

УДК 339.9

Tom Coupe

ON THE EVALUATION OF THE PERFORMANCE OF ACADEMIC ECONOMISTS

Universities, governments, grant making institutions and students are more and more interested in evaluating the performance of scientists and scientific departments. This essay focuses on the evaluation of the performance of academic economists to explain some methods that are often used to judge the quality of economists' research and teaching and to show some results of recent ranking exercises.

Introduction

These days, universities, governments, grant making institutions and students are more and more interested in evaluating the performance of scientists and scientific departments. In this essay, I will focus on the evaluation of the performance of academic economists. I will explain some methods that are often used to judge the quality of economists' research and teaching, and show some results of recent ranking exercises. In addition, I will point out the benefits and dangers of these performance evaluations.

The Evaluation of Research

A specific characteristic of research is that it is cumulative: researchers use the results of previous research, either they use those previous results to develop new theories, or they try to test previously developed theories. As a consequence, a person will only be able to evaluate research if that person is aware of what research has been done previously. This explains why 'peer review' — the evaluation of one's scientific performance by other scientists - has become the main quality control mechanism in academe.

In economics, peer review is mainly done through the refereeing process. A paper is submitted to a

journal; the editor, with the help of referees, decides whether the article is of sufficient quality to warrant publication. A rejection by a journal, however, does not mean that the paper was bad. As different journals have different quality standards, it just means that the quality was not high enough to be published in that journal. The author then has the choice to either improve the paper or to submit it to a journal with lower standards.

Each published article thus can be seen as a proof of some kind of quality research. A straightforward way to measure research performance is then to count the number of articles an economist has published in a given time period. Because of the time it takes between submitting a paper and seeing the result in printed form, and because of the importance of luck in the acceptance/rejection decision, one generally counts publications in a given time period, spanning most often five or ten years. The more articles published in that given period, the 'better' an economist performed. This method is for example used by Miracki (1999) to study the publication performance of Ukrainian economists.

Of course, this is quite a rough method to evaluate performance. As mentioned above different journals have different standards so ideally one should weigh for these differences in standards and quality. Two methods have been used in economics to create rankings of journals, one based on questionnaires,

the other based on citation counts. The questionnaire method basically consists of asking a (large) number of economists to grade journals in terms of quality and then to take an average grade. The top 5 of Mason et al (1997), based on the responses of a sample of US economists, consists of the American Economic Review, the Journal of Political Economy, Econometrica, the Quarterly Journal of Econometrics and the Journal of Economic Literature. Brauning and Haucap (2003)'s top 5, based on the responses of a sample of German-speaking economists, consists of the same five journals (though in somewhat different order).

The second method consists of ranking journals in terms of the number of times they get cited. Since the sixties, the Institute of Scientific Information (ISI) develops a database of 'references' - they collect from each article the reference list. The fact that an article is cited can be interpreted as a sign that somebody used this article for subsequent research. Thus, the more citations, the more an article is used by other researchers and hence the higher the quality of this article is supposed to be. On the basis of this database, the ISI annually constructs the 'Journal Citation Reports' (JCR). This is a list with for each journal the number of times it gets cited (referenced) by other journals and an 'impact factor', which basically gives the average number of citations per article in a period of two years. The higher the impact factor of a journal, the more it is used by researchers and hence the more important it is. The 5 economics journals that make the top 5 in terms of reputation are all in the top 10 of JCR¹. In economics, there have been attempts to make 'better' rankings: Kalaitzidakis et al. (2001) only take into account citations by other economics journals and attach a bigger value to citations from better (more cited) journals. These citation rankings, rather than the rankings based on questionnaires, are most often used to make rankings of economics departments and of economists (for example see, Combes and Linnemer (2002), Coupe (2003), Kalaitzidakis et al (2001), Lubrano et al. (2003)).

Rather than weighing articles by the quality of the journals in which they are published, the ISI database makes it also possible to weigh articles by their own quality. Indeed, citations are recorded per article, so a 'citation count' becomes possible: for each economist or each economics department we can count the number of cites generated in a given period. Given that even within a journal, there is a large variation in the number of times an article gets

cited, citation counts are clearly superior to article counts that are weighted by a journal quality indicator. Examples of the use of citations to rank economists are Garfield (1990), Medoff (1996) and Coupe (2003). For departments, examples can be found in Davis and Papanek (1984) and Coupe (2003).

While citations counts are superior to article counts, this does not mean that they are without flaws. First, citations can be 'negative' - a paper can be cited as an illustration of how not to analyze a specific problem. Second, some ideas can have become common knowledge which implies that the papers in which they were developed will not be cited anymore (for example, one might use White heteroscedasticity-corrected standard errors without citing White' 1980 paper). Third, subfield-specific practices can influence citation counts - some subfields have a habit of referring to more articles than others². Fourth, people can be biased towards citing their friends.

In addition, there are a number of problems common to citation and article counts. First, there's the question of whether one should also count books, working papers, papers published in collective volumes, conference proceedings and other forms of output. One reason not to include them is that these forms of output are not subject to peer review, hence there is no quality-check. Rankings based on article counts focus typically on articles published in refereed journals. In contrast, citation counts, which, by definition, take into account quality, often include citations to these types of output. Article counts, however, have the advantage that Econlit, a bibliographical database often used to make article counts covers a lot more journals than does the ISI citation database.

Second, there is the question of coauthorship. Economists coauthor more and more, over half of all paper (in a sample of 8 top journals) are coauthored (Hudson, 1996). The question then is how to control for this coauthorship - is one two-authored paper equal to a solo-article or is it only worth half of that solo-paper. Rankings most often divide the value of an article over the coauthors, though some argue that, for example, in tenure decisions there is no such discounting. Sauer (1988) presents some evidence for the former, Moore et al. (2001) for the latter³.

Third, there is the question of stock versus flows. Should one look at the output since the start of one's career or rather in say, the last 5 year. The first has

¹ http://www.in-cites.com/research/2003/february_3_2003-2.html

² This might explain for example the presence of health economics journals among the most cited journals.

³ Both studies look at this in the context of the determinants of academic salaries.

Table 1. Rankings of Departments

| | CL - Europe | C - World | KMS - World | LBKP- Europe |
|----|--------------|-----------------|----------------|---------------------|
| 1 | LSE | Harvard | Harvard | LSE |
| 2 | Toulouse I | Chicago | Chicago | KUB, Tilburg |
| 3 | Oxford | U Ca, Berkeley | MIT | Oxford |
| 4 | KUB, Tilburg | Stanford | Northwestern | Cambridge |
| 5 | U Nottingham | U Penn | UPenn | Erasmus U Rotterdam |
| 6 | UCL, London | MIT | Yale | UCL, Louvain |
| 7 | Warwick | Yale | Princeton | U Amsterdam |
| 8 | Stockholm SE | U MI, Ann Arbor | Stanford | Warwick |
| 9 | UCL, Louvain | Northwestern | U CA, Berkeley | U Toulouse |
| 10 | Cambridge | Princeton | New York | U Paris I |

CL: based on table 1, column 1 from Combes and Linnemer (2002) - European ranking, total output of the university, quality weighted publications, 1996-2000. C: based on Table 4 from Coupe (2003) - World Ranking, total output of the university, citations, 1990-2000. KMS: based on Table 2 from Kalaitzidakis et al. (2001) - World Ranking, total output of the university, quality weighted publications, 1995-1999. LBKP: based on Table 7 from Lubrano et al. (2003) - European ranking, total output of the university, quality weighted publications, 1991-2000.

the disadvantage that older economists have an advantage over younger economists, the second has the disadvantage that young economists have an advantage over older economists as older economists often have higher teaching loads and more administrative duties. The stock versus flow problem has an extra dimension when looking at departmental rankings - if a scholar moves from one department to another, do the publications and the citations of that scholar follow to that new university or not. In other words, is what important 'who' generated these outputs or 'where' these outputs were generated. Scott and Mitias (1996) make a comparison between these two methods. Not surprisingly, they give somewhat different results.

Probably the most important criticism to both article and citation counts is that they do not take into account the inputs used to produce the articles. This is problematic as there is quite some variation in teaching loads⁴. So people that publish a lot are not necessarily the better researchers, they maybe just spend more time on research, time that those that published less, had to spend on teaching (or other duties)⁵. Similar, different economics departments have different sizes (the number of profs they employ) and most often departmental rankings do not control for size.

One potential way out of this problem is to rely on a more pure form of peer-review and ask (panels of) scientists to judge the overall research quality of an economists or an economics department. In principle, these panels could then 'control' for the inputs. The British Research Assessment Exercise

for example uses such 'panel-peer-review' to judge the quality of scientific departments. These panels get information on the composition of the faculty and their most important papers but also on the number of students in the department and the number and amount of research grants obtained. Based on this information, they are asked to rate the department. Studies (for example, Taylor, 1995), however, show that, keeping output constant, size in terms of budget or staff correlates positively with these ratings. While at first sight strange, this can be explained by the fact that obtaining a research budget is itself a proof of being of high-quality. Also the US National Research Council's ranking exercises (NRC, 1995) provide information on the number of graduates and the faculty composition.

This panel-peer-review is also used to award scientific prizes and awards (like the Nobel prize, the Clark Bates Medal, The Yrjo Jahnsson Award, the Fellowships of the Econometric Society). However, also in those cases, 'inputs' are not really taken into account. In other words, for scientific performance it might seem that what counts is product, not productivity. This should, however, be qualified: what counts is past product - universities when recruiting their faculty will look at future productivity, that is, at output controlled for wage, teaching and administration.

Recently, the European Economic Association has sponsored four evaluation studies of economics departments (Combes and Linnemer (2003), Coupe (2003), Kalaitzidakis et al. (2003), Lubrano et al. (2003)). Below are some of their results.

⁴The Universal Academic Questionnaire organized by the American Economic Association shows that, in 1998-1999, the average teaching load at Ph.D. institutions is 3.5, at M.A. institutions 5.2 and 6 at B.A. institutions (Siegfried and Scott, 1999).

⁵The assumption behind this is that economists are not correctly sorted into jobs, that there are some economists with high teaching loads that would do top research if only they had been in a research institution.

Table 2. Rankings of individuals

| | Citations | Publications |
|----|------------------------|------------------------|
| 1 | Johansen, -Soren | Phillips, -Peter-C.-B. |
| 2 | Barro, -Robert-J. | Tirole, -Jean |
| 3 | Krugman, -Paul-R. | Heckman, -James-J. |
| 4 | Andrews, -Donald | Krueger, -Alan-B. |
| 5 | Phillips, -Peter-C.-B. | Stiglitz, -Joseph-E. |
| 6 | Romer, -Paul-M. | Andrews, -Donald-W.-K. |
| 7 | Fama, -Eugene-F. | Viscusi, -W.-Kip |
| 8 | Juselius, -Katarina | Laffont, -Jean-Jacques |
| 9 | Levine, -Ross | Sen, -Amartya |
| 10 | Shleifer, -Andrei | Smith, -Bruce-D. |

Based on Table 9 and 10 from Coupe (2003) - citations and quality weighted publications, 1990-2000.

From the above table (and the underlying articles), two observations emerge. First, the top universities are located in the United States. Comparisons over time show however that Europe is slowly catching up. Second, while different methods give different results, the same universities tend to show up at the top of ranking.

Coupe (2003) also contains rankings of economists. The top 10 based on citations and publications for the period 1990-2000 are given in table 2.

The Evaluation of Teaching

Rankings of the research performance of economists are regularly published in economics journals. Rankings of teaching or educational performance however tend to be published in more popular journals (US News and World Reports, Der Spiegel, etc.). Der Spiegel ranks university departments on the basis of results of a questionnaire of professors and students, US News uses a combination of questionnaire-results and statistical indicators (measuring graduation rates, financial resources etc). Next to these teaching evaluation by newspapers, there are large scale teaching evaluations by nonprofit institutions (like National Research Council in the US) and by governments. In the UK, the Quality Assurance Agency organizes what is sometimes called Teaching Assessment Exercises. These basically consist of on-site visits by professors who evaluate the quality of teaching. No worldwide rankings of the teaching quality of economics departments have been made so far.

Next, to these evaluations of the teaching quality of departments, almost all US universities replying to a questionnaire by Becker and Watts (1999) indicated that they organize individual-level teaching assessments, by giving their students an opportunity to assess their professors. But these assessments remain local, no such thing exists as a country/region wide teaching ranking of individuals, like it is the case for research.

Why is Evaluation Useful?

For a long time no one thought it to be necessary to evaluate the performance of scientists. Scientists, being driven by intrinsic motivations, were trusted to teach high quality lectures and do frontier research⁶. Still, many economists think performance evaluation is useful. A statement on the website of the European Economic Association illustrates this: '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 toward improving higher education in economics in European Universities'.

The most obvious reason for evaluating academic performance is the incentives it provides. If universities and scholars know that their performance will be monitored, and if this monitoring leads to rewards for good performance and penalties for bad performance, then they have a reason to do their best and put effort. These rewards and penalties take different forms. We first will look at rewards and penalties for scholars, then for universities.

There are several papers that show the link between wages and measures of research output. Hamermesh et al. (1982) find that economics professors earn 1 % extra wage for each citation. Hamermesh (1989) estimates are somewhat lower at 0.2 %. Similarly, Broder (1993) finds that a publication in a top journal adds 3.4 % to the salary. Research performance does not only influences wages, it also influences the prestige of the university where one will work. According to estimates by Grimes and Register (1997), each article published increases the rank of the university where one will end-up by about 2.5 places. Ault et al. (1979 and 1982) look at the effect of publications on mobility and find that people with more publication are slightly more likely to move to better institutions. Finally, research performance also

*Adam Smith (1759) wrote "Mathematicians and natural philosophers, from their independency upon the public opinion, have little temptation to form themselves into factions and cabals, either for the support of their own reputation, or for the depression of that of their rivals. They are almost always men of the most amiable simplicity of manners, who live in good harmony with one another, are the friends of one another's reputation, enter into no intrigue in order to secure the public applause, but are pleased when their works are approved of, without being either much vexed or very angry when they are neglected" (The Theory of Moral Sentiments, part III, Chapter 2).

influences the likelihood of getting a scientific award: Hamermesh and Schmidt (2003) provide estimates of the impact of academic quality on the chance of being elected as a fellow of the Econometric Society⁷.

As far as the rewards for teaching are concerned, Tuckman et al. (1978) and Moore et al. (2001) find evidence that high-quality teaching also pays-off. Estimating a salary-equation, they estimate a significant positive effect of receiving a teaching award. Becker and Watts (1999)'s questionnaire also shows that many universities use teaching evaluations when making decisions about tenure and wage-rises.

Performance matters for individuals, it also matters for departments and universities. Students, both at the bachelor level and at the PHD-level, use rankings to decide where to apply. And when students are admitted by several universities, these rankings help them to decide which offer to accept. Ehrenberg and Monks (1999) investigate the effect of the US News and World Report rankings, the most popular American rankings. Their findings illustrate the impact of evaluations. First, if a university succeeds in improving its rank, the ratio of admitted students to applications will fall (hence it can be more selective). Second, the ratio of matriculants on admitted student will increase, meaning that more students prefer that university. Finally, it allows the university to ask students to finance a larger part of the costs of their studies⁸. Hence, doing well in rankings is important for the university's reputation and financial situation. To score high on the US News rankings a university must score high on the indicators on which these rankings are based - academic reputation, faculty resources, financial resources, graduation rates etc. So a way to improve the ranking is by improving the quality of teaching, as this should influence both academic reputation and graduation rates, two indicators that determine the US news ranking.

Just like students, also PHD graduates can use rankings to help them to decide where to apply for a job. And if they get offers of several universities, they can use again these rankings to help them to decide which offer to accept. So the quality of the future faculty might also be affected. While there are no formal studies on this issue, there is some

circumstantial evidence: Job Openings for Economists, the main job-site of the American Economic Association has a link to a paper (Cawley, 2003) with advice on how to find a job as academic economist. This paper contains several links to ranking websites (p. 14).

Finally, governments and funding organizations are interested in evaluations to help them in their funding decisions. Some governments organize large-scale assessments of academic performance and make the distribution of government funds depending on its result. The UK Research Assessment Exercise is just one example. Also the European Commission has started some exploratory studies⁹.

Some caveats about rankings

While performance evaluation has some advantages, it also has some disadvantages. One problem of performance is that those evaluated might 'game' the evaluation indicator (Milgrom and Roberts, 1992). In our case, this means that, rather than improving teaching quality, universities could try to manipulate the perception of academic quality by printing more glossy publicity material (see Ehrenberg, 2003). Similarly, one way a professor can get good teaching evaluations by students, is to give high grades to these students (Nichols and Soper, 1972).

A second problem is due to the multitask nature of academic jobs: an economics professor is supposed to do research, to teach, to do some administration and some other 'services to the community'. In such a context, there is the danger of unbalanced incentives (Holmstrom and Milgrom, 1991): if one gives incentives for one task but not for others, then these other tasks are likely to be neglected. That is, if a department attaches a lot of importance to journal publications, the professors are likely to start paying less attention to teaching, with reduced teaching quality and unsatisfied students as a consequence.

Conclusions

In this essay, I have surveyed some methods to evaluate the performance of scientists. While

⁷ This paragraph borrows heavily from Coupe (2003). More examples can be found in that article.

⁸ One caveat here: it is difficult to say whether the cause of these changes is a change in the US news ranking rather than the change in quality that caused the change in ranking.

⁹ The aim of the European Research Area is to strengthen the coherence of research activities and policies conducted in Europe. It is recognised that excellence exists in practically all areas and disciplines in Europe but that these competencies are not always sufficiently well known across national borders, notably by companies. In order to foster the dissemination of excellence, a methodology on "how to map excellence" was developed by the Commission.' http://www.cordis.lu/indicators/projects_cra.htm#strengththen

sometimes controversial, such evaluations have become widespread in academia, both for teaching and research, and for individuals and institutions.

When done correctly, such evaluations provide incentives and information, thus leading to a better performing academia.

1. *Ault, D., Rutman, G. and Stevenson, T.* (1979), 'Mobility in the Labor Market for Academic Economists', *American Economic Review*- V. 69.- P. 148-153.
2. *Ault, D., Rutman, G. and Stevenson, T.* (1982), 'Some Factors Affecting Mobility in the Labor Market for Academic Economists', *Economic Inquiry*, - V. 20.- P. 104-132.
3. *Becker, W. and Watts, M.* (1999), 'How Departments of Economics Evaluate Teaching', *American Economic Review*.— P. 344-349.
4. *Brauninger, M. and Haucap, J.* (2002), 'What Economists Think of Their Journals and How They Use Them: Reputation and Relevance of Economics Journals', University of the Federal Armed Forces Hamburg Discussion Paper 111.
5. *Broder, I.* (1993), 'Professional Achievements and Gender Differences among Academic Economists', *Economic Inquiry*. - V. 31.-P. 116-127.
6. *Cawley, J.* (2003), 'A Guide (and Advice) for Economists on the U. S. Junior Academic Job Market', Cornell Working Paper.
7. *Combes, P. and Linnemer, L.* (2001), 'Measuring and Ranking Economics Throughout Europe', CREST working paper.
8. *Coupe, T.* (2003), 'Revealed Performances. Worldwide Rankings of Economics Departments and Economists', *Journal of the European Economic Association*, forthcoming.
9. *Ehrenberg, R. and Monks, J.* (1999), 'The Impact of U.S. News and World Report College Rankings on Admissions Outcomes and Pricing Policies at Selective Private Institutions', NBER working paper 7227.
10. *Ehrenberg, R.* (2003), 'Method or Madness? Inside the USNWR College Rankings', CHERI working paper.
11. *Grimes, P. and Register, C.* (1997), 'Career Publications and Academic Job Rank: Evidence from the Class of 1968', *Journal of Economic Education*, winter 1997,— P. 82-91.
12. *Hamermesh, D., Johnson, G. and Weisbrod, B.* (1982), 'Scholarship, Citations and Salaries.: Economic Rewards in Economics', *Southern Economic Journal*.— V. 49.-N 2.- P. 472-481.
13. *Garfield, E.* (1990), 'Who will win the Nobel Prize in Economics? Here's a Forecast Based on Citation Indicators', *Current Contents*.- V. 11.- P. 3-7.
14. *Hamermesh, D.* (1989), 'Why Do Individual-Effects Models Perform So Poorly? The Case of Academic Salaries.', *Southern Economic Journal*.
15. *Hamermesh, D. and Schmidt, P.* (2003), 'The Determinants of Econometric Society. Fellows Elections', *Econometrica*.- V. 71 (1).- P. 399-407.
16. *Holmstrom B. and Milgrom P.* (1991), 'Multitask Principal-Agent Analysis: Incentive Contracts, Asset Ownership and Job Design', *Journal of Law Economics and Organization*.-V. 7,- P. 24-51.
17. *Hudson, J.* (1996), 'Trends in Multi-Authored Papers in Economics', *Journal of Economic Perspectives*.-V. 10.-N 3,- P. 153-158.
18. *Kalaitzidakis P., Mamuneas T. and Stengos T.* (2001), 'Rankings of Academic Journals and Institutions in Economics', University of Cyprus working paper.
19. *Lubrano M., Bauwens L., Kirman A. and Protopopescu C.* (2002) 'Ranking European Economic Departments: a Statistical Approach', GREMAQ working paper.
20. *Masson, P., Steagall, J. and Fabritius, M.* (1997), 'Economics Journal Rankings by Type of School: Perceptions Versus Citations', *Quarterly Journal of Business and Economics*.- V. 36.-N 1.-P. 69-79.
21. *Medoff M.* (1996), 'A Citation-Based Analysis of Economists and Economics Programs', *The American Economist*, - V. 40.— P. 46-59.
22. *Milgrom, P. and Roberts, J.* (1992), 'Economics, Management and Organization', Prentice Hall, New Jersey.
23. *Mirucki, J.* (1999) 'A Visibility Analysis of the Scientific Production of Ukrainian Economists: 1969-94', *Journal of Socio-Economics*,-V. 28.-P. 185-196.
24. *Moore W., Newman R. and Turnbull G.* (2001), 'Reputational Capital and Academic Pay', *Economic Inquiry*, - V. 39.-N4.- P. 663-671.
25. *Nichols, A. and Soper, J.* (1972), 'Economic Man in the Classroom', *Journal of Political Economy*- V. 80.- P. 1069-1073.
26. *Sauer R.* (1988), 'Estimates of the Returns to Quality and Coauthorship in Economic Academia', *Journal of Political Economy*.- V. 96.- N 4.- P. 855-866.
27. *Taylor, J.* (1995), 'A Statistical Analysis of the 1992 Research Assessment Exercise', *Journal of the Royal Statistical Society A*, p. 241-261.
28. *Siegfried, J. and Scott, C.* (1999), 'American Economic Association Universal Academic Questionnaire', *American Economic Review*.- P. 499—502.
29. *Scott, L. and Mitias, P.* (1996), 'Trends in rankings of Economics Departments in the U. S.: an Update', *Economic Inquiry*, - V. 34,- P.- 378-400.
30. *Tuckman H., Gapinski J. and Hagemann R.* (1977), 'Faculty Skills and the Salary Structure in Academe: A Market Perspective Author', *American Economic Review*.- V. 67.— P. 692-702.

Том Купе

ОЦІНКА ЕФЕКТИВНОСТІ РОБОТИ ВЧЕНИХ-ЕКОНОМІСТІВ

У статті розглядаються питання оцінювання результатів роботи вчених-економістів академічної сфери, розкриваються сучасні методи, які використовуються для атестації якості досліджень і викладання, а також формування відомих рейтингів університетів, кафедр, окремих учених.