2nd INTERNATIONAL CONFERENCE
" DEVELOPMENTS IN ECONOMIC THEORY AND POLICY"
Bilbao, 7th- 8th July, 2005

THE RELEVANCE OF THE NUMBER OF PAGES AND OF COAUTHORS IN THE ASSESSMENT OF ECONOMIC ARTICLES' VALUE – A BOOTSTRAP NON-LINEAR MODEL ESTIMATION

Pedro Cosme Costa Vieira
Faculdade de Economia do Porto

Abstract

In this article I investigate which is the effect that the number of co-authors and pages have in the value of economics articles. I address these questions by estimating numerically a non-linear model with data from the 100 top ranking economics journals covered by the ISI-Web of Knowledge database throughout 1986-1996 (39895 articles). The data reinforces the conjecture that published articles value is increasing with the number of co-authors and the pages but these variables only explain 0,46% and 4,89% of the articles value, respectively. Additionally, the data reinforces the conjecture that reviewers maximize journals citation potential and that there are not significant differences between upper-quantile ranking economics journals and theirs lower-quantile counterpart.

Keywords: Assessment of academics, Co-authorship, Number of pages, Value of articles

JEL: J24, J31

1. Introduction

OECD countries devote an enormous quantity of resources to scientific activities, being an important proportion of these activities performed by academics. Being acknowledged that in order to promote an efficient resources allocation the more productive individuals or work groups must be encouraged to the detriment of theirs counterpart, it is essential to assess the relative academics' output. As the produced scientific items have dissimilar value, the assessment method must use a scale that permits a transversal comparison. As a rule, in decentralised economic system, market prices are the aggregation weights. However, academics are primarily devoted to basic science investigation that the market is unable to price (e.g., Freeman and Soete, 1997). Being so, it is essential to develop alternative ways of scaling scientific output.

For long time, OECD countries' universities use a panel of judges in the comparison of the candidates to a job position or funding opportunity. But economic science has numerous different areas of expertise, being difficult that a panel of judges includes experts in all candidates' areas of specialisation. Additionally, human beings are positively biased in the evaluation of persons similar to them (e.g., Webster's, 1964, studies on the employment interview identifies that the similarity between interviewer and interviewee is the determinate factor in the hiring). To overpass both this difficulties and due to the USA positive influence, in the assessment of academics articles published in scientific journals that implement blind peer evaluation are more and more important.

Starting with Moore (1972), it is an empirical regularity that there is a rank in economic journals, being that top-ranking journals contain a very higher percentage of influential articles than low-ranking ones. Since this work, the ranking of economics journal is abundant in the literature, based on reputation surveys (e.g., Hawkins *et al.*, 1973) or on counting of citations (e.g., Liebowitz and Palmer, 1984, and Laband and Piette, 1994). Mason *et al* (1997, table 2) show that journal rankings based on counting of citations do correlate very positively with rankings based on reputation surveys (correlation proximal to 0.8).

It is also an empirical regularity that economics journals position in the ranking is stationary over time, Vieira (2004). This regularity strength the hypothesis that the reviewers' screening guarantees that, on average, the value of future articles is higher in journals that are top-ranking at present.

Related to the reward policy, Hamermesh *et al* (1982) estimated that each citation increases the salary by about 1%. Sauer (1988) estimated "that an individual's return from a co-authored paper with *n* authors is approximately 1/*n* times that of a single-authored paper". By contrary, McDowell and Kiholm-Smith's (1992) show that for promotion, departments make no difference between co-authored and single-authored papers that is in accordance with Ragan *et al* (1999) that estimated that each published article increases the authors salary by about 1 %.

Being that there is no unanimous opinion about the relevance on the articles' value of the number of co-authors, in this work I intend to test if this variable is statistically significant and quantify, when significant, the magnitude of its influence on the articles' value. Additionally I consider the influence of the articles' number of printed pages. Although those issues have been partially addressed in the literature (e.g., Hudson, 1996, Heck and Zaleski, 1991, Johnson, 1997, Laband and Tollison, 2000; Hollis, 2001, Coupé, 2004) my perspective and methodology are new. First, I use an extended panel data sample (with approximately 40000 articles). Second, I estimate the explicative magnitude of the number of co-authors and the number of pages simultaneously. Third, I use a non-linear model (iso-elastic) where parameters are estimated using *OLS* and Bootstrapping (Efron, 1979) implemented in MS Visual Basic ™. This last issue, although technical, seems to me important because *Ordinary Least Squares* estimators are centric and efficient, and non-linear models are increasingly required. Being *Ordinary Least Squares* method the conceptually simplest and more often used, its estimations are best understood by those that need to estimate a model but are not econometric experts.

As conclusions, data reinforces the conjecture that the number of co-authors and pages is significant in explaining the value of articles. In quantitative terms, I estimate that, on average, the value of co-authors articles is 16% higher than single-authored articles. This supports the conjecture that each co-author introduces in the article a different point of view, interdisciplinarity, that enlarges, although slightly, the applicability of its results. Data reinforces too that the number of pages is significant on the explaining of articles value. In quantitative terms, I estimate that the value of an article with twice the number of pages has, on average, twice the value (the influence is proportional). This supports the conjecture that, due to space limitations, reviewers are exigent on the relevance of each page, expurgating the articles from all non-essential text. Additionally the data reinforces the conjecture that reviewers maximize journals citation potential as if they were unique (monopolistic) and that there are not significant differences between upper-quantile ranking and lower-quantile ranking economics journals.

2. The empirical model

I intend to validate (or invalidate) the hypotheses that the number of co-authors and pages has a positive effect in the articles value. These hypotheses result from the conjectures that each author introduces in the article a different point of view, interdisciplinarity, that enlarges the applicability of its results and that the reviewers, due to space limitations, are exigent on the relevance of each page expurgating the articles from all non-essential text.
Given that the primary objective of publication is the diffusion of knowledge, ex-post it is acceptable to credit more value to articles that were cited more often than others (see, Laband and Sophocleus, 1985). Reversing this correspondence, ex-ante it is more probable that a high value article will be cited in a posterior article than a low value one. Being so, the hypotheses I intend to validation/invalidate may be based on historical data collected by the ISI Web of knowledge.

Panel data was downloaded in the beginning of May 2005 from the ISI Web of knowledge site isi4.newisiknowledge.com. I selected all articles published in the 11 years period between 1986 and 1996 in the 100 top ranking economics journals (ISI Journal Citation Reports' classification).

The data includes 39895 articles that, on average, have 15.83 pages, 1.601 co-authors and were cited 12.83 times in the time span between the day of publication and May 2005. From these articles, 13.9% of them were never cited and 74.9% of them were cited less than average. In the data, the journal with higher average number of citation is *Econometrica* (67.38 citations per article) and the one with lower average number of citation is *Journal of World Trade* (2.09 citations per article).
Similar to Hudson (1996) results, during this 11 years time span, there is an increasing tendency in the number of co-authors, NC, and pages, NP (t statistics in parentheses):

$$\underset{}{\hat{NC}} = \underset{(426.4)}{1.601} + \underset{(17.2)}{0.022}(t-\bar{t}), R^2 = 0.74\%; \quad \hat{NP} = \underset{(358.7)}{15{,}834} + \underset{(16.5)}{0.251}(t-\bar{t}), R^2 = 0.67\% \qquad (1)$$

I selected the time span between 1986 and 1996 (being Jun 1991 the midpoint) because "approximately 2/3 of all citations occur 13 years after the paper being published", Vieira (2004: 946).

After the article being published, other academics may find it, read it and, when they judge that it is valuable, they cite it in the scientific papers they are elaborating for possible publication. This seems an "arriving process" that may be properly modelled as a count data model. Basically, in the count data model the independent variable C is discrete and has

Poisson distribution function with parameter λ that is dependent on a vector of explicative variables $\boldsymbol{X}$.

$$C \sim Po(\lambda) \Rightarrow P(C = c) = \frac{\lambda^c}{c!} e^{-\lambda}, \text{ being } \lambda = h(\boldsymbol{X}) \tag{2}$$

As λ quantify the expected value of the variable C, it formalises the systematic part of the model, while $\varepsilon = C - \lambda$ formalises its random part.

In the Poisson distribution the variance is linear increasing with the expected value. This seems in accordance with the data although in a quadratic form. Computing for each journal the citation average and standard deviation, in Fig. 1 it can be seen a positive linear relation between this two measures. The linear regression shows there is a significant increase of the standard variation with the average (t statistics in parentheses):

$$\begin{array}{llll} StDev = -0.609 & +1.668 \cdot Average & & R^2 = 0.753 \\ \quad\;\; (-0.36) & (17.27) & & \end{array} \tag{3}$$

For the sake of clearness, in the Fig.1 I omitted 5 observations that lie distant from the centre of the figure.

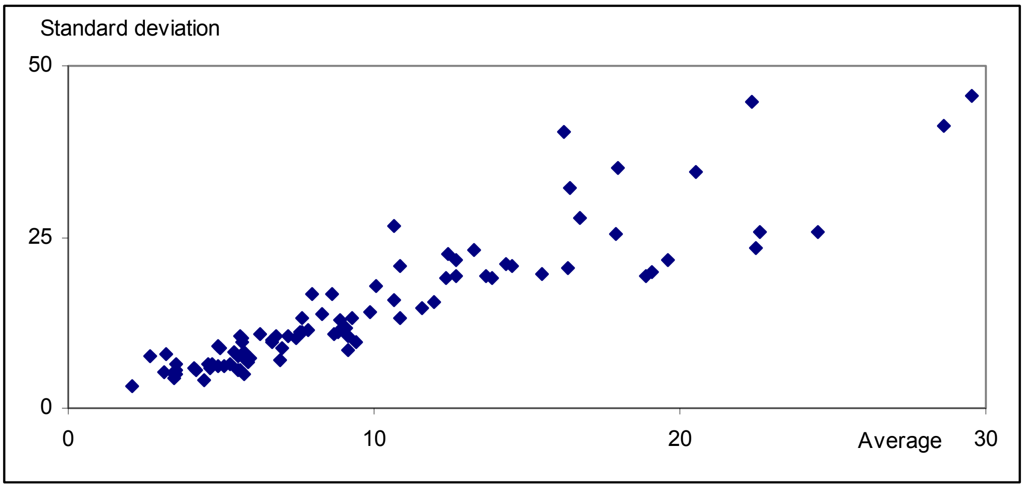

Fig. 1 – Relation between citations standard deviation and average

I speculate that the variance/average quadratic relation is due to authors' asymmetric behaviour: as high-value papers are not submitted to low-ranking journals and numerous low-value papers are submitted to high-ranking journals, almost certainly the reviewers error magnitude is higher in high-ranking journals than in its counterparts. This adds extra variance to the data.

Known the explicative variables, one must yet propose the model functional form. It is straightforward to accept that, when there is an increase in one co-author or one page in an article, the deviation in its value is much lower for those in a low-ranking journal than those

in a high-ranking journal. But in percentage, most likely its effect is the identical. Being assumed that, the adequate model functional form would be the exponential (iso-elastic):

$$\lambda(a_i, p_i \mid \alpha, \beta) = g(j,t) \cdot a_i^{\alpha} \cdot p_i^{\beta} \tag{4}$$

This functional form will permit to test if journals maximize the average citation potential. In this case, the parameter β will be 1 (and decreasing).

The function $g(j, t)$ condenses the fixed-effect of the journal j and the time instant t where the article is published. Assuming independence amongst the journals, theirs fixed effect might be estimated with 100 dummy variables (without autonomous term). Being known that in an arriving process the probability of an occurrence increases proportionally with the time span duration, the time fixed effect is this duration. Being that using a Least Squares minimization (*LS*) the estimated model integrates the "average" point, the fixed-effects may be estimated using this property:

$$g(j,t) = \frac{\bar{c}_j}{\bar{a}_j^{\alpha} \cdot \bar{p}_j^{\beta} \cdot \bar{t}_j} \tag{5}$$

The fixed-effect estimated in this way guarantees that substituting the average value of the explicative variables in the function (4), it results the journal average impact per article.

3. Estimation procedure

Being known which are the explicative variable and the functional form of the function $\lambda(a_i, b_i)$, one need to know the magnitude of the parameters of the model (α and β). Being collected a sample of observations (c_i, a_i, b_i), statistically it is possible to estimate the magnitude of the parameters α and β and test if they are significant. The statistical procedure of obtaining an estimative of the parameters being given a sample of observations is known as regression. Being e_i the deviation from the observed to the estimated model:

$$e_i(\alpha, \beta) = c_i - \lambda(a_i, p_i \mid \alpha, \beta) = c_i - \bar{c}_j \cdot \left(\frac{a_i}{\bar{a}_j}\right)^{\alpha} \cdot \left(\frac{p_i}{\bar{p}_j}\right)^{\beta} \cdot \frac{t_i}{\bar{t}_j} \tag{6}$$

The parameters α and β may be estimated by the minimization of the sum of squared deviations. This method is named *Least Squares* and from its use it results estimators to the unknown parameters that have two good characteristics: they are efficient and centric. Ads to these that *Ordinary Least Squares* method is conceptually simple and it is the method mostly used, being its estimations best understood an accepted by those that need to estimate a model but are not econometricians. Being R the sum of squared residues, one must select values for α and β that minimizes it:

$$R(\alpha,\beta) = \sum_i \left[c_i - \lambda(a_i, p_i \mid \alpha, \beta)\right]^2 \tag{7}$$

The model whose parameter I intend to estimate is non-linear and impossible to linearise because in the data there are articles with zero impact (13.9% of them) that cannot be censored. Being this difficulty, I will use a computational procedure implemented in MS Visual Basic 6.0™ to minimize the expression (7). Then I use bootstrapping re-sampling to compute the variance of the estimators.

The result of the estimation is (the inverse of the coefficient of variation in parentheses):

$$\hat{\alpha} = 0.212\ (5.05) \quad \hat{\beta} = 1.024\ (30{,}56) \quad R^2 = 18.72\% \tag{8}$$

In the computing of the coefficient of variation I used 2500 bootstrapping re-samplings because, increasing from 2000 to 2500, it changes less that 0.001 (both from α and β).
I represent in the Fig. 2 the estimators' frequency density distribution:

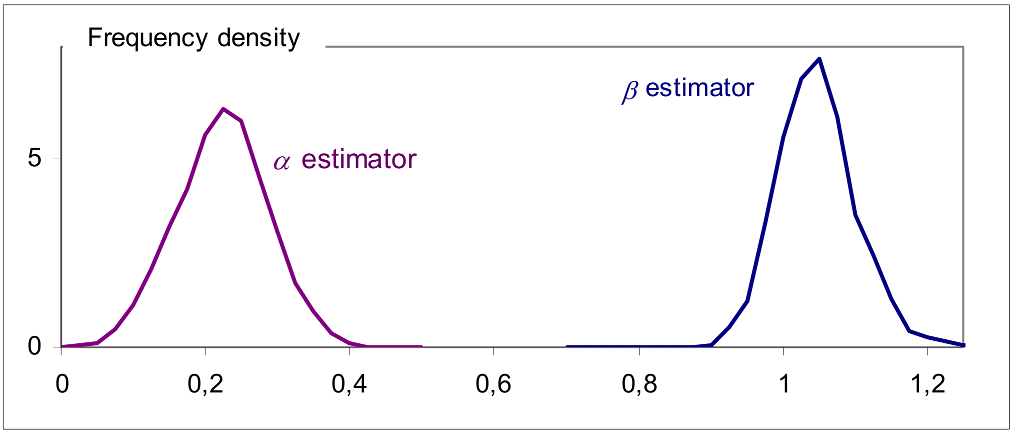

Fig. 2 – Frequency density distribution of the estimator of α and β

In 2500 bootstrapping re-samplings, one do not observes any estimative for α or β smaller than or equal to zero or greater than or equal to twice the average.
Assuming "Ho: the parameter is zero" and that the estimator distribution is normal, the parameter is significant at a certain level when the inverse of the coefficient of variation is greater than the t - distribution critical value.

The normality of α estimator hypotheses may be evaluated using the Kolmogorov-Smirnov test. Even for a 10% level of significance (where the Kolmogorov-Smirnov critical value is 0.0244 = 1.22/√2500), one cannot reject the hypothesis that the α estimator distribution is normal (because the observe value, 0.0188, is smaller than this critical value). The distribution of the β estimator is not so "normal" (the computed value is 0.0345) being rejected the hypothesis that it is normal for a 1% level of significance.

Testing the bilateral significance of the α parameter for a 0.1% level of significance, the critical value is 3.29. For this level of significance, it can be rejected the hypothesis that the α parameter is equal to zero. Although the normality of the β estimator is not guaranteed, its significance is not at danger because the inverse of the coefficient of variation is very larger than the *t*-student 0.1% critical value. Being so, it is rejected the hypothesis that the β parameter is equal to zero. It cannot be rejected the hypothesis that the β parameter is equal to one (the value to test, $(\hat{\beta}-1)/S$, is equal to 0.73 that is very smaller than the critical value for the 5% bilateral test, 1.96). Notice than 63.9% of the β bootstrapping re-sampling estimations are outside the interval]1, 1.0488[.

One may estimate the importance of the variables of the model by computing the percentage of the data variance that is reduced by each variable of de model. The time span reduces the variance by 0.19%; the journal and time fixed-effects reduce the variance by 13,37%; the number of co-authors reduces the variance by 0,46%; and the number of pages reduces the variance by 4,89%.

4. Non-lirearities in the effect of articles' number of pages

I estimated that, on average, one could reject the hypothesis that the value of an article is not proportional to its number of pages (the elasticity is one). This reinforces the idea that reviewers are identically exigent on the relevance of each page, expurgating the articles from all non-essential text. Remembering that, when elasticity is decreasing, a monopolistic maximizes its profit by setting the price where the elasticity of demand is equal to one, this empirical fact reinforces that reviewers maximize the journals citation potencial. That is truth if the citations elasticity is decreasing with the relative number of pages. Assuming an extended model where β evolves linearly with the relative increase of the article number of pages, one may test this conjecture:

$$\beta = \beta_0 + \beta_1 \cdot \left(\frac{p_i}{\bar{p}_j} - 1 \right) \tag{9}$$

The result of the estimation is (the inverse of the coefficient of variation in parentheses):

$$\hat{\alpha} = 0.210\,(3.29) \quad \hat{\beta}_0 = 1.194\,(18{,}33) \quad \hat{\beta}_1 = -0.144\,(-3{,}15) \quad R^2 = 18.89\% \tag{10}$$

The fact that the parameter β_1 is statistically significant for a bilateral significant level of 1% (the critical value is 2.36), validates the conjecture that, on average, reviewers maximize the journals citation potencial. This means that, maintaining the journal total number of pages, an increase in the extension of an article would imply a decrease in the journal citations potential.

5. Are upper-quantile journals diferent from lower-quantile ones?

Although I control the journal where the article is published by estimating a fixed-effect, maybe the influence of the number of co-authors and pages overpass that fixed-effect, being different in upper-quantile journals comparing to lower-quantile ones. To evaluate this

hypothesis I estimate the model (9) for 33% upper-quantile (*U*) and for 33% lower-quantile (*L*) and I compare the estimative:

$$\begin{array}{lll} \hat{\alpha}_L = 0.343\ (3.09) & \hat{\beta}_{0,L} = 1.373\ (6{,}22) & \hat{\beta}_{1,L} = -0.016\ (-0{,}11) \\ \hat{\alpha}_U = 0.098\ (0.81) & \hat{\beta}_{0,U} = 1.116\ (13{,}51) & \hat{\beta}_{1,U} = -0.159\ (-3{,}20) \\ \hline \Delta\hat{\alpha} = 0.245\ (1.50) & \Delta\hat{\beta}_0 = 0.172\ (0.72) & \Delta\hat{\beta}_1 = -0.143\ (-0.94) \end{array} \quad (11)$$

Although it seems that the effect of the number of co-authors is higher in lower-quantile journals, for a significance level of 5%, statistically it cannot be rejected the hypothesis that there are no differences (the critical value for the 5% bilateral test is 1.96). I represent in the Fig. 2 the frequency density distribution of α estimators from 2500 bootstrapping re-samplings where the intersected area is the counterpart of the non-significance result:

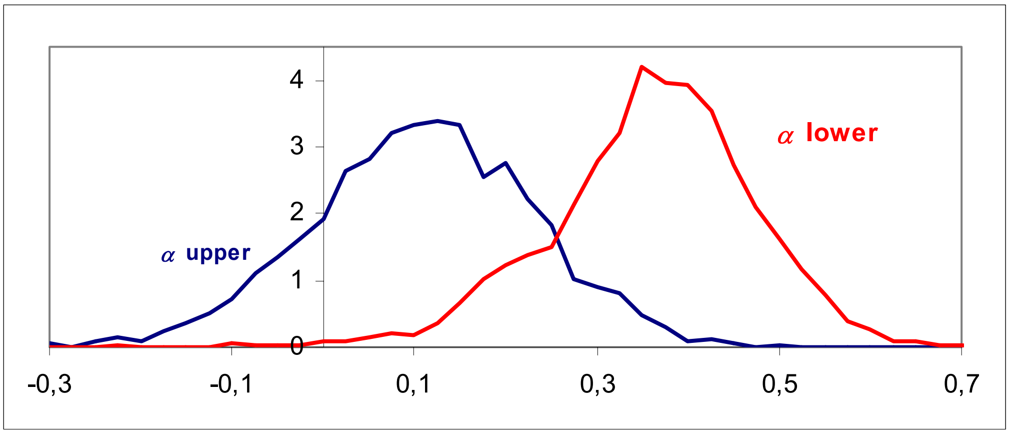

Fig. 3 – Frequency density distribution of the estimator of "α upper" and "α lower"

6. Conclusion

In this work I intend to validate the conjecture that articles' number of co-authors and pages are relevant in the assessment of published economics articles value. Downloaded panel data from the ISI-Web of Knowledge database (39895 articles), I estimated a non-linear model by minimizing the sum of squared deviations. Being that an important percentage articles have zero impact, it was impractical to linearise the model. Being so, I estimated the non-linear model using a minimization computer algorithm and bootstrapping re-sampling.
The data reinforces the conjecture that published articles value is increasing with the number of co-authors and the number of published pages. In quantitative terms, the value of a article

is proportional increasing with the number of pages and a two co-authors article value is 16% higher than a single-authored one. Adds that articles value is increasing with the journal ranking. Nonetheless the statistical significance, the number of co-authors only explains 0,46%; the number of pages 4,89%; and the journal fixed-effect 13,37% of the variance of articles value.

Additionally, the data reinforces the conjecture that reviewers maximize journals citation potential (average elasticity is in the neighbourhood of one and it is decreasing).

Finally, The data reinforces the conjecture that there are no significant structural differences between upper-quantile and lower-quantile ranking economics journals. Nonetheless, qualitatively it seems that the importance of the number of co-authors is higher in lower-ranking journals and that its reviewers are less capable of maximizing the journal citation impact (the elasticity seems higher than one: on average the articles extension is smaller than optimal).

Upon request, the author provides used data and computer programs.

References

Coupé, Tom (2004), "What Do We Know about Ourselves? On the Economics of Economics", *Kyklos* **57**, 197-216

Efron, B. (1979), "Bootstrap methods: Another look at the jackknife", *Annals of Statistics* **7**, 1-26.

Freeman, C. and L. Soete (1997), *The Economics of Industrial Innovation*. The MIT Press: Cambridge, third ed.

Hamermesh, D., G. Johnson and B. Weisbrod (1982), 'Scholarship, Citations and Salaries,: Economic Rewards in Economics', Southern Economic Journal **49**, 472-481.

Hawkins, R. G., L. S. Ritter and I. Walter (1973), "What Economists Think of Their Journals", *Journal of Political Economy* **81**, 1017-32.

Heck, J. L. and P. Zaleski (1991), 'Trends in Economic Journal Literature', 1969- 1989', Atlantic Economic Journal **19**, 27-32.

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

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

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

Laband, D. and J. P. Sophocleus (1985), "Reveals Preference for Economic Journals. Citations as Dollar Votes" Public Choice **46**, 317-24.

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**, 632-662.

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

Mason, P., J. Steagall and M. Fabritius (1997), “Economics Journal Rankings by Type of School: Perceptions Versus Citations “, *Quarterly Journal of Business and Economics* **36**, 69-79.

McDowell, J. and J. Kiholm-Smith (1992), ‘The Effect of Gender-Sorting on Propensity to Co-author: Implications for Academic Promotion’, *Economic Inquiry* **30**, 68-82.

Moore, W.J. (1972), “The relative quality of economics journals: a suggested rating system”, *Western Economic Journal* **10**, 156-69.

Ragan, James F., Jr. and John T. Warren and Bernt Bratsberg (1999), “How Similar Are Pay Structures in 'Similar' Departments of Economics?”, *Economics of Education Review* **18**, 347-60

Sauer, Raymond D. (1988), “Estimates of the returns to quality and coauthorship in economic academia”, *Journal of Political Economy* **96**, 855–66.

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

Webster, E. C. (1964), *Decision Making in the Employment Interview*. Montreal: Eagle.