PHOENIX
CENTER
FOR ADVANCED LEGAL & ECONOMIC PUBLIC POLICY STUDIES
www.phoenix-center.org

PHOENIX CENTER POLICY BULLETIN NO. 64

T. Randolph Beard
George S. Ford, PhD

June 2023

CAN SELF-PREFERENCING BY AN ONLINE RETAILER BE DETECTED?

A MONTE CARLO SIMULATION

Abstract: Internet search results must be presented in a sequential manner—there is a first result, a second, a third, and so forth. For some platforms, an independent party may pay a fee for a higher ranking in the sequence. And when a platform offers its own products or services, the platform may favor its offerings over other sellers in its search returns. A paper by Farronato, Fradkin, and MacKay (2023) (hereinafter "*FFM Study*") recently evaluated whether Amazon favors its own brands in search returns, reporting a mild preference, though middling rank, for Amazon brands. It is an interesting paper, but unfortunately the results of the *FFM Study* are invalid. In this POLICY BULLETIN, we demonstrate that the use of Ordinary Least Squares regression when looking at the sequence of search returns provides systematically biased coefficients and is prone to find preferences that do not exist and miss preferences that do. Also, the results of hypothesis tests are incorrect. The problem of omitted variables, which are certain to exist in the study, makes matters worse. A suitable alternative for detecting preference in ranked data has proven elusive.

I. Background

House-branded products are common in retailing and offer consumers lower prices for quality goods. As such, they are generally regarded as pro-competitive and desirable. Yet, the provision of house-branded products by Amazon is frowned upon by some government officials, though the online-availability of house brands offered by more traditional retailers such as Walmart and Target receive little-to-no scrutiny. Estimates suggest that Amazon's house brands account for only 3% of sales.[1] In contrast, more than 77% of sales for grocers ALDI and 59% of Trader Joe's are house brands, while house brands at Costco, Sam's Club, Walmart, and Target are about a quarter to a third of their sales.[2] Amazon evidently faces special scrutiny for reasons other than the share of sales made up by their house brands.[3]

The reasons to condemn house brands in online settings are somewhat unclear but seem to be related to the requirement that any online retailor must present products to consumers in some sort of sequence.[4] It is claimed that if Amazon promotes its own products in its store and thus they appear higher in the presented order, then it may encourage sales of its own brand to the detriment of other sellers and brands Amazon voluntarily allows to operate on its platform.[5] Of course, the same is true for any other online or brick-or-mortar outlet that promotes its house

---

1 *Aldi, Target, Amazon Lead with Fastest-Growing Private Label Brands, Numerator Reports*, Numerator (August 1, 2022) (available at: https://www.globenewswire.com/news-release/2022/08/01/2489496/0/en/ALDI-TARGET-AMAZON-LEAD-WITH-FASTEST-GROWING-PRIVATE-LABEL-BRANDS-NUMERATOR-REPORTS.html); H. Lalley, *In Private Label, Walmart Towers Over Everybody Else*, WINSIGHT GROCERY BUSINESS (August 1, 2022) (available at: https://www.winsightgrocerybusiness.com/cpg/private-label-walmart-towers-over-everybody-else); Target Annual Report (2021), at p. 18 (available at: https://corporate.target.com/_media/TargetCorp/annualreports/2021/pdfs/2021-Target-Annual-Report.pdf). Amazon's house brands may have a higher market share in certain commodities, for example batteries, due to Amazon's lower prices.

2 *Id*.

3 C. Newton, *The Tech Backlash is Real, and It's Accelerating*, THE VERGE (September 17, 2019) (available at: https://www.theverge.com/interface/2019/9/17/20869495/tech-backlash-nyt-rob-walker-antitrust-privacy); L. Hurley and D. Ingram, *Biden And Republican Senators Join Forces In Attack On Big Tech At Supreme Court*, NBC NEWS (February 18. 2023) (available at: https://www.nbcnews.com/politics/supreme-court/biden-republican-senators-join-forces-attack-big-tech-supreme-court-rcna69353); L.J. Spiwak, *Why Does Congress Want to Break Amazon Prime?*, NOTICE & COMMENT – YALE JOURNAL ON REGULATION (February 18, 2022) (available at: https://www.yalejreg.com/nc/why-does-congress-want-to-break-amazon-prime-by-lawrence-j-spiwak).

4 *See, e.g.*, G.S. Ford and M. Stern, *Retail Platform Bias?* PHOENIX CENTER POLICY PERSPECTIVE NO. 22-02 (February 10, 2022) (available at: https://www.phoenix-center.org/perspectives/Perspective22-02Final.pdf); T.R. Beard, G.S. Ford and M. Stern, *Product Guidance By Digital Platforms: A Welfare Analysis*, PHOENIX CENTER POLICY BULLETIN NO. 57 (May 2022) (available at: https://www.phoenix-center.org/PolicyBulletin/PCPB57Final.pdf).

5 Concerns over Amazon's alleged self-preferencing led to the introduction last Congress of a bill to prohibit such practices. As the bill contained many flaws, the legislation never received a floor vote. *See* Spiwak, *Why Does Congress Want to Break Amazon Prime?*, *supra* n. 3; GS. Ford, *The American Innovation and Choice Online Act is an "Economics-Free Zone"*, NOTICE & COMMENT – YALE JOURNAL ON REGULATION (June 10, 2022) (available at: https://www.yalejreg.com/nc/american-innovation-and-choice-online-act-economics-free-zone); L.J. Spiwak, *The Third Time is Not the Charm: Significant Problems Remain With Senator Klobuchar's Antitrust Reform Bill*, FEDERALIST SOCIETY BLOG (June 7. 2022) (available at: https://fedsoc.org/commentary/fedsoc-blog/the-third-time-is-not-the-charm-significant-problems-remain-with-senator-klobuchar-s-antitrust-reform-bill).

brands (Dube 2022).[6] Like most house brands, Amazon's products offer high quality for lower prices—often much lower prices—than the major independent brands, saving consumers money and enhancing competition, so their promotion, even if present, need not be detrimental to consumer interests.

Does Amazon unduly favor its own products in search returns? Amazon presumably ranks products to maximize the profitability of its long-term relationships with its customers. Providing poor guidance, even toward its own products, is unlikely to serve that purpose.[7] Amazon has stated that it does not favor its own brands over other brands but instead prioritizes relevance, customer satisfaction, and other consumer concerns.[8] Amazon's products may rank high in search results, however, if its own brands have high sales volumes, low prices, good reviews, and high conversion rates, among other relevant factors.

As part of our own research on the sequential ordering of search returns, we considered how one might detect a preference for own-brands. While we learned a great deal from the effort, the empirical methods to accurately quantify a preference were elusive. The central difficulty is that the order in which products are returned is a ranked variable, and we were unable to find a statistical procedure suitable to such data (though the search continues). Also, omitted variables, which are inevitable, introduced meaningful bias on the coefficients of product attributes, making a *ceteris paribus* analysis implausible. Since the ranked order returns are not a choice-problem, even models suitable to ordered data seem inappropriate.

Related to (though independent of) our own research, a recent study by Farronato, Fradkin, and MacKay (2023) (hereinafter "*FFM Study*") attempts to quantify any preference given to Amazon's house brands using data collected from the search activity of a convenience sample of

---

[6] J. Dubé, *Amazon Private Brands: Self-Preferencing vs Traditional Retailing*, Working Paper (September 11, 2022) (available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4205988).

[7] *Id*. *See also* J. Carlson, *Don't Bite the Hand That Feeds You: Amazon's Self-Preferencing Paradox*, Information Technology & Innovation Foundation (May 2, 2022) (available at: https://itif.org/publications/2022/05/02/dont-bite-hand-feeds-you-amazons-self-preferencing-paradox); F. Scott Morton, F. Etro, O. Latham, and C. Caffarra, *Designing Regulation for Digital Platforms: Why Economists Need to Work on Business Models*, VOXEU (June 4, 2020) (available at: https://cepr.org/voxeu/columns/designing-regulation-digital-platforms-why-economists-need-work-business-models).

[8] Responses to Questions for the Record following the July 16, 2019, Hearing of the Subcommittee on Antitrust, Commercial, and Administrative Law, Committee on the Judiciary, Entitled "Online Platforms and Market Power, Part 2: Innovation and Entrepreneurship," Amazon (October 11, 2019) (available at: https://docs.house.gov/meetings/JU/JU05/20190716/109793/HHRG-116-JU05-20190716-SD038.pdf) ("Amazon's algorithms do not take into account whether [a product is private label sold by Amazon] when ranking shopping results.").

184 volunteers; a sample that includes 228,281 search results on 3,019 unique searches.[9] The study, only a few pages long, appears to be a preliminary reporting of results given the lack of relevant details and the discussion of ongoing data collection. The *FFM Study* does not suggest Amazon is preferencing its products, at least by much. Rarely did Amazon-branded products appear in the set of top 10 listings and, on average, Amazon house-branded products ranked fewer than four positions above the mean rank of 38.9 on a mean number of 76 products per search, a small difference and moderate rank that do not appear to support an inference of explicit preferencing. Also, Amazon branded products are shown to have high customer ratings, low prices, and faster shipping, which the *FFM Study* acknowledges are factors "that organically push them to the top of the page,"[10] and the *FFM Study* admits that Amazon's products may have other favorable attributes unaccounted for in the *FFM Study*. Considering the parsimonious data on product attributes (among other concerns), such a small rank difference and middling average rank is fairly strong evidence of a neutral ranking process in relation to house brands. The authors of the *FFM Study* conclude their findings "do not necessarily imply that consumers are hurt by Amazon brands' position in search results."[11]

While the *FFM Study* will be taken by some critics as demonstrating a mild preference for Amazon brands, the results of the study are questionable for several reasons, including especially the empirical method used. Due to our own research into this issue, we were hopeful the *FFM Study* had devised a suitable empirical method for analyzing ranked data in a non-choice context. Yet, the *FFM Study* relied on Ordinary Least Squares ("OLS"), an approach we have determined to be a poor choice due to its severely biased coefficients and incorrect standard errors.[12]

As the *FFM Study* may have some policy relevance, and other researchers may seek to study this same issue, in this BULLETIN we use a Monte Carlo Simulation to show the regression model

---

[9] C. Farronato, A. Fradkin, and A. MacKay, *Self-Preferencing at Amazon: Evidence from Search Rankings*, 113 AEA PAPERS AND PROCEEDINGS 239-243 (2023) (available at: https://www.aeaweb.org/articles?id=10.1257/pandp.20231068). A convenience sample is based on individuals that are ready and able to participate in a survey and often consists mostly of volunteers. As such, a convenience sample may not be representative of the population, so the results of any quantitative or qualitative analysis may not be generalizable. While convenience samples are often necessary, Galloway (2005) notes "the enormous disadvantage of convenience sampling that stems from an inability to draw statistically significant conclusions from findings obtained." A. Galloway, *Non-Probability Sampling*, ENCYCLOPEDIA OF SOCIAL MEASUREMENT 859-864 (2005); M. Stommel and C. Wills, CLINICAL RESEARCH CONCEPTS AND PRINCIPLES FOR ADVANCED PRACTICE NURSES (2004) at Ch. 19.

[10] *Id.* at pp. 4, 7 ("[a]fter controlling for many observable characteristics, Amazon brands remain about 30% cheaper and have 68% more reviews than other similar products.").

[11] *Id.* at p. 7.

[12] *See*, *e.g.*, G.G. Judge, W.E. Grifiths, R.C. Hill, H. Lütkepohl, and T. Lee, THE THEORY AND PRACTICE OF ECONOMETRICS (1985) at p. 757.

of the sort used in the *FFM Study* offers biased estimates of the coefficients, including possibly producing coefficients of the wrong sign, and incorrect hypothesis tests. OLS is systematically prone to indicate a preference where none exists; for products with favorable attributes, like many of Amazon's house brands, a result indicating a preference is nearly guaranteed even in the absence of an own-brand bias. The residuals of an OLS regression on rank data are also autocorrelated, which may lead to poor estimates of the standard errors. Omitted variables bias, as usual, is a serious problem.

At present, we are unable to offer any meaningful solutions to these concerns. It is a difficult problem. While there are empirical approaches for ordered choice data in most major statistical packages, such as the Ordered Logit Model ("OLM"), that model is also not satisfactory in this case.[13] The data analyzed in the *FFM Study* are not a "choice" problem (e.g., the choice of transportation modality) and the number of product returns is very large. Absent a full understanding of Amazon's algorithm, or else data on all relevant features and knowledge of the specific functional form of algorithm, the prospects for reliably detecting a preference for Amazon-branded products (or any other preference in search returns) seem grim.

II. A Review of the *FFM Study*

The objective of the *FFM Study* is to quantify what, if any, rank preference Amazon gives to its own house brands. Data in the *FMM Study* include the order of product returns and (short of) a handful of variables that presumably enter Amazon's unobserved ranking algorithm (*e.g*., price, review score, and number of reviews). With these data, the *FMM Study* tests for a ranking preference for Amazon's house brands using the OLS regression,

$$R_{ij} = \delta A_{ij} + \beta S_{ij} + \gamma X'_{ij} + \mu_j + e_{ij}\ , \tag{1}$$

where R_{ij} is the rank for product i in search j, A is a dummy variable for Amazon's house brand, S is a dummy variable for a "sponsored product" (involving a payment for a higher ranked result), the X'_{ij} include variables such as price and buyer reviews, μ_j is a product class (or search class) fixed effect, and e_{ij} is the econometric disturbance term. Standard errors are clustered on the product class j. The δ coefficient is taken to be an estimate of the difference in rank for

[13] The methods developed over time, but the commonly cited source for such models is R.D. McKelvey and W. Zavoina, *A Statistical Model for the Analysis of Ordinal Level Dependent Variables*, 4 JOURNAL OF THE MATHEMATICAL SOCIOLOGY 103-120 (1975). Any modern econometrics textbook covers the topic.

Amazon's house brands, other things constant (at least in models where the X_{ij} are included). (We demonstrate below it is not.)[14]

Ignoring the X_{ij}, the Amazon products are found to rank 6.12 positions higher on an average rank of about 32.7 on an average of 76 product returns per search, a relatively small unconditional difference given the favorable attributes of Amazon's house brands.[15] Sponsored products, in contrast, rank 7.09 positions above the average rank. Including the $X_{ij,}$ and thus accounting for differences in rank based on a few observed product attributes, but they find Amazon's house brands rank only 2.89 to 3.87 positions higher than non-Amazon products, depending on the particular sample used. So, even a few potentially mis-specified covariates leads to a sizable decline in the δ coefficient (about a 40-50% reduction), while the β coefficient on sponsored products is hardly affected, falling from -7.09 to -6.55, which we suspect is due the disconnect between the rank and the product attributes. Conditioned on the X_{ij}, Amazon's house brand averages a rank of about 34.65 on a median rank of 38.5, a small difference that does not indicate that Amazon is engaged in preferencing its own brands in search returns, at least by much. Amazon products rank about average in product search returns, despite their favorable attributes, many of which do not appear in the *FFM Study*'s data.

The variables X_{ij} are few, including, it appears, the price, the product rating, the number of reviews, and some delivery speed indicators. These variables are, at best, a sparse representation of inputs used by Amazon's algorithm. The authors admit as much, stating that their findings may "simply be due to other characteristics that we cannot observe," which are presumably numerous given the complexity and likely nonlinearity of Amazon's algorithm.[16] Public, business-focused discussions on how to receive higher rankings in Amazon's algorithm suggest many other important, yet unobserved, variables including sales history, the conversion rate (*i.e.*, successful delivery), the quality of images, the thoroughness of the product description, the

---

[14] It appears that the authors of the *FFM Study* may understand the problems with their empirical model, noting "to make the size of the coefficient estimate [of δ] interpretable, we compare it to the size of the coefficient for the sponsored dummy, which we expect should increase prominence," implying that the interpretation of the δ coefficient must be based on a comparison to the β coefficient. However, no hypothesis tests of the relationship between the coefficients are conducted, so it is unclear exactly what was meant by this statement.

[15] *FFM Study*, *supra* n. 9 at Table 2.

[16] *Id.* at p. 7. For instance, important factors in Amazon's algorithm include things like text match, availability, selection, sales history, conversion rates (*i.e.,* successful delivery), the quality of images, the thoroughness of the product description, the frequency of responses to questions and comments, and so forth, none of which appear in the FFM Study's data. *See*, *e.g*., L. Baker, *Amazon's Search Engine Ranking Algorithm: What Marketers Need to Know*, SEARCH ENGINE JOURNAL (August 14, 2018) (available at: https://www.searchenginejournal.com/amazon-search-engine-ranking-algorithm-explained/265173); S. Breslin, *21 Ways to Rank Your Products Higher on Amazon*, REPRICER.COM (last visited April 10, 2022) (available at: https://www.repricerexpress.com/rank-your-products-higher-on-amazon).

frequency of responses to questions and comments, and so forth.[17] Even the treatment of price is not simple, since low prices may signal low quality or other issues. Also, products that may be viewed as near-commodities (e.g., a HDMI cable or toilet paper) often come in varying sizes, quantities, qualities, and so forth. There is no accounting for such variety in the *FFM Study*. Recent review trends and other factors that are undetectable in the averages of these variables may also be relevant to Amazon's algorithm. Moreover, the few product attributes included in the *FFM Study* do little to explain the rank.[18] The R^2 of the unconditional model is 0.489, nearly all of which is likely explained by the fixed effects. With the X_{ij} included, the R^2 of the model increases to only 0.516, a difference of only 0.027. Thus, the X_{ij} do very little to explain the dependent variable (and, as shown below, the coefficients on these variables tend to be downward biased). The regression model appears to be a poor representation of Amazon's algorithm, leading to a poor estimate of the counterfactual rank for Amazon-branded and sponsored products. Omitted variables and blindness as to the functional form of the algorithm makes obtaining estimates of a rank preference difficult if not impossible.

Moreover, and perhaps most fatal to the analysis, the statistical approach used in the *FFM Study* (OLS) is ill-suited to the rank dependent variable. Using OLS with rank data presents many serious problems including biased coefficients and invalid inferences. Furthermore, the "treatment" in the *FFM Study* (an Amazon-branded product designation) is not randomly assigned, possibly leading to selection bias. Unobserved and relevant attributes, nonlinearity, and selection bias lead to the classic *omitted variables problem* (in addition to the omission of variables in the algorithm) which causes biased and inconsistent estimates of the parameters and incorrect statistical inferences. In the following analysis, we demonstrate those problems and find they are severe in the type of data used in the *FFM Study*. At present, we are unable to offer an alternative estimation strategy that is suited to this sort of data.

III. Monte Carlo Simulation

To explore the properties of the OLS estimator in the context of ranked data and omitted variables, we carry out a limited Monte Carlo Simulation ("MCS").[19] MCS is a stochastic simulation procedure in which a dataset of known properties is created and then studied using empirical methods. MCS is useful for detecting problems with an empirical method since

---

[17] *See*, *e.g.*, S. Breslin, *21 Ways to Rank Your Products Higher on Amazon*, REPRICER.COM (last visited April 10, 2022) (available at: https://www.repricerexpress.com/rank-your-products-higher-on-amazon).

[18] *FFM Study*, *supra* n. 9 at Table 2.

[19] All analysis is done using Stata 17.

shortcomings are exposed when the results of that method are inconsistent with the known properties of the data.

The data generation process ("DGP") is as follows. Let y^*_{ij} be a latent index for product *i* in search *j* calculated by an online retailer based on product attributes. This latent index is not observed by the public, but its rank y_{ij} is evident. Suppose that,

$$y^*_{ij} = \beta_1 + \beta_2 P_{ij} + u_j + e_{ij} = 20 + 10P_{ij} + u_j + e_{ij}. \tag{2}$$

In Expression (1), the β_k are parameters, u_j is an individual heterogeneity term for search group *j* with an assumed $N(0, 25)$ distribution. The random error is $e_{ij} \sim N(0, 100)$. The variable P_{ij} (*e.g.*, price) is assumed distributed as a Gamma random variable with shape and scale parameters $P_{ij} \sim \Gamma(9, 0.5) > 0$. The model in Expression (2) does not include any preference for Amazon products in the latent index. To add such a preference, let A_{ij} denote an indicator variable that simulates the online retailer's branded product. Including this preference in Expression (2) we obtain,

$$y^*_{ij} = \beta_1 + \beta_2 P_{ij} + \delta A_{ij} + u_j + e_{ij} = 20 + 10P_{ij} - 6A_{ij} + u_j + e_{ij}. \tag{3}$$

In our simulation A_{ij} is randomly assigned to be the j^{th} listing in each product class. We generate 1,000 samples with 300 product classes with 50 products in each. A traditional 95% interval estimator will successfully cover the true parameter with probability 0.95. With 1,000 samples a 95% interval estimate of the proportion of Monte Carlo successes is [0.9365, 0.9635], which we deem sufficiently accurate for this experiment.[20] Several models of varying complexities are estimated, though we do not exhaust all relevant and noteworthy cases. This analysis should be considered a preliminary examination of this interesting problem.

A. *Model 1: The Simple Case*

Model 1 is the simplest case wherein the latent DGP is,

$$y^*_{ij} = 20 + 10P_{ij} + e_{ij}, \tag{4}$$

which excludes the heterogeneity term and omitted variables. This latent index is ranked (1 to 50) in each class for the empirical analysis. There are 15,000 observations in each simulated sample. The means of the interesting parameters from the 1,000 simulations are summarized in Table 1.

---

20 *Id.* at 149.

While we report the usual OLS standard errors, they are incorrect for this model because the ranked data are used as the dependent variable. The use of robust or clustered standard errors cannot resolve this problem. Since the standard errors (and thus hypothesis tests) are incorrect, no hypothesis testing is performed. The standard deviation and range of the β_2 parameter are also provided.

Table 1. Estimates from Model 1

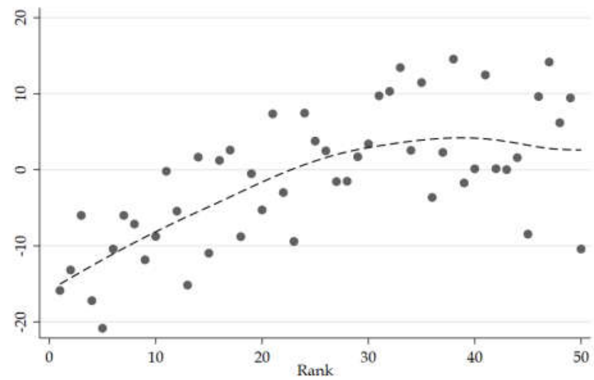
Intercept	-8.3726			
	(0.2319)			
		Stan. Dev.	Min	Max
P	7.5276	0.04838	7.36522	7.68287
	(0.0489)	(0.00052)	(0.04719)	(0.05048)
Fixed Effects	No			
N	15,000			
R^2	0.612			
df	14,998			
AR1 (Pooled)	0.510			
	(0.007)			

Standard errors in parentheses.

The model "fits" the data reasonably well ($R^2 = 0.61$). Recall from Expression (2) that the true value of the β_2 coefficient is 10, but the estimated coefficient is 7.527. Across the 1,000 samples the slope estimates range between 7.365 and 7.683, so that no estimate comes close to the true parameter value. Even in this simple case, the β_2 coefficient is biased and well below its true value. We conclude that the OLS estimator using ranked data is severely downward biased in applications of this sort. Somewhat surprising is that the mean of the nominal standard errors is close to the standard deviation of the estimates of β_2.[21] Unfortunately, the estimate of β_2 is so biased that having a standard error that is informative does not change the fact that interval estimates, or hypothesis tests, will not be correct.

[21] Recall that under standard assumptions the estimator of the coefficient variance is an unbiased estimator of the true estimator variation. The standard error is not unbiased for the estimator's standard deviation, but it is consistent. The consistency holds in the stochastic regressor case as well.

Figure 1. OLS Residuals



In analyzing the findings, a surprising result is obtained. The latent DGP random errors are normally and independently distributed, but the residuals from the ranked data in Model 1 are autocorrelated, which leads to standard errors that are too small and a rejection rate of the null hypothesis that the parameter is zero is too frequent. As is standard, we compute the least squares residuals and estimate the AR(1) model,

$$\hat{e}_{ij} = \rho\hat{e}_{i,j-1} + v_{ij} . \tag{5}$$

The average estimate of the autocorrelation parameter ρ is 0.510 with minimum and maximum values [0.484, 0.534]. Figure 1 illustrates a plot of the residuals over the ranking showing a pattern typical of serially correlated errors. As autocorrelation is not a property of the DGP, we conclude that the serial correlation we observe in the residuals is induced by the ranking of the latent index. Serial correlation is observed in all subsequent models.

B. *Model 2: Omitted Variables Case*

For Model 2, we generate a random variable $Z_{ij} = P_{ij} + \varepsilon_{ij}$, $\varepsilon_{ij} \sim N(0, 25)$ and include it in the latent variable DGP,

$$y^*_{ij} = 20 + 10P_{ij} + \left(5Z_{ij} + e_{ij}\right) , \tag{6}$$

PHOENIX CENTER FOR ADVANCED LEGAL & ECONOMIC PUBLIC POLICY STUDIES
5335 Wisconsin Avenue, NW, Suite 440
Washington, D.C. 20015
Tel: (+1) (202) 274-0235 • Fax: (+1) (202) 318-4909
www.phoenix-center.org

but exclude it from the OLS estimation. Thus, P is endogenous, being correlated with the omitted variable Z.[22] This model is misspecified (an omitted variable). Again, the heterogeneity term is excluded. Results are summarized in Table 2.

Table 2. Estimates from Model 2

Intercept	-0.6809 (0.2967)			
		Stan. Dev.	Min	Max
P	5.8182 (0.0625)	0.04674 (0.00059)	5.66062 (0.06069)	5.96157 (0.06444)
Fixed Effects	No			
N	15,000			
R^2	0.366			
df	14,998			
AR1 (Pooled)	0.655 (0.006)			

Standard errors in parentheses.

The fit declines ($R^2 = 0.366$) because of the additional noise in the random error. The β_2 coefficient is again too small and smaller than in Model 1. The β_2 coefficient is biased downward and no estimator comes close to its true value, having a maximum value of 5.962. The average estimator of the autocorrelation coefficient ρ is 0.655. Due to this bias, including product attributes in an OLS regression on ranked data does not properly adjust for the influence of these attributes.

C. *Model 3*

Model 3 includes both the omitted variable and the heterogeneity term, so the latent DGP is,

$$y_{ij}^* = 20 + 10P_{ij} + \left(5Z_{ij} + u_i + e_{ij}^*\right). \tag{7}$$

The estimates for Model 3 include fixed effect estimation for the individual searches (as in the *FFM Study*). Results are summarized in Table 3. As before, we report the OLS estimates of the misspecified model and corrected standard errors for 300 clusters (which remain invalid), hence the reduction in the reported degrees of freedom (*df*).

---

22 The R^2 of a regression on P_{ij} and Z_{ij} is about 0.08 and the correlation coefficient between them is about 0.28.

Table 3. Estimates from Model 3

Intercept	-1.2113 (0.2225)			
		Stan. Dev.	Min	Max
P	5.9361 (0.0495)	0.04781 (0.00227)	5.77251 (0.04206)	6.08559 (0.05621)
A	…			
Fixed Effects	Yes			
N	15,000			
R^2	0.373			
df	299			
AR1 (Pooled)	0.637 (0.007)			

Standard errors in parentheses.

These results are comparable to those in Model 2, which is unsurprising. The R^2 of the model is 0.373. Again, the β_2 coefficient is well below its true value across all 1,000 simulations with a maximum value of 6.086, and the residuals are autocorrelated (ρ = 0.637).

D. *Model 4: Looking for a Preference*

For Model 4, the A_{ij} variable is included to simulate the preference for a product in each product class. Using Expression (3), the latent variable DGP for Model 4 in the preferencing case is,

$$y_{ij}^* = 20 + 10P_{ij} - 6A_{ij} + u_i + e_{ij}^*, \tag{8}$$

where the A_{ij} indicator was randomly chosen among the top fifteen values as ranked based on the observed values of the latent index and the explicit preference in indicated by the addition of -6A_{ij} in calculating the rank. In this scenario there is no omitted variable. Note that the average rank of the A_{ij} across the simulations is 7.4 without a preference (by design) and 4.5 with an explicit preference, so the mean of the β_3 coefficient across simulations should be equal to approximately -3.0. To determine whether OLS correctly indicates no preference we also *exclude* the preference so that the A_{ij} product ranks "organically" based on its attributes. In this case, the β_3 should be zero as there is no preference.

Table 4. Estimates from Model 4, Preferencing Case

Intercept	-8.2082			
	(0.2213)			
		Stan. Dev.	Min	Max
P	7.5413	0.05687	7.37557	7.70882
	(0.0490)	(0.00265)	(0.04054)	(0.05738)
A	-11.2824	0.78957	-12.85745	-8.68357
	(0.3405)	(0.02064)	(0.27825)	(0.39622)
Fixed Effects	Yes			
N	15,000			
R^2	0.636			
df	299			
AR1 (Pooled)	0.457			
	(0.007)			

Standard errors in parentheses.

We report the OLS estimates for the case where the A_{ij} receives a preference in Table 4, including fixed effects. Note that the regression using of the ranked dependent variable produces an estimate on *P* with the correct sign, though still too small is magnitude (7.54 versus 10) with a maximum value of 7.71. More importantly, the coefficient on the A_{ij} variable (-11.2824), while of correct sign, is far from its true value (-3.0 positions) being severely downward biased. OLS fails to reasonably approximate the rank preference, for the reasons explained below.

Next, consider the case where there is no preference and the A_{ij} product ranks organically given its attributes (which are favorable). The β_3 coefficient should be zero since there is no preferencing. Results are summarized in Table 5. These results are a potent indictment against using OLS with rank data. The β_3 coefficient averages -8.23 across the simulations while it should be zero. The maximum value of β_3 is -4.49, so in no case does the coefficient come near its true value. From Tables 3 and 4 we see that OLS regression on ranked data cannot measure a preference nor accurately indicate a lack of a preference. We conclude that even ignoring omitted variables, the results of the *FFM Study* is highly unlikely to produce valid conclusions.

Table 5. Estimates from Model 4, No Preferencing

		Stan. Dev.	Min	Max
Intercept	-8.4298 (0.2227)			
P	7.5769 (0.0493)	0.06527 (0.00266)	7.39418 (0.04065)	7.76032 (0.05739)
A	-8.2251 (0.3365)	1.64013 (0.01853)	-10.94421 (0.27825)	-4.49180 (0.39066)
Fixed Effects	Yes			
N	15,000			
R^2	0.631			
df	299			
AR1 (Pooled)	0.460 (0.007)			

Standard errors in parentheses.

Further analysis reveals an interesting property of OLS with ranked outcomes. In the prior examples, the A_{ij} product was assumed to have favorable attributes and was chosen randomly from ranks 1 to 15. Instead, we could choose the rank from the bottom rankings (35 to 50) or from the middle ranks (17 to 32). The sample mean changes in rank in these cases due to the assigned preference are -4.2 and -6.4, respectively. Table 6 summarizes the results.

Table 6. Estimates from Model 4, Alternate Rank for A_{ij}

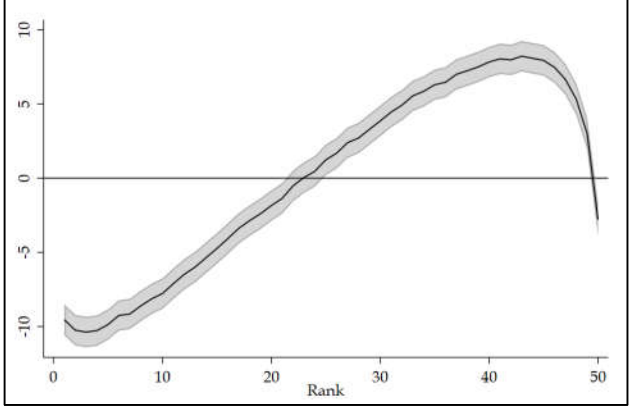
	A_{ij} from rank 35-50		**A_{ij} from rank 17-32**	
	Preference	**No Pref.**	**Preference**	**No Pref.**
Intercept	-8.9228 (0.2229)	-8.8616 (0.2211)	-8.7968 (0.2227)	-9.0485 (0.2231)
P	7.6390 (0.0498)	7.6060 (0.0493)	7.6491 (0.0495)	7.6758 (0.0496)
A	2.4535 (0.4233)	6.8185 (0.4200)	-6.1155 (0.3887)	0.4606 (0.3751)
Fixed Effects	Yes	Yes	Yes	Yes
N	15,000	15,000	15,000	15,000
R^2	0.624	0.629	0.627	0.625
df	299	299	299	299
AR1 (Pooled)	0.470 (0.007)	0.464 (0.007)	0.470 (0.007)	0.471 (0.007)

Standard errors in parentheses.

Consider first the case where the A_{ij} product is chosen from the bottom portion of rankings. The β_3 coefficient is 2.45 in the preference case and 6.82 in the non-preference case. Neither coefficient is anywhere near the sample mean change in rank (-4.2 and -6.4), and, despite the preference, the coefficient in the preferencing case is positive. The coefficient on A_{ij} has the wrong sign. In the non-preference case, the coefficient is 6.82 when it should be zero. The difference in the coefficients between the preference and non-preference case approximates the true difference (-4.37), but both coefficients cannot be observed.

When A_{ij} is selected from the middle part of the ranking, the coefficient for the preference is -6.12, which is very close to the sample mean increase in rank from the preference (-6.4). Also, the coefficient in the non-preference case (0.625) is close to zero with a large relative standard error (which is inaccurate in any case). The coefficients on A_{ij}, therefore, depend on the organic rank of the product based on its attributes. If the attributes are favorable and there is no preference, then OLS finds a preference nonetheless.

Figure 2. Dummy Coefficient by Rank



This bias in the A_{ij} coefficient depending on its organic rank may be illustrated by cycling through all ranks to assign the A_{ij} product with *no preference* given. We use the simplest model (Model 1). Figure 2 illustrates the results and their 95% confidence interval. This figure reveals that products with favorable attributes will have negative coefficients even in the absence of preferencing, and products with unfavorable attributes will have positive coefficients (if not too unfavorable). OLS performed on the rank order produces a systematically biased coefficient on product-specific indicator A_{ij}. As illustrated in Table 6 and Figure 2, if promoted products have average attributes, then the coefficient on a promoted product may be estimated with some accuracy, but this result is mere coincidence. (Note that the shape of this curve will depend on the distribution of the underlying latent variable behind the rank; in practice, that distribution is unobservable.) However, when the attributes of a product are worse or better than average, the coefficient for any specific product in a search may indicate an upward-or-downward preference even in the total absence of a preference. Also, the coefficient may be wrongly signed, as is the case in Table 6.

E. *Model 4: Including a Promoted Product*

The authors of the *FFM Study* suggest that the difference in the Amazon and promoted product coefficients may say something about preferencing. To test the claim, we consider if a comparison to a promoted product may reveal something about a preference for A_{ij}. To add such a preference, let W_{ij} denote an indicator variable that simulates a preference a single product in each product class. Including this preference in Expression (2) we obtain,

$$\begin{aligned} y_{ij}^* &= \beta_1 + \beta_2 P_{ij} + \delta A_{ij} + \lambda W_{ij} + u_j + e_{ij} \\ &= 20 + 10P_{ij} - 6A_{ij} - 6W_{ij} + u_j + e_{ij} \end{aligned}, \tag{9}$$

where the promoted product W_{ij} receives the same adjustment as A_{ij}. Here A_{ij} is randomly assigned from the first fifteen items while the W_{ij} is randomly assigned from the 17th to the 32nd ranked products (so it has middling product attributes). With a preference, the change in rank will not be the same for the two scenarios: the A_{ij} increases in rank by 3.3 positions while the W_{ij} increases by 6.3 positions. As another alternative, we select the A_{ij} as being adjacent to (one position above or below) the W_{ij}, so both A_{ij} and W_{ij} have middling attributes. The sample mean changes in rank are -6.5 for A_{ij} and -5.9 for W_{ij}.

Table 7. Estimates from Model 4, Incl. Promoted Product

	Product Attributes A_{ij} High, W_{ij} Middling		**Product Attributes A_{ij} Middling, W_{ij} Middling**	
	Preference	**No Pref.**	**Preference**	**No Pref.**
Intercept	-7.9431	-8.1746	-8.5651	-8.8048
	(0.2207)	(0.2221)	(0.2222)	(0.2225)
P	7.5110	7.5480	7.6206	7.6462
	(0.0488)	(0.0491)	(0.0494)	(0.0495)
A	-11.4265	-8.1935	-5.7423	0.8574
	(0.3402)	(0.3358)	(0.3879)	(0.3735)
W	-6.2924	-6.2773	-5.7036	-6.0889
	(0.3815)	(0.3841)	(0.3873)	(0.3878)
Fixed Effects	Yes	Yes	Yes	Yes
N	15,000	15,000	15,000	15,000
R^2	0.638	0.634	0.628	0.628
df	299	299	299	299
AR1 (Pooled)	0.451	0.456	0.470	0.467
	(0.007)	(0.007)	(0.007)	(0.007)

Standard errors in parentheses.

5335 Wisconsin Avenue, NW, Suite 440
Washington, D.C. 20015
Tel: (+1) (202) 274-0235 • Fax: (+1) (202) 318-4909
www.phoenix-center.org

Table 7 summarizes the results for cases where A_{ij} receives a preference and when it does not across the two scenarios. While the W_{ij} coefficient is comparable to its change in rank, the A_{ij} coefficient is distant from its true value. This difference is related to the larger bias in the coefficient as the organic rank departs from the mean (see Figure 2). The difference in the coefficients in the preference case (-5.10 positions) is also far from the sample mean change in rank for A_{ij} (-3.3). Even when A_{ij} receives no preference, its coefficient is more negative than that on the promoted product W_{ij}. For detecting a preference—or the absence of one—OLS is not up to the task.

In the second case, where both A_{ij} and W_{ij} are of middling rank and adjacent (A_{ij} is randomly either ±1 position of W_{ij}), the coefficients are of similar size. When the A_{ij} receives no preference, the A_{ij} coefficient is near zero and the difference in the A_{ij} and W_{ij} coefficient is -5.23, which is close to but less than the true difference (-6 positions). While these results may seem encouraging, this is a unique situation that seems practically unrealistic. In all, comparing the W_{ij} and A_{ij} coefficients appears uninformative, absent the rare coincidence that both products are of middling attributes and adjacent in rank, and then their equivalence is unreliably testable given the inaccurate standard errors.

F. *Model 4: Including a Promoted Product and an Omitted Variable*

Thus far, we have ignored omitted variables when including the A_{ij} and W_{ij}. Here we add the omitted variable Z that is correlated with P, but exclude it from the regression. With a preference included in the DGP, the A_{ij} increases in rank by 1.6 positions while the W_{ij} increases by 3.3 positions. As before we also select the A_{ij} as being adjacent to (one position above or below) the W_{ij}, so both A_{ij} and W_{ij} have middling attributes. The sample mean changes in rank are -3.3 for A_{ij} and -2.8 for W_{ij}.

Table 8. Estimates from Model 4, Incl. Promoted Product and an Omitted Variable

	Product Attributes A_{ij} High, W_{ij} Middling		Product Attributes A_{ij} Middling, W_{ij} Middling	
	Preference	**No Pref.**	**Preference**	**No Pref.**
Intercept	-0.1946	-0.2912	-0.9624	-1.0545
	(0.2222)	(0.2227)	(0.2232)	(0.2231)
P	5.7902	5.8042	5.9095	5.9170
	(0.0490)	(0.0491)	(0.0494)	(0.0494)
A	-13.9853	-12.3236	-3.3171	0.0499
	(0.3466)	(0.3420)	(0.3989)	(0.3893)
W	-4.0125	-3.9859	-3.2633	-3.7056
	(0.3932)	(0.3942)	(0.4007)	(0.4012)
Fixed Effects	Yes	Yes	Yes	Yes
N	15,000	15,000	15,000	15,000
R^2	0.394	0.390	0.377	0.376
df	299	299	299	299
AR1 (Pooled)	0.612	0.615	0.633	0.632
	(0.006)	(0.006)	(0.006)	(0.006)

Standard errors in parentheses.

Table 8 summarizes the results. With the omitted variable, the coefficients on A_{ij} and W_{ij} when chosen from different parts of the rank distribution are very different despite receiving the same preference in the latent index. Again, this difference is attributable to the organic rank of the products being from different parts of the rank order (see Figure 2). Between the preference and no-preference cases, the A_{ij} coefficient changes little despite the absence of a preference. The OLS model cannot detect a preference. When both A_{ij} and W_{ij} are chosen from the middling ranks, the coefficients are closer to their sample means, but again this is a special and unrealistic case, and it depends on the distribution of the latent index. Also, the *FFM Study* notes the favorable attributes of the Amazon-branded products, so it is reasonable to suppose, absent more information, that the coefficient on their Amazon variable is downward biased.

IV. Conclusion

Analyzing the order in which products are presented to customers by online retailers (or the returns presented by search engines) is likely to receive considerable attention in the future as the government scrutiny of large, online vendors intensifies. It is an interesting problem and one we have studied in recent months. In the *FFM Study,* the authors analyze data from a convenience sample on the order in which products are returned by Amazon to detect a preference, if any, for Amazon's house brands. This study finds a small increase in rank and a middling average rank for Amazon brands that do not support a conclusion of any meaningful systematic preference for Amazon's own products.

As we discovered early in our research and detail here, however, the results of the *FFM Study* are obtained using a methodology that is very ill-suited to the task. The application of Ordinary Least Squares regression on the rank of product returns renders severely biased results, possibly of the wrong sign, and invalid hypothesis tests. If Amazon's house brands have favorable attributes, as the *FFM Study* reports, then the OLS regression is very likely to indicate a preference even when none exists. While the *FFM Study* considers an interesting problem and represents significant effort in data collection, it seems improbable that these data can offer an even roughly accurate indication of preferencing. Our own investigation of the problem suggests that a meaningful solution to this empirical quagmire is elusive. Without observing all the variables that enter Amazon's algorithm and its functional form, the ability to detect a preference in Amazon's product returns seems, for now, hopeless.