

Online Appendix

A.1 Estimating Sales

We assume that sales ranks follow a Pareto distribution so that the natural logarithm of sales for a product is a linear function of the natural logarithm of the product’s sales rank minus one.

$$\ln(\text{sales}_{cim}) = a_{cm} + b \ln(\mathbf{x}_{cim}) + u_{cim} \quad (1)$$

We estimate Equation (1) with $\mathbf{x}_{cim} = (\text{rank}_{cim} - 1)$ for each category and market, provided there are at least three distinct sales observations and excluding products with a sales rank of one. After excluding category-market pairs with a poor model fit (i.e., $R^2 \leq 0.1$), the mean R^2 across 354 regressions is 0.42. To improve our prediction, we include the number of reviews as a covariate and re-estimate Equation (1) using $\mathbf{x}_{cim} = (\text{rank}_{cim} - 1, \text{review count}_{im})$. Since reviews accumulate from past sales, they are predictive of sales. As a result, the mean R^2 (across 466 category-country pairs) improves to 0.52. We use these predicted sales values for our analysis. Figure A.1 provides a graphical illustration of how the estimates fit the data, using a specific category-market example.

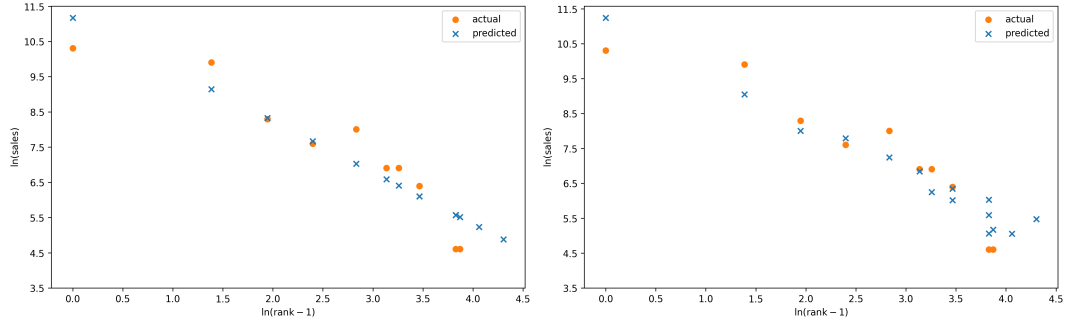


Figure A.1: Model fit for category 'Rotating Power Toothbrushes' on Amazon.com using $\ln(\text{rank} - 1)$ (left) and $\ln(\text{rank} - 1)$ as well as $\ln(\text{review count})$ (right) as predictors

A.2 Additional Tables and Figures

Category	Product Count (All)	Product Count (AB)
Automotive	619	46
Baking & Cookie Sheets	264	34
Bedding Comforter Sets	341	66
Cell Phone Screen Protectors	477	14
Clothing	375	42
Electronics	929	94
Extension Cords	320	44
Floor Lamps	306	4
LED Bulbs	253	54
Mixing Bowls	268	16
Mouse Pads	276	24
Office Products	2236	460
Patio	918	56
Pet Supplies	545	108
Self-Stick Note Pads	241	60
Sports & Outdoors	1306	158
Throw Pillow Covers	278	18
Tools & Home Improvement	2530	222
USB Flash Drives	248	28
Video Games	256	4

Table A.1: (AB) product counts for the top 20 largest categories

Dependent Variables: Model:	In Box		Has Box	
	(1)	(2)	(3)	(4)
<i>Variables</i>				
Amazon Basics	0.70*** (0.02)	0.70*** (0.02)	-0.20*** (0.04)	-0.28*** (0.03)
Amazon's Choice	0.06*** (0.005)	0.06*** (0.006)	-0.03*** (0.01)	-0.03*** (0.008)
1P	-0.04*** (0.007)	-0.04*** (0.007)	-0.08*** (0.02)	-0.07*** (0.01)
log(price)	-0.008*** (0.003)	-0.008*** (0.003)	-0.010 (0.01)	0.009 (0.01)
log(review count)	0.06*** (0.02)	0.05** (0.02)	-0.05 (0.05)	-0.05 (0.04)
rating	0.06** (0.02)	0.05** (0.02)	0.02 (0.05)	0.008 (0.04)
rating \times log(review count)	-0.01** (0.005)	-0.010** (0.005)	0.01 (0.01)	0.01 (0.009)
Constant	-0.24** (0.10)		0.51** (0.24)	
<i>Fixed-effects</i>				
category	No	Yes	No	Yes
<i>Fit statistics</i>				
Observations	28,810	28,810	28,810	28,810
R ²	0.287	0.313	0.018	0.144
Within R ²		0.277		0.025

Clustered (category) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table A.2: Determinants of box assignment (US data only)

Dependent Variables:	log(rank)		log(sales)	
		<i>non-AB only</i>		
Model:	(1)	(2)	(3)	(4)
<i>Variables</i>				
In Box (US)	-1.3*** (0.23)	-0.43*** (0.07)	0.56*** (0.06)	0.21*** (0.05)
Substitutes	-0.95*** (0.24)	-0.38*** (0.08)	0.31*** (0.05)	0.14*** (0.02)
CA	-0.21 (0.69)		-0.91*** (0.04)	
In Box (US) \times CA	0.68** (0.33)	0.39*** (0.12)	-0.27*** (0.07)	-0.25*** (0.06)
Substitutes \times CA	0.20 (0.34)	0.05 (0.11)	-0.12* (0.07)	-0.10*** (0.04)
Amazon's Choice		-0.84*** (0.04)		0.25*** (0.02)
1P		0.15*** (0.04)		-0.04*** (0.01)
log(price)		0.16*** (0.03)		-0.02*** (0.006)
rating		0.09 (0.14)		-0.15* (0.07)
log(review count)		-0.18 (0.12)		0.09 (0.06)
rating \times log(review count)		-0.04 (0.03)		0.05*** (0.01)
Constant	6.7*** (0.50)		5.4*** (0.03)	
<i>Fixed-effects</i>				
category-domain	No	Yes	No	Yes
<i>Fit statistics</i>				
Observations	54,644	54,644	44,982	44,982
R ²	0.011	0.702	0.160	0.587
Within R ²		0.219		0.421
<i>Clustered (category-domain) standard-errors in parentheses</i>				
<i>Signif. Codes: ***: 0.01, **: 0.05, *: 0.1</i>				

Table A.3: Effect of self-preferencing where non-exposed products are split into more and less substitutable to exposed products

Dependent Variables:	log(rank)		log(sales)	
		<i>non-AB only</i>		
Model:	(1)	(2)	(3)	(4)
<i>Variables</i>				
Have Box (US)	0.12 (0.36)	-0.04 (0.10)	-0.08 (0.07)	-0.001 (0.03)
CA	-0.12 (0.69)		-0.95*** (0.04)	
Have Box (US) \times CA	-0.43 (0.50)	-0.55*** (0.13)	0.11 (0.11)	0.17*** (0.05)
Amazon's Choice		-0.87*** (0.04)		0.27*** (0.02)
1P		0.16*** (0.04)		
log(price)		0.16*** (0.02)		-0.02*** (0.007)
rating		0.08 (0.14)		-0.14* (0.08)
log(review count)		-0.18 (0.12)		0.10 (0.06)
rating \times log(review count)		-0.04 (0.03)		0.05*** (0.01)
Constant	6.5*** (0.50)		5.5*** (0.03)	
<i>Fixed-effects</i>				
category-domain	No	Yes	No	Yes
<i>Fit statistics</i>				
Observations	54,644	54,644	44,982	44,982
R ²	0.0007	0.701	0.148	0.586
Within R ²		0.217		0.419
<i>Clustered (category-domain) standard-errors in parentheses</i>				
<i>Signif. Codes: ***: 0.01, **: 0.05, *: 0.1</i>				

Table A.4: Effect of self-preferencing on the sales of non-AB products having the box in the US