

Fall 9-1-2011

ESTIMATING APP DEMAND FROM PUBLICLY AVAILABLE DATA


Rahul Telang

Carnegie Mellon University, rtelang@andrew.cmu.edu

Rajiv Garg

Carnegie Mellon University, rgarg@andrew.cmu.edu

Follow this and additional works at: <http://repository.cmu.edu/heinzworks>

 Part of the [Industrial Organization Commons](#), [Management Information Systems Commons](#),
and the [Management Sciences and Quantitative Methods Commons](#)

This Article is brought to you for free and open access by the Heinz College at Research Showcase @ CMU. It has been accepted for inclusion in Heinz College Research by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

ESTIMATING APP DEMAND FROM PUBLICLY AVAILABLE DATA

Rajiv Garg, Rahul Telang
{rgarg, rtelang}@andrew.cmu.edu

School of Information Systems and Management, Heinz College
Carnegie Mellon University

ABSTRACT

With the abundance in variety of products available online, many online retailers provide sales rankings for available products to make it easier for the consumer to find popular products. Successful implementation of product rankings on online platform was done a decade ago by Amazon and more recently by Apple's App store. However, none of these market providers provide actual downloads data, a very useful statistics for both practitioners and researchers. To address similar issues, researchers in the past developed strategies to estimate sales from product rank. Almost all of that work is based on either doing some experiments to shift sales or partnering with a vendor to get actual sales data. In this research, we present an innovative method to use purely public data to infer this relationship for Apple's iTunes App store. We provide various validations to show our method provides highly accurate estimates on downloads if rank data is available for a given app.

1 INTRODUCTION

The growth of cellular phones and Smartphones over the last few years has been spectacular. Based on recently published reports, there are about 82 million users of smartphones¹ in US with over 100,000 new units being sold every quarter². As more countries deploy high speed wireless networks, users are spending increasing amounts of time on their phones. Based on AdMob's 2010 report³ smartphone users spend about 80 minutes/day using applications downloaded from the app stores that are becoming ubiquitous on almost all mobile operating systems. A significant reason for this growth has been the availability of large number of phone applications – estimates suggest that downloaded applications are used by over 40% of mobile phone users¹.

Apple iPhone ushered an era where developers were able to sell their innovative products (or applications) to a large consumer base through the sales platform provided by iTunes app store. Many of these apps are free and they cover wide variety of domains including games, location services, productivity, health care etc. According to appshopper.com, in June 2011 the total number of apps available for iPhone was 367,473 and the number for iPad was 103,095. Total apps for Android platform in May 2011 were approximately 200,000 (Barra 2011). Apple recently announced that more than 15 billion apps have been downloaded till date⁴. The role of app market cannot be overemphasized. These apps make the cellular device very attractive to buyers and create strong potential for future revenues from diverse services. More importantly, as users start valuing these apps more and more, app stores have an opportunity to engender strong externality. For a firm, the large number of developers writing diverse applications for its platform is a critical and valuable asset. Therefore, it is not surprising that

¹

http://www.comscore.com/Press_Events/Press_Releases/2011/8/comScore_Reports_July_2011_U.S._Mobile_Subscriber_Market_Share

² <http://www.gartner.com/it/page.jsp?id=1764714>

³ <http://metrics.admob.com/wp-content/uploads/2010/06/May-2010-AdMob-Mobile-Metrics-Highlights.pdf>

⁴ <http://www.apple.com/pr/library/2011/07/07Apples-App-Store-Downloads-Top-15-Billion.html>

firms are luring developers, sometimes by providing deep subsidies, to write applications for their platform. App stores are also a great opportunity for armature and upcoming developers to generate revenue. Mobile application development has been a very dynamic market with entry of many large and small firms.

The growth of applications also provide a great opportunity for researchers to examine various issues like innovation, entry and exit strategy, platform leadership, externality and so on (Boudreau 2011) However, a key detriment to understanding this market is lack of demand data. Similar to Amazon's book market, Apple or Google or Nokia do not provide information on downloads for any application. In fact, even the vendors get somewhat aggregated data from these platforms. For example, Apple may not provide details on application downloads on iPad and iPhone separately to the developer.⁵ Moreover, due to competitive reasons, app vendors also do not provide any details on demand. Thus, only the aggregate numbers are available to most individuals. As is true with any market, without appropriate metrics, one cannot do proper valuations. Without reasonable demand data, firms cannot ascertain the profitability of app markets. As an entrant, it is hard to determine whether a particular niche is worthy of entering or how to set various marketing mix variables. While many stores publish ranks of most downloaded apps, without good demand data, it is also hard to access what the value of a particular rank is. Thus it goes without saying that without good demand data, it is difficult answer many interesting questions.

While information on the download of any app is unavailable, most market providers provide details on the rank of an application. For example, Apple's iPad App store provides a list of top 200 ranked applications along with the rank of each individual application. Using publicly available rank data to infer demand has been a tradition in the literature since the growth of Amazon. Researchers have

⁵ Based on our communication with a developer.

investigated the relationship between sales rank and actual sales (Brynjolfsson et al. 2003; Chevalier & Goolsbee 2003; Chevalier & Mayzlin 2006). In all these papers, rank and sales are assumed to be related via the power law (or Pareto distribution) implying that a small number of products capture a large share of the market. As a matter of fact, Pareto distribution (Pareto 1896) is also the most common version of power law used by information systems and marketing literature. Typical Pareto distribution that has been estimated in extant research is:

$$sales = b * (\text{rank})^{-a} + \varepsilon \quad (1)$$

Where 'b' is the scale parameter and 'a' is the shape parameter.

However, to fit a distribution to predict sales from rank requires that researchers have access to demand (or sales) data; the rank data being publicly available. To get demand, the researchers either conducted experiments or collaborated with publishers. Chevalier and Goolsbee (2003) conducted a creative experiment to infer demand. They selected low selling books and purchased a large quantity (relative to already low demand) from Amazon. As the ranks for those books changed, they could infer the relationship between the sales rank and experimented demand. The downside of this approach is that such experiments can only be performed with very low selling books otherwise one cannot manipulate the rank readily. Also, since most books are low selling, inferring the relationship at top ranks is not accurate. In some other studies (Brynjolfsson et al. 2003), researchers collaborated with book publishers to get access to demand data. In both approaches, it is imperative to get access to demand data.

Table 1 below summarizes the shape parameter values as estimated in prior studies.

Source	Shape Parameter Estimate
(Chevalier & Goolsbee 2003) Data Source: (Poynter 2000)	1.199
(Chevalier & Goolsbee 2003) Data Source: (Weingarten 2001)	1.05
(Chevalier & Goolsbee 2003) (evidence from various experiments suggesting a value between 0.9 and 1.3)	1.2
(Brynjolfsson, et al. 2003)	0.871
(Ghose et al. 2006) Data Source: (Weingarten 2001)	0.952
(Chevalier & Mayzlin 2006)	0.78
(Brynjolfsson et al. 2010)	0.613

Table 1: Pareto Shape Parameters Estimated by Existing Literature

Notice that the shape parameter is decreasing in magnitude over the last decade. The smaller value of 'a' suggests a flatter curve for (1). For comparison purpose, we have also plotted the shape of Equation (1) for different estimates for 'a'.

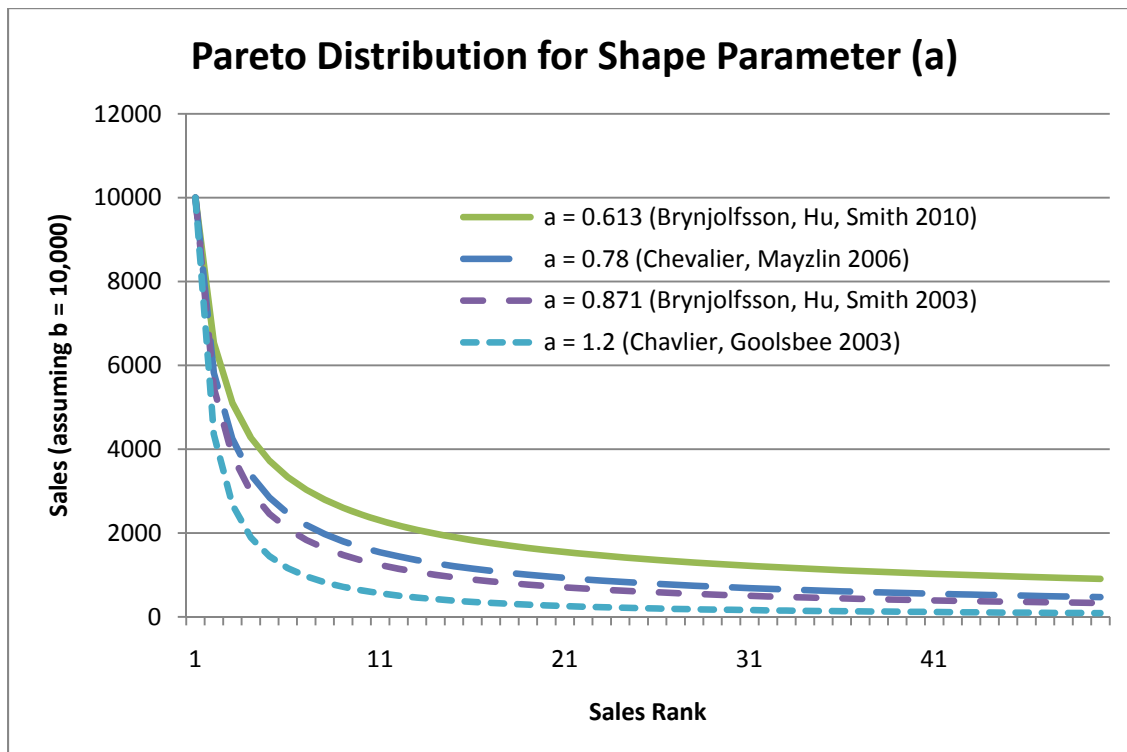


Figure 1: Relationship between sales rank and actual sales as estimated in the extant research

An important aspect of estimating demand is that it paves the way for other interesting work. For example, Ghose et al. 2006 estimated the elasticity of substitution between new and used goods. Ghose & Sundararajan 2006 used it to study software security product marketplace. Brynjolfsson, et al. 2003 and Anderson 2004 used this to establish the long tail phenomenon on the Internet. In particular, they showed how Internet is making it easier for users to find niche products on the Internet and users are more likely to buy lower ranked products. Also, notice from Table 1, the estimates of 'a' have decreased over time. Thus one can infer dynamics of demand over time. In summary, use of sales rank to compute actual sales or use in lieu of sales has become common in academic research because of the unavailability of actual sales data.

In this paper, we provide a methodology to link rank with actual downloads for mobile apps using publicly available data. While our method is similar to methods proposed in prior studies, it is different along two very important dimensions. First, to calibrate the relationship between rank and sales, the prior work needed access to demand data (either from an experiment or from a book publisher). Getting demand data is quite challenging, much more so in the App market, rendering this work quite difficult. Second, in prior studies one is more likely to collect data on books with very low rank/ baseline sales and use it to fit the relationship to top rank books making the results less accurate. Since we are using publicly available data for high ranking apps as well as using it to find a relationship for high rank apps, this is a non-issue in our methodology.

In our paper, we will show the one can calibrate this relation using publicly available data alone. Therefore, one does not need access to any demand data at all. Furthermore, we know of no paper which has tried to calibrate the relationship between rank and sale for the mobile apps - an important exercise in itself. The next section discusses our estimation theory and data, section 3 presents our results and section 4 concludes with discussion for future work.

2 MODEL

Our focus will be Apple's iPad and iPhone app store. A key feature of this store is that three different rank lists are publicly available: top-free applications, top-paid applications, and top-grossing applications. Top-free list shows the most downloaded applications that have no upfront purchase price. Top-paid list shows the most downloaded applications that have non-zero price. In addition, both free and non-free apps may include additional features inside the application that users may purchase. Top-grossing list shows the rank of most downloaded applications that generated maximum combined revenue from both upfront costs and in-app purchases. We do not have data on in-app revenues (only prices for apps are visible). However, a cursory look at the applications show that in-app option is not particularly common for paid apps. Therefore, we assume that the top-grossing list orders applications based on the total downloads and the upfront price. We will use data from two of these lists: top-paid list and top-grossing list. This forms the basis for our estimation model.

Like the extant literature we also assume that sales follow the power law distribution in sales rank. When browsing for apps, individuals can acquire apps from recommendation system, top rank lists, keyword search, or casual browsing within each app category with sorting based on popularity or newness. Overall the consumption behavior of individuals defines the total downloads of each application, which is reflected in the top rank lists. As discussed earlier, we assume the Pareto distribution for estimating downloads from an app rank.

Assuming number of downloads of an application at rank r_p in the top-paid list is given by d_{r_p} , the Pareto distribution could be written as:

$$d_{r_p} = b_p * r_p^{-\alpha_p} \quad | \quad 1 \leq r_p \leq 200 \quad (2)$$

Here b_p defines the scale factor that is dependent on the total number of applications' downloads during a given time period for either iPad or iPhone, and a_p defines the shape of the Pareto curve suggesting the change in number of downloads for applications ranked differently. Recall that we have data for only top 200 ranks.

Similarly we define the Pareto distribution of the apps in the top-grossing list where pd_{r_g} is the revenue generated by the application at rank r_g in the list. This revenue as defined by the Pareto for top-grossing list could also be defined as the product of price (p) and number of downloads (d_{r_p}) of same application in the top-paid list. Thus we can write the distribution for the top-grossing apps as:

$$pd_{r_g} = p * d_{r_p} = b_g * r_g^{-a_g} \quad (3)$$

From the publicly available data we know p , r_p , r_g . The unknown parameters that we need to estimate are b_p , b_g , a_p , a_g . We can re-write (3) after taking logs as:

$$\log(r_g) = \frac{1}{a_g} * \log\left(\frac{b_g}{b_p}\right) + \frac{a_p}{a_g} * \log(r_p) - \frac{1}{a_g} * \log(p)$$

Or,

$$\log(r_g) = \beta_0 + \beta_1 * \log(r_p) + \beta_2 * \log(p) \quad (1)$$

Where

$$a_g = -1 * (1/\beta_2) \quad (2)$$

$$a_p = -1 * (\beta_1/\beta_2) \quad (3)$$

$$b_g/b_p = \exp(-1 * (\beta_0/\beta_2)) \quad (4)$$

This could be estimated using a simple truncated ordinary least square regression with various values of r_g , r_p , and p from different time periods and for different apps.

Notice, we can only recover the ratio of scaling factors (b_g/b_p). Estimating the individual values of the scaling factors (b_p and b_g) requires additional information. Since the knowledge for actual downloads for an individual app is not readily available, we use aggregate downloads in a day to recover (b_p and b_g). To see this, notice that if we know aggregate downloads (D_t) then

$$D_t = \sum d_{r_p} = b_p \sum_{r=1}^N r_p^{-a_p}$$

Thus, with the knowledge of total number of downloads of top ranked apps we can recover b_p and b_g from the formula above as:

$$b_p = \left(\sum_{r_p=1}^N d_{r_p} \right) / \left(\sum_{r_p=1}^N r_p^{-a_p} \right) \quad (5)$$

$$b_g = \exp(-1 * (\beta_0/\beta_2)) * \left(\sum_{r_p=1}^N d_{r_p} \right) / \left(\sum_{r_p=1}^N r_p^{-a_p} \right) \quad (6)$$

In the equation above, the shape parameter (a_p) is estimated from prior equations and integral of individual app downloads (d_{r_p}) defines the total downloads associated with all top ranked apps.

3 DATA

As we mentioned earlier, the top-paid list rank and top-grossing rank data is readily available publicly from various websites like Apple, Appshopper, AppAnnie, and so on. Our data period was from April 2011 to May 2011. The information collected contained top-paid and the top-grossing app ranking for each day during this period for both iPad and iPhone. We also collected data on prices. The summary statistics of the downloaded apps are given in table 1.

	iPad (Paid)	iPad (Grossing)	iPhone (Paid)	iPhone (Grossing)
N	23471	19857	21855	18008
Average Rank	99.63 (57.91)	100.04 (57.66)	99.98 (57.78)	99.94 (57.73)
Average Price (\$)	4.31 (4.44)	12.18 (51.03)	1.73 (1.68)	6.76 (44.68)
Average File Size (MB)	73.55 (127.44)	83.98 (142.14)	51.58 (110.08)	68.26 (131.49)

Table 2: Summary Statistics (from April 2011 to May 2011)

From table 2 it is evident that the average price of applications in top-paid list is lower when compared to the applications in top-grossing list. This suggests that apps that gain higher position on the top-grossing lists generally are more expensive than apps positioned higher on the top-paid lists. We also observe that iPhone apps are cheaper; possibly because of smaller screen size makes it difficult for more sophisticated apps to run on iPhones. This is also evident from the difference in the average app file size, which is smaller for iPhone, suggesting fewer features leading to a lower price. iPad apps have bigger file size relative to iPhone apps. The lower prices of apps on iPhone could also be attributed to perceived maturity and competitiveness of iPhone market relative to iPad.

However, the top-paid apps may not necessarily be the top-grossing apps. The overlap between these two lists is provided in table 3 below. We would use the apps that are on both lists for estimation.

	iPad	iPhone
N	10709	8164
Average Rank on Top Paid List	74.00 (54.22)	64.46 (52.87)
Average Rank on Top Grossing List	88.67 (57.50)	93.72 (57.32)
Average app price	6.41 (5.11)	2.45 (2.22)
Average app size (MB)	94.39 (149.80)	78.94 (149.49)
Average number of app downloads per day (Distimo 2011)	78,000	430,000

Table 3: Summary Statistics of Apps Ranked in both Top Paid and Top Grossing Lists

We plot the ranks of apps in the top-paid and top-grossing list. Generally there is strong correlation that suggests higher rank in paid list is correlated with higher rank in top grossing. However, there is a fair bit of distribution as well.

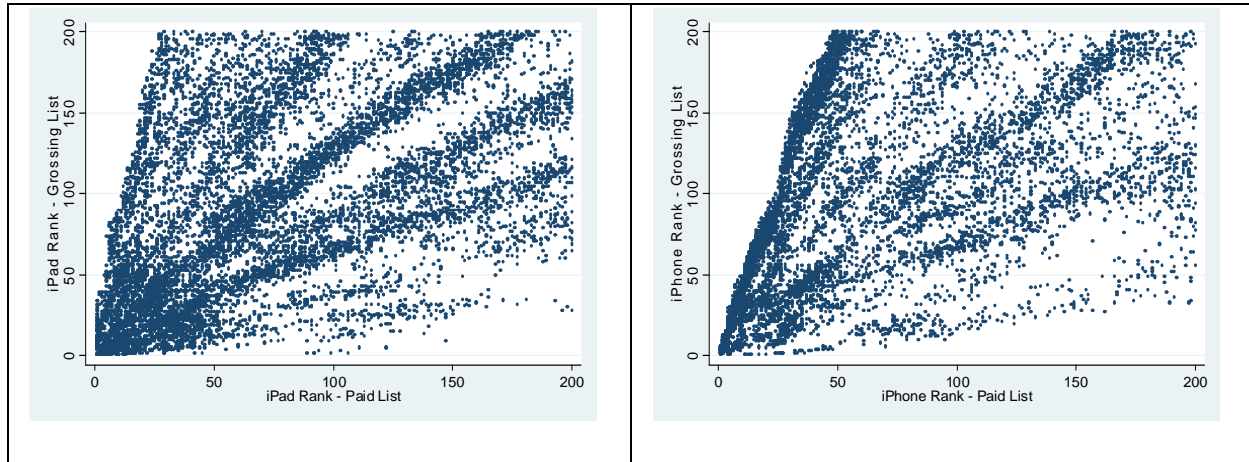


Figure 2: Graphs of App Rank in Top Paid List vs Rank in Top Grossing List (iPad and iPhone)

In the next section we will use the model developed in the prior section and use the data to estimate the shape and scale parameters for both iPad and iPhone apps.

4 RESULTS

4.1 SHAPE PARAMETER (α)

Solving the previously mentioned linear function using truncated regression (Hausman & Wise 1977; Wooldridge 2009), we estimate Equation (4) to recover the following coefficients:

Log(rank_GROSSING)	iPhone Coef. (robust std. err.)	iPad Coef. (robust std. err.)
Log(rank_PAID) - β_1	1.098 (0.008)***	1.02 (0.009)***
Log(price) - β_2	-1.163 (0.011)***	-1.129 (0.01)***
Constant - β_0	1.014 (0.024)***	2.131 (0.025)***
R²	0.820	0.815
N	8164	10709

*** p-value < 0.001

Table 4: Coefficients from Truncated OLS

An increase in the rank on top-paid list increases the rank on top-grossing list as well which is consistent with the summary statistics. Similarly, higher prices reduce the demand and hence reduce the rank on top grossing list. In fact, the relationship between rank and price is highly elastic. This suggests that increasing prices leads to a rapid decline in rank (and hence sales). This is not surprising. The mobile apps markets are highly competitive and lots of substitutes are available.

The coefficients in the regression are highly significant and a high value of R^2 suggests confidence in the model to fit the data.

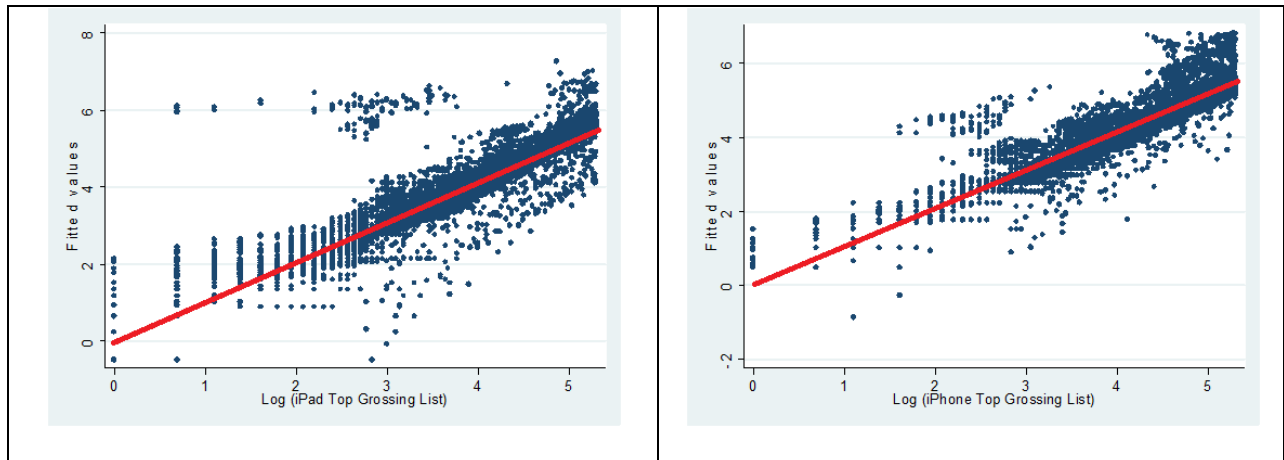


Figure 3: Graphs of Fitted and Actual of Log of App Rank in Top Grossing List (iPad and iPhone)

Looking at the graphs in figure 3, we see that the estimated values of rank in the top grossing lists are very close to the actual values of the ranks.

As we showed in 6, 7, and 8, once we estimate (4), we can recover the shape parameter for both the top-grossing and top-paid apps readily. They are produced in the table below:

	iPad	iPhone
a_p	0.903	0.944
a_g	0.886	0.860

Table 5: Estimated Pareto Shape Parameters

Both of these parameters seem sensible and plausible. This gives us confidence in our method. While the Amazon book market is not readily comparable the estimates seem to be in similar range (see Table 1).

Thus from the above shape parameters, we can estimate the ratio of the number of downloads of two different apps that are ranked differently during any given day on Apple's app store.

$$\text{for iPad: } d_{p1}/d_{p2} = (r_{p1}/r_{p2})^{-0.903} \quad (7)$$

$$\text{for iPhone: } d_{p1}/d_{p2} = (r_{p1}/r_{p2})^{-0.944} \quad (8)$$

The shape parameter allows us to estimate the fraction of downloads or revenues received for a lower ranked app compared to the app with a higher rank. Therefore, an important finding of (10) and (11) is that one can compare the value of different ranks. For example, the number of downloads enjoyed by an iPad app placed at rank 1 on top-paid list is 120 times higher than the app positioned at rank 200. Similarly an iPhone app at rank 1 enjoys 150 times more downloads than the app ranked at 200. From the shape parameters for revenue (based on top grossing app list), a rank 1 application is will gross 1.86 times more that a rank 2 application on iPad. This relative valuation is an important factor for firms when they are investing their marketing dollars in promoting their applications.

4.2 SCALE PARAMETER (b)

Usually, the shape parameter is all we are interested in. Estimates of the shape parameters allow ready relative comparison between two ranks. Moreover, the shape parameters tend to remain more stable over time. However, we now provide a method to estimate the scale parameter as well.

To estimate scale parameter we need additional information. Note from (7) that we have only recovered the ratio of scale parameter from the data so far. To estimate scale parameter, we would need access to actual sales volume from a vendor. However, as we showed in equations (9) and (10), even aggregate total number of downloads is enough to recover these parameters.

Fortunately, the total number of downloads is presented in the industry report provided by (Distimo, Feb 2011) - the total number of downloads per day for top 300 iPad apps is approximately 78,000 and total number of downloads per day for top 300 iPhone apps is approximately 430,000. Given these statistics and the coefficients from the regression above, we estimated the scale parameters using equations (9) and (10) is shown in the first row of table 6 below.

	iPad	iPhone
b_p	9525	52511
b_g	62866	125596

Table 6: Estimated Pareto Scale Parameters

Thus we can now write down the relationship that links the total downloads of each app for any given rank. Similarly, we can specify the function that links revenues with rank. Using our estimated parameters these functions are as follows:

$$d_{iPAd} = 9,525 * r_p^{-0.903} \tag{9}$$

$$d_{iPhOne} = 52,511 * r_p^{-0.944} \tag{10}$$

Obviously, the estimate of scale parameter is likely to change with time as more and more people buy apps from these markets. Even so, we show that with aggregate numbers (which are usually readily available), we can readily estimate the scale parameter. The graph in figure 4 shows the Pareto distribution (equation 12) for app sales as a function of app rank in the top paid list for iPad. We see that

the number of downloads quickly drops for the top 20 ranked apps and is low for the rest 180 apps. Thus the relationship exhibits a long tail.

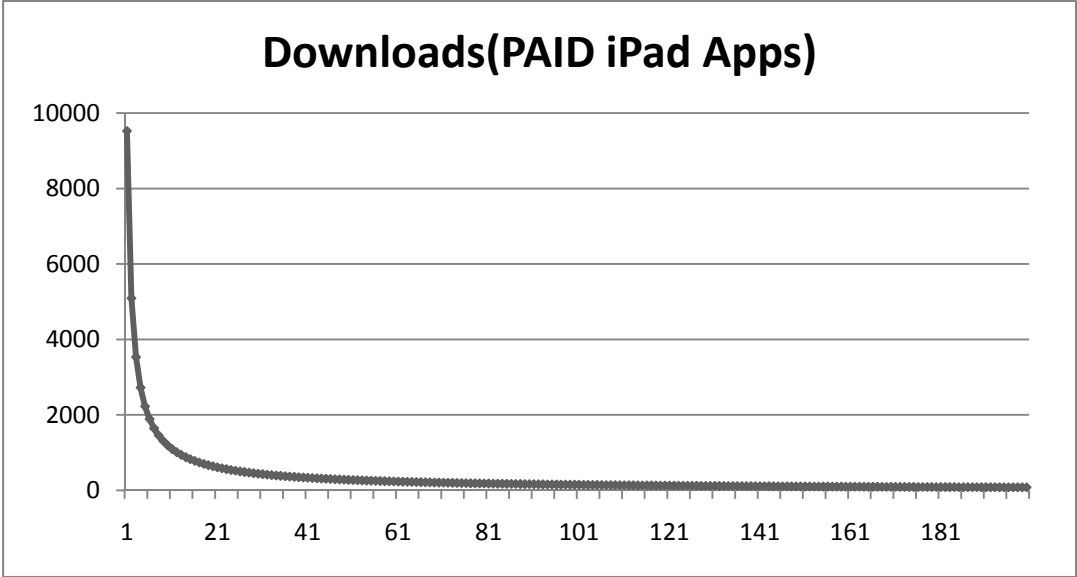


Figure 4: Number of App Downloads vs. App Rank on Top Paid List (iPad)

5 MODEL VALIDITY

One challenge in testing these models is lack of available data on downloads and app revenue. We provide three different ways to test the validity of our model.

First, recall that we estimate two different models (i) one estimating total downloads and second, (ii) estimating total revenue. The difference in the two is the second model is prices multiplied by the demand coming from the first model. Thus, we can estimate the downloads in the first model and multiply by the app price to get the predicted revenues. We can then compare these revenue numbers to the estimated revenues from the second model. This will give us some confidence in the estimated models.

From the tables 7 and 8, we see that the values of estimated revenue are very close in both cases suggesting some confidence in the accuracy of the computed models.

iPhone	Obs	Mean (std. dev.)
Log (Estimated Downloads * Price)	8164	7.97 (0.877)
Log (Estimated Revenue)	8164	8.10 (0.813)

Table 7: Summary Statistics of Estimated Revenue and Price * Estimated Downloads for iPhone

iPad	Obs	Mean (std. dev.)
Log (Estimated Downloads * Price)	10709	7.23 (0.999)
Log (Estimated Revenue)	10709	7.39 (0.930)

Table 8: Summary Statistics of Estimated Revenue and Price * Estimated Downloads for iPad

Second, we partnered with an app developer who shared data on an application, its rank and total downloads (iPad + iPhone) for a month. If we had separate data for each platform (iPad and iPhone), we would have been able to estimate parameters ‘a’ and ‘b’ and compare it to our estimates. However, we can still compare whether the predicted total downloads from our model match with the download number provided by the developer. Recall that total predicted downloads are

$$d_{total} = d_{iPad} + d_{iPhone} = 9,525 * r_p^{-0.903} + 52,511 * r_p^{-0.944}$$

The mean value of the actual downloads of the app is 1737 (std. dev. = 666) and the mean value for the estimated downloads based on rank information was 1620 (std. dev. = 554). From the paired t-test, we find the p-value = 0.145 (p-value for unpaired t-test is 0.6434). Thus, we can safely conclude that the estimated model is predicting the values that do not have significantly different mean from the actual download numbers of the iTunes app store apps.

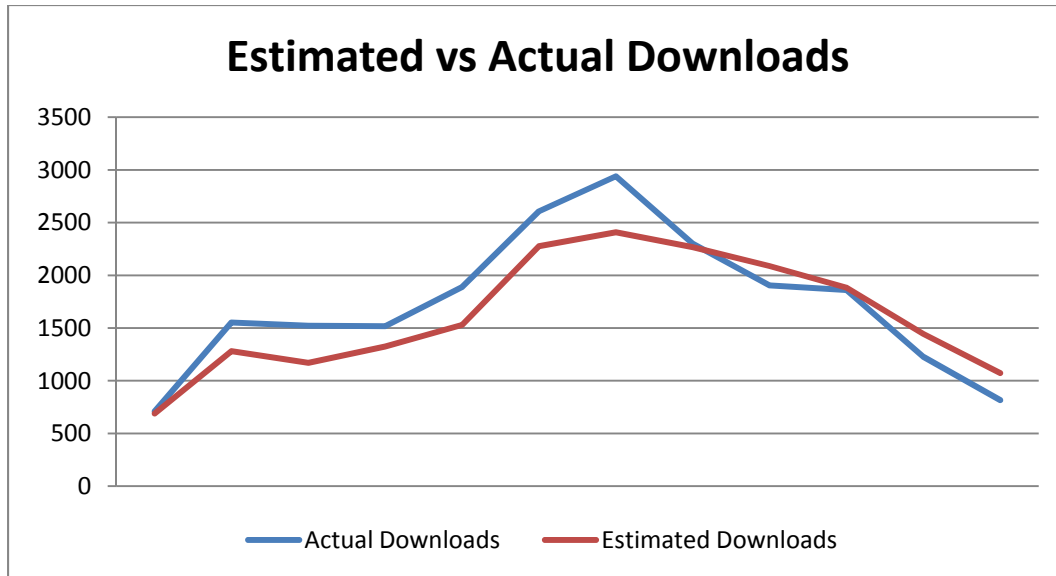


Figure 5: Number of App Downloads vs. App Rank on Top Paid List (iPad)

We plot the actual and predicted (from our model) downloads per day for this app over the sample period in the figure above. It is quite evident from the plot that our model is doing a very good job of predicting the sale of any give app if the rank is known to us. This provides additional evidence regarding the robustness of our method.

Finally, we estimate the shape and scale parameters from the available data assuming some of the parameter estimates were accurate. To estimate the scale parameter while assuming the scale parameters are known, we used ordinary least squared regression model as shown in equation below. Here the coefficients estimated represent the scale parameters and all other variables (downloads, ranks, and shape parameters) are assumed to be known.

$$d_{total} = (d_{iPad} + d_{iPhone}) = b_{iPad} * r_{iPad}^{-a_{iPad}} + b_{iPhone} * r_{iPhone}^{-a_{iPhone}}$$

When estimating the shape parameter while assuming that scale parameters are know, we see that the above equation will require a non-linear estimation approach. Thus we used a simulated set of shape parameter values (ranging from 0.8 to 1.2) and estimated downloads for each value combination (for

iPad and iPhone). We finally averaged all values that were within the 0.1% range of the actual sales figures. Estimated shape and scale parameters are shown in the table below.

App Store	Actual Data (b)	Model Estimate (b)	Actual Data (a)	Model Estimate (a)
iPad	16540 (7667) ***	9525	0.901 (0.076)	0.903
iPhone	60482 (8164) ***	52511	0.934 (0.012)	0.944

Although, this test assumes that the shape or scale parameters are known, we find that the estimated parameters are close to what we observe from the actual sales data as provided by the app developer. One difference is that the estimated scale parameter values from actual sales data are slightly higher than the estimates from the model. We believe that this difference could be attributed to the deviation of the app from the average app on the Apple’s app store. From the rank data collected for the single app, we do observe that the average sales-rank on iPad was slightly better than that for iPhone, but the rank for iPhone stayed in top 200 for 2 additional days. Although, this validation approach is not the cleanest and most reliable approach, it does show that the estimated values are close to what we found from a data from an app publisher. This gives us some additional confidence in the estimated models.

6 DISCUSSION

We have used publicly available rank data and presented a methodology to accurately estimate the product sales from the ranks of apps listed in top-200 lists on Apple’s app store for both iPhone and iPad. From our analysis, we find that number of downloads enjoyed by an iPad app placed at rank 1 on top paid lists is 120 times higher when compared to the app positioned at rank 200. Similarly an iPhone app at rank 1 enjoys 150 times more downloads when compared to the app ranked at 200. We also showed that the iPhone app positioned at rank 1 on top grossing list earns 95 times more revenue compared to a rank 200 app. The number for iPad app store is 110, suggesting that app at rank 1 is generating 110 times more revenue when compared to app at rank 200.

We are also able to find the scale parameter in the assumed Pareto distribution, which suggests that apps positioned in the last 10 ranks in iPad app store have a difference of only 3 downloads between them with rank 200 app being downloaded 80 times and rank 190 app downloaded 83 times. The difference in the downloads for apps for iPhone at rank 190 and 200 is 19 with the app ranked at 200 downloaded 354 times. Rank 1 on iPad is downloaded 9,525 times and rank 1 app on iPhone is downloaded 52,510 times. These estimates should help app producers and marketing professionals in better directing their marketing efforts to improve the app ranks on Apple's app store and potentially gaining more sales in the future time periods.

Being able to infer demand data from rank is highly valuable information for practicing managers as well as researchers. Mobile apps are a fast growing and an important technology market. Understanding this market and the opportunities it offers is important for different stakeholders. We believe our research method and results have many important implications and hence make a very useful contribution not only to academic literature but also to managers and entrepreneurs.

Our research can be improved in many different dimensions. Our dataset is limited to top 200 apps. Having access to a larger rank pool will provide a better fit. Similarly, while we verify our estimates with data from an app provider, there are significant opportunities to expand the scope of this data collection by collecting data from more apps across different categories. Also, we assumed that both download and revenue based rankings follow a power law, but alternate and more precise distributions could also be developed if more data is available from app producers thus extending the scope of this research.

ACKNOWLEDGEMENTS

We would like to thank Michael D Smith for his ongoing support and advice on our research project. We also thank Vanlal Peka who served as a research assistant on this project and helped us during the data acquisition phase. We would also like to thank our confidential source for sharing the application sales data for the Apple's app store.

REFERENCES

- Anderson, C., 2004. The Long Tail. *Wired Magazine*, (12.10). Available at: <http://www.wired.com/wired/archive/12.10/tail.html> [Accessed June 22, 2011].
- Barra, H., 2011. Android: momentum, mobile and more at Google I/O. Available at: <http://googleblog.blogspot.com/2011/05/android-momentum-mobile-and-more-at.html> [Accessed June 21, 2011].
- Boudreau, K. Let a Thousand Flowers Bloom? An Early Look at Large Numbers of Software “Apps” Developers and Patterns of Innovation. *Organization Science*. (forthcoming)
- Brynjolfsson, E., Hu, Y. (Jeffrey) & Smith, M.D., 2003. Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers. *MANAGEMENT SCIENCE*, 49(11), pp.1580-1596.
- Brynjolfsson, E., Hu, Y.J. & Smith, M.D., 2010. The Longer Tail: The Changing Shape of Amazon’s Sales Distribution Curve. *SSRN eLibrary*.
- Chevalier, J. & Goolsbee, A., 2003. Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics*, 1(2), pp.203-222.
- Chevalier, J. & Mayzlin, D., 2006. The effect of word of mouth on sales: online book reviews. *Journal of Marketing Research*, 43(3), pp.345-354.
- Distimo, 2011. *Distimo February 2011 Report*, Available at: <http://www.prepaidmvno.com/2011/03/18/distimo-february-2011-report/> [Accessed June 21, 2011].
- Ghose, A. & Sundararajan, A., 2006. Evaluating Pricing Strategy Using e-Commerce Data: Evidence and Estimation Challenges. *Statistical Science*, 21(2), pp.131-142.
- Ghose, A., Smith, M.D. & Telang, R., 2006. Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact. *INFORMATION SYSTEMS RESEARCH*, 17(1), pp.3-19.
- Hausman, J.A. & Wise, D.A., 1977. Social Experimentation, Truncated Distributions, and Efficient Estimation. *Econometrica*, 45(4), pp.919-38.
- Poynter, D., 2000. Publishing Poynters. *Para Publishing*. Available at: <http://www.parapublishing.com/files/newsletter/News0400.cfm> [Accessed September 1, 2011].
- Weingarten, G., 2001. Below the Beltway. *Washington Post*. Available at: <http://www.washingtonpost.com/wp-dyn/content/article/2008/10/14/AR2008101400789.html> [Accessed August 31, 2011].
- Wooldridge, J.M., 2009. *Introductory econometrics: a modern approach*, South-Western Cengage Learning. Available at: <http://books.google.com/books?id=64vt5TDBNLwC>.

