
Pierre Dubois,* Adriana Hernandez-Perez,* Marc Ivaldi*

November 2006

Abstract

This article analyzes the demand and cost structure of the French market of academic journals. After merging several databases, we estimate an aggregated nested logit demand system combined simultaneously with a pricing equation, taking into account the evolution of impact factors of journals. We identify the structural parameters of this market and find that price elasticities of demand are relatively high, indicating that this industry experiences competitive constraints. We conjecture that the two-sidedness feature of academic journals could be a rationale for our results.

* University of Toulouse (INRA, IDEI) and CEPR.
* University of Toulouse (GREMAQ).
* Toulouse School of Economics and CEPR.
1. Introduction

This article analyzes the demand and cost structure of the French market of academic journals. Taking academic journals as differentiated products, we rely on the recent developments of the empirical Industrial Organization literature to estimate an aggregated nested logit demand combined simultaneously with a pricing function under the assumption of Bertrand competition. In other words, we assume publishers adopt the readers-pay business model and that the subscription pricing policies are determined oligopolistically. However the public reach of a journal is also affected by its reputation which could change through network effects over time. Hence we introduce an additional equation for recognizing that the evolution of the impact factor - our measure of journal’s number of citations – that characterizes the reputation of a journal and affects its price level and its number of subscriptions, could be endogeneously determined.

Based on the merging of two important price databases, EBSCO and SWETS, together with the Journal of Citations Report edited by ISI, we collected data covering the yearly subscription of journals by French universities from several domains of sciences and social sciences and their characteristics for the period 1994 to 2004. We fit our model to this data and we are able to test different hypotheses, in particular the endogeneity of the impact factor variable. Our main result is that French universities’
subscriptions are substantially elastic to the price of journals, with publishers seizing a relatively low mark-up. Our model allows us to conjecture that our results could be supported by the two-sidedness feature of academic journals.

2. The Data

Our database combines several sources. The exhaustive annual levels of subscription per journal are obtained from the information network of all French university libraries, ABES (Agence Bibliographique de l’Enseignement Supérieur), from 1994 to 2004. Other journals’ characteristics are obtained from the merger of two annual publications of the Journal of Citation Reports (JCR herein), (1) the Sciences Edition from 1994 to 2003 and (2) the Social Sciences Editions from 1994 to 1997, 1999 and 2003. They include the total number of citations, impact factor, number of issues and articles, publisher and its nationality and fields covered. The journals are selected according to their fields’ importance in the database. The covered fields are Business, Chemistry, Computer Science, Economics, Engineering, Mathematics, Medicine, Probability and Statistics, Physics and Psychology.

The price variable is the combination of listed institutional subscription prices given by the two main firms distributing journals in France, EBSCO and Swets Information Services. Notably, the available price schedules are
based on per journal subscription, which abstracts from any of the quantity discounts publishers usually offer libraries or the types of media (printed and/or electronic) included.\textsuperscript{1}

An important drawback of this work is the lack of data on the costs of publishers. As an alternative, we construct some proxy cost variables based on the journals and publishers’ information. These variables include field and editor dummies, number of subscribed journals proposed by the publisher, the nationality of the publisher, number of issues per year, and some interactions between these variables. We also include a dummy for not-for-profit publishers.\textsuperscript{2}

The following tables provide summary descriptive statistics of variables that are used in the specifications we discuss below. These variables include prices (in real dollars 2000), market shares, price per total number of citations and per impact factor, and number of journals.

Table 1 gives the mean and median market shares and prices of the journals. Although our definition of market share and our price adjustment algorithm have inflated the statistics for the period 1998-2000, some trends are obvious. The median market share of a journal has been decreasing steadily in the last ten years and prices have not changed significantly. The

\textsuperscript{1} We control for the type of media using a step dummy for the year main publishers started offering the journal’s electronic version. See Case (2004)

\textsuperscript{2} Not For Profit journals are all journals that belong to a University Press or Society type of publisher.
median ratio between price and journal’s citation and between price and unit impact factor have remained constant (if not decreasing) during the sampled period.

The journals listed in JCR are characterized by up to 5 subfields of science. From a total of 219 subfields, we have selected the 10 most frequent domains of science by grouping the subfields into its respective major field.

Table 2 provides per field medians over some important journal’s characteristics. We find that some fields have a different citation dynamics than others and that median prices differ considerably across fields. Given these characteristics, field specificity seems relevant to properly capture the network effect of academic journals. We address it in our estimates.

We finally turn to the publishers characteristics. Because the Academic Publishing industry has considerably seized the mergers and acquisitions wave of the nineties, we have chosen the most representative of the sample and controlled for their merger activity. From a total of 262 publishers, we selected the ten largest ones.
Table 1: Descriptive statistics

<table>
<thead>
<tr>
<th>Year</th>
<th>Market Shares Mean (a)</th>
<th>Price Mean</th>
<th>Price per citation Mean (c)</th>
<th>Median</th>
<th>Price per impact factor Median</th>
<th>Number of journals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td>0.0007188</td>
<td>1076.3</td>
<td>2.3573</td>
<td>0.3474</td>
<td>2.3573</td>
<td>772</td>
</tr>
<tr>
<td>1996</td>
<td>0.0006314</td>
<td>1119.6</td>
<td>1.8538</td>
<td>0.3370</td>
<td>1.8538</td>
<td>872</td>
</tr>
<tr>
<td>1998</td>
<td>0.0016411</td>
<td>1648.2</td>
<td>3.0818</td>
<td>0.3929</td>
<td>3.0818</td>
<td>303</td>
</tr>
<tr>
<td>2000</td>
<td>0.0024651</td>
<td>1517.9</td>
<td>2.4624</td>
<td>0.4053</td>
<td>2.4624</td>
<td>253</td>
</tr>
<tr>
<td>2002</td>
<td>0.0005142</td>
<td>964.08</td>
<td>1.0248</td>
<td>0.2744</td>
<td>1.0248</td>
<td>1037</td>
</tr>
<tr>
<td>2004</td>
<td>0.0004421</td>
<td>1105.4</td>
<td>1.1782</td>
<td>0.4575</td>
<td>1.1782</td>
<td>1254</td>
</tr>
</tbody>
</table>

Notes:  
(a) The market share of a journal j at time t is defined as the number of universities that have at least one of its libraries subscribing to j at time t divided by the total number of journals available at time t. (b) The impact factor is a measure of the importance of citations of a journal which is the ratio of total cites in a current year of articles published in a given journal the previous two years over the total number of articles published the previous two years. (c) Computed averages over all journals for 1998 and 2004.

Table 2: Subscribed journals’s characteristics per field (medians)

<table>
<thead>
<tr>
<th>Fields</th>
<th>Impact Factor</th>
<th>Number of Articles</th>
<th>Price in 2000 dollars</th>
<th>Market Share %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics</td>
<td>0.55294</td>
<td>60</td>
<td>552.61</td>
<td>9</td>
</tr>
<tr>
<td>Economics</td>
<td>0.58974</td>
<td>39</td>
<td>321.33</td>
<td>6</td>
</tr>
<tr>
<td>Engineering</td>
<td>0.61716</td>
<td>103</td>
<td>569.07</td>
<td>12</td>
</tr>
<tr>
<td>Computer Science</td>
<td>0.62000</td>
<td>54</td>
<td>648.07</td>
<td>4</td>
</tr>
<tr>
<td>Probability and Statistics</td>
<td>0.65809</td>
<td>49.5</td>
<td>210.49</td>
<td>3</td>
</tr>
<tr>
<td>Business</td>
<td>0.68493</td>
<td>38</td>
<td>306.49</td>
<td>3</td>
</tr>
<tr>
<td>Psychology</td>
<td>1.25022</td>
<td>41</td>
<td>280.79</td>
<td>4</td>
</tr>
<tr>
<td>Physics</td>
<td>1.37632</td>
<td>194</td>
<td>1379.14</td>
<td>5</td>
</tr>
<tr>
<td>Chemistry</td>
<td>1.68600</td>
<td>240</td>
<td>1394.37</td>
<td>7</td>
</tr>
<tr>
<td>Medicine</td>
<td>1.84971</td>
<td>150.5</td>
<td>443.54</td>
<td>47</td>
</tr>
</tbody>
</table>

Note: Computed medians over all journals.

3. Econometric Specification and Estimation

This model borrows from the recent developments on the empirical Industrial Organization of differentiated products. The representative consumer is a university library, which decides for buying one of the

---

3 See Berry (1994) and Nevo (1997) among many others.
available academic journals, based on the researchers it represents, which varies according to field and to the quality of their research output. The library might also buy an outside alternative or not buy any journal at all.

The nested demand framework assumes products are classified in $G$ different groups plus the group corresponding to the outside alternative. In the context of academic journals, journals are classified according to different fields of science. In this framework, journals of the same field are closer substitutes than journals outside the field. The utility of subscribing a journal $j$ by consumer $i$ is the sum of the average utility of journal $j$, $\delta_j$, which is common to all consumers, and a composite random component $\varepsilon_i + (1 - \sigma)\varepsilon_p$. The average utility is defined as

$$\delta_j = X_j \beta + \rho I_j - \alpha p_j + \xi_j,$$  \hspace{1cm} (1)

where the time subscript $t$ is omitted for the sake of simplicity. The vector, $X_j$, includes the journal’s characteristics such as number of articles, number of issues, field, dummies for major publishers, their nationality, year dummies and some interactions between them. The second variable, $I_j$, is the quality of the journal, that is, the scientific importance of its published papers. The third variable is the price of the journal, where the parameter $\alpha$ represents the disutility of price of a journal and should be positive. Finally $\xi_j$ represents the unobserved components of quality. Parameters $\beta$ and $\rho$ must be estimated.
The error component of the utility function, $\varepsilon_{ig}$ and $\varepsilon_{ij}$, are random variables that reflect the difference between the consumer’s individual appraisal over the journal and the average payoff it delivers, represented by $\delta_j$. Notably, $\varepsilon_{ig}$ is common to all journals belonging to the same field $g$ and $\varepsilon_{ij}$ is specific to the journal $j$ itself. The multiplicative parameter $\sigma$ ranges between 0 and 1 and denotes the degree of intragroup correlation, which measures the correlation of the consumer’s utility from journals that belong to the same field. The closer this parameter is to one, the higher the chance the consumer will switch to another journal within the same field when its price increases. The closer $\sigma$ is to zero, the consumer does not make distinction between fields when subscribing a journal, and it approximates the standard logit model, where all journals are symmetric.

The representative library $i$ subscribes the journal $j$ that maximizes her utility. In order to obtain a closed form probability that a library subscribes a journal $j$, the nested model assumes that both $\varepsilon_{ig}$ and $\varepsilon_{ij}$ are such that its composite term $\varepsilon_{ig} + (1-\sigma)\varepsilon_{ij}$ follows an extreme value distribution. The average utility of the outside alternative is normalized to zero, that is, $\delta_0 = 0$. At the aggregate level, such probability $s_j$ coincides with the market share of the journal $j$. The total number of subscriptions of journal $j$, say $q_j$, is directly given by expression $q_j = s_j N$. Following Berry (1994), we can write the demand equation as
\[ \ln s_j - \ln s_0 = X_j \beta + \rho I_j - \alpha p_j + \sigma \ln s_{jg} + \xi_j, \] (2)

where \( s_{jg} \) is the market of journal \( j \) in group \( g \) and \( s_0 \) is the market share of the outside good. Given our definition of market share of a journal, we include as outside good all the journals that were not subscribed at period \( t \), though they were available in the previous years.

We assume that each publisher \( f \) produces a set of journals \( F_f \). Its net profit is the sum of its operational profits minus a fixed cost \( K \). The operational profit of journal \( j \) is equal to the product of its total subscriptions and the margin, that is, the price \( p_j \) minus the marginal cost \( c_j \) of journal \( j \).

Assuming that publishers compete in prices à la Nash-Bertrand and given the nested logit specification of the demand, the pricing equation for each journal \( j \) is given by

\[ p_j = c_j - \frac{1 - \sigma}{\alpha(1 - \sigma s_{jg} - (1 - \sigma)s_f)}, \] (3)

where \( s_{jg} \) is the publisher \( f \)'s market share in field \( g \); \( s_f \) is the publisher \( f \)'s overall market share and \( c_j \) is the (constant) marginal cost of journal \( j \). The marginal cost of a journal \( j \) is parameterized as: \( c_j = \exp(w_j \gamma + \omega_j) \), where \( w_j \) is the vector of the deterministic part of the journal’s characteristics, \( \gamma \) is the technological parameters to be estimated and \( \omega_j \) is an unobserved random part. The deterministic part includes a constant term, number of issues per year, number of journals subscribed per publisher, dummies for fields, years, major publishers, nationality of the publisher, for non for profit
journals and some interactions between them. We also include the impact factor as cost characteristic. We expect its effect to be negative on costs, because more reputed journals have more efficient editorial board and easier access to good articles which lower total economic costs.

Note that the impact factor is not necessarily exogenous. It could be affected by past decisions in terms of public reach, pricing, and reputation. For these reasons, we aim at approximating the value of a journal by the lagged impact factor. More precisely, we assume that

$$I_{jt} = \lambda I_{jt-1} + Z_j \theta + u_{jt},$$

where $I_{jt}$ is the impact factor of journal $j$ at time $t$, $I_{jt-1}$ is the lagged impact factor of the journal $j$ and $Z_j$ is a vector of journals’ characteristics, which includes the number of articles and issues per year, a dummy for non-for-profit journals, some field dummies, a step dummy indicating a change of publisher. The parameter $\lambda$ would be capturing the network effect of the past readership of the journal.

3.1. Estimation Results

We now estimate different models, from the simplest model where the impact factor is treated as an exogenous variable to the most complete model where, in addition to the nested logit demand (Equation 2) and the pricing Equation (3), we take into account the impact factor equation (4). This last model is called the Full Model. To estimate these models we use different
set of instruments. By running these different models, we have observed that not recognizing the endogenous character of the impact factor yields a zero correlation between impact factor and the demand for journals. This result is counterintuitive since it implies that subscriptions would be unrelated with the impact factor, which is a celebrated measure of the quality of a journal. Once we allow for the impact factor to be endogenous, its effect on demand becomes statistically significant and with the expected positive sign. Notably, in the Full Model, which captures the network effect through the journal’s previous impact factor, this effect is positive and highly significant.

While the parameter $\alpha$ varies very little across the model specifications from 0.002298 (with a $t$-value of 25) to 0.0093 (26), the same is not true for the estimated parameter, $\sigma$, our measure of intragroup correlation, which ranges from 0.91 (48) to 0.95 (33) under the full model. The associated first $R$-squares, which roughly speaking measures the fraction of the variation of prices ($\alpha$) and market shares ($\sigma$) that are explained through the instruments, also changes considerably and for both parameters.

Some side results from the estimates are worth mentioning. The number of issues of a journal is weakly and positively correlated with the quality of the journal and the NFP journals usually have higher impact factor. Interestingly, the impact factor is not statistically affected by changes in the

---

4 A complete set of estimation results, as well as the set of instruments, is available from the authors.
ownership structure of journals.\textsuperscript{5} The field dummies indicate the importance of accounting for the specificities of a field citation pattern. In particular, Medicine and Chemistry have a quite distinct citation pattern when compared to other fields.

The estimates for the major publishers’ dummies indicate that they have a statistically negative effect on the impact factor when compared to the other publishers and their effects are quite similar between each other. At last, the electronic dummy, which is a step dummy for the year these major publishers started offering the electronic version of the journal (bundled or not with the printed version) is positively correlated with the impact factor. This indicates that the publishers were successful in using the electronic format to increase the quality of their own journals.

The cost-side parameters have the expected signs for all the model specifications. The estimated coefficients for number of issues and the constant are positive and significantly different from zero. The coefficient for the impact factor is significantly positive and, once taken as an endogenous variable, it becomes negative and statistically different from

\textsuperscript{5} One would expect that the merging activity of the last years has been concentrated among the journals with the highest quality/prestige in academics, which would indicate that the publishers’ choice towards buying a journal could not be considered deterministic. If we add this dummy in the list of endogenous variables of the model, its effect is positive and highly significant.
zero. If one takes such variable as a measure of the talent of journals, we verify that journals indeed differ with respect to their talent to select articles.

We find that as we improve the specification of the model, the sign of the estimate of the non-for-profit (NFP) dummy changes from negative but insignificant when we include impact factor as exogenous, to finally become positive and at 10% significance level when it is taken as endogenous. The empirical literature on academic journals usually finds that NFP journals price lower than for-profit (FP) ones. Our result implies that such pricing policy would not be due to lower costs but to the benefit of higher impact factors everything being equal.

The role of the dummies for nationality of the journal is very relevant for the estimates of both the impact factor and the NFP dummy estimates on the cost function. If we do not control for nationality, the effect of NFP dummy on the cost becomes negative and the effect of the impact factor becomes positive. However, a close look at our merged database yields average prices that are quite similar across countries for both FP and NFP publishers, except for UK. There, the average price of its FP journals is the double of NFP journals, even though their quality is lower. Therefore, in a context where UK data is responsible for almost half of the observations (2026 FP journals and 65 NFP journals out of 3956 observations), we find relevant to control for such idiosyncrasies with some country dummies.

Nonetheless, the lack of more precise information on costs, such as wage paid by publisher, number of pages, surface of the journal, expenses
with material are taking its toll. Our estimates on the (short run) costs are certainly not as precise as one would like it to be.

### 3.2. Empirical Analysis

Table 3 compares the key structural estimates derived from the model where impact factor is exogenous (Model 1) with the results from our Full Model (Model 2). We have captured a remarkable feature with the available data: Our estimates reveal that the demand for academic journals is highly elastic, under both scenarios. Moreover, the estimated elasticity increases once we introduce the impact factor equation. In other words, an increase in prices creates a multiplicative effect since it directly reduces the number of (paying) readers which in turn reduces the expected gain from researchers to publish in the journal.

The estimates for the marginal cost do not vary much from one specification to the other. However the pricing policy and therefore the mark-up changes considerably, decreasing by 43%, when we use the Full Model. Nevertheless, given that the estimated elasticities are already very high under the Model 2, the estimated average mark-ups are low, around 9.8%, and reduce to 5.5%. The median of the annual marginal cost of a journal is around $668 (2000 USD) and its average is close to $1081 under the Full Model.\(^6\) Furthermore, the aggregated elasticity, almost does not

\(^6\) We can also provide estimates at the publisher level. For our selected publishers, the median ranged from 1.7% to 7.3%, while the overall median economic margin is around 4%. 
change from one model to the other since $\alpha$ does not change significantly. Its estimated value is 0.52.

We can also provide estimates at the publisher level. For our selected publishers, the median ranged from 1.7% to 7.3%, while the overall median economic margin is around 4%.

<table>
<thead>
<tr>
<th>Table 3: Estimated demand and cost parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Own Elasticity</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Model 1</td>
</tr>
<tr>
<td>Model 2</td>
</tr>
</tbody>
</table>

4. Conclusion

In this article, we fit a structural model of the market for academic journals to a data set of French libraries. Our main findings are the following. Firstly, we find that library subscriptions are substantially elastic to the price of journals. Although the lack of some information on prices of a number of journals and the lack of better information about the editing and publishing costs per journal prevent us to identify more precisely the structural parameters on this market, we are able to identify the crucial

Note that we estimate long run marginal costs which could be very different from (i.e., much higher than) short run accounting variable costs. This could explain that our margins are low.
structural parameters estimates. We find that both own and cross price elasticities of demand are quite large and margins relatively low, indicating that this industry experiences competitive constraints. Secondly, we find that journals differ across their ability to select good articles. A high quality journal normally enjoys lower costs of hiring a high level editorial board and motivated referees. It is expected that these journals require lower effort to publish a good selection of articles. We cannot confirm the common perception that NFP journals have lower costs than FP journals but we confirm that NFP journals have a higher reputation, which finally could explain why these journals could have lower prices.

This last result points to a strong interrelation between readership and reputation of journals. Moreover, we have been able to identify that an increase in prices creates a multiplicative effect since it directly reduces the number of (paying) readers which in turn reduces the expected gain from researchers to publish in the journal. All these results stress out the intermediary role of academic journal between researchers, who are both producers and consumers of knowledge. Readers do not buy a journal if they do not expect its articles to be academically relevant and researchers do not submit to a journal with either limited readership or weak reputation. While on the consumer side, journals compete for subscriptions, on the producer side, journals compete for papers that would maximize the expected number of citations. This two-sidedness feature of academic
journals should be explored in further econometric research.\textsuperscript{7} At this point we could just remark that our results, in particular the endogeneity of the impact factor variable and the higher estimates of demand elasticities when we account for the endogeneity of the impact factor variable, echo the theory’s prediction that the market becomes more competitive when two-sidedness is accounted.\textsuperscript{8,9}

References


\textsuperscript{7} To our knowledge, the first theoretical models that recognize the two-sidedness feature of the academic journals market were developed by McCabe and Snyder (2005) and Jeon and Rochet (2006)

\textsuperscript{8} See Rochet and Tirole (2003 and 2005) for a presentation of this new paradigm and latest developments. See also Evans (2002).

\textsuperscript{9} Note that, if the two-sidedness of the market was precisely specified, the computation of mark-ups could be different as explained by Evans (2002). In this context our estimates of margins are probably biased, but it is hard to tell in which directions at this point.
AEI-Brookings Joint Center of Regulatorsies Studies, Related Publication
02-13.


McCabe, Mark J. and Christopher M. Snyder (2005): “Open Access and
Academic Journal Quality,” AER, 95(2), 452-458.

for price competition, merger analysis and valuation of new goods,” PhD
Dissertation, Harvard University.

Rochet, Jean-Charles and Jean Tirole (2003): “Platform Competition in
Two-sided Markets,” Journal of the European Economic Association,
1(4), 990-1029.

Rochet, Jean-Charles and Jean Tirole (2005): “Two-Sided Markets: A