The Effect of Bestseller Rank on Demand

Evidence from the App Market

Octavian Carare

University of Maryland

IDEI Toulouse - January 13, 2011

Outline

Outline

Research Objectives

- Measure the effect of bestseller rank information on demand
- Use rank data
- Ensure robustness

- Experience goods: qualities are ex-ante uncertain
- Best selling products are conspicuously displayed (itunes, Amazon, etc.)
- Bestseller rank may be used to predict quality: affects willingness to pay
- Bestselling products are more visible: they are more likely to be chosen by consumers
- How does tomorrow's willingness to pay for a product change with today's bestseller rank of that product?

Research Objectives

- Measure the effect of bestseller rank information on demand
- Use rank data
- Ensure robustness

- Practical difficulty: only daily bestseller rank is known (not the actual market shares)
- Can bestseller ranks be used to estimate a discrete-choice model of demand?

Research Objectives

- Measure the effect of bestseller rank information on demand
- Use rank data
- Ensure robustness

- I estimate today's demand system with yesterday's bestseller rank as a RHS variable
- Results may not reflect a causal effect of rank information on demand (e.g., autocorrelated demand shocks, or missing variables like product ratings)
- How to show that the relationship between ranks and demands is not spurious?

Outline

Literature

Non-additivity

Demand is not "additive": fads, fashions, etc.

Research

Morgenstern (1948), Liebenstein (1950)

Observational learning

Learning from the actions of others

Research

Cai, Chen, Fang (*AER* 2009)
Salganik, Dodd, Watts (*Science* 2006)

Bestseller lists

Summarize past purchases, but may also convey quality information

Research

Sorensen (*JIndEc* 2007): making the *NY Times* bestseller list causes an increase in sales

Literature

Non-additivity

Demand is not "additive": fads, fashions, etc.

Research

Morgenstern (1948), Liebenstein (1950)

Observational learning

Learning from the actions of others

Research

Cai, Chen, Fang (*AER* 2009)
Salganik, Dodd, Watts (*Science* 2006)

Bestseller lists

Summarize past purchases, but may also convey quality information

Research

Sorensen (*JIndEc* 2007): making the *NY Times* bestseller list causes an increase in sales

Literature

Non-additivity

Demand is not "additive": fads, fashions, etc.

Research

Morgenstern (1948), Liebenstein (1950)

Observational learning

Learning from the actions of others

Research

Cai, Chen, Fang (*AER* 2009)
Salganik, Dodd, Watts (*Science* 2006)

Bestseller lists

Summarize past purchases, but may also convey quality information

Research

Sorensen (*JIndEc* 2007): making the *NY Times* bestseller list causes an increase in sales

Outline

Data

- App Store
- Paid apps data

- Went online in July 2008
- Sold billions of apps for the iphone and the ipod touch platforms (more recently, also for the iPad).
- Yesterday's bestsellers are conspicuously displayed on the download interfaces
- Daily data collected between January 1 and June 16 2009

Data

- App Store
- Paid apps data

- Data contain about 14,000 rank-app-day tuples from the top 100 paid apps by unit sales
- Price is observed in the data; price changes are commonplace

Summary Statistics

Variable	Description	Count	Mean	Stdev	Median
ID	Unique app ID number	452			
Rank	App rank on top 100 list	13,996	47.86	28.35	47
L(Rank)	Lagged rank	13,996	46.76	27.50	46
L_2(Rank)	Second lag	13,996	46.07	26.96	45
L_3(Rank)	Third lag	13,996	46.09	27.06	45
Price	Price in US Dollars	13,996	2.72	2.25	1.99
Age	Days since first release	13,996	109.40	83.15	85
VAge	Age of current version	13,996	92.63	75.85	66
Size	App size (mega bytes)	13,996	17.20	29.11	7.8
NewVer	New Version Dummy	13,996	0.0162	0.1263	0

Table: Summary Statistics

Sales-Rank Correspondence for Joel Comm

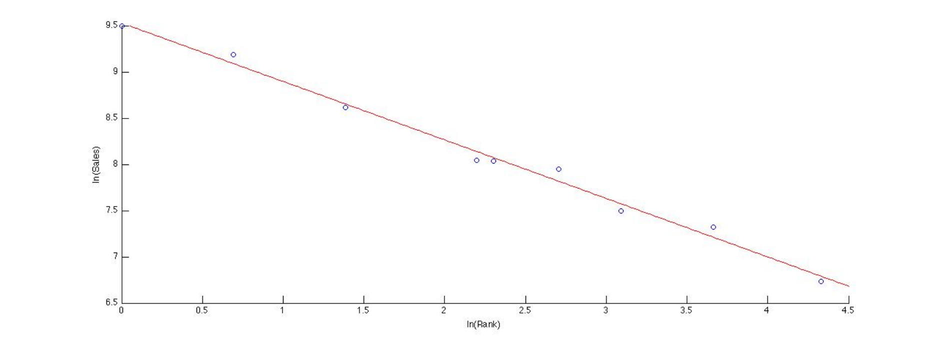


Figure: Sales-Rank Relationship for Available Sales Data

Outline

Challenges and Strategies

Challenge

Bestseller rank data: intertemporal volume comparisons are not possible. Today's rank 20 app may have more sales than tomorrow's rank 10 app.

Strategy

- Assume unit sales Pareto distributed (need not be iid)
- Observe that log-differences of successive order statistics of Pareto RVs, scaled using rank, are iid exponential

Challenge

Data issues: autocorrelation, missing variables

Strategy

- Instrumental Variables estimation
- Are there any good instruments for previous-day ranks?

Challenges and Strategies

Challenge

Bestseller rank data: intertemporal volume comparisons are not possible. Today's rank 20 app may have more sales than tomorrow's rank 10 app.

Strategy

- Assume unit sales Pareto distributed (need not be iid)
- Observe that log-differences of successive order statistics of Pareto RVs, scaled using rank, are iid exponential

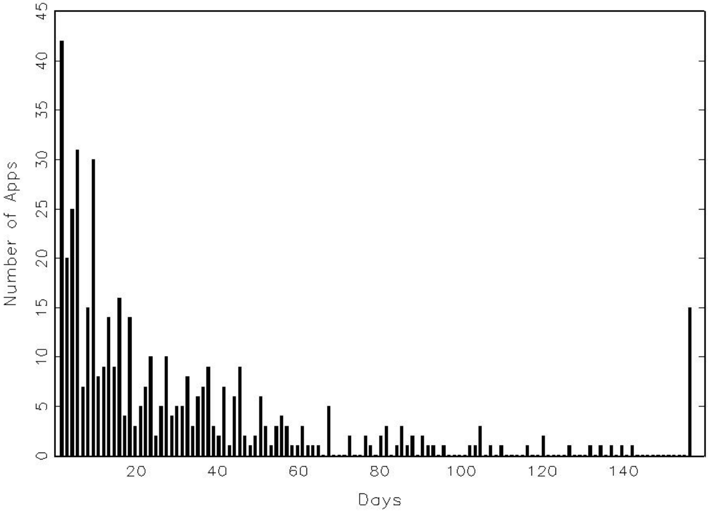
Challenge

Data issues: autocorrelation, missing variables

Strategy

- Instrumental Variables estimation
- Are there any good instruments for previous-day ranks?

App survival on the top 100 list



Outline

Building an Estimating Equation

- Mean Utility
- Random Utility
- Berry Inversion
- Differencing
- Simulation

Specification

Mean utility of app j at time t:

$$\delta_{jt} = X_{jt}\beta - \alpha\left(P_{jt} - M_{r(j,t-1)}\right)$$

P is price, $r(j, t-1)$ is bestseller rank of app j at $t-1$, M is rank-specific value (in \$). X_{jt} are characteristics of app j at t like age and version age.

Building an Estimating Equation

- Mean Utility
- Random Utility
- Berry Inversion
- Differencing
- Simulation

Specification

$$u_{ijt} = \delta_{jt} + \epsilon_j + \varepsilon_{jt} + \xi_{ijt}$$

ξ_{ijt} is distributed iid Type-I EV. App fixed effects soak up unobserved product quality that may be correlated with price.

Building an Estimating Equation

- Mean Utility
- Random Utility
- Berry Inversion
- Differencing
- Simulation

Specification

Relationship between market shares and mean utilities for product with market share rank k (Berry, 1994):

$$\ln s_{kt} - \ln s_{0t} = X_{kt}\beta - \alpha \left(P_{kt} - M_{r(k,t-1)} \right) + \epsilon_k + \varepsilon_{kt}$$

Building an Estimating Equation

- Mean Utility
- Random Utility
- Berry Inversion
- Differencing
- Simulation

Specification

$$\ln s_{kt} - \ln s_{k+1,t} =$$

$$\hat{X}_{kt}\beta - \alpha\left(\hat{P}_{kt} - M_{r(k,t-1)} + M_{r(k+1,t-1)}\right)$$

$$+\hat{\epsilon}_k + (\varepsilon_{kt} - \varepsilon_{k+1t})$$

If unit sales are Pareto distributed, $\ln s_{kt} - \ln s_{k+1,t}$ is iid exponential with constant (but unknown) mean. Mean is invariant: not a function of the parameters to be estimated.

Building an Estimating Equation

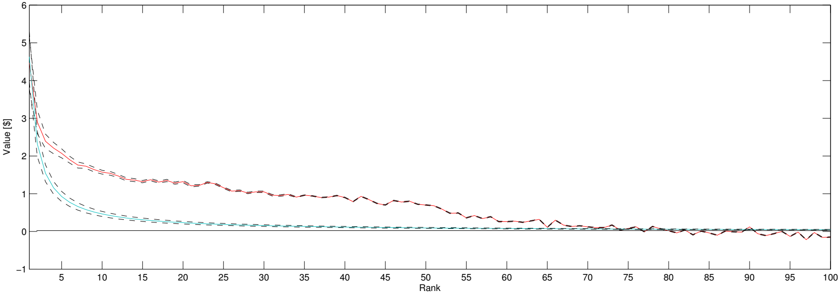
- Mean Utility
- Random Utility
- Berry Inversion
- Differencing
- Simulation

Specification

Generate iid exponential RVs to replace the unobserved LHS of the previous equation. Estimate parameters of interest using OLS for each draw of the LHS. Replicate the process many times.

Outline

Estimation Results



WHAT DO THE RESULTS TELL?

DROPS

Estimates of M appear to drop around ranks 25 and 50

POTENTIAL EXPLANATION

Ranks 25 and 50 are the last displayed on the two pages of the mobile App Store interface

LARGE Ms

Estimated Ms for the top 5 apps are large relative to price; decline steeply

INTERPRETATION

- Popularity begets popularity
- Effect of popularity information much larger for top 25 apps than for other apps

PARAMETERS

- The Ms are identified relative to M_{101} (unranked app)
- Other parameters identified up to scale

IDENTIFICATION

What do the results tell?

Drops

Estimates of M appear to drop around ranks 25 and 50

Potential Explanation

Ranks 25 and 50 are the last displayed on the two pages of the mobile App Store interface

Large Ms

Estimated Ms for the top 5 apps are large relative to price; decline steeply

Interpretation

- Popularity begets popularity
- Effect of popularity information much larger for top 25 apps than for other apps

Parameters

- The Ms are identified relative to M_{101} (unranked app)
- Other parameters identified up to scale

Identification

What do the results tell?

Drops

Estimates of M appear to drop around ranks 25 and 50

Potential Explanation

Ranks 25 and 50 are the last displayed on the two pages of the mobile App Store interface

Large Ms

Estimated Ms for the top 5 apps are large relative to price; decline steeply

Interpretation

- Popularity begets popularity
- Effect of popularity information much larger for top 25 apps than for other apps

Parameters

- The Ms are identified relative to M_{101} (unranked app)
- Other parameters identified up to scale

Identification

Outline

Robustness

Challenge

Central challenge: results may not reflect a *causal* effect of rank information on demand

Why?

- Demand shocks may be autocorrelated
- Missing variables (e.g., ratings) possibly correlated with ranks

Solution

Instrument past ranks

Implementation

Instrumenting all 100 past rank dummies impractical. Instead, instrument reciprocal of the rank.

Robustness

Challenge

Central challenge: results may not reflect a *causal* effect of rank information on demand

Why?

- Demand shocks may be autocorrelated
- Missing variables (e.g., ratings) possibly correlated with ranks

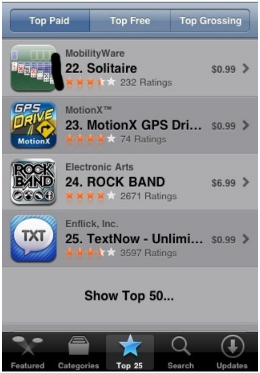
Solution

Instrument past ranks

Implementation

Instrumenting all 100 past rank dummies impractical. Instead, instrument reciprocal of the rank.

Instruments?



Endogenous Price

Even with app fixed effects, price changes may be triggered by changes in bestseller rank. Used higher-order rank lags as instruments for price. IV estimates close to OLS, indicating that autocorrelation does not significantly affect the results.

Behavioral Mechanisms

Observational Learning & Saliency

What do these mean?

- Consumers use bestseller rank to infer app quality: better rank implies better app (*observational learning*).
- Alternatively, they only learn that an app exists (*saliency*).

Mechanisms...

Policy implications

Different when observational learning or when saliency is the dominant mechanism .

Empirically isolating the two mechanisms

Additional regressions

- Observational learning more likely for the paid apps than for free ones. Confirmed in data.
- Transitions from ranks 26-25 and 25-24 inform about saliency & observational learning. Latter appears to dominate.

Outline

Conclusions

Bestseller rank information is an important determinant of demand. On average bestseller rank beats price. Outcomes may be inefficient, as for informational cascades.

Learning about quality or saliency? Analysis indicates that learning is the most likely explanation

Crowd-driven pricing? Why not! Works for music (AmieStreet.com)