From first submission to citation: 
An empirical analysis*

Philip Hans Franses**

Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 
NL-3000 DR, Rotterdam, The Netherlands

The academic publication process consists of two stages. The first stage covers for example the conception of a paper, its submission to a journal, possible revisions due to comments made by (anonymous) reviewers, and acceptance of the manuscript. The second stage concerns the eventual publication of the paper and the second part of its academic life-cycle, which is usually measured by a citation score.

Next to describing this process in some detail, this paper describes the results of an empirical analysis of my personal database with detailed records of 66 published papers. Descriptive statistics give insights as to how long it takes (on average) before the editor returns to the author with the reviews, and also how long it takes for the editor to make a final decision on acceptance, based on a revised manuscript. Econometric models are used to see if, for example, the number of pages, the number of authors, and the number of previous rejections have an impact on these durations. Also, it is examined if a special issue makes a difference. It is found that it does, in the sense that special issue papers get cited more often. Finally, it is studied whether the editorial process and observable properties of the paper have any effect on the number of citations, which can be seen as a measure of quality.

Key Words and Phrases: publications, citations score, editorial process.

1 Introduction

A key activity of academic researchers concerns the publication of their research findings in scientific journals. There are many researchers affiliated with universities...
and research institutes, and their scattered research interests are reflected by a sheer infinite number of scientific journals. Each scientific discipline seems to have established a set of journals that are considered its best, and a larger set of journals that are second best or otherwise. It is also widely accepted that academic researchers should once in a while publish their findings in these high quality journals. This is seen as a measure of scientific achievement, and it is usually met with respect. This holds in particular for young academics, like graduate students, postdoctoral researchers and assistant professors, whose subsequent academic careers often depend on their publications in high quality journals. Whether this is good or bad is beyond the scope of this paper, although some of the descriptive statistics below may shed some light on the time frame within which one can evaluate junior faculty. For example, the period between first submission and eventual publication (at least for the cases at hand) appears to cover 2 years on average, which matches closely with the findings in Ellison (2002).

As the publication process is key to the careers of academic researchers, it is no surprise that there are many theoretical and empirical studies on this matter and that there are also topical journals. A quick scan of the relevant literature shows that there are mainly two types of studies. Those of the first type consider the process starting with the conception of the paper until final acceptance. For example, it might be interesting to see how classic papers (as these turn out to be so, after many years or even decades) proceeded through the editorial process. The study of Gans and Shepherd (1994) is a nice and fascinating example of this. Another frequently encountered research issue concerns the editorial process and the decisions involved, see Trivedi (1993), Hamermesh (1994), Street et al. (1998), and Ellison (2002). Many journals publish editorial statements and some journals also evaluate their editorial behavior, often to see if there has been some bias or some form of particularism, see for example Beyer, Chanove and Fox (1995). Note that the editorial board of Statistica Neerlandica aims to keep track of these matters too, and we are compiling statistics on the editorial process. Finally, there are also papers which explicitly address the issue of how one can better write a paper, see for example Thomson (1999).

The second type of studies is concerned with what happens with the academic results once they have been published. One way to measure the quality of scientific output amounts to measuring the number of times other academics cite a paper and perhaps build their research activity on its content, see for example van Dale and Henkens (2001). Very successful papers in economics and econometrics can get something like 150 citations per year for a period of a decade, see Franses (2002) for some illustrations. However, as Klamer and van Dale (2002) report, for non-exceptional research an average of five citations per author per year is a more common figure. These citations are not only of importance for authors, but also for the journals. Indeed, when a journal publishes papers which get cited more often, the journal gets a higher rating, and in turn may attract better submissions in the future. There are now various rankings of journals, and in several countries there are lists
which rate (economic) departments and researchers on their scientific output, based on these rankings.

Interestingly, except for perhaps Gans and Shepherd (1994), there are no studies that follow the full life cycle of individual papers over time. If journals report on their editorial policy, and investigate the trajectories of papers in their journals, they do not analyze the trajectories of the papers they rejected. This is indeed rather difficult, as papers might get different names, other authors or even topics. Additionally, if one studies citations, then usually there is no information available on how the paper eventually got published. Hence, it may be that it was rejected first by many other journals.

It is the aim of my, predominantly empirical, paper to study the full life cycle of papers, ranging from first submission to citations. The analysis will be based on a database, which contains information on various relevant variables, like dates of submission, revision and acceptance (or previous rejection) and annual citations. It concerns 66 (of my own) published papers for the period of 1990–1999. Of course, this database concerns a single individual, and in that sense it is not representative. However, as I will indicate below, in various other dimensions it seems reasonably representative, as it concerns different journals, topics and journal ratings, as well as that it covers 9 years. I have had many discussions with colleagues and asked whether they had kept all their records. It turned out that nobody did. However, if there is anyone who reads this, and does have such files, please contact me. The database, by the way, concerns publications which (on average) get cited close to the average, found by Klamer and van Dalen (2002). The papers were published in a wide range of journals, although these all related to economics, statistics and econometrics. There are a couple of insights that might be gained from the empirical analysis. For example, does a better review process, measured for example by the number and size of referee reports, lead to more citations? How long do editorial processes take, on average? Do specific features of the paper make a difference for the editorial process? And, do special issues make a difference?

The outline of this paper is as follows. In Section 2, I briefly discuss various stages in the publication process. In Section 3, I first discuss various features of the database, and next I provide various descriptive and model-based results. Finally, in Section 4, I conclude with some remarks.

2 The publication process

In this section the focus is on a few features of the publication process. The discussion is narrowed down such that it matches with the empirical discussion in the next section. For example, there is no discussion on how new ideas get shaped and how they arise, even though such a discussion could be illuminating, simply as these features are difficult to measure quantitatively. Also, as the database concerns an academic researcher in economics/econometrics, the discussion is based on personal
experiences concerning journals in these areas. In other disciplines, matters may run much differently, but I am simply not aware of that.

From first submission to acceptance

The first stage concerns the submission of a manuscript to a journal, the receipt of a letter from the editor with attached referee reports and an associated decision to reject or an allowance to revise, and final acceptance.

Naturally, before one can submit a paper for possible publication, there first should be a paper. In some areas one can shape the paper beforehand towards the guidelines of the target journal. This means that one follows the style files, but it can also mean that one explicitly aims to contribute to a range of papers that have already appeared in the same journal on the very same topic. This strategy implies that the author is up-to-date with the published literature, and also with the unpublished working papers from other academics who happen to have the same strategy. This strategy can be noticed in disciplines such as marketing and finance, where the number of top quality journals is limited.

In other scientific disciplines, one first thinks about an issue or a problem, collects data, designs a new theory, evaluates related theories, or perhaps derives a new methodological tool, and after that, one considers which journal could perhaps be the best outlet. This strategy is more often followed in econometrics and statistics, where there are plenty good quality journals. Indeed, if a manuscript gets rejected by one journal, one can always send it to another journal.

Before a manuscript is sent out to a journal, the author usually presents his or her work at departmental seminars or international conferences. In most disciplines there are many of these conferences, and there is a tendency to present a paper at at least one or two of these each year. The smaller scale seminars can be more fruitful in terms of feedback, as there is an opportunity to have a closer contact with the audience and hence there is perhaps more room for discussion. During the lectures, the authors themselves also get to know their own work better, and this can be fruitful in the process of writing the final draft of the paper. First rough drafts usually appear in working paper series, and these can be markedly different from the final version. Hence, one may want to be a bit cautious with the working paper versions. Usually, these working papers serve as a kind of patent, implying that the author’s rights are claimed.

The notion of “patent” is quite important, in particular for younger researchers, as it sometimes happens that colleagues “suddenly appeared to have the same idea as you did”. The ultimate case in which the notion of patent did not work out, is when one finds out that one’s paper has been copied to a large extent and published elsewhere. As an example, the paper in Wei (2002) shows a remarkable overlap with Franses and van Dijk (1996).

Key features of a manuscript that can be measured are its length (in number of pages), the number of references and the number of co-authors. The quality is difficult to measure, and an indicator of a topic is too. It can happen that a
manuscript is submitted to a special issue of a journal, and hence this usually implies that the topic of the paper matches with the interest of the (special issue) editors.

By the time the author is confident enough to submit the manuscript and to undergo the scrutiny of fellow researchers in the area, it can be sent off to the editorial address. If things run by regular mail, it is my personal experience that it may take a month on average before the editor has sent it to the relevant associate editor, and we strive for the same time frame at *Statistica Neerlandica*. Usually, the first response of the editor is that he or she has done so.

Once the manuscript has been sent off, the most uncertain period commences, that is, the author has to await the response from the editor and the associate editor, where their judgement is usually based on what the referees have to say. In the meantime one is better off working on one or a few new papers, as simply waiting for the postman would lead to a very low average productivity. Indeed, this uncertain period may take a long time, as we will see below.

The key variables here are the number of referees and the average size of their reports. One may believe that more comments, at least if these are constructive and helpful, can make a paper better. So, if there would be a measure of eventual quality of a paper, these two variables should have a positive impact. Additionally, the time between submission and first response can be of interest. More time may suggest that the manuscript is perhaps difficult to understand or that referees do not like it at all and postpone their refereeing duty. Ellison (2002) investigates various possible reasons for differing delays.

The response of the editor can involve a few decisions. The first is the recommendation that the manuscript should be sent to another journal, that is, a rejection. The second is that the editor invites the submission of a revised manuscript, where this revision should be based on what the referees say. This can involve much additional work, and sometimes one may feel that the referees ask for this work just to slow you down. In the case where one feels able to respond adequately, the revision gets sent back to the editor, with a set of accompanying letters to the referees, which outline how their comments were handled.

Then, a second uncertain period commences. Below we will see that this period can be about equally as long as the first period. The second letter of the editor can involve a decision of acceptance, of rejection (meaning one has to start all over again by sending it to another journal) and a request for a further round of revision. There are journals that have a policy that one round of revision should do it, while there are also journals that ask you to revise perhaps 4(!) or even more times. Most certainly, it rarely happens that a paper gets accepted outright.

The key variables here are the number of prior rejections by other journals, before acceptance at the current journal, and the number of revisions. One could think that more revisions would make a paper better. But, if the paper got rejected frequently, it is perhaps not a very good one, and it might not turn out to be effective in the end. Obviously, the discussion in Gans and Shepherd (1994) indicates differently, but it
should be borne in mind that they consider only classic studies. One might argue that papers became classics because of their innovation content, and this may scare off referees. The database to be considered below does not contain such classic papers (I am quite confident to state this!), so perhaps the results are a little more representative. Finally, one may be interested in the time between revision and decision. The shorter this period, perhaps, the more likely it was that the revised manuscript met the referees’ suggestions.

Before I proceed to the second stage of the publication process, there are a few remarks to be made about special issues. Such a journal issue usually concerns a specific theme, and the special issue editors invite certain specific people to submit, while also allowing interested individuals to submit. Hence, some form of reputation effect is at stake here. It may be that only those researchers from whom one might expect a paper to be cited well get invitations to submit. Also, it may be that the editorial process is different, in the sense that there are smaller sized referee reports, there is a faster response, a smaller number of referees and so on. In Section 3, I will examine whether this is the case, at least for the personal database at hand.

From publication to citation

Once a paper has been accepted, it will be put in the queue for eventual publishing. And, once it has been published, everybody can access it, read it, use it, cite it, apply it, and so on.

As rankings of academic researchers are based on their published papers, the timing of the actual publication is not unimportant. The queue for publication can be rather long. For the record, for *Statistica Neerlandica* it is usually around 6 to 9 months. This may be due to an a priori allocated special issue or simply to a large amount of papers in the pipeline. The latter may be due to a high acceptance rate of the editorial board, and to a large number of submitted manuscripts. Hence, the time between the decision of acceptance and the final publication is not very informative for the scientific quality of the editor, the managing editor, or the journal. On the other hand, one should of course take this time into account when evaluating resumes of young faculty.

A word of caution is needed here for those authors who were familiar with the working paper version of the eventually published paper. In many cases, these two versions differ substantially, which is simply due to the fact that the author has tried to meet the comments from the referees. Hence, once a paper is published, one better discard the working paper version, although there are always some exceptions. Indeed, for historical and more qualitative analysis, it would be interesting to see to what extent the working paper and the final version differ.

Once a paper has been published, it is uncommon to present it again at seminars and conferences, and hence, the paper has to do the attracting itself, see Klamer and van Dalen (2002). A key measure of attraction is the number of citations. When a paper gets cited often, it can be considered as more relevant. Of course, there are
various reasons why in exceptional situations this suggested relevance may not hold (the author’s students have to cite his or her own work, or, the paper is so wrong that everybody wants to stress that), but in general the citation score is believed to be a rather reliable measure of quality.

An interesting aspect is the time it takes before papers get cited and before papers do not get cited anymore. Hence, it is of interest to see what the diffusion of publications look like. In Franses (2002), I illustrate that this process can be highly nonlinear, and it is found for one volume of one journal (Econometrica, 1987) that peak citations occur after about 6 years. Without going into details here, it seems relevant to measure citations over time (not in just a single year), and that some average measure might be used. In this paper, I will use the number of citations (including self-citations) and an annual average. The latter variable may be improved, as it assumes linearity, but at the time of writing I do not know how.

3 An empirical analysis

This section contains two parts. In the first, I give an outline of the contents of my database. In the second part, I give various descriptive statistics and two econometric modeling results.

Data collection

The data concern 66 published papers in 28 different journals, for the period 1991–1999. This period effectively starts in the year of my graduation. In the Netherlands, these journals get rated as A (top), B, C and D. There are 8 publications in A journals, 19 in B journals, 25 in C journals and 14 in D journals, and this suggests a reasonably balanced and acceptable spread over the journal qualities. By the way, Statistica Neerlandica is qualified as a B journal. The journals, their rating and the number of papers published in these journals appear in Table 1.

Table 2 gives the empirical distribution of the number of publications per year. Again, this distribution does not show obvious peaks or asymmetries, although there is a moderate trend upwards.

For the empirical analysis, the following variables were constructed. First, there are the number of citations and the number of citations when standardized. This standardization entails that the data are averaged over the years since the actual publication. Upon doing this, care has been taken of the timing within a year that the journal issue with the paper appeared. Hence, the data are scaled back from months to years.

Next, for all 66 papers, there is information on the number of co-authors, the number of published pages, the number of cited references and whether the paper appeared in a special issue or not. The number of published pages is of course not
necessarily equal to the number of initial manuscript pages, but these were difficult to collect and standardize.

As measures of the editorial process, I use the amount of prior rejections by other journals (before final acceptance by the last one), the number of referees, the average size of the referee reports and the number of revisions.

Finally, there is information on the number of months between first submission and first response, between submission of the revision and the decision of acceptance (hence the total time that the manuscript has been with the journal), and the time between the decision of acceptance and final publication.

In the next subsections, I will first provide some descriptive statistics and then discuss some modeling outcomes.

© VVS, 2002
Descriptive statistics

Table 3 provides key statistics, such as mean, median, maximum, minimum, and standard deviation of the papers themselves.

It can be seen that on average each paper has about 4 citations. Several papers never got cited (18 of the 66) and one paper has 27 citations. The empirical distributions of these variables appear in Figure 1 and Figure 2. Note that the citations also include self-citations.

The maximum number of co-authors is two, and there are 21 single-authored papers. The largest paper contains 47 pages (and also happens to have 97 references as it is a survey paper). Deleting this observation in the modeling below does not lead to very different results. The average number of references is about 18.5, and this is close to the average of all articles in the Social Science Citation Index, see Klamer and van Dalen (2002).

Table 3. Descriptive statistics concerning the published papers.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citations</td>
<td>3.955</td>
<td>2</td>
<td>27</td>
<td>0</td>
<td>6.072</td>
</tr>
<tr>
<td>Citations, standardized</td>
<td>0.630</td>
<td>0.405</td>
<td>3.290</td>
<td>0</td>
<td>1.904</td>
</tr>
<tr>
<td>Co-authors</td>
<td>0.818</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0.654</td>
</tr>
<tr>
<td>Pages</td>
<td>13.045</td>
<td>10</td>
<td>47</td>
<td>2</td>
<td>8.204</td>
</tr>
<tr>
<td>References</td>
<td>18.492</td>
<td>15</td>
<td>97</td>
<td>3</td>
<td>14.743</td>
</tr>
<tr>
<td>Special issue</td>
<td>0.121</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.329</td>
</tr>
<tr>
<td>Prior rejections</td>
<td>0.500</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0.965</td>
</tr>
<tr>
<td>Referees</td>
<td>1.379</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0.973</td>
</tr>
<tr>
<td>Average report size</td>
<td>1.284</td>
<td>1</td>
<td>3.5</td>
<td>0</td>
<td>0.933</td>
</tr>
<tr>
<td>Revisions</td>
<td>1.061</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0.802</td>
</tr>
</tbody>
</table>

Fig. 1. Frequency (vertical axis) of amount of citations, cumulative up to October 2001 (horizontal axis).
Next, I turn to features of the editorial process. Some papers were rejected three times before they were eventually published. When the number of referees is zero, this can mean that the paper got accepted outright or that the editor him- or herself wrote the reports. This is the case for 14 of the 66 papers.

The average report size is one page, with exceptions of size 3.5 pages. These extensive reports, by the way, also correspond with the journal which endorsed 4 revisions before final acceptance. Clearly, the number of revisions is on average equal to 1.

Finally, I turn to the statistics on the time between various important events. Table 4 provides some key statistics, and Figure 3 and Figure 4 give graphs of the time between first submission and response, and between the revision and final acceptance. The time the paper has been with the author for revisions is of course very case-specific, although for this database I found it to be equal to about 6 months.

In various respects, the numbers and graphs above give interesting insights. Note that many journals publish these time frames too, like for example the time the paper has been with the editor and the referees. Here, the numbers are computed across
journals. In some cases, the editorial process can be shorter than a month, but it may also happen that a paper is with a journal for 27 months. On average, this number is 11 months, with a median of 10 months. Additionally, it can happen that a paper is in the publication queue for 24(!) months.

In sum, taking an average time that the paper is with a journal, that is, 11 months, take another 6 months for the author(s) to do the revisions, and then take another 9 months for eventual publication, there can be an average time difference
of about 2 years between first submission and final publication. Note that this holds true for the papers submitted to and published by the same journal. Indeed, some papers took five years to get published, as they first got rejected a few times. Hence, these outcomes are not discounted for the time it took to handle earlier rejections.

**Econometric models**

The next question concerns the impact of aspects of the paper, and of features of the editorial process, on the quality of the paper, as measured by citations.

It is commonly believed that special issues of journals are different, not only in topical interest, but particularly in terms of treating certain researchers differently (that is, favorably) and perhaps have differing refereeing procedures. Also, it is said that special issues might have more impact, in terms of citations, than regular issues have. Note that a casual glance at the list of most frequently downloaded papers published in *Statistica Neerlandica* also features many special issue papers in the top part of the list.

In Table 5, I report on a regression of a variable of interest on an intercept and a 1/0 dummy in the case where the paper appeared in a special issue of a journal. This dummy variable takes a value of 1 in 9 of the 66 cases. The results in this table clearly suggest that nearly all aspects of the paper and all features of the editorial process are not significantly different for special issue papers. The only variables for which the special issue makes a significant contribution concern the citations and standardized citations. Hence, indeed, special issue papers seem to get cited more often.

The second type of model analysis concerns the possible explanatory value of the average report size, the number of co-authors, the number of pages, the number of referees, the number of references, the number of previous rejections, the type of journal (A, B, C or D), and whether it is a special issue or not, on the various time

<table>
<thead>
<tr>
<th>Variable</th>
<th>Intercept</th>
<th>Dummy for special issue (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citations</td>
<td>3.379</td>
<td>4.746 (0.037)</td>
</tr>
<tr>
<td>Citations, standardized</td>
<td>0.544</td>
<td>0.709 (0.022)</td>
</tr>
<tr>
<td>Co-authors</td>
<td>0.776</td>
<td>0.349 (0.143)</td>
</tr>
<tr>
<td>Pages</td>
<td>13.086</td>
<td>-0.336 (0.915)</td>
</tr>
<tr>
<td>References</td>
<td>18.741</td>
<td>-2.491 (0.655)</td>
</tr>
<tr>
<td>Prior rejections</td>
<td>0.569</td>
<td>-0.444 (0.225)</td>
</tr>
<tr>
<td>Referees</td>
<td>1.362</td>
<td>0.138 (0.710)</td>
</tr>
<tr>
<td>Average report size</td>
<td>1.220</td>
<td>0.530 (0.133)</td>
</tr>
<tr>
<td>Revisions</td>
<td>1.052</td>
<td>0.073 (0.811)</td>
</tr>
<tr>
<td>Submission–first response</td>
<td>5.379</td>
<td>1.371 (0.254)</td>
</tr>
<tr>
<td>Revision–decision</td>
<td>5.483</td>
<td>-2.483 (0.202)</td>
</tr>
<tr>
<td>With journal</td>
<td>10.862</td>
<td>-1.112 (0.661)</td>
</tr>
<tr>
<td>Decision–publication</td>
<td>8.328</td>
<td>2.297 (0.179)</td>
</tr>
</tbody>
</table>
variables. Each of these models also includes a variable which measures the year, that is, 1 for 1991, 2 for 1992 and so on. This is done to see if things might have changed over the years, which is convincingly suggested by Ellison (2002) and others.

The models for this analysis are the proportional hazard model and the accelerated lifetime model, for which the Eviews code is given in Chapter 8 of Franses and Paap (2001), and models for the natural logarithmic transformed time variables. However, for all models it turned out that none of the above variables is significant, except for the year, which obtains a positive sign in all models. Hence, over time, the editorial process appears to take longer and longer, which is in agreement with the findings in Ellison (2002).

As there are 18 papers that never got cited, the analysis of citations might need a tobit-type regression model. With the number of standardized citations as the explanatory variable, and all other variables as potential explanatory variables, the final tobit-1 model (where insignificant variables at the 10% have been deleted) implies the following effects. A significant and positive effect (t-statistic values in parentheses) is found for the number of pages (4.447), the number of referees (1.864), and the 1/0 dummy variable for a special issue (2.723). A significant and negative effect gets the year (−4.409). This last finding indicates that my work gets less and less cited over the years. Of course, one reason may be the quality of my scientific work, another reason may be the increasing abundance of authors and journals.

4 Concluding remarks

This paper has provided a coherent empirical, but personal, analysis of various features of the publication process. Additional to various interesting descriptive statistics, it was found that special issues are not treated differently by authors and editors, but that papers in these issues get cited more often. Also, it was found that the editorial process takes longer and longer, and that longer papers which had more referees get more citations.

This study has of course various limitations. First of all, it concerns the publications of a single individual. It would be of interest to combine various databases, although the reluctance of academics to cooperate, as noted in Trivedi (1993), may make this a non-trivial exercise. Additionally, if one includes data for different individuals, one somehow has to account for observed and unobserved heterogeneity. A second limitation is that papers that never got published were not included, simply as these files have been thrown away. Hence, databases which also include detailed records of rejected papers would be informative.

An important conclusion to be drawn from this empirical analysis, which in fact enhances the findings in Ellison (2002), is that the editorial process now takes longer and that only a few papers get frequently cited.
References


Received: May 2002. Revised: July 2002.