Electronic Commerce Research and Applications 16 (2016) 1–6

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

journal homepage: www.elsevier.com/locate/ecra

Rank-sales relationship in electronic commerce: Evidence from publicly available data on 11 product categories

Evgeny A. Antipov *, Elena B. Pokryshevskaya

National Research University Higher School of Economics, Kantemirovskaya St. 3, Saint-Petersburg 194100, Russia

ARTICLE INFO

Article history:
Received 24 June 2015
Received in revised form 17 October 2015
Accepted 18 November 2015
Available online 11 January 2016

Keywords:
Sales rank
Pareto law
Online merchant
Demand estimation

ABSTRACT

Unlike actual sales figures, sales ranks are widespread in the field of electronic commerce, which motivates economists and marketing scholars to look for the avenues of converting sales ranks into actual sales or market shares that are needed for demand estimation. In this study the relationship between actual sales and sales ranks is calibrated using a large online store's unique data on 11 product categories, for which this relationship has never been calibrated before. By allowing the shape parameter of the power law to vary with the sales rank we managed to increase a traditionally used model's fit for most of the product categories. Our parameter estimates can be used by researchers that would like to get a reasonably good approximation of market shares based on sales ranks. We also validated and modified Garg and Telang's (2013) approach to inferring market shares using data on product price, sales rank and revenue rank. The approach, especially its modified version, was shown to lead to a reasonably low market shares prediction error, making it possible for researchers to infer the shares of sales based solely on sales and revenue rankings from companies that prefer not to disclose actual sales data.

1. Introduction

Actual sales data is usually kept confidential by firms, while sales ranks are often available to researchers. This motivated economists and marketing scholars to use ranks to proxy actual sales or to infer sales based on publicly available data. In previous studies researchers were able to infer demand from the rank data in the case of Amazon's book sales (Brynjolfsson et al. 2003, 2010; Chevalier and Goolsbee 2003; Chevalier and Mayzlin 2006). Garg and Telang (Garg and Telang 2013) proposed an ingenious approach to inferring app demand from publicly available data using both sales and revenue rankings available from App Store. Even though their approach is of great practical value, it requires not only a sales ranking, but also a revenue ranking. The latter is typically available only for some of the mobile apps stores.

Scholars used the rank-sales relationship to establish the long tail phenomenon on the Internet, i.e. the fact that niche products account for an unusually high proportion of sales in online markets compared to traditional markets (Brynjolfsson et al. 2006, 2011; Peltier and Moreau 2012). Despite a number of arguments that justify that the phenomenon is persistent over time (Brynjolfsson et al. 2006), some researchers claim that it may be short-lived (Fleder and Hosanagar 2009). Continuous empirical studies of "rank-sales" relationship are needed to see, whether the long-tail phenomenon is really persistent over time or not.

One of the most important aspect of calibrating the rank-sales relationship is that it paves the way for other interesting work related to demand estimation, studying sales dynamics, pricing strategies, etc. For example, in one of the studies the elasticity of substitution between new and used goods was estimated (Ghose et al. 2006). In another study researchers used the sales-rank relationship to study software security product marketplace (Ghose and Sundararajan 2006). Market shares data is also necessary for various demand estimation studies (Ackerberg et al. 2007, Berry and Pakes 2007). Researchers may also be interested in studying the dynamics of demand.

The purposes of this paper are to estimate the parameters of the traditional power law functional form for a large number of product categories, suggest a more flexible alternative with a varying shape parameter and compare the accuracy of the two models in predicting product sales when only sales and revenue ranking data is available.

So far the rank-sales relationship $sales_i = b \cdot rank_i^a$ has been calibrated only for a few markets. In prior studies the shape parameter a varied from -0.613 (Brynjolfsson et al. 2010) to -1.2 (Chevaier and Goldsbee 2003). We obtained parameter estimates for 11 product categories using a unique source of publicly

* Corresponding author.
E-mail address: eantipov@hse.ru (E.A. Antipov).

http://dx.doi.org/10.1016/j.elerap.2015.11.005
1567-4223/

available data – *Utinet.ru* – one of the largest and innovative Internet stores in Russia, which for some reason provides users with sales data. We are not aware of any other store that discloses their sales data publicly. The richness of *Utinet.ru* data allowed us to test the link between the sales and the sales rank for multiple markets, not just a single one as it was done previously (usually, the book market). To the best of our knowledge, this is the first study that calibrates the rank-sales relationship for such products as CPUs, electric shavers, external HDDs, headphones, internal HDDs, memory cards, smartphones, tablet PCs, TVs, USB flash drives and washing machines. These categories are among the top-selling at *Utinet.ru* and seem to be important for academics and practitioners studying differentiated product markets. This interest is reflected in a great number of hedonic studies based on the information on high-tech product prices and characteristics (Chwelos et al. 2008, Dewenter et al. 2007). While some influential demand estimation studies (Ackerberg et al. 2007, Berry et al. 1995, Berry and Pakes 2007) deal with products for which sales data is relatively easy to collect (e.g. automobiles), there seem to be unmet demand for sales data from the markets for computers, consumer electronics and appliances, for which sku-level sales data is either unavailable due to confidentiality issues, or extremely expensive.

Although absolute sales figures (measured in units) are store-specific, our estimation results can be useful for converting sales ranks to market shares, which are used as dependent variables in demand estimation (e.g. nested logit model). The fact that sales in units vary across stores and countries does not matter in most cases as market share, not actual sales, is usually used as the dependent variable in structural demand modeling approaches, such as random-coefficients discrete-choice modeling (Ackerberg et al. 2007, Nevo 2001). We believe that data from a large online store can often serve as a proxy for market shares data that is very difficult to obtain.

In the second section of the paper we introduce two functional forms describing the sales-rank relationship: a typically used power law and a more flexible alternative with a varying shape parameter. We also infer the corresponding formulas for the share of sales. Section 3 describes the dataset used in our study. In Section 4 we assess how well the competing functional forms fit the data. We also validate and modify Garg and Telang's approach (Garg and Telang 2013) that allows inferring market shares using data on product price, sales rank and revenue rank. The predictions obtained with the original and the modified approaches are then compared with the actual sales data and the accuracy is reported. Section 5 concludes.

2. Model

In most papers (Garg and Telang 2013) sales and sales rank are assumed to be related via the power law, implying that a small number of products account for a large market share. The ith product sales are given by the following formula:

$$s_i = b_s \cdot r_i^{a_s} \tag{1}$$

In Eq. (1) s_i are the sales of the ith product measured by the number of items sold, r_i is the sales rank of the ith product, b_s is the scale parameter and a_s is the shape parameter of the "sales-sales rank" relationship. The share of the ith product can be computed as follows:

$$share_i = \frac{s_i}{\sum_{j=1}^{N} s_j} = \frac{b_s \cdot r_i^{a_s}}{\sum_{j=1}^{N} b_s \cdot r_j^{a_s}} = \frac{r_i^{a_s}}{\sum_{j=1}^{N} r_j^{a_s}} \tag{2}$$

Therefore market share is independent of the scale parameter b_s and estimating the shape parameter a_s is enough to infer market shares if sales ranks are known.

The parameters of Eq. (1) can be estimated by running the following log-linear regression:

$$\ln s_i = \beta_0 + a_s \ln r_i \tag{3}$$

The traditional log-linear regression method assumes that the coefficient on $\ln r_i$ does not vary as a product's sales rank increases. However, a_s is likely to vary with r_i indicating that the rank elasticity of sales is not necessarily constant. To test this hypothesis and make sales modeling more accurate we suggest estimating the following modified log-linear equation:

$$\ln s_i = a_0 + (a_{1s} + a_{2s} r_i) \ln r_i \tag{4}$$

Then the share of the ith product can be computed as follows:

$$share_i = \frac{r_i^{a_{1s}+a_{2s}r_i}}{\sum_{j=1}^{N} r_j^{a_{1s}+a_{2s}r_i}} \tag{5}$$

Eqs. (3) and (4) were estimated using OLS regressions with robust standard errors (Wooldridge 2010).

3. Data

The data was collected from an online merchant *Utinet.ru* – one of the largest, fastest growing and innovative online stores in Russia and the only one that did the IPO (as of the end of 2014). This online merchant has a unique mini-statistics feature that allows users to see how many units of each product were sold within the last month.

We obtained monthly sales data for products from 11 categories, in which the number of models with non-zero sales exceeded 25 at the time of data collection (Table 1).

4. Empirical results

4.1. Estimation of rank-sales equations

First, we estimated the parameters of the simple log-linear model ((3)) for each product category (Table 2).

The lowest absolute values of the coefficient on $\ln r_i$ are observed for TVs and washing machines which are characterized by a relatively high price, technical complexity and a large size. The smaller absolute value of a_s corresponds to a flatter curve (longer tail) for Eq. (1). Absolute values of a_s are the highest for external HHDs and memory cards.

Parameter estimates of a more flexible Eq. (4) are presented in Table 3.

The hypothesis that the slope parameter varies with the sales rank was supported for 7 out of 11 product categories (at a 1% or a lower significance level). The direction of the moderating effect appeared to be positive for electric shavers, headphones, tablet PCs, TVs and washing machines, meaning that the higher the rank,

Table 1
Product categories included in the analysis.

Product category	Number of models with non-zero sales
CPUs	37
Electric shavers	26
External HDDs	26
Headphones	37
Internal HDDs	72
Memory cards	40
Smartphones	120
Tablet PCs	64
TVs	47
USB flash drives	45
Washing machines	87