

Apportioning the Cost of a Full-Text Database Among the Journals in the Database: A Comparison of Six Methods

William H. Walters

Estimates of the price or value of the individual journals within a full-text database may be useful to librarians engaged in serials reviews or other collection development projects, to scholars investigating the determinants of journal prices, and to publishers seeking to rationalize their pricing strategies. This paper evaluates six methods of apportioning the cost of a full-text database among the individual journals in the database—methods based on variables such as journal size, total citations, Journal Impact Factor (JIF) percentile, and single-journal list price. Each method is evaluated based on how well the resulting prices can be predicted by the determinants of journal prices identified in previous research. Although the six methods yield similar results, the single best option is to use price estimates that account for JIF percentile. If citation data are not available and cannot be estimated, the best alternative is to rely on the equal-value assumption—to split the total price equally among the wanted journals in the database.

Introduction

Although nearly 20 studies have examined the determinants of scholarly journal prices since 1989, virtually all of them have focused exclusively on the prices of single-journal subscriptions.¹ The single-journal approach to price analysis remains common even today, when academic libraries acquire most of their journals through full-text databases.²

Just a few large-scale price studies have accounted for the journals available through online databases or collections. One approach to evaluating the cost of these journals is to treat each database as an indivisible entity, calculating statistics such as price per article and price per citation for each database.³ A second approach is to estimate the cost of each individual journal by apportioning the total database price among the journals in the database.⁴ The first approach has the advantage of relying on authoritative data; no price estimation is required. However, the second approach may be more useful when the goal is to evaluate journal-specific determinants of price (e.g., subject area and scholarly reputation) or when the prices of individual journals are required for library collection development decisions—

* William H. Walters is Executive Director of the Library at Manhattan College; email: william.walters@manhattan.edu. Acknowledgements: I am grateful for the advice and assistance of Susanne Markgren, Esther Isabelle Wilder, Brendon Ford, and Helen White. ©2024 William H. Walters, Attribution-NonCommercial (<https://creativecommons.org/licenses/by-nc/4.0/>) CC BY-NC.

when determining whether to bundle or unbundle subscriptions to individual journals and full-text collections, for instance.⁵

When estimates of individual journal prices are required, the total cost of each full-text database must be apportioned among the journals in the database. This can be done on the basis of

1. The equal-value assumption (total cost split equally among the wanted journals in the database)
2. Journal size (articles per year)
3. Total citations for the journal as a whole
4. Journal Impact Factor (JIF) percentile (average citations per article)
5. Single-journal list price, representing the publisher's own assessment of the journal's relative value
6. A composite indicator that accounts for variables 2–5.

There are other possibilities, of course, but these are the journal-level variables identified in previous research as the most consistent correlates of journal prices.⁶

This paper first estimates journal prices based on each of the six criteria. Each price variable is then used as the dependent variable in a regression with independent variables representing resource provider type (scholarly society, university, other non-profit, commercial publisher, or library vendor), subject field (engineering, physical sciences, life sciences, business, social sciences, or education), publisher size, JIF percentile, and journal size. The study evaluates one primary research question: Which method of estimating prices results in a dependent variable that is most fully explained by the combination of independent variables? That is, which method results in the highest R^2 value? The assumption is that an effective method of estimating price is one for which variations in price are (a) systematic rather than random, and (b) closely linked to the variables that might reasonably be expected to contribute to variations in price.

A secondary question is whether the results support or challenge an earlier finding—that for a typical U.S. master's university, the journals available through commercial publishers' databases cost substantially less than those available through the databases of non-profit publishers and library vendors. Previous research shows that while commercial databases are especially expensive for the major research universities, they are especially *inexpensive* for American bachelor's and master's universities.⁷ This study investigates whether the same finding can be seen when several different methods of price estimation are used.

Methods

The data used in this analysis were compiled for a recent Manhattan College serials review. Specifically, we attempted to acquire 2,717 *wanted journals*—those identified by the faculty as the most important titles for their teaching and research—while minimizing cost per wanted journal. Manhattan College, a 4,000-student university in the Bronx, offers bachelor's and master's degrees in engineering, business, arts and sciences, education, health, and professional studies. The college is typical of U.S. universities in the Carnegie *master's—larger* category except for the size of its engineering school, which accounts for 30% of undergraduate students.

The price data used here are actual 2019 or 2020 invoice prices (or, in some cases, price quotes) obtained by Manhattan College for the 236 full-text databases considered as possible means of gaining access to the 2,717 wanted journals. Unlike list prices, they represent the amounts actually paid or payable. The details of the data compilation process are described in an earlier study.⁸

Journals in the arts and humanities (A&H) were excluded from the study due to data limitations—specifically, because citation data were unavailable for a relatively high proportion of those journals. That is, the A&H journals selected by the faculty include quite a few that are not indexed in Web of Science. Open Access (OA) journals were also excluded from the analysis since they are freely accessible without a subscription. Consequently, *number of wanted journals in the database* refers to the number of wanted journals that are neither A&H nor OA. Likewise, *total database price* refers to the total price times the proportion of journals in the database that are neither A&H nor OA.

It is important to keep in mind that any one journal may be acquired through several different subscriptions or databases. Consequently, price is an attribute not of a particular journal, but of a particular *acquisition opportunity*.⁹ To gain current access to *Northeastern Naturalist*, for instance, a library might choose a single-journal subscription from the publisher or subscribe to any of 13 full-text databases offered by BioOne, EBSCO, or ProQuest. That's 14 acquisition opportunities with annual prices ranging from \$105 to \$545. The data file for this investigation has 4,529 cases that correspond to 4,529 acquisition opportunities—4,529 instances in which a particular wanted journal was included in a particular full-text database. For each case, there are 6 dependent variables (price estimates) and 14 independent variables that represent 5 constructs: resource provider type, subject field, publisher size, JIF percentile, and journal size.

Price Estimates (Dependent Variables)

Five of the six price estimates—all but the composite indicator—were calculated using similar methods.

1. For *price (equal value)*, the total database price was split equally among the wanted journals in the database. This calculation is based on the assumption that the value of each journal (relative to that of the other journals in the same database) does not vary systematically on the basis of size, scholarly impact, or list price.
2. For *price (journal size)*, each wanted journal was assigned a value equal to the total database price times the proportion of the wanted-journal articles in the database that appeared in the journal. (*Wanted-journal articles* are simply articles that appeared in the wanted journals. No differentiation between *wanted* and *not wanted* status was made at the article level.) This calculation is based on the assumption that price is determined mainly by the number of articles in each journal—specifically, the number of Web of Science citable items published in 2019. Citable items include empirical articles, review articles, research notes, and other substantive contributions but not items such as announcements, editorials, and letters to the editor.
3. For *price (total citations)*, each wanted journal was assigned a value equal to the total database price times the proportion of the database's wanted-journal citation total (number of citing articles) that could be attributed to the journal. With this variable, price is proportional to the number of times the journal (all articles combined) was cited in 2019.¹⁰
4. For *price (JIF percentile)*, each wanted journal's 2019 Impact Factor was first expressed as the average of the journal's percentile ranks in all the Web of Science subject categories in which the journal was classified. Each journal was then assigned a value equal to the total database price times the proportion of the database's wanted-journal

percentile-rank total that could be attributed to the journal. *Price (JIF percentile)* is based on the assumption that price is proportional to the average number of times each article in the journal was cited in 2019. It is therefore different from *price (total citations)* in two important ways. First, it makes use of data on average citations per article rather than total citations per journal; it is therefore not influenced by the number of articles published in the journal. Second, it is based on percentile ranks rather than raw scores; it represents each journal's impact relative to that of the other journals in the same subject category. With *price (JIF percentile)*, a top-tier political science journal is assigned the same price as a top-tier biochemistry journal in the same database. This method disregards the fact that the average citation rate is higher in biochemistry than in political science.

5. For *price (single-journal price)*, each wanted journal was assigned a value equal to the total database price times the proportion of the database's single-journal list price total (the sum of the single-journal list prices of the wanted journals) that could be attributed to the journal. This calculation assumes that the publishers themselves have a good idea of the value of each of their journals, and that their assessments of value are incorporated into the journals' list prices. *Price (single-journal price)* is consistently lower than actual list price, but proportional to it. With just a few exceptions, the single-journal list prices used in this analysis are 2019 or 2020 prices from EBSCO or from the publishers' web sites.
6. A different method was used to arrive at *price (composite)*, a composite indicator that incorporates dependent variables 2–5, above. First, unweighted least squares extraction—the initial step in factor analysis—was used to calculate communality values, which represent the extent to which each price variable contributes to the shared variance within the set of four variables (i.e., the extent to which each variable can be represented by the other three).¹¹ Communalities of 0.89, 0.76, 0.67, and 0.72 were obtained for variables 2–5, respectively, revealing that *price (journal size)* best captures the variance common to the set of four variables. Because the eigenvalues of the extracted factors showed that all four variables could be represented well by a single composite indicator, a composite score for each journal was calculated as the sum of the four (*communality * estimated price*) values. That is, each of the four component variables was weighted in proportion to its contribution to the shared variance.¹² Finally, each wanted journal was assigned an estimated price equal to the total database price times the proportion of the database's composite-score total that could be attributed to the journal.

Three of the six price estimates require the use of citation data. Because the A&H journals—those most likely to have missing values for the citation variables—were excluded from the analysis, just 5.7% of the remaining 4,529 cases have one or more missing values. For those cases, *total citations* and *JIF percentile* were estimated.¹³

The correlations among the six price variables are shown in Table 1. As described earlier, each price variable was used as the dependent variable in a regression that included the independent variables identified in earlier research as effective predictors of journal prices. (See below.) The dependent variables were entered in natural log form in order to maintain linearity.

TABLE 1
Correlations Among the Dependent Variables (the Six Price Variables)

Variable	Price (equal value)	Price (journal size)	Price (total citations)	Price (JIF percentile)	Price (single-journal price)	Price (composite)
Price (equal value)	—	0.75	0.66	0.92	0.79	0.84
Price (journal size)	0.75	—	0.85	0.74	0.81	0.95
Price (total citations)	0.66	0.85	—	0.73	0.70	0.92
Price (JIF percentile)	0.92	0.74	0.73	—	0.72	0.86
Price (single-journal price)	0.79	0.81	0.70	0.72	—	0.89
Price (composite)	0.84	0.95	0.92	0.86	0.89	—

Correlates of Price (Independent Variables)

All six regressions used the same set of independent variables:

1. Resource provider type (five categories): scholarly society, university, other non-profit, commercial publisher, or library vendor. The resource provider is almost always the publisher, except for the databases provided by library vendors such as EBSCO and ProQuest. The *university* category includes both university presses and academic departments/centers.
2. Subject field (six categories): engineering, physical sciences, life sciences, business, social sciences, or education, based on the Manhattan College department(s) that identified the journal as a wanted journal. Because some journals were wanted by more than one department, about 10% of the journals have more than one subject designation.
3. Publisher size: number of wanted journals published by the publisher (not always the resource provider), including those of subsidiary imprints.
4. JIF percentile: 2019 JIF, expressed as a percentile within the relevant Web of Science subject category. If the journal appeared in multiple subject categories, the percentile scores were averaged. Because JIF is independent of journal size, it represents the average citation impact of an article in the journal rather than the impact of the journal as a whole.
5. Journal size: number of citable items published in 2019.

Although two of the independent variables were used in the construction of the dependent variables, this is not a problem, since the dependent and independent variables do not represent the same constructs. Moreover, because characteristics *not* represented within the set of independent variables (e.g., total database price and the number of wanted journals) figure heavily in each price estimate, the correlations between the dependent variables and the independent variables are modest. The correlation between *price (journal size)* and *journal size* is 0.27, for instance, indicating that just 7% of the variation in *price (journal size)* can be

explained by *journal size* ($r^2 = 0.07$). Likewise, the correlation between *price (JIF percentile)* and *JIF percentile* is just 0.17 ($r^2 = 0.03$).

Results and Discussion

The independent variables, taken together, are more closely associated with some price estimates than with others (Table 2). The highest R^2 value is that for *price (JIF percentile)*. This indicates that the independent variables are most effective at explaining variations in price when the total database price is allocated among the wanted journals based on the average citation impact of an article in each journal (JIF), expressed as a percentile score (i.e., relative to the other journals in the same Web of Science subject category). If we want the price variable that is most sensitive to the characteristics that might reasonably be expected to influence price, then *price (JIF percentile)* is the best of the options shown in Table 2.

If price estimates are needed for journals for which citation data are unavailable, then *price (equal value)* is a good alternative to *price (JIF percentile)*. As noted earlier, three of the six price estimation methods require actual or estimated citation data for every journal. For journals not included in Web of Science, three options are available: (1) use a price estimation method that does not rely on citation data, such as the equal-value method; (2) use a data source that includes citation data for a broader range of journals (e.g., Scopus rather than Web of Science, and CiteScore rather than JIF); or (3) estimate the citation values for the journals with missing data before calculating price estimates. Fortunately, the regression results suggest that the first of these options is entirely reasonable. Based on the R^2 and SEE values shown in Table 2, the equal-value method is a good alternative to the JIF percentile method. Moreover, the two methods result in price estimates that are very closely related ($r = 0.92$; see Table 1).

Comparing the Results for Particular Price Variables

The fact that *price (JIF percentile)* has a higher R^2 value than *price (equal value)*, *price (composite)*, and *price (total citations)* is surprising for at least two reasons. First, we might expect a higher R^2 value for the composite indicator since it incorporates the shared variance common to all four of its component variables. In fact, however, the composite indicator produces less satisfactory results than either *price (JIF percentile)* or *price (equal value)*.

Second, we might expect a higher R^2 value for *price (total citations)* than for *price (JIF percentile)* since *total citations* represents the scholarly impact of the journal as a whole rather than the average impact of a single article in the journal. For instance, if there are two journals with equal JIF percentile scores but one publishes twice as many articles as the other, *price (total citations)* will account for the difference in journal size while *price (JIF percentile)* will not. One explanation for the lower R^2 value for *price (total citations)* is that the price or value of a journal is not closely related to the number of articles it publishes. This first explanation is not unreasonable, especially considering the relatively low R^2 value associated with *price (journal size)*.

Variable	Adj. R ²	SEE
Price (JIF percentile)	0.43	0.91
Price (equal value)	0.33	0.81
Price (composite)	0.33	0.97
Price (total citations)	0.29	1.24
Price (single-journal price)	0.21	1.21
Price (journal size)	0.20	1.46

There is a second and perhaps more likely possibility, however; the high R^2 value for *price (JIF percentile)* may be related to the use of percentile scores. If this is the case, it suggests that the price of a journal is tied to its relative standing within its subject area—not to its actual citation rate—and that we ought to use percentile scores to account for the differences in average citation rates across disciplines. A price variable based on JIF raw scores can be used to test this assertion. If the assertion is valid, then *price (JIF raw score)* will have a lower R^2 value than *price (JIF percentile)*—and it does. A regression with *price (JIF raw score)* as the dependent variable results in a low R^2 value (0.22) and an error (SEE) value of 3.12, far higher than any of the values shown in Table 2. We can therefore conclude that *price (JIF percentile)* is probably effective due to the use of percentile scores rather than actual JIF values.¹⁴

As Table 2 shows, *price (single-journal price)* and *price (journal size)* are associated with the lowest R^2 values. Notably, the price variable with the most shared variance, *price (journal size)*, has the lowest R^2 value of all. Conversely, the price variable with the least shared variance, *price (JIF percentile)*, yields the highest R^2 value. The reasons for this are not clear. These results do suggest two related findings, however. First, combining multiple dimensions of price into a single variable (the composite variable) does not increase the extent to which the estimated prices can be explained by the independent variables in the regression. Second, the price estimates that can be predicted most effectively are not necessarily those with the most shared variance.

Correlates of Price

Because the dependent variables were entered in natural log form, the unstandardized regression (*B*) coefficients cannot be interpreted as dollar amounts. Table 3 shows the *effect coefficients*, which are more intuitively meaningful. Each represents the percentage change in price asso-

TABLE 3
Effect Coefficients for the Six Regressions (the Six Price Variables)*

Variable	Price (equal value)	Price (journal size)	Price (total citations)	Price (JIF percentile)	Price (single-journal price)	Price (composite)
Scholarly society	226	169	128	199	319	177
University	216	331	176	212	466	263
Other non-profit	124	439	252	106	473	236
Commercial publisher	—	—	—	—	—	—
Library vendor	189	303	225	200	233	258
Engineering	-12	23	ns	-19	31	ns
Physical sciences	-9	-12	-14	-14	13	-8
Life sciences	23	66	99	33	78	55
Business	-10	ns	ns	-10	12	ns
Social sciences	—	—	—	—	—	—
Education	ns	ns	-13	ns	-12	-13
Publisher size	ns	0.1	0.1	ns	0.1	0.0
JIF percentile	ns	ns	1.7	2.2	0.3	1.0
Journal size	0.0	0.1	0.1	ns	0.1	0.1

*Each effect coefficient is equal to $(\exp(B)-1) * 100$. *Commercial publisher* and *social sciences* are the reference categories. Values of “ns” are not significant at the 0.05 level, two-tailed.

ciated with a one-unit change in the independent variable—or, for categorical variables, the percentage change in price associated with inclusion in the indicated category rather than the reference category. (The complete regression results can be found in the Appendix.)

As Table 3 reveals, the results for resource provider type are similar across all six regressions. Moreover, all six confirm earlier reports that for a typical master's university, the journals available through commercial publishers' databases cost less, all else equal, than those available through the databases of library vendors and nonprofit providers.¹⁵ The publisher-type differentials do vary in magnitude, however. All else equal, the journals acquired from scholarly societies may cost from 128% to 319% more than those acquired from commercial publishers, depending on which price variable is used.

Earlier investigations also identified two subject variables, *life sciences* and *physical sciences*, as important determinants of journal prices. Those same findings can be seen in Table 3. The very modest effects of publisher size, JIF percentile, and journal size are also consistent with previous research.¹⁶

Conclusion

Because there is no definitive way to determine the correct market price of each journal included in a full-text database, the results presented here cannot be regarded as authoritative. If there is a strong theoretical or methodological reason for estimating prices based on a particular construct, such as journal size or single-journal list price, then that construct should determine the method by which prices are estimated.

In the absence of a strong rationale for a particular price estimation method, however, it seems reasonable to use price estimates that make intuitive sense—estimates that can be explained in terms of the variables most consistently associated with price. By that criterion, the best approach is to use the JIF percentile method described here—to apportion the total database price in accordance with the JIF percentile scores of the wanted journals included in the database. If citation data are unavailable, then *price (equal value)* is a good alternative to *price (JIF percentile)*.

The results for all six price variables are consistent with earlier reports that for a typical master's university, the journals acquired through commercial publishers' databases cost less than those acquired through the databases of scholarly societies, universities, other nonprofits, and library vendors.

Application of These Findings

There are several contexts in which the findings of this investigation may be useful. First, recent studies suggest that the acquisition of full-text journal resources for library collections should involve two separate steps: (1) the selection of individual journals on a title-by-title basis and (2) the identification of the full-text databases that can provide access to those journals in the most cost-effective way.¹⁷ If the serials review or evaluation procedure requires price estimates for every acquisition opportunity—every wanted journal within each full-text database—then a defensible method of apportioning database prices among journals will be needed.

Second, scholarly investigations of the determinants of journal prices are also likely to require the allocation of total database cost among the journals in each database. Some determinants of price (e.g., publisher's market share and for-profit/non-profit status) are attributes of particular publishers or databases rather than individual journals, while others (e.g., subject area and scholarly reputation) are specific to each journal and therefore require

the estimation of prices for individual acquisition opportunities. Recent journal price studies have relied on *price (equal value)* and *price (journal size)*,¹⁸ but this investigation shows that at least one indicator, *price (JIF percentile)*, is likely to be a better choice.

Third, publishers and library vendors may find it useful to disaggregate database prices in order to assess their own pricing strategies, to identify anomalies in the list prices of particular journals, or to demonstrate to libraries that their products are cost-effective—to show, for instance, that their own journals are a good value in comparison with similar titles from other vendors. Because single-journal subscriptions account for relatively few of the titles held by libraries,¹⁹ the most meaningful comparisons involve not single-journal prices, but the prices that would be paid if each journal were acquired through the most cost-effective full-text database offered by the vendor or publisher.

Further Research

Further research using data for a range of institutions might help extend or clarify the findings presented here. Nonetheless, these results, based on Manhattan College price data, are likely to be useful to other universities as well. For one thing, Manhattan College is typical of many U.S. bachelor's and master's institutions with regard to its size, mission, reputation, selectivity, student characteristics, teaching/research focus, and library budget. The curriculum is not unusual except for the size of the engineering program, and the wanted journals selected by the faculty include nearly all the high-impact journals in the subjects typically taught at U.S. undergraduate colleges.²⁰ Moreover, most of the library's journal budget is devoted to resources acquired through WALDO and LYRASIS, two of the largest library consortia in the United States. The consortial price schedules that apply to Manhattan College also apply to more than 1,400 other member libraries.

Research on journal prices would also benefit from greater transparency and more widespread dissemination of price information. Even today, many investigations rely on list prices, which often bear little relationship to the prices actually paid by libraries. A broader, and perhaps insurmountable, challenge lies in the disconnect between the end user's desire for particular scholarly works and the publisher's (and librarian's) focus on *information products*. While researchers need access to particular journals—or, more accurately, particular articles—publishers and librarians tend to think of cost or revenue in terms of the journal databases or packages that are marketed and acquired as indivisible units. The main analytical problem stems not from the sale or acquisition of full-text databases, but from the fact that their associated costs cannot be readily disaggregated. As long as this remains true, price estimation methods such as those described here are likely to remain useful despite their limitations.

Appendix

Each of the six price variables was used as the dependent variable in a separate regression (Tables A1–A6). *B* is the unstandardized regression coefficient, *Beta* is the standardized coefficient, *n* = 4,529, and the significance levels are two-tailed. Each effect coefficient is equal to $(\exp(B)-1) * 100$. *Commercial publisher* and *social sciences* are the reference categories for resource provider type and subject field.

	Effect	B	SE	Beta	Sig.
Scholarly society	226	1.182	0.056	0.347	0.00
University	216	1.149	0.065	0.241	0.00
Other non-profit	124	0.807	0.080	0.135	0.00
Commercial publisher	—	—	—	—	—
Library vendor	189	1.061	0.037	0.484	0.00
Engineering	-12	-0.123	0.043	-0.039	0.00
Physical sciences	-9	-0.092	0.034	-0.037	0.01
Life sciences	23	0.210	0.037	0.073	0.00
Business	-10	-0.103	0.033	-0.041	0.00
Social sciences	—	—	—	—	—
Education	-3	-0.035	0.044	-0.010	0.42
Publisher size	0.0	0.000	0.000	-0.035	0.05
JIF percentile	-0.1	-0.001	0.000	-0.015	0.25
Journal size	0.0	0.000	0.000	0.031	0.02
Y-intercept		5.124	0.047		
Adj. R ²		0.33			
SEE	125	0.812			

	Effect	B	SE	Beta	Sig.
Scholarly society	169	0.990	0.101	0.176	0.00
University	331	1.462	0.117	0.186	0.00
Other non-profit	439	1.685	0.144	0.171	0.00
Commercial publisher	—	—	—	—	—
Library vendor	303	1.395	0.066	0.385	0.00
Engineering	23	0.204	0.077	0.039	0.01
Physical sciences	-12	-0.127	0.062	-0.031	0.04
Life sciences	66	0.507	0.067	0.107	0.00
Business	-2	-0.020	0.059	-0.005	0.74
Social sciences	—	—	—	—	—
Education	-4	-0.044	0.079	-0.008	0.58
Publisher size	0.1	0.001	0.000	0.130	0.00
JIF percentile	0.1	0.001	0.001	0.026	0.07
Journal size	0.1	0.001	0.000	0.284	0.00
Y-intercept		3.856	0.085		
Adj. R ²		0.20			
SEE	332	1.463			

TABLE A3
Regression Results for Price (Total Citations)

	Effect	B	SE	Beta	Sig.
Scholarly society	128	0.826	0.085	0.162	0.00
University	176	1.016	0.099	0.143	0.00
Other non-profit	252	1.258	0.123	0.141	0.00
Commercial publisher	—	—	—	—	—
Library vendor	225	1.178	0.056	0.360	0.00
Engineering	10	0.100	0.065	0.021	0.13
Physical sciences	-14	-0.153	0.052	-0.041	0.00
Life sciences	99	0.691	0.057	0.161	0.00
Business	7	0.067	0.050	0.018	0.18
Social sciences	—	—	—	—	—
Education	-13	-0.141	0.067	-0.028	0.04
Publisher size	0.1	0.001	0.000	0.096	0.00
JIF percentile	1.7	0.017	0.001	0.329	0.00
Journal size	0.1	0.001	0.000	0.237	0.00
Y-intercept		3.008	0.072		
Adj. R ²		0.29			
SEE	246	1.242			

TABLE A4
Regression Results for Price (JIF Percentile)

	Effect	B	SE	Beta	Sig.
Scholarly society	199	1.095	0.063	0.263	0.00
University	212	1.137	0.073	0.195	0.00
Other non-profit	106	0.725	0.090	0.099	0.00
Commercial publisher	—	—	—	—	—
Library vendor	200	1.100	0.041	0.410	0.00
Engineering	-19	-0.211	0.048	-0.055	0.00
Physical sciences	-14	-0.150	0.038	-0.049	0.00
Life sciences	33	0.282	0.042	0.081	0.00
Business	-10	-0.106	0.037	-0.035	0.00
Social sciences	—	—	—	—	—
Education	-7	-0.074	0.049	-0.018	0.13
Publisher size	0.0	0.000	0.000	0.023	0.17
JIF percentile	2.2	0.021	0.000	0.510	0.00
Journal size	0.0	0.000	0.000	0.016	0.21
Y-intercept		3.573	0.053		
Adj. R ²		0.43			
SEE	149	0.913			

TABLE A5
Regression Results for Price (Single-Journal Price)

	Effect	B	SE	Beta	Sig.
Scholarly society	319	1.432	0.083	0.306	0.00
University	466	1.734	0.097	0.266	0.00
Other non-profit	473	1.746	0.119	0.213	0.00
Commercial publisher	—	—	—	—	—
Library vendor	233	1.202	0.054	0.400	0.00
Engineering	31	0.273	0.064	0.063	0.00
Physical sciences	13	0.126	0.051	0.037	0.01
Life sciences	78	0.575	0.055	0.146	0.00
Business	12	0.113	0.049	0.033	0.02
Social sciences	—	—	—	—	—
Education	-12	-0.129	0.065	-0.028	0.05
Publisher size	0.1	0.001	0.000	0.247	0.00
JIF percentile	0.3	0.003	0.001	0.054	0.00
Journal size	0.1	0.001	0.000	0.168	0.00
Y-intercept		3.659	0.070		
Adj. R ²		0.21			
SEE	235	1.208			

TABLE A6
Regression Results for Price (Composite)

	Effect	B	SE	Beta	Sig.
Scholarly society	177	1.017	0.067	0.249	0.00
University	263	1.289	0.077	0.227	0.00
Other non-profit	236	1.212	0.095	0.170	0.00
Commercial publisher	—	—	—	—	—
Library vendor	258	1.276	0.044	0.486	0.00
Engineering	10	0.094	0.051	0.025	0.06
Physical sciences	-8	-0.088	0.041	-0.029	0.03
Life sciences	55	0.438	0.044	0.128	0.00
Business	-2	-0.022	0.039	-0.007	0.58
Social sciences	—	—	—	—	—
Education	-13	-0.139	0.052	-0.034	0.01
Publisher size	0.0	0.000	0.000	0.111	0.00
JIF percentile	1.0	0.010	0.001	0.247	0.00
Journal size	0.1	0.001	0.000	0.222	0.00
Y-intercept		3.819	0.056		
Adj. R ²		0.33			
SEE	163	0.967			

Notes

1. William H. Walters, "Can Differences in Publisher Size Account for the Relatively Low Prices of the Journals Available to Master's Universities Through Commercial Publishers' Databases? The Importance of Price Discrimination and Substitution Effects," *Scientometrics* 127 (Feb. 2022): 1065–97, <https://doi.org/10.1007/s11192-021-04205-5>.
2. Stephen Bosch, Barbara Albee, and Sion Romaine, "The New Abnormal: Periodicals Price Survey 2021," *Library Journal* 146 (Apr. 27, 2021): 20–25, <https://www.libraryjournal.com/story/The-New-Abnormal-Periodicals-Price-Survey-2021>; Oliver T. Coomes, Tim R. Moore, and Sébastien Breaux, "The Price of Journals in Geography," *The Professional Geographer* 69 (2017): 251–62, <https://doi.org/10.1080/00330124.2016.1229624>; Rob Johnson, Anthony Watkinson, and Michael Mabe, *The STM Report: An Overview of Scientific and Scholarly Publishing*, 5th ed. (The Hague: STM: International Association of Scientific, Technical, and Medical Publishers, 2018), https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf; Lewis G. Liu and Harold Gee, "Determining Whether Commercial Publishers Overcharge Libraries for Scholarly Journals in the Fields of Science, Technology, and Medicine, with a Semilogarithmic Econometric Model," *Library Quarterly* 87 (Apr. 2017): 150–72, <https://doi.org/10.1086/690736>; Karla L. Strieb and Julia C. Blixrud, "Unwrapping the Bundle: An Examination of Research Libraries and the "Big Deal,"" *Portal: Libraries and the Academy* 14 (Oct. 2014): 587–615, <https://doi.org/10.1353/pla.2014.0027>.
3. Theodore C. Bergstrom, Paul N. Courant, R. Preston McAfee, and Michael A. Williams, "Evaluating Big Deal Journal Bundles," *PNAS* 111 (June 16, 2014): 9425–30, <https://doi.org/10.1073/pnas.1403006111>.
4. Walters, "Can Differences in Publisher Size"; William H. Walters and Susanne Markgren, "Comparing the Prices of Commercial and Nonprofit Journals: A Realistic Assessment," *portal: Libraries and the Academy* 21 (Apr. 2021): 389–410, <http://doi.org/10.1353/pla.2021.0021>.
5. Khue Duong, Carol Perruso, and Hema Ramachandran, "Content Overlap and Replacement Cost Analyses: Tools to Evaluate Abstracting/Indexing (A&I) and Full-Text Databases in Science and Engineering," *Science & Technology Libraries* 32 (2013): 84–94, <https://doi.org/10.1080/0194262X.2012.758461>; Asen O. Ivanov, Catherine Anne Johnson, and Samuel Cassady, "Unbundling Practice: The Unbundling of Big Deal Journal Packages as an Information Practice," *Journal of Documentation* 76 (2020): 1051–67, <https://doi.org/10.1108/JD-09-2019-0187>; Elizabeth Parang and Jeremy Whitt, "When to Hold Them, When to Fold Them: Reassessing "Big Deals" in 2020," *The Serials Librarian* 80 (2021): 147–52, <https://doi.org/10.1080/0361526X.2021.1877083>.
6. Walters, "Can Differences in Publisher Size."
7. See Bergstrom et al., "Evaluating Big Deal Journal Bundles," appendix table SI 17; Walters, "Can Differences in Publisher Size"; Walters and Markgren, "Comparing the Prices of Commercial and Nonprofit Journals." Although most journal price studies have reported that commercially published journals are more expensive than those of non-profit publishers, these three papers are more realistic than the others in several respects: (a) unlike other journal price studies, they present data for all types of U.S. colleges and universities rather than just the major research universities; (b) they include only the journals that have met libraries' selection criteria—the journals that faculty and librarians actually want for their collections; (c) they evaluate not just single-journal subscriptions, but the acquisition opportunities available through full-text databases and other online resources; and (d) they are based not on list prices, but on the prices actually paid by academic libraries—the prices negotiated by library consortia, university systems, and individual institutions.
8. William H. Walters and Susanne Markgren, "Zero-Based Serials Review: An Objective, Comprehensive Method of Selecting Full-Text Journal Resources in Response to Local Needs," *Journal of Academic Librarianship* 46 (Sept. 2020), article 102189, <https://doi.org/10.1016/j.acalib.2020.102189>; William H. Walters and Susanne Markgren, "A Two-Stage Approach to Serials Review: Minimizing Journal Costs Through Title-by-Title Selection with Package-Based Acquisition," *Insights: The UKSG Journal* 34 (July 21, 2021a): article 18, <http://doi.org/10.1629/uksg.550>.
9. Walters and Markgren, "Zero-Based Serials Review."
10. The data on journal size, total citations, and JIF percentile are all 2019 data from Clarivate Analytics, *Journal Citation Reports* (London: Clarivate Analytics, 2019).
11. Leandre R. Fabrigar and Duane T. Wegener, *Exploratory Factor Analysis* (New York: Oxford University Press, 2012); Jae-On Kim and Charles W. Mueller, *Factor Analysis: Statistical Methods and Practical Issues* (Beverly Hills, CA: SAGE Publications, 1978); Jae-On Kim and Charles W. Mueller, *Introduction to Factor Analysis: What It Is and How To Do It* (Beverly Hills, CA: SAGE Publications, 1978); Paul Kline, *An Easy Guide to Factor Analysis* (New York: Routledge, 1994).
12. Although factor scores might have been used in the creation of the composite variable, this procedure seemed more appropriate. Factor scores represent departures from the mean for all the journals in the entire data set, but the goal here is to represent the relative prices of the journals within each particular database.

13. The estimation procedure is based on the assumption that the journals not listed in Web of Science are similar to the lower-impact journals listed in Web of Science. Each missing value was replaced with the average value for the wanted journals in the lowest 20% of the Web of Science distribution within the appropriate subject area: engineering, physical sciences, life sciences, business, social sciences, or education.

14. In the calculation of *price (JIF raw score)*, each wanted journal was assigned a value equal to the total database price times the proportion of the database's wanted-journal JIF total that could be attributed to the journal. The regression for *price (JIF raw score)* used the same independent variables as the other analyses. In all seven regressions, *JIF percentile* was used as an independent variable because, overall, (a) it was more closely related to the various price variables than either JIF raw score or total citations, and (b) its use resulted in lower levels of multicollinearity among the independent variables.

15. Bergstrom et al., "Evaluating Big Deal Journal Bundles"; Walters, "Can Differences in Publisher Size"; Walters and Markgren, "Comparing the Prices of Commercial and Nonprofit Journals."

16. Walters, "Can Differences in Publisher Size."

17. Walters and Markgren, "Zero-Based Serials Review"; Walters and Markgren, "A Two-Stage Approach."

18. Walters, "Can Differences in Publisher Size"; Walters and Markgren, "Comparing the Prices of Commercial and Nonprofit Journals."

19. Johnson, Watkinson, and Mabe, *The STM Report*; Strieb and Blixrud, "Unwrapping the Bundle."

20. William H. Walters and Susanne Markgren, "Do Faculty Journal Selections Correspond to Objective Indicators of Citation Impact? Results for 20 Academic Departments at Manhattan College," *Scientometrics* 118 (Jan. 2019): 321–37, <https://doi.org/10.1007/s11192-018-2972-7>.