years after the paper being published" (Vieira, 2004b: 946).

2. THE EMPIRICAL MODEL

For each selected paper I collected from ISI database the number of papers that cite it (from journals also in the ISI database). That number is a measure of the paper's impact.

Statistically, the conjectures I intend to evaluate imply that the impact $C_{i,j}$ of the paper i published in journal j results from summing a deterministic component with a stochastic component. The deterministic component is a function of the instant of publication T_i (a fractional number that includes the month), the number of authors A_i and the number of pages P_i. The stochastic component e_i is assumed to have a known distribution function (normal).

Being that I intend to estimate a proportional relation (when there is a duplication in the number of co-authors or pages, how does it increase the papers impact?), I use the iso-elastic model, also known as the Cobb-Douglas model.

Representing small letters the natural logarithm of the variables, $x = Ln(X)$, and $\overline{x}_j = \sum_{i=1}^{n} x_{i,j} / n$ the average value of the logarithm variable x in journal j, it is the elastic model that encompasses a "relative" relation between the variables:

$$c_{i,j} = \overline{c}_j + b_1 \cdot (t_i - \overline{t}) + b_2 \cdot (a_i - \overline{a}_j) + b_3 \cdot (p_i - \overline{p}_j) + e_i \qquad (1)$$

To overpass the non-existence of the logarithm of zero, I assume zero impact papers as 0.1.

The magnitude of the deterministic component can be computed comparing the in the sample deterministic variance with the total variance of the impact (excluding the time trend).

3. RESULTS AND CONCLUSIONS

From the data, pooling all the papers together, statistically the effect in the papers' impact of both the number of authors and the number of pages is significant and positive (in brackets it is the *t statistics*), being $R^2 = 14.78\%$:

$$\hat{c}_i = \overline{c}_j - 134.2 \cdot (t_i - \overline{t}) + 0.4237 \cdot (a_i - \overline{a}_j) + 1.083 \cdot (p_i - \overline{p}_j) \tag{2}$$
$$\phantom{\hat{c}_i = \overline{c}_j} (-14.2) (12.8) (37.4)$$

This validates the conjecture that the expected impact of a published paper with more pages or more authors is higher that its counterpart.

Assuming the Bayesian viewpoint where a parameter that is unknown is assumed as a stochastic variable whose distribution may be estimated with the information collected from data, controlled the journal and the instant of publication, when a paper increases one percent in size, in a 99% confidence interval, its expected impact increases between 1.009% and 1.158%. That means the expected impact of a paper is approximately proportional to its size.

When one, two, three or four people author a paper, to each one it must be assumed an equivalence of 0.84, 0.57, 0.45, 0.38 or 0.33 papers, respectively.

Conduction the analysis of variance (ANOVA) on the average logarithm of the impact to test if journals expected impact is equal, the treatment sum of squares (*TSS*) quantifies 3535.7, the residual sum squares (*RSS*) quantifies 22439.4, and Snedecor's *F* quantifies $(TSS/20)/(RSS/10233) = 80.6$. Being the critical value $F_{(20,\ 10233)}$ for a 1% significance 1.88, it is rejected the statistical hypothesis that the expected logarithm of impact for all 21 journals is equal.

Correcting the papers' impact to the same time instant by summing up to them the time trend $134.2 \cdot (t_i - \overline{t})$ and delogarithmed the model, the in the sample deterministic component variance considering the journal average impact, this later plus the number of pages and these two plus the number of authors, is 62.2, 122.1 and 141.1, respectively, from a total variance of 2491.7. Being so, to explicit the journal where the paper is published only explains 2.5% of the total variability of papers impact; the variables number of authors and number of pages considered together increase the explanatory power of the model merely 2.81% of the total variability of papers impact.

Concluding, although the effect in the papers' impact of both the number of authors and the number of pages is statistically significant, validating therefore the conjectures, its magnitude is very small as these variables jointly considered only explain 2.8% of the total variability of papers' impact. Marginal to my analyses but an interesting result is that to explicit the journal where the paper is published only explains 2.5% of the total variability of papers impact. From this, it is statistically acceptable to consider one of the two following evaluation

procedures. First, regardless the journal where the paper is published (in one of the 21 top ranking considered) to assume that each paper is an equally top quality piece of work that must be divide proportionally by the number of authors. Second, to take account of the average journal impact where the paper is published, that the work is proportional to the number of pages and that it is equivalent to 0.84, 0.57, 0.45, 0.38 and 0.33 equivalent papers when it is authored by one, two, three, four or five people, respectively.

Upon request, the author provides the data that was processed using a Microsoft Excel 2000 ™ datasheet.

4. FURTHER WORK

Two anonymous referees selected by the editors of *Economics Bulletin* (that may be accessed at www.economicsbulletin.com) proposed, wisely, that more work needs to be done in this working paper so that it be suitable for publication. Namely, that there are several papers that look at this issue of co-authorship in economics and that regress the number of citations on number of pages and/or number of coauthors that have not been mentioned (e.g., Hudson, 1996; Johnson, 1997; Laband and Tollison, 2000; Hollis, 2001 and Coupe, 2004).

Furthermore, that it is not clear why it is used the iso-elastic specification that forces the author to assume 0.1 citations to papers that are never cited, without any proper diagnostic testing that this specification is adequate. Why not to explore the addition of a number of polynomial terms in the variables of interest and multiplicative, interaction, terms between these variables and testing for the significance of these extra terms by simple F-type tests? Why not apply count data techniques?

Finally, that it would be interesting to see how the analysis shapes up if it is based on a different set of journals or perhaps a subset of the chosen list of 21, where journals like the Journal of Economic Literature and the Journal of Economic Perspectives are excluded since they include a number of solicited state of the art surveys that are highly cited. Also a division between more empirical journals and theoretical ones might shed some additional light into the possible impact effects paper size and numbers of authors.

As I'm not a English native speaker, the text is often hard to read, being needed a serious editorial overhaul.

REFERENCES

Coupe, T. (2004), "What Do We Know about Ourselves? On the Economics of Economics", *Kyklos* **57**, p. 197-216

Hollis, A. (2001), "Co-authorship and the Output of Academic Economists", *Labour Economics* **8**, p. 503-530.

Hudson, J. (1996), "Trends in Multi-Authored Papers in Economics", *Journal of Economic Perspectives* **10**, p. 153-158.

Johnson, D. (1997), "Getting Noticed in Economics: The Determinants of Academic Citations", *American Economist* **41**, nr. 1, p.42-53.

Kalaitzidakis, P. and T. P. Mamuneas and T. Stengos (2003), "Rankings of Academic Journals and Institutions in Economics", *Journal of the European Economic Association* **1**, 1346-66.

Laband, D. and M. Piette (1994), "The Relative Impact of Economic Journals", *Journal of Economic Literature* **32**, 640-66.

Laband, D. and R. Tollison (2000), "Scientific Collaboration", *Journal of Political Economy* **108**, p. 632-662.

Liebowitz, S.J. and J.P. Palmer (1984), "Assessing the relative impacts of economics journals", *Journal of Economic Literature* **22**, 77-88.

Vieira, P.C.C. (2004a), "Top ranking economics journals impact variability and a ranking update to the year 2002", *FEP Working Papers* **149**, 1-16.

Vieira, P.C.C. (2004b), "Statistical variability of top ranking economics journals impact", *Applied Economics Letters* **11**, 945-948.

Recent FEP Working Papers

Nº 168	Leonor Vasconcelos Ferreira, *Social Protection and Chronic Poverty: Portugal and the Southern European Welfare Regime*, March 2005
Nº 167	Stephen G. Donald, Natércia Fortuna and Vladas Pipiras, *On rank estimation in symmetric matrices: the case of indefinite matrix estimators*, February 2005
Nº 166	Pedro Cosme Costa Vieira, *Multi Product Market Equilibrium with Sequential Search*, February 2005
Nº 165	João Correia-da-Silva and Carlos Hervés-Beloso, *Contracts for uncertain delivery*, February 2005
Nº 164	Pedro Cosme Costa Vieira, *Animals domestication and agriculture as outcomes of collusion*, January 2005
Nº 163	Filipe J. Sousa and Luís M. de Castro, *The strategic relevance of business relationships: a preliminary assessment*, December 2004
Nº 162	Carlos Alves and Victor Mendes, *Self-Interest on Mutual Fund Management: Evidence from the Portuguese Market*, November 2004
Nº 161	Paulo Guimarães, Octávio Figueiredo and Douglas Woodward, *Measuring the Localization of Economic Activity: A Random Utility Approach*, October 2004
Nº 160	Ana Teresa Tavares and Stephen Young, *Sourcing Patterns of Foreign-owned Multinational Subsidiaries in Europe*, October 2004
Nº 159	Cristina Barbot, *Low cost carriers, secondary airports and State aid: an economic assessment of the Charleroi affair*, October 2004
Nº 158	Sandra Tavares Silva, Aurora A. C. Teixeira and Mário Rui Silva, *Economics of the Firm and Economic Growth. An hybrid theoretical framework of analysis*, September 2004
Nº 157	Pedro Rui Mazeda Gil, *Expected Profitability of Capital under Uncertainty – a Microeconomic Perspective*, September 2004
Nº 156	Jorge M. S. Valente, *Local and global dominance conditions for the weighted earliness scheduling problem with no idle time*, September 2004
Nº 155	João Correia-da-Silva and Carlos Hervés-Beloso, *Private Information:Similarity as Compatibility*, September 2004
Nº 154	Rui Henrique Alves, *Europe: Looking for a New Model*, September 2004
Nº 153	Aurora A. C. Teixeira, *How has the Portuguese Innovation Capability Evolved? Estimating a time series of the stock of technological knowledge, 1960-2001*, September 2004
Nº 152	Aurora A. C. Teixeira, *Measuring aggregate human capital in Portugal. An update up to 2001*, August 2004
Nº 151	Ana Paula Delgado and Isabel Maria Godinho, *The evolution of city size distribution in Portugal: 1864-2001*, July 2004
Nº 150	Patrícia Teixeira Lopes and Lúcia Lima Rodrigues, *Accounting practices for financial instruments. How far are the Portuguese companies from IAS?*, July 2004
Nº 149	Pedro Cosme Costa Vieira, *Top ranking economics journals impact variability and a ranking update to the year 2002*, June 2004
Nº 148	Maria do Rosário Correia, Scott C. Linn and Andrew Marshall, *An Empirical Investigation of Debt Contract Design: The Determinants of the Choice of Debt Terms in Eurobond Issues*, June 2004
Nº 147	Francisco Castro, *Foreign Direct Investment in a Late Industrialising Country: The Portuguese IDP Revisited*, May 2004
Nº 146	Óscar Afonso and Álvaro Aguiar, *Comércio Externo e Crescimento da Economia Portuguesa no Século XX*, May 2004

Nº 145	Álvaro Aguiar and Manuel M. F. Martins, *O Crescimento da Produtividade da Indústria Portuguesa no Século XX*, May 2004
Nº 144	Álvaro Aguiar and Manuel M. F. Martins, *Growth Cycles in XXth Century European Industrial Productivity: Unbiased Variance Estimation in a Time-varying Parameter Model*, May 2004
Nº 143	Jorge M. S. Valente and Rui A. F. S. Alves, *Beam search algorithms for the early/tardy scheduling problem with release dates*, April 2004
Nº 142	Jorge M. S. Valente and Rui A. F. S. Alves, *Filtered and Recovering beam search algorithms for the early/tardy scheduling problem with no idle time*, April 2004
Nº 141	João A. Ribeiro and Robert W. Scapens, *Power, ERP systems and resistance to management accounting: a case study*, April 2004
Nº 140	Rosa Forte, *The relationship between foreign direct investment and international trade. Substitution or complementarity? A survey*, March 2004
Nº 139	Sandra Silva, *On evolutionary technological change and economic growth: Lakatos as a starting point for appraisal*, March 2004
Nº 138	Maria Manuel Pinho, *Political models of budget deficits: a literature review*, March 2004
Nº 137	Natércia Fortuna, *Local rank tests in a multivariate nonparametric relationship*, February 2004
Nº 136	Argentino Pessoa, *Ideas driven growth: the OECD evidence*, December 2003
Nº 135	Pedro Lains, *Portugal's Growth Paradox, 1870-1950*, December 2003
Nº 134	Pedro Mazeda Gil, *A Model of Firm Behaviour with Equity Constraints and Bankruptcy Costs*, November 2003
Nº 133	Douglas Woodward, Octávio Figueiredo and Paulo Guimarães, *Beyond the Silicon Valley: University R&D and High-Technology Location*, November 2003.
Nº 132	Pedro Cosme da Costa Vieira, *The Impact of Monetary Shocks on Product and Wages: A neoclassical aggregated dynamic model*, July 2003.
Nº 131	Aurora Teixeira and Natércia Fortuna, *Human Capital, Innovation Capability and Economic Growth*, July 2003.
Nº 130	Jorge M. S. Valente and Rui A. F. S. Alves, *Heuristics for the Early/Tardy Scheduling Problem with Release Dates*, May 2003.
Nº 129	Jorge M. S. Valente and Rui A. F. S. Alves, *An Exact Approach to Early/Tardy Scheduling with Release Dates*, May 2003.
Nº 128	Álvaro Almeida, *40 Years of Monetary Targets and Financial Crises in 20 OECD Countries*, April 2003.
Nº 127	Jorge M. S. Valente, *Using Instance Statistics to Determine the Lookahead Parameter Value in the ATC Dispatch Rule: Making a good heuristic better*, April 2003.
Nº 126	Jorge M. S. Valente and Rui A. F. S. Alves, *Improved Heuristics for the Early/Tardy Scheduling Problem with No Idle Time*, April 2003.
Nº 125	Jorge M. S. Valente and Rui A. F. S. Alves, *Improved Lower Bounds for the Early/Tardy Scheduling Problem with No Idle Time*, April 2003.

Editor: Prof. Aurora Teixeira (ateixeira@fep.up.pt)

Download available at:

http://www.fep.up.pt/investigacao/workingpapers/workingpapers.htm

also in http://ideas.repec.org/PaperSeries.html

FEP 2005