

e - c o m p a n i o n

ONLY AVAILABLE IN ELECTRONIC FORM

Electronic Companion—"Goodbye Pareto Principle, Hello Long Tail:
The Effect of Search Costs on the Concentration of Product Sales"
by Erik Brynjolfsson, Yu (Jeffrey) Hu, and Duncan Simester,
Management Science, DOI 10.1287/mnsc.1110.1371.

Online Appendix

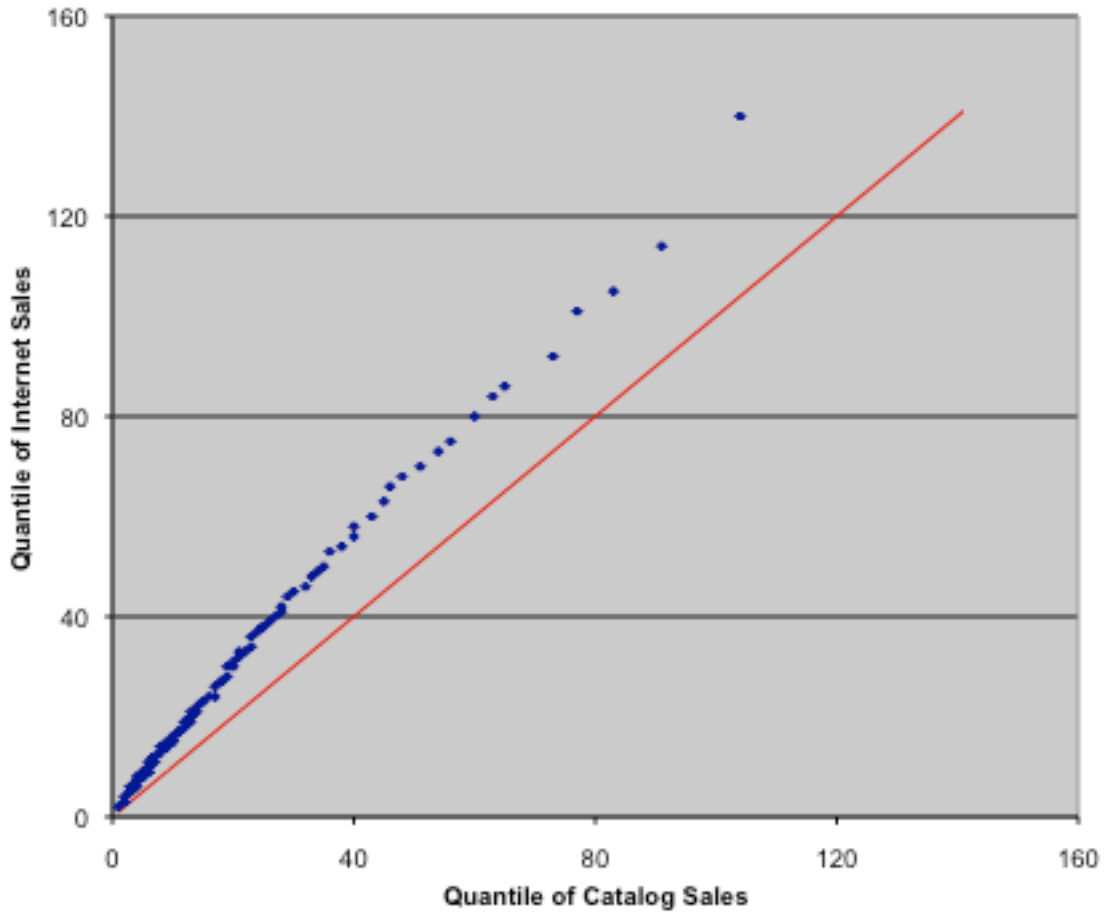
Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales

This version: January 2011

Erik Brynjolfsson*, Yu (Jeffrey) Hu **, Duncan Simester***

* MIT Sloan School of Management; email: erikb@mit.edu; web: <http://digital.mit.edu/erik>
** Purdue University, Krannert School of Management; email: yuhu@purdue.edu
*** MIT Sloan School of Management; email: simester@mit.edu

Figure A1: Quantile-Quantile Plot of Internet Sales and Catalog Sales



This figure plots the quantile of Internet sales of each product on the Y-axis and the quantile of catalog sales of each product on the X-axis. The solid red line is a 45 degree line. The plot starts as an almost straight line and then curves downward.

Table A1: Correlation Matrix

	Unit Sales of Bottom 50% Products	Total Unit Sales	(Usage of) Directed Searches	(Usage of) Non-directed Searches	(Usage of) Recommendation System	Population Density	Median Household Income	Percentage with Bachelor's Degree	Percentage of Female	Median Age
Unit Sales of Bottom 50% Products	1.00									
Total Unit Sales	0.48	1.00								
(Usage of) Directed Searches	0.06	0.15	1.00							
(Usage of) Non-directed Searches	0.02	0.02	0.03	1.00						
(Usage of) Recommendation System	0.05	0.04	-0.09	0.01	1.00					
Population Density	0.02	0.00	0.00	0.00	-0.00	1.00				
Median Household Income	0.03	0.02	0.02	0.03	0.03	0.02	1.00			
Percentage with Bachelor's Degree	0.03	0.03	0.02	0.02	0.03	0.17	0.77	1.00		
Percentage of Female	0.01	-0.02	0.03	0.03	0.02	0.13	0.62	0.48	1.00	
Median Age	0.02	-0.01	0.03	0.02	0.02	0.07	0.59	0.50	0.90	1.00

The table presents Pearson correlation coefficients. The sample size is 11,648.

Table A2: Sales of Niche Products and Consumers' Use of Search and Recommendation Tools
Using % of Dollar and Unit Sales

	% of Unit Sales	% of Dollar Sales
Directed Search	-0.125 (0.078)	-0.108 (0.080)
Non-directed Search	0.414** (0.165)	0.409** (0.168)
Recommendation System	0.299*** (0.070)	0.248*** (0.072)
Population Density	0.001 (0.002)	0.001 (0.002)
Median Household Income	0.002 (0.012)	0.002 (0.012)
% with Bachelor's Degree	0.022 (0.077)	0.032 (0.078)
Percent Female	0.186 (0.192)	0.136 (0.195)
Median Age	-0.016 (0.024)	-0.012 (0.024)
No Demographic Information	0.017 (0.159)	0.022 (0.162)
Total Unit Sales	0.009*** (0.002)	0.010*** (0.002)
Intercept	0.084 (0.159)	0.083 (0.162)
R ² (or Pseudo R ²)	0.006	0.006
Sample Size	11,648	11,648

The table reports the coefficients when estimating Equation 3 using % of unit and dollar sales of products through the Internet channel. Standard errors are in parentheses; ***Significantly different from zero, $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

Table A3: Matching Using Historical Transactions

	Catalog Sample	Internet Sample	Matched Internet Sample
Matching Variables			
Recency	3.147	2.668*** (0.000)	3.131 (0.715)
Frequency	2.198	1.510*** (0.000)	2.199 (0.985)
Monetary Value	2.186	1.948*** (0.000)	2.187 (0.965)
No RFM Information	0.334	0.401*** (0.000)	0.334 (0.976)
Historical Niche Percentage	0.098	0.115*** (0.000)	0.101 (0.218)
No Historical Niche Information	0.376	0.440*** (0.000)	0.383 (0.380)
Percentage of Total Sales Generated by Each Sample of Niche Products			
Unit Sales			
Bottom 40% (294 products)	8.2%	10.0%*** (0.000)	9.5%*** (0.000)
Bottom 50% (367 products)	12.7%	15.2%*** (0.000)	14.7%*** (0.000)
Bottom 60% (440 products)	18.9%	21.7%*** (0.000)	21.0%*** (0.000)
Dollar Sales			
Bottom 40% (294 products)	8.2%	10.3%*** (0.000)	10.3%*** (0.000)
Bottom 50% (367 products)	12.7%	15.4%*** (0.000)	15.3%*** (0.000)
Bottom 60% (440 products)	18.9%	22.0%*** (0.000)	21.7%*** (0.000)
Sample Size	6,905	12,081	6,905

The numbers in parentheses are p-values, measuring the probability that the difference between the Internet and Catalog sample averages will be larger than the observed difference, under the null hypothesis that the true averages are identical. *** p < 0.01; ** p < 0.05; * p < 0.10.

**Table A4: Pareto Curve Estimates
September Catalog**

	Model 1: Internet Data	Model 2: Catalog Data	Model 3: Pooled Data, Linear Regression	Model 4: Pooled Data, Quantile Regression
Constant	8.415*** (0.109)	8.778*** (0.108)	8.778*** (0.109)	9.527*** (0.129)
Sales Rank	-0.963*** (0.021)	-1.019*** (0.021)	-1.019*** (0.021)	-1.134*** (0.025)
Internet			-0.363** (0.154)	-0.923*** (0.183)
Internet * Sales Rank			0.056* (0.029)	0.167*** (0.035)
R ²	0.828	0.846	0.838	0.641
Sample Size	438	439	877	877

Models 1 and 2 present the coefficients from Equation 2 estimated using OLS. Model 1 uses sales in the Internet channel and Model 2 uses sales in the catalog channel. Model 3 presents the OLS coefficients when estimating Equation 3 using the pooled data from Models 1 and 2. Model 4 re-estimates Model 3 using Quantile Regression rather than OLS. Standard errors are in parentheses; ***Significantly different from zero, $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

**Table A5: Results Using the Matched Samples
September Catalog**

	Catalog Sample	Internet Sample	Matched Internet Sample
Matching Variables			
Population Density (00s)	28.0	25.9 (0.136)	28.0 (0.977)
Median Household Income (\$000s)	45.7	47.2*** (0.001)	45.7 (0.979)
Percent with Bachelor's Degree	10.3%	11.0%*** (0.000)	10.2% (0.797)
Percent Female	46.9%	46.3%** (0.048)	46.9% (0.859)
Median Age	33.8	33.3** (0.018)	33.8 (0.806)
No Demographics Information	8.2%	9.1% (0.164)	8.3% (0.934)
Percentage of Total Sales Generated by Each Sample of Niche Products			
Unit Sales			
Bottom 40% (294 products)	5.4%	8.6%*** (0.000)	9.1%*** (0.000)
Bottom 50% (367 products)	9.6%	13.7%*** (0.000)	14.5%*** (0.000)
Bottom 60% (440 products)	16.4%	20.1%*** (0.000)	21.2%*** (0.000)
Dollar Sales			
Bottom 40% (294 products)	5.7%	8.9%*** (0.000)	9.4%*** (0.000)
Bottom 50% (367 products)	10.2%	14.2%*** (0.000)	15.1%*** (0.000)
Bottom 60% (440 products)	17.4%	20.9%*** (0.000)	22.0%*** (0.000)
Sample Size	2,749	11,308	2,749

The numbers in parentheses are p-values, measuring the probability that the difference between the Internet and Catalog sample averages will be larger than the observed difference, under the null hypothesis that the true averages are identical. *** p < 0.01; ** p < 0.05; * p < 0.10.

**Table A6: Matching Using Historical Transactions
September Catalog**

	Catalog Sample	Internet Sample	Matched Internet Sample
Matching Variables			
Recency	3.238	2.905*** (0.000)	3.229 (0.877)
Frequency	2.584	2.134*** (0.000)	2.572 (0.816)
Monetary Value	2.503	2.068*** (0.000)	2.499 (0.916)
No RFM Information	0.245	0.369*** (0.000)	0.246 (0.943)
Percentage of Total Sales From Each Sample of Niche Products			
Unit Sales			
Bottom 40% (294 products)	5.4%	8.6%*** (0.000)	7.6%*** (0.000)
Bottom 50% (367 products)	9.6%	13.7%*** (0.000)	11.9%*** (0.001)
Bottom 60% (440 products)	16.4%	20.1%*** (0.000)	25.4%*** (0.000)
Dollar Sales			
Bottom 40% (294 products)	5.7%	8.9%*** (0.000)	7.9%*** (0.000)
Bottom 50% (367 products)	10.2%	14.2%*** (0.000)	12.3%*** (0.004)
Bottom 60% (440 products)	17.4%	20.9%*** (0.000)	26.0%*** (0.000)
Sample Size	2,749	11,308	2,749

The numbers in parentheses are p-values, measuring the probability that the difference between the Internet and Catalog sample averages will be larger than the observed difference, under the null hypothesis that the true averages are identical. *** p < 0.01; ** p < 0.05; * p < 0.10.

**Table A7: Sales of Niche Products and Consumers' Use of Search and Recommendation Tools
September Catalog**

	Bottom 50%
Directed Search	0.243 (0.435)
Non-directed Search	1.664 (1.382)
Recommendation System	1.376*** (0.428)
Population Density	0.016 (0.012)
Median Household Income	0.093 (0.071)
Percent with Bachelor's Degree	0.363 (0.467)
Percent Female	-0.618 (1.293)
Median Age	0.138 (0.158)
No Demographic Information	1.362 (1.001)
Total Unit Sales	1.238*** (0.024)
Intercept	-3.539** (1.000)
Pseudo R ²	0.138
Sample Size	11,131

The table reports the coefficients when estimating Equation 3 using unit sales of products through the Internet channel. Standard errors are in parentheses; ***Significantly different from zero, $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.