# EVALUATING ECONOMICS RESEARCH IN EUROPE: AN INTRODUCTION

**J. Peter Neary**
University College Dublin and CEPR

with
**James A. Mirrlees**
Trinity College Cambridge

**Jean Tirole**
University of Toulouse and Ecole des Ponts et Chaussées, Paris

**Abstract**
This paper introduces a symposium of EEA-funded studies that evaluate economics research in Europe. The paper considers some general issues in evaluations, paying special attention to the problem of selecting journal weights, and notes some special features of the individual studies. Despite their very different approaches, the same group of institutions tend to appear at the top of all lists, though individual ranks are sensitive to the choice of more or less elitist journal weights. All the studies show that the gap between economics research in Europe and the United States is narrowing, but remains very wide. (JEL: A10, J44)

## 1. Introduction

This issue of the *Journal of the European Economic Association* contains the output of a project on evaluating economics research in Europe. The project was initiated by the Council of the Association in 1999 and the following call for proposals was issued:

> The European Economic Association is eager to promote a scientific evaluation of research and education in economics carried out in Europe. The EEA views this evaluation as a key step towards improving higher education in economics in European universities. For this purpose, the EEA invites bids for carrying out the task of ranking economics departments throughout Europe and comparing them with the top U.S. departments. The bids should be sent before the end of January 2000 to the Secretary of the EEA: Henry Tulkens. They should include the scope and the methodology of the planned study as well as the budget requested. The EEA hopes to be able to finance three to four studies. The Selection Committee will be chaired by the President of the EEA.

Note: The views expressed in this introduction are the authors' and not those of the European Economic Association.

E-mail addresses: Neary: peter.neary@ucd.ie; Mirrlees: jmirrlees@econ.cam.ac.uk; Tirole: tirole@cict.fr.

In response to this call, eight teams applied, and four were selected for funding by a committee of the EEA consisting of Jim Mirrlees, Peter Neary, and Jean Tirole (EEA President, Vice-President, and President-Elect in 2000, respectively). All four teams submitted full reports, which are available on their web sites (which in turn may be accessed from the EEA web site, ⟨www.eeassoc.org⟩). The next four papers in this issue—by Combes and Linnemer; Coupé; Kalaitzidakis, Mamuneas, and Stengos; and Lubrano, Bauwens, Kirman, and Protopopescu—are edited versions of these reports and have been refereed in the normal way.[1]

Many sets of research rankings have been published for the United States and for individual European countries. However, only two have appeared to date for Europe as a whole. (See Kalaitzidakis, Mamuneas, and Stengos 1999 and Kirman and Dahl 1994.) The Council of the EEA took the view that it was very desirable to encourage more studies of this kind, given the poor governance structures and inappropriate incentives that still characterize so many European universities. The results should help all those attempting to evaluate and develop research capacity, including officials in charge of overseeing and funding universities, and university officers trying to assess the quality of their economics departments.

From the beginning, the focus of this project was on ranking institutions by research in economics as a whole, rather than on other goals such as: evaluating teaching; explaining (as opposed to just measuring) research output; or calculating output measures for different subfields. These and other related topics all merit exploration. However, research is the easiest activity to compare across national boundaries; it seems preferable to first reach consensus on the desired outputs before exploring how efficiently different institutions or different national systems produce them; and research excellence in a subfield of economics is best evaluated by the standards of the worldwide profession as a whole.

Why four studies rather than just one? In principle, the ideal may be a single widely accepted index of every department's research output. However, given the many legitimate areas of disagreement on how an index should be computed, this ideal seems unattainable for the present. The committee took the view that, while there are many pitfalls to be avoided, there is no single best way to carry out studies like these, so the EEA should not endorse a single ranking. In the event, the four studies take quite different approaches (in spite of the fact that all authors had the benefit of seeing first drafts of each others' papers), and they illustrate the diversity of approach in what is rapidly becoming a minor research field in itself.

The remainder of this introduction considers some general issues that arise in any evaluation study, paying special attention to the problem of selecting journal weights, and notes some special features of the individual studies.

1. Each paper was refereed by two outside referees and then considered by the same committee which selected the successful proposals. Authors of submitted papers were not used as referees of other papers. The committee thanks the eight referees for their invaluable assistance.

## 2. General Issues

Research is disseminated in many different ways, from books and journal articles to the Internet. However, only published journal articles undergo a widely accepted process of peer review which is the essence of quality control in any scientific discipline. It is true that an economist's reputation can be enhanced by a well-received book. However, textbooks apart, books in economics are often based on previously published journal articles. In any case, the heterogeneity of book quality makes it nearly impossible to base an objective ranking on them. Hence, like almost all previous evaluation studies, the four papers in this symposium concentrate on evaluating the quality of research by considering only published journal articles.

Even limiting the definition of research to journal articles, there are many choices that need to be made. Alternative approaches can be classified using a formula due to van Damme (1996). This determines the "score" $S$ of an individual researcher $r$ as follows:

$$S = \sum_r \frac{\beta(p_r)\omega(p_r)}{\alpha(p_r)} \tag{1}$$

where for each publication $p_r$ by that individual, $\beta(p_r)$ denotes its length, $\omega(p_r)$ denotes a quality weight, and $\alpha(p_r)$ denotes a correction for coauthorship. The score of an institution is then the total of the scores of its individuals, corrected for multiple affiliations.

While this is simple in principle, there are many areas of disagreement in applying this formula. Probably the most controversial issue is the choice of weights $\omega(p)$, which is considered in Section 3. This section considers the remaining issues in turn, referring to Table 1 for a summary of how the four studies have dealt with them.

*Correcting for Publication Length.* Longer articles need not be better, but length is correlated with importance, at least as perceived ex ante by editors and referees. So it makes sense to correct for the length of each paper. Most authors

TABLE 1. SCHEMATIC COMPARISON OF THE STUDIES

| | Publication length: $\beta(p)$ | Coauthorship: $\alpha(p)$ | Stocks/ Flows | # of journals | Journal weights: $\omega(p)$ |
|---|---|---|---|---|---|
| Combes-Linnemer | # of pages | $n$ | Stocks | 855 | Subjective |
| Coupé | # of pages | $n$ | Flows | 273 | Eclectic: citations and articles |
| Kalaitzidakis et al. | # of pages, corrected for page size | $n$ | Flows | 30 | Eigenvector |
| Lubrano et al. | $\beta(p) = 1$ | $n^{0.5}$ | Flows | 505 | Subjective and objective |

Notes: See text for details. Where a study gives alternative approaches, the table indicates the one used in its preferred specification.

in the literature do so. A refinement is to recognize that journals differ in their page sizes, so a correction should be made for the number of characters per page. Among the four papers in this issue, only Lubrano et al. depart from this approach, and give an equal weight to each article in a given journal irrespective of its length. They justify this on the grounds that journals differ in the average length of the articles they publish, and that notes are sometimes more cited than full articles. While these points are well taken, it is not clear that they justify setting $\beta(p)$ equal to one in all cases.

*Correcting for Coauthorship.* Some correction for coauthorship seems essential. Most authors adopt the simple $\alpha(p) = n$ rule, where $n$ is the number of coauthors. Lubrano et al. argue that this penalizes small institutions (whose members are less likely to have in-house coauthors) and choose $\alpha(p) = n^{0.5}$ instead. However, any value for $\alpha(p)$ less than $n$ implies that multiple authorship per se increases the aggregate value of an article.[2] If widely adopted, it would also encourage strategic behavior, as authors would face incentives to "swap" coauthorship with colleagues in order to raise their total score.

*Stocks versus Flows.* How should authors be allocated to institutions: by their affiliations at the time of publication (flows) or at the end of the period (stocks)? The stock approach seems clearly preferable in principle for most purposes (other than historical). Students choosing where to do their Ph.D., or officials overseeing departments, want to be informed about its current state, not about past performance. However, it is harder to measure accurately and requires a lot of extra work. Dusansky and Vernon (1998), in a much-quoted study of U.S. departments, found that a software-compiled first-pass list taken from EconLit was less than 50 percent as accurate as a more careful (and so more labor-intensive) compilation of authors' current affiliations. Since they looked at only eight journals and eighty institutions in the relatively homogeneous U.S. market, this suggests that even more care is needed with the extremely diverse range of European institutions. Combes and Linnemer is the only study of the four included here that attempts this approach.

*Correcting for Multiple Affiliations.* The logic of the flow approach is to take authors at their word when they list their affiliations in their papers. Thus each institution to which an author gives an affiliation is credited with an equal share of the author's score for that paper. (An uncontroversial exception is that network affiliations such as CEPR, CNRS, NBER, etc., should be ignored.) By contrast, the logic of the stock approach is that every author's current affiliation needs to be independently verified, perhaps by consulting the author's web site

2. Coupé quotes mixed evidence on this. On the one hand, multiauthored papers appear to be more cited on average. On the other hand, wage equations suggest that academics are paid on average according to the $\alpha(p) = n$ rule.

or their head of department. This increases the work load enormously, and of the four papers, only Combes and Linnemer attempt to implement it.

*Aggregating from Individuals to Departments.* This should be a straightforward exercise, but even here pitfalls arise. Whereas most of the studies count all publications by the members of a department, Lubrano et al. set a threshold level, equal to the equivalent of one *AER* paper with a single coauthor over a ten-year period, below which an individual's publications are not counted. A different issue is what counts as a department, and in particular how to treat subcenters. This issue also arises in the United States. For example, the study of Dusansky and Vernon (1998) was criticized because it excluded economists working in business schools, even in cases where the business schools were closely integrated with the economics departments in the university in question. In Europe the problem is even more severe. Funding and other constraints have led to a proliferation of subcenters in many countries, making it harder to attain the economies of research scale that typify the best U.S. departments. Most of the studies treat all economists on one campus as belonging to the same institution, whereas Combes and Linnemer devote considerable attention to distinguishing between different centers and subcenters.

*Correcting for Department Size.* A final issue is whether research output should be calculated for an institution as a whole or on a per member basis. The former measures overall strength, though it is clearly biased in favor of large departments. The latter is a crude measure of productivity, though a full study of the efficiency of research production should take account of all the inputs and outputs of each institution. A case can be made for evaluating institutions either by total output or by output per member (bearing in mind that the latter is likely to be measured with more error). What is less defensible is the practice, adopted by Dusansky and Vernon (1998) and followed here by Combes and Linnemer, of presenting a final ranking based on the arithmetic mean of the total and per member rankings.

## 3. Choosing Weights

Weighting articles for quality is perhaps the most controversial issue in the evaluation process. Either subjective or objective weights could be used. In the case of subjective weights, it is not practical to compute weights for each article, so weights for journals are typically devised. This could be done by means of a survey: for example, Axarloglou and Theoharakis (2003) present a recent evaluation of economics journals based on an e-mail survey of 2,103 economists. Of the present studies, Combes and Linnemer in their preferred specification and Lubrano et al. in all their rankings evaluate journals on the basis of their own subjective opinions. Lubrano et al. use a refinement known as the Delphi method (in which respondents are given feedback to their initial ranking

and invited to revise it) with one coauthor (Alan Kirman) playing the role of the Delphic oracle.

Probably only a large and carefully administered survey is likely to produce a subjective ranking which will command widespread acceptance in the profession. Since one of the main points of the ranking exercise is to devise objective measures of research output, it seems preferable to use objective weights to rank each article. Here there are two possible approaches, based on either the citations it attracts or the journal in which it appears. Citations have the attraction of being article-specific, but the great disadvantage that they are subject to long and variable lags. Coupé attempts to correct for this by dividing each paper's citations by the number of years since it was published. However, this implicitly assumes that the average time-path of citations to an article is linear, which is very far from the truth. Citations can also be criticized for privileging surveys and expository papers, and for the fact that, to some extent, citation practices differ between subfields. In light of these problems, most authors fall back on using objective weights based on the average citations to all articles published in a given journal.

The most easily available objective journal weights, and the only ones that are regularly updated, are the impact factors published in the *Journal Citation Reports of the Social Science Citation Index*. These give a weight to each journal based on the number of citations the average article in it receives from all journals in a specified period. For example, the series denoted "JCR" in Figure 1 (taken from Kalaitzidakis et al.), gives the impact factors based on the number of citations received in 1998 by articles published in the previous ten years. Such unweighted impact factors can be expressed as follows:

$$\omega_i = \sum_j \gamma_{ij}, \quad \gamma_{ij} = C_{ij}/C, \quad C \equiv \sum_i \sum_j C_{ij} \tag{2}$$

where $C_{ij}$ is the number of citations by articles in journal $j$ to articles in journal $i$, so $\gamma_{ij}$ is the number of citations which an average article in journal $i$ receives from journal $j$. (The $C_{ij}$ may be adjusted for the average number of papers in journal $i$ or its average page size.) An obvious problem with this approach is that the citing journal itself should be weighted: a citation is more valuable if it comes from a prestigious journal. This suggests that it would be preferable to weight each citing journal by the citations it receives from other journals. Thus, the weight attached to journal $i$ equals:

$$\omega_i = \sum_j \gamma_{ij}\omega_j \tag{3}$$

which can be written in matrix notation as follows:

$$\omega = \Gamma\omega \tag{4}$$

This is an eigenvector equation, so the weights $\omega$ can be calculated as the largest

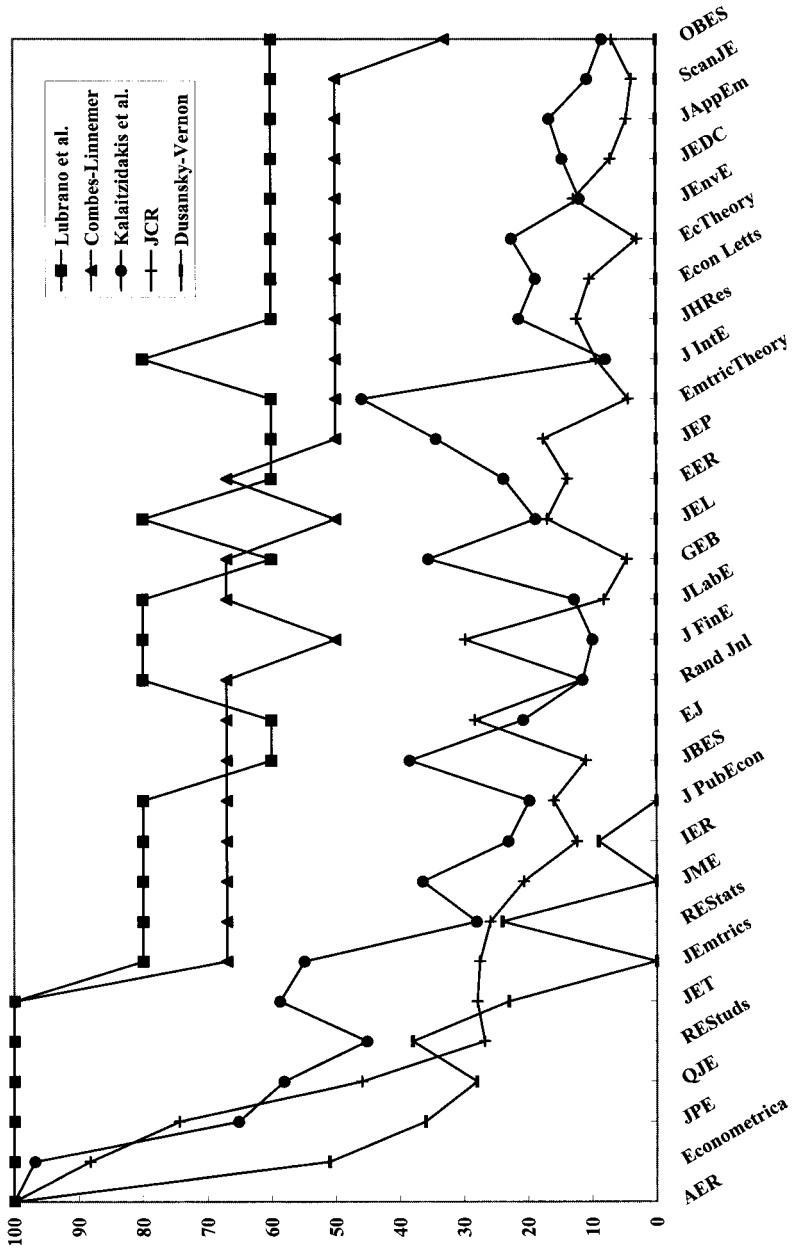


FIGURE 1. Alternative Journal Weighting Schemes

eigenvector of the known matrix $\Gamma$.[3] Palacios-Huerta and Volij (2002) propose extending this approach to correct for the fact that journals differ in their citation intensity, and so a citation from a journal that cites fewer papers on average should get a higher weight. In the present studies, the estimates of equation (4) by Laband and Piette (1994) are the basis for the weights used by Dusansky and Vernon (1998), which are used in turn (along with others) by Combes and Linnemer and by Coupé; while equation (4) is recalculated by Kalaitzidakis et al.

In practice, the choice of weights makes a considerable difference. Figure 1 illustrates, for a group of thirty top journals, the principal weighting schemes used in the four studies, as well as the JCR scheme already mentioned and that used by Dusansky and Vernon (1998). (Of course, highlighting these thirty journals involves an implicit weighting in itself. They correspond to those selected by Kalaitzidakis et al.) All the weighting schemes are normalized by setting the weight for the *American Economic Review* equal to *100.00.*[4]

No great statistical sophistication is required to see from Figure 1 that the different weighting schemes can be roughly classified along a scale from "egalitarian" to "elitist." Both the subjective weighting schemes used by Combes and Linnemer and Lubrano et al. are highly egalitarian, valuing (for example) two articles in the *Scandinavian Journal of Economics* as at least as good as one in the *American Economic Review.* Subjectivity and egalitarianism need not go together however, since the weights used by Dusansky and Vernon (which are a subjectively truncated subset of the objective weights devised by Laband and Piette (1994)) are the most elitist of all.

Perhaps just as important as the weights given to top journals and illustrated in Figure 1 are the weights that each scheme assigns to lesser journals. For the elitist weighting schemes these are typically zero. However, for the Lubrano et al. weighting scheme the minimum weight in their sample of 505 journals is *10.00,* while for the Combes and Linnemer scheme the minimum equals *8.33* for *all* 855 journals in EconLit.[5] Table 17 in Combes and Linnemer and Section 4.4 of Lubrano et al. throw valuable light on this. They show clearly that in most European countries (even including the largest such as France, Germany, and the United Kingdom), the most popular publication outlets are local ones, most of which, for language or other reasons, have relatively small

3. In practice, following Liebowitz and Palmer (1984) and Laband and Piette (1994), the eigenvector is calculated by an iterative process, so its estimate at iteration $t$ is given by $\omega_t = \Gamma\omega_{t-1} = \Gamma^t\omega_0$, for some starting vector $\omega_0$ (usually set equal to $\Gamma\iota$, where $\iota$ is a vector of ones, as in equation (2)). Convergence to a unique solution is guaranteed because the $\Gamma$ matrix has a unique positive eigenvector: see Palacios-Huerta and Volij (2002) or any text in matrix algebra for discussion. Kalaitzidakis et al. also give weights based on a small number of iterations, but it is not clear what interpretation can be given to these.

4. Journals are ordered in Figure 1 by the unweighted average of all five weighting schemes.

5. Lubrano et al. also present results for a list of "top" journals, which is their full list truncated at *69,* with a minimum score of *60.00.* Combes and Linnemer also present results for a number of weighting schemes, including an index they label "*E1n*" which is a simple unweighted total of articles published in all journals.

international circulation. The egalitarian weighting schemes value ten or twelve articles in such local journals as equivalent to at least a single article in the *American Economic Review* (or, even more strikingly, three to four sole-authored articles in such journals as equivalent to an article with two coauthors in the *American Economic Review*). It seems unlikely that this weighting corresponds to those used by most European economists to rank their colleagues in other countries, or to the valuation that the profession worldwide places on contributions in different journals.

## 4. Overview of the Studies

Some further special features of the individual studies are worth mentioning. (The papers are ordered alphabetically by first author.)

Combes and Linnemer present rankings based on both a five-year and a thirty-year window, and on both total publications and publications per head. The latter are supplemented by measures of "career" output, which attempt to correct for the average age of an institutions's researchers by deflating their score by the number of years since their first publication. As already noted, they also give prominence to the subcenters in which researchers work. Here, differences in the degree of coverage of different countries reflect the authors' detailed knowledge of the maze of French institutions (including their own), but less familiarity with those of other countries. (So, for example, they count the University of Wales as a single institution with 302 economists, whereas it has many campuses.)

The study by Coupé makes no methodological innovations, but is the most wide-ranging in that it calculates rankings for all institutions in the world and for eleven different article weighting schemes (both citation- and journal-weighted). This paper's preferred ranking is based on an unweighted average of all eleven rankings. Since the latter include both highly elitist and highly egalitarian schemes, it might be hoped that the law of large numbers will apply and the extremes will cancel, though there is no formal basis for this.

As already noted, the principal contribution of Kalaitzidakis et al. is to calculate a new set of journal weights, using the iterative procedure described in Section 3. These therefore represent the most up-to-date set of objective journal weights available. However, in calculating these weights they omitted all "self-citations," not just between authors but between journals. (Thus, in applying equation (4), they set all the diagonal elements of $\Gamma$ equal to zero.) Further investigation is needed to determine whether this feature explains the relatively higher weight that their scheme gives to econometrics journals, as Figure 1 shows.

Lubrano et al. make a notable effort to investigate the statistical robustness of their rankings, calculating *t*-statistics to test the significance of the differences between their top twenty institutions. They also compare the top institutions from each European country with those of California, and find that the top

California institutions are far better, though the distribution of the state's institutions is more skewed than that of Europe's.

## 5. Conclusion

Despite the very different approaches used in the different studies, some common features emerge from the rankings they present. The same group of institutions tend to appear at the top of all lists, with the leading institutions from the United Kingdom (LSE, UCL, Oxford, Cambridge, and Warwick), France (Toulouse), Israel (Tel Aviv), and the Netherlands (Rotterdam and Tilburg) featuring in most.[6] Lower down, an institution's place in the rankings can vary greatly between the different studies, and seems particularly sensitive to the choice of more or less elitist journal weights. Even here, many institutions tend to appear on more than one list, though, of course, many others appear on only one or do not appear at all. Finally, all the studies show that the gap between economics research in Europe and the United States is narrowing, but still remains very wide. No European economics department can claim to be in the top ten in the world (except using the highly egalitarian weights in Combes and Linnemer's Table 18). The studies thus paint a well-known but distressing picture of relatively inferior performance in Europe, reflecting the poor governance of most European universities and the limited role given to research criteria in their funding. More positively, the studies as a whole give a reasonably coherent picture of where good economics research is being carried out in Europe.

It is tempting to end by summarizing the results of all four papers in a single table. However, it would be impossible to present such a table without implicitly endorsing a particular ranking (perhaps the average or the principal component of the preferred rankings from each paper). This ranking in turn would risk being interpreted as one endorsed by the European Economic Association. It was never the objective of this EEA project to produce a single unambiguous ranking. Rather, it is to be hoped that the papers presented here will encourage debate and stimulate further research, both on the methodology of rankings themselves and on wider issues such as the continuing gap between European and American economics departments.

## References

Axarloglou, Kostas and Vasilis Theoharakis (2003). "Diversity in Economics: An Analysis of Journal Quality Perceptions," *Journal of the European Economic Association*, 1(6), pp. 1402–1423.

Combes, Pierre-Phillipe and Laurent Linnemer (2003). "Where Are the Economists Who

6. Necessary and sufficient conditions for an institution to appear in this list of candidates for the European "top five" are that they appear at least once in the top five and at least twice in the top ten of the preferred rankings from each of the four studies.

Publish? Publication Concentration and Rankings in Europe Based on Cumulative Publications." *Journal of the European Economic Association,* 1(6), pp. 1250–1308.

Coupé, Tom (2003). "Revealed Performances: Worldwide Rankings of Economists and Economics Departments, 1990–2000." *Journal of the European Economic Association,* 1(6), pp. 1309–1345.

Dusansky, Richard and Clayton J. Vernon (1998). "Rankings of U.S. Economics Departments." *Journal of Economic Perspectives,* 12, pp. 157–170.

Kalaitzidakis, Pantelis, Theofanis P. Mamuneas, and Thenasis Stengos (1999). "European Economics: An Analysis Based on Publications in the Core Journals." *European Economic Review: Journal of the European Economic Association,* 43, pp. 1150–1168.

Kalaitzidakis, Pantelis, Theofanis P. Mamuneas, and Thenasis Stengos (2003). "Rankings of Academic Journals and Institutions in Economics." *Journal of the European Economic Association,* 1(6), pp. 1346–1366.

Kirman, Alan and Mogens Dahl (1994). "Economic Research in Europe." *European Economic Review: Journal of the European Economic Association,* 38, pp. 505–522.

Laband, David N. and Michael J. Piette (1994). "The Relative Impacts of Economic Journals: 1970–1990." *Journal of Economic Literature,* 32, pp. 640–666.

Liebowitz, Stanley J. and John P. Palmer (1984). "Assessing the Relative Impacts of Economics Journals." *Journal of Economic Literature,* 22, pp. 77–88.

Lubrano, Michel, Luc Bauwens, Alan Kirman, and Camelia Protopopescu (2003). "Ranking Economics Departments in Europe: A Statistical Approach." *Journal of the European Economic Association,* 1(6), pp. 1367–1401.

Palacios-Huerta, Ignacio and Oscar Volij (2002). "The Measurement of Intellectual Influence," mimeo, Brown University, Providence, Rhode Island.

van Damme, Eric (1996). "Measuring Quality of Academic Journals and Scientific Productivity of Researchers," mimeo, CenTer, Tilburg University, Tilburg, Netherlands.