The Market of Academic Journals: Empirical Evidence from Data on French Libraries

Pierre Dubois,[*] Adriana Hernandez-Perez,[⋆] Marc Ivaldi[⋆]

September 2006

Abstract

This paper analyzes the demand and cost structure of the French market of academic journals, taking into account its intermediary role between researchers, who are both producers and consumers of knowledge. This two sidedness feature echoes similar problems such as the chicken and egg problem Here readers don't 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 public reach or weak reputation. After the merging of several databases, we estimate the aggregated nested logit demand system combined simultaneously with a cost function. We identify the structural parameters of this market and find that price elasticities of demand are quite large and margins relatively low, indicating that this industry experiences competitive constraints.

[*] Université de Toulouse 1 (INRA, IDEI) and CEPR.
[⋆] Université de Toulouse 1 (GREMAQ).
[⋆] Université de Toulouse 1 (EHESS, IDEI) and CEPR.

1. Introduction

This paper analyzes the demand and cost structure of the French market of academic journals, taking into account its intermediary role between researchers, who are both producers and consumers of knowledge. This two sidedness feature echoes similar problems, such as the chicken and egg problem already observed in other markets: Readers do not buy a journal if they do not expect its articles to be academically relevant and researchers, who live under the mantra "Publish or Perish", do not submit to a journal with either limited readership or weak reputation. Therefore, 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. In this context, journals have their price settled according to their ability to attract academically relevant articles.[1]

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. Furthermore, by recognizing that the impact factor - our measure of journal's number of citations – is determined by the public reach and the reputation of the journal, we introduce an additional equation that is able to capture the two-sidedness feature of the industry.

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 can show that French universities' subscriptions are substantially elastic to the price of journal, with publishers seizing a relatively low mark-up. The data also strongly supports the two-sidedness feature of academic journals.

Academic Journals as Two Sided Platforms

The standard two-sided market models share a common feature.[2] All stress the fact that platforms that link two types of traders by some means are valued by the potential sizes of these two groups of users. As such, the

[1] 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). Our empirical study relates to McCabe and Snyder's theoretical model with regard to the journal's technology.

[2] See Rochet and Tirole (2003 and 2005) for a presentation of this new paradigm and latest developments.

platforms' pricing policy take network effects into account, adjusting its price structure according to the willingness-to-pay of each side of the market. An increase in prices on one side directly reduces its users' participation, which, in turn, reduces the expected gain from the other side of the platform.

One of our goals here is to assess the extent to which this paradigm applies to the market for academic journals. Academic researchers are both producers and consumers of knowledge. As a platform, an academic journal not only plays the traditional role of information flow management (i.e.; research output dissemination), but it also certifies the articles they publish. Usually, the larger the number of citations a journal obtains, the stronger its attractiveness among readers. For that reason, journals compete for the best articles through referee committees, which screen the submitted articles.

Now, the academic research community has strongly benefited from the decreasing costs of information processing and telecommunications, which boosted up the research output proportionally. Such phenomena, combined with the need for a faster access to scientific literature, has raised tensions on the functioning of the traditional printed journals.

In this context, observers remark a lasting movement towards a reorganization of the academic journals. The publishers reacted in twofold ways: On the researchers' side, established publishers have launched new academic journals; on the readers and librarians' side, the publishers proposed several value-added services and special journal packages. We briefly comment them in what follows.

Concerning the researchers' side, the reply to the increasing research output, driven by the sharp decrease of research costs, has been the creation of new journals, especially among for profit journals, in order to adjust to the differentiation of new research fields and the increased variance of the research quality.[3]

As regard the readers' side, a typical contract between a library and a publisher would entail simple print and/or electronic versions, with some added value services to libraries. In particular, one initiative known as the Big Deal, which differs from publisher to publisher, consists of a multi-year contract that bundles the journal's printed version to its electronic one and has a limiting policy regarding cancellation of subscriptions. Generally, this contract would be tied to the library, but also, publishers frequently would propose clauses based on the library's previous subscriptions record.[4]

In summary, the two-sidedness nature of academic journals, that is to say, dissemination on one side towards researchers/consumers and certification on the other side to researchers/producers have been responsible for the impact of the electronic revolution on the dynamics of the market of

[3] See Case, 2004, p.2 and Dewatripont *et alii*, 2005, p.44.

[4] See Edlin and Rubinfeld (2004) for a discussion on the Big Deal as a strategic barrier to entry.

academic journals, i.e., creation of new journals and the design of new contracts.

The Data

Our database combines several sources. The 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.[5]

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.[6]

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 I 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 median ratio between price and journal's citation and between price and unit

[5] 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)

[6] Not For Profit journals are all journals that belong to a University Press or Society type of publisher.

impact factor have remained constant (if not decreasing) during the sampled period.

Table I: Descriptive statistics

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

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 for 1998 and 2004.

Table II: Subscribed journals's characteristics per field (medians)

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

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 II 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 90's, 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. (See below.)

Econometric Specification and Estimation

This model borrows from the recent developments on the empirical Industrial Organization of differentiated products.[7] The representative consumer is a university library, which decides for buying one of the 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*, δ_j, which is common to all consumers, and a composite random component $\varepsilon_{ig} + (1-\sigma)\varepsilon_{jt}$. The average utility is decomposed in three parts (the time subscript *t* is omitted for the sake of simplicity):

$$\delta_j = X_j\beta + \rho I_j - \alpha p_j + \xi_j . \quad (1)$$

The first part, $X_j\beta$, includes the journal's characteristics in $X_j\beta$ such as number of articles, number of issues, field, dummies for major publishers, their nationality, year dummies and some interactions between them. The second part, I_j, includes the quality of the journal, that is, the scientific importance of its published papers. The third part is the price of the journal, where the parameter α represents the disutility of price of a journal and should be positive. Finally ξ_j represents the unobserved components of quality.

The error component of the utility function, ε_{ig} and ε_{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 δ_j. Notably, ε_{ig} is common to all journals belonging to the same field *g* and ε_{ij} is specific to the journal *j* itself. The multiplicative parameter σ 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 σ is to zero, the consumer does not

[7] See Berry (1994) and Nevo (1997) among many others.

make distinction between fields when subscribing a journal, and it approximates the standard logit model, where all journals are symmetric.

The 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 ε_{ig} and ε_{ij} are such that its composite term $\varepsilon_{ig} + (1-\sigma)\varepsilon_{jt}$ 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 follows:

$$\ln s_j - \ln s_0 = X_j \beta + \rho I_j - \alpha p_j + \sigma \ln s_{jg} + \xi_j , \qquad (2)$$

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:

$$p_j = c_j - \frac{1-\sigma}{\alpha(1-\sigma s_{fg} - (1-\sigma)s_f)} , \qquad (3)$$

where s_{fg} 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, γ is the technological parameters to be estimated and ω_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. For the reasons already discussed, we also include the impact factor as cost characteristic and we expect its effect to be negative on costs.

In the context of two-sided markets, we aim at approximating the value of a journal by the lagged impact factor. The full model is characterized by the addition of the following equation:

$$I_{jt} = \lambda I_{jt-1} + Z_j\theta + u_{jt}, \quad (4)$$

where I_{it} the impact factor of journal *j* at time *t*, I_{jt-1} is the lagged impact factor of the journal *j*. The parameter λ would be capturing the network effect of the past readership of the journal. 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.

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.[8] 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 α varies very little across the model specifications, the same is not true for the estimated σ, our measure of intragroup correlation, which ranges from 0.91 to 0.95 under the full model. The associated first R-squares, which roughly speaking measures the fraction of the variation of prices (α) and market shares (σ) that are explained through the instruments, also changes considerably and for both parameters. The implications of such change on our structural estimates for demand and costs echo the theory's prediction that the market becomes more competitive when two-sidedness is accounted.

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

[8] A complete set of estimation results, as well as the set of instruments, is available from the authors.

Empirical Analysis

Table III 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. On the top of that, the estimated elasticity increases once we introduce the impact factor equation, corroborating the results obtained in the two-sidedness theoretical literature. As discussed in a previous section, 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.[9] Furthermore, the aggregated elasticity, almost does not changed from one model to the other since α 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 III: Estimated demand and cost parameters

	Own Elasticity		Cross Elasticity		Marginal Cost		Mark Up (%)	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Model 1	-29.02	-18.28	0.526	0.108	1058.64	645.10	9.80	6.92
Model 2	-52.27	-32.94	0.987	0.203	1081.26	668.69	5.56	3.85

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

[9] 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%.

publishing costs per journal prevent us to identify more precisely the structural parameters on this market, we are able to identify the crucial 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. Also, our data confirm the relevance of the two-sidedness on the industry. 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.

The results obtained in this paper are striking and original. They have implications not only on the way the competition analysis of the industry should be carried out but also on the way publishers affect the scientific output as a certification vector and as a dissemination one.

References

Berry, S. (1994): "Estimating Discrete Choice Models of Product Differentiation," *RAND Journals of Economics*, 25(2), 242-262.

Case, Mary M. (2004): "A Snapshot in Time: ARL Libraries and Electronic Journal Resources," *Association of Research Libraries,* Bimonthly Report 235.

Dewatripont, M., V. Ginsburg, P. Legros, A. Walckiers, J-P. Devroey, M. Dujardin, F. Vandooren, P. Dubois, M. Ivaldi, J. Foncel, M-D. Heusse (2005): "Study on the economic and technical evolution of the scientific publication markets in Europe," Final Report, December 2005.

Edlin, Aaron and Daniel Rubinfeld (2004): "Exclusion or efficient pricing? The "Big Deal" bundling of academic journals," *ABA: Antitrust LJ*, 72(1), 128-159.

Jeon, Doh-Shin and Jean-Charles Rochet (2006): "The Pricing of Academic Journals: A Two-Sided Markets Perspective," *Mimeo*.

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

Nevo, Aviv (1997): "Demand for the ready-to-eat cereal and its implications 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 Progress Report", *IDEI Working Papers*, 275.