Electronic Commerce Research and Applications

Inferring market shares of products sold online from publicly available proxies: empirical evidence from a large marketplace

--Manuscript Draft--

Manuscript Number:	ECRA-D-20-00908
Article Type:	Research Paper
Keywords:	sales rank; Pareto law; online sales; demand estimation; reviews; market share; consumer electronics
Corresponding Author:	Evgeny A. Antipov National Research University Higher School of Economics Saint-Petersburg, RUSSIAN FEDERATION
First Author:	Evgeny A. Antipov
Order of Authors:	Evgeny A. Antipov
	Elena B. Pokryshevskaya
Abstract:	In this study we show the potential of proxying actual market shares of products sold online with features often available publicly: sales ranks, number of reviews, number of questions asked at the product's page, and the number of sellers. Using actual sales data on 4873 SKUs from 19 categories of consumer electronics and appliances from a large marketplace we calibrate all pairwise relationships, as well as assess the joint power of various combinations of proxies. While sales ranks are naturally the strongest proxy of market shares, the number of questions asked about the product turns out to be the second-best alternative, which is surprisingly more powerful than the number of reviews. The number of sellers offering the product is a moderately strong predictor of sales in most categories, but there is a substantial heterogeneity in category-level effects of this feature.

- Sales rank elasticity of sales is consistent with the long tail phenomenon
- The number of questions asked about a product is a strong proxy of its sales
- The number of reviews is a weaker proxy than the number of questions
- Number of offers is a moderately strong proxy for most product categories
- Log-transformation of all rank-based proxies can be recommended

Inferring market shares of products sold online from publicly available proxies: empirical evidence from a large marketplace

Abstract

In this study we show the potential of proxying actual market shares of products sold online with features often available publicly: sales ranks, number of reviews, number of questions asked at the product's page, and the number of sellers. Using actual sales data on 4873 SKUs from 19 categories of consumer electronics and appliances from a large marketplace we calibrate all pairwise relationships, as well as assess the joint power of various combinations of proxies. While sales ranks are naturally the strongest proxy of market shares, the number of questions asked about the product turns out to be the second-best alternative, which is surprisingly more powerful than the number of reviews. The number of sellers offering the product is a moderately strong predictor of sales in most categories, but there is a substantial heterogeneity in category-level effects of this feature.

Keywords: sales rank, Pareto law, online sales, demand estimation, reviews, market share, consumer electronics

1. Introduction

Inferring actual sales based on factors publicly disclosed by individual stores or marketplaces is important both from the academic research perspective (reliable proxies of actual sales are needed in the absence of actual sales data) and from a practitioner's perspective (to work out which products are likely to be particularly well-sold at a particular point in time and how market shares of products compare to one another). Most studies investigating online demand have typically relied on either sales ranks (Sun, 2012) or the volume of reviews (Öugüt and Onur Tacs, 2012; Yang et al., 2018) as proxies of actual sales due to the unavailability of real sales data, especially from more than a single store. However, there have been few efforts aimed at calibrating the relationships between proxies and actual sales outcomes. In this paper we investigate several proxies of actual sales, including not only widely used sales ranks, which are not available in most cases, but also ranks based on the number of reviews, the number of questions asked about the product and the number of sellers offering the product at the marketplace. We investigate which transformations of proxy variables have the highest correlation with sales performance, calibrate the corresponding relationships and explore how their parameters vary across product categories.

Considering the context-dependence of actual unit sales scale (sales are higher for larger marketplaces compared to small marketplaces, countrywide sales are higher than those from a single store, sales in developed countries are larger than in developing countries, etc.), universal formulas that allow predicting exact sales figures from rank data alone can hardly be inferred. Instead, our goal is to provide researchers and practitioners with evidence on which mathematical transformations of some publicly available metrics produce the best proxies, i.e. variables strongly correlated with market shares or log-transformed market shares. The fact that the market share metric is relative and scale-independent allows us to provide parameter estimates which can be used to convert sales proxies to market shares. Market shares are widely used as dependent variables in demand estimation in the context of random-coefficients discrete-choice modeling using aggregate data (Ackerberg et al., 2007; Bhuyan, 2020; Nevo, 2001), as well as compositional and Dirichlet models (Morais et al., 2018).

Existing research studies calibrated sales-rank relationships for one or a few categories either using now outdated online data from the early 2000s (Brynjolfsson et al., 2006; Chevalier and Mayzlin, 2006), or using data on predominantly offline stores from the late 1990s (Bae et al., 2020). The most recent data (collected in 2015) was used in a study of the relationship between sales and sales ranks and covered only 11 categories sold at a single online-store (Antipov and Pokryshevskaya, 2016). While the issue of market share modeling based on sales ranks has been considered previously (Antipov and Pokryshevskaya, 2016), we extend existing research by providing evidence on a larger number of product categories sold through a large marketplace, using country-wide instead of single-store data, and, most importantly, by considering under-investigated proxies beyond sales ranks alone. The use of the hierarchical Bayesian framework allows model parameters to vary across product categories and to fully quantify the uncertainty around the elasticities of sales ranks to changes in each proxy, providing useful prior estimates for practitioners wanting to infer market shares but having no or very limited data on actual sales to calibrate these relationships themselves. Finally, we use recently collected data from 2020, and thus provide up-to-date estimates that account for the growing role of niche products in electronic commerce, known as the long-tail phenomenon (Brynjolfsson et al., 2011).

Our study addressed the following research questions.

- What are optimal transformations of ranks based on sales, the number of reviews, number of offers and number of questions that allow linearizing relationships between them and market shares or log-transformed market shares?
- Can publicly available ranks based on sales, the number of reviews, number of offers and number of questions be individually helpful in inferring actual market shares of products?
- Can ranks based on the number of reviews, offers and the number of questions be combined to proxy market shares better than each of these predictors individually?
- How much do parameters of the relationships between market share (log market share) and proxy variables vary across categories and what is the magnitude of these parameters?

2. Data

We use a rich novel dataset collected in 2020 and containing both reviews and unique sales data from Yandex Market - one of the largest online marketplaces in Russia and the whole Europe, owned by Yandex – a publicly traded company (NASDAQ: YNDX). We use data on 4873 products from 19 product categories that represent popular types of search goods with the largest number of items (at least 30 per category) and ratings. The number of reviews was systematically more correlated with the number of reviews in the last two months, which is why we use the total number of reviews in our analysis.

A product's market share in its category (*share*) and its log-transformed version (*log_share*) were used as the dependent variables (Table 1). The final choice of the better version of the dependent variable was data-driven (based on the strength of the linear association provided) and is described in the Methods section of the paper. Modeling market shares is sufficient to infer the sensitivity of sales to changes in each of the proxies, because a product's market share is obtained simply by dividing the number of its units sold by a category-specific constant, reflecting the total number of units sold. Therefore, correlation of any independent variable with unit sales is the same as with the share.

Table 1. Variables included in the analysis

Variable name	Variable description	Variable role in the analysis
share	Product's share in the total number of units sold in its category in the last 2 months	dependent
log_share	Natural logarithm of *share*	dependent
rank	Sales rank within product category, ties were treated using the "random" method	predictor
rank_reviews	Rank by the number of reviews	predictor
rank_offers	Rank by the number of offers	predictor
rank_questions	Rank by the number of questions	predictor
category	Product category (19 levels)	clustering

The number of SKUs was larger than 30 for all product categories involved in our analysis. (Table 2).

Table 2. Product categories included in the analysis

Product category	**Number of SKUs with non-zero sales**
Mobile phones	650
Headphones	592
TVs	570
Refrigerators	463
Washing machines	421
Laptops	371
Vacuum cleaners	264
Microwave ovens	231
Printers	229
Blenders	222
Smart watches	219
Coffee machines	204
Tablet PCs	134
Food processors	71
Cameras	70
Toasters	52
Teapots	44
E-book readers	35
Music players	31
Total	*4873*

3. Methods

1. In order to find optimal transformations of independent variables that provide the highest correlation with dependent variables we used Box-Cox power family of transformations of rank variables (x) acting as predictors in our analysis:

$$x(\lambda) = \begin{cases} \dfrac{(x+1)^{\lambda} - 1}{\lambda}, \text{ if } \lambda \neq 0 \\ log(x+\lambda), \text{ if } \lambda = 0 \end{cases}$$

Using a grid search we computed absolute correlations between transformed proxies and dependent variables by considering all combinations determined by λ from -5 to +5 incremented by 0.5, 2 dependent variables, 4 independent variables, and 19 product categories.

2. For each predictor its own transformation and the version of the dependent variable (untransformed or log-transformed) were selected so as to maximize the average absolute value of the linear correlation coefficient. A series of mixed regression models was estimated using the Bayesian approach with non-informative priors using Stan-based R package *brms* (Bürkner, 2018; Carpenter et al., 2017) to obtain the parameters of the relationships between each of the rank predictors (x_{ij}) and the dependent variable (y_{ij}) allowing both the intercept and the slope to vary across product categories (j=1,…,19):

$$y_{\mathrm{ij}} = (\beta_0 + u_j) + (\beta_1 + v_j)x_{\mathrm{ij}} + \varepsilon_{ij} \qquad (1)$$

Simple (bivariate) hierarchical regression models were further extended by including various linear combinations of rank regressors in the right-hand side of model 1. The MCMC estimation method partially pooled information across respondents, allowing estimates for different categories to be more or less similar to one another based on the data. For replicability purposes we report the settings of the MCMC algorithm used when estimating all models:

- Number of market chains: *chains* = 4
- Number of total iterations per chain (including warmup): *iter* = 2000
- Warmup (burn-in) iterations: *warmup* = 1000
- Thinning rate: *thin* = 1
- The seed for random number generation to make results reproducible: *seed*=100

4. Results

4.1. Optimal power transformations of predictors

Figure 1 illustrates the results of the grid search for an optimal lambda for the best power transformations of proxy variables that provide the highest absolute linear correlation with each of the 2 dependent variables (*share* and *log_share*). According to Figure 1 in the case of relationships between *share* and each predictor there is always a segment of data points that clearly stand out. These are observations related to the smart watch product category, where one model of Xiaomi watches occupied an unusually high share (31% of unit sales in the category). This problem was substantially relaxed after the log-transformation of the market share. Another advantage of the share's log-transformation is that the resulting dependent variable is almost perfectly normally distributed. From the upper 4 plots in Figure 1 it can be seen that for all *log_share* - *predictor* relationships the optimal power transformation parameters λ are 0 (corresponding to the log-transformation) or 0.5 (corresponding to the square root transformation). For consistency and ease of interpretation we will estimate log-log relationships between shares and various types of ranks.

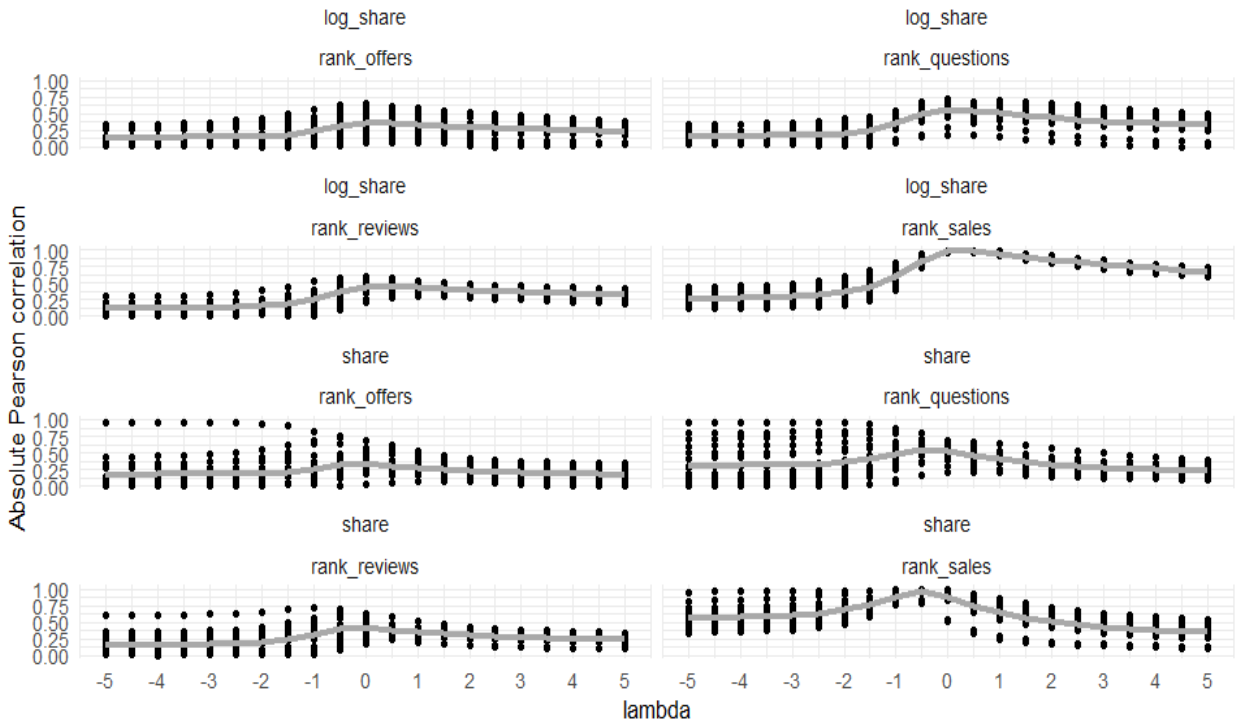

Figure 1. Absolute Pearson correlations associated with various values of λ parameter of Box-Cox transformations

4.2. Parameter estimates of simple regressions with random intercepts and slopes

Fixed (population-level) parameter estimates of hierarchical log-log regressions of market share on each of the rank proxies are presented in Table 3. The marginal R^2 considers only the variance of the fixed effects, while the conditional R^2 takes both the fixed and random effects into account (Gelman et al., 2019). While almost perfect fit between market shares and sales rank was expected, among three other proxies the conditional R^2 turned out to be the highest for the rank based on the number of questions asked about the product (Conditional R^2=0.638). The rank based on the number of offers has about the same explanatory power (Conditional R^2=0.560) as the rank based on the number of reviews (Conditional R^2=0.542). This ordering of models was confirmed to agree with leave-one-out cross-validation aimed at figuring out the expected log predictive density (ELPD) of each model for a new dataset (Vehtari et al., 2017).

Table 3. Two-level bivariate regressions of log market shares on log rank predictors: population-level (fixed) effects

	(1.1) ***log_share***	***(1.2)*** ***log_share***	***(1.3)*** ***log_share***	***(1.4)*** ***log_share***
Intercept	2.998 (2.675 – 3.320)	0.766 (0.486 – 1.083)	1.325 (1.055 – 1.606)	0.589 (0.108 – 1.103)
log_rank_sales	-0.969 (-1.082 – -0.865)			
log_rank_reviews		-0.449 (-0.519 – -0.369)		

log_rank_questions			-0.568 (-0.671 – -0.463)	
log_rank_offers				-0.402 (-0.538 – -0.272)
Observations	4873	4873	4873	4873
Marginal R^2 / Conditional R^2	0.833 / 0.972	0.172 / 0.542	0.289 / 0.637	0.137 / 0.560

95% CI in parenthesis

Even though we do not report p-values common for the frequentist approach, we can still pay attention to whether zero is within the credible interval or not. The 95% credibility interval allows for the intuitively attractive interpretation that there is 95% chance that the true population value of the elasticity parameter falls within this interval.

While the fixed parts of the estimated models give us an idea of some average prevalence of the long tail in the distribution of market shares in today's electronic commerce, these estimates can vary substantially from sample to sample depending on which product categories prevail. The role of the random component is clear from the differences between the marginal and the conditional R^2 values reported for each model. In the case of ranks based on reviews, questions, and offers the difference between the two coefficients of determination is especially pronounced, implying a substantial variation of the elasticity parameter across categories.

Table 4 presents a summary of market share elasticities with respect to each rank predictor for each product category inferred from the posterior distributions of the sum of population-level effects and corresponding group-level effects. Parameter estimates vary substantially across product categories. The highest absolute elasticity of market shares with respect to sales rank is observed for Mobile phones, while the lowest – for teapots and cameras. The highest absolute elasticity with respect to the rank by the number of reviews was observed in the case of headphones and smart watches, while the lowest – in the case of cameras and teapots. Market shares of mobile phones and smart watches were by far the most sensitive to changes in how products rank on the number of questions asked about them, while teapots and cameras were the least sensitive. A 1% change in the rank by the number of offers was associated with the largest change in the market share in the case of mobile phones, while market shares of blenders, food processors, toasters, teapots, and music players were the least sensitive (the 95% CI even contains zero).

Table 4. Market share elasticities with respect to each predictor (based on the sum of population-level and group-level effects of hierarchical log-log regressions)

	Predictor			
	rank_sales	*rank_reviews*	*rank_questions*	*rank_offers*
Blenders	-0.859 (-0.890 – -0.828)	-0.413 (-0.507 – -0.311)	-0.434 (-0.537 – -0.328)	-0.082 (-0.212 – 0.046)
Cameras	-0.646 (-0.707 – -0.583)	-0.287 (-0.444 – -0.121)	-0.285 (-0.458 – -0.105)	-0.293 (-0.504 – -0.072)
Vacuum cleaners	-0.965 (-0.994 – -0.935)	-0.485 (-0.578 – -0.393)	-0.598 (-0.693 – -0.503)	-0.352 (-0.464 – -0.239)
Coffee machines	-0.979 (-1.011 – -0.945)	-0.481 (-0.586 – -0.381)	-0.600 (-0.708 – -0.496)	-0.317 (-0.446 – -0.189)
E-book readers	-1.145 (-1.234 – -1.055)	-0.358 (-0.559 – -0.161)	-0.581 (-0.840 – -0.343)	-0.630 (-0.933 – -0.316)
Food processors	-0.822 (-0.882 – -0.761)	-0.353 (-0.508 – -0.206)	-0.500 (-0.671 – -0.323)	-0.151 (-0.366 – 0.066)
Headphones	-1.178 (-1.197 – -1.158)	-0.650 (-0.724 – -0.579)	-0.765 (-0.830 – -0.701)	-0.570 (-0.648 – -0.492)
Laptops	-1.065	-0.447	-0.570	-0.665

	(-1.089 – -1.041)	(-0.529 – -0.360)	(-0.651 – -0.488)	(-0.762 – -0.570)
Microwave ovens	-0.892	-0.461	-0.522	-0.212
	(-0.924 – -0.86)	(-0.556 – -0.367)	(-0.619 – -0.417)	(-0.328 – -0.095)
Mobile phones	-1.450	-0.559	-0.915	-0.983
	(-1.468 – -1.432)	(-0.624 – -0.491)	(-0.980 – -0.854)	(-1.056 – -0.909)
Music players	-0.741	-0.298	-0.367	-0.282
	(-0.839 – -0.641)	(-0.513 – -0.087)	(-0.629 – -0.100)	(-0.602 – 0.040)
Printers	-0.952	-0.499	-0.661	-0.421
	(-0.983 – -0.921)	(-0.601 – -0.404)	(-0.758 – -0.563)	(-0.537 – -0.298)
Refrigerators	-0.728	-0.406	-0.464	-0.212
	(-0.750 – -0.707)	(-0.485 – -0.323)	(-0.538 – -0.388)	(-0.297 – -0.129)
Smart watches	-1.210	-0.618	-0.840	-0.649
	(-1.241 – -1.177)	(-0.727 – -0.520)	(-0.950 – -0.736)	(-0.769 – -0.527)
Tablet PCs	-1.216	-0.452	-0.734	-0.682
	(-1.260 – -1.173)	(-0.569 – -0.337)	(-0.870 – -0.603)	(-0.842 – -0.523)
Teapots	-0.620	-0.266	-0.264	-0.201
	(-0.699 – -0.538)	(-0.453 – -0.069)	(-0.482 – -0.04)	(-0.477 – 0.078)
Toasters	-0.951	-0.411	-0.403	-0.206
	(-1.020 – -0.883)	(-0.594 – -0.241)	(-0.606 – -0.203)	(-0.453 – 0.039)
TVs	-1.121	-0.588	-0.747	-0.436
	(-1.141 – -1.102)	(-0.660 – -0.518)	(-0.814 – -0.683)	(-0.512 – -0.360)
Washing machines	-0.861	-0.443	-0.609	-0.271
	(-0.883 – -0.838)	(-0.519 – -0.363)	(-0.684 – -0.536)	(-0.360 – -0.180)

95% CI in parenthesis

4.3.Parameter estimates of multiple regressions with random intercepts and slopes

Various combinations of proxies were considered to make conclusions about their joint explanatory power. When sales ranks are available, any other ranks are essentially useless (Model 2.1): their parameter estimates are near-zero, and their inclusion does not increase the explanatory power compared to Model 1.1 (Table 3). When sales ranks are not available, the most parsimonious combination of predictors is comprised of ranks based on the number of questions and the number of offers (Table 5, Model 2.3). The improvement when the number of reviews was added turned out to be negligible.

Table 5. Two-level multiple regressions of log market shares on log rank predictors: population-level (fixed) effects

	(2.1) ***log_share***	***(2.2)*** ***log_share***	***(2.3)*** ***log_share***	***(2.4)*** ***log_share***	***(2.5)*** ***log_share***
Intercept	2.986 (2.680 – 3.291)	2.391 (2.042 – 2.755)	2.286 (1.844 – 2.708)	1.466 (1.293 – 1.684)	1.956 (1.433 – 2.467)
log_rank_sales	-0.977 (-1.092 – -0.870)				
log_rank_reviews	0.002 (-0.010 – 0.015)	-0.084 (-0.162 – -0.010)		-0.085 (-0.195 – 0.030)	-0.399 (-0.455 – -0.333)
log_rank_questions	0.010 (-0.012 – 0.030)	-0.465 (-0.570 – -0.348)	-0.510 (-0.585 – -0.428)	-0.520 (-0.669 – -0.365)	
log_rank_offers	-0.002 (-0.019 – 0.017)	-0.279 (-0.376 – -0.186)	-0.291 (-0.385 – -0.205)		-0.331 (-0.464 – -0.202)
Observations	4873	4873	4873	4873	4873
Marginal R^2 / Conditional R^2	0.833 / 0.972	0.450 / 0.710	0.437 / 0.704	0.311 / 0.650	0.344 / 0.644

95% CI in parenthesis

Conclusion

Yandex Market platform systematically offers unique sales data allowing to conduct research related to online sales modeling without introducing measurement errors that commonly occur when proxies like sales ranks or the number of reviews are used. However, researchers interested in using data from other platforms can use our calibration results to transform metrics available on most other platforms so that the derived proxies have the highest correlation with actual sales, log-sales, market share or log market share depending on the desirable dependent variable in their study.

The fixed part of the shape parameter of the sales-sales rank relationship is insignificantly different from -1 (95% CI: [-1.082 – -0.865] and is larger (in absolute terms) than the weighted average of category-specific parameter estimates reported by Antipov and Pokryshevskaya (2016) using data from a single online seller (-0.783), but significantly smaller (in absolute terms) compared to estimates based on 1999 data (95% CI: [−1.136, −1.229]) reported in Bae et al. (2020). However, in all studies involving many product categories there was a high heterogeneity across categories. Conducting a more formal meta-analysis of all estimates accumulated to date can be a direction for future research aimed at the investigation of the long tail phenomenon in online commerce.

Log-transformation of all variables was shown to be the optimal linearizing transformation both for the dependent variable (market shares) and for all predictors. While a product's sales rank is by far the best proxy, it is not always publicly available. The rank by the number of questions asked about the product is the second-best predictor, the rank by the number of offers and the rank by the number of reviews have a weaker, but still moderately high predictive power. However, while the number of offers (sellers) is, on average, a good proxy, for 5 categories out of 19 the 95% high density interval of the corresponding elasticity includes zero.

When sales ranks are used, other proxies do not provide incremental explanatory power when added to the model. When sales ranks are not available, a linear combination of log-transformed ranks of the number of question and the number of offers provided the best fit. The rank by the number of reviews contributes to model improvement only if either the rank by the number of offers or the number of questions is not available.

We noticed that limiting the number of reviews by those left exactly in the period for which sales are reported (2 months) did not not provide a better proxy of sales compared to the total number of reviews, which may be explained by the high inertia in online sales of consumer appliances and electronics, as well as by higher propensity to purchase products with a large total number of reviews. However, analysis of panel data may offer new insights into how to weight individual reviews depending on their timing and valence to come up with the best measure of sales for a given period. We have shown that the number of questions asked about the product is a useful proxy of sales. Even though, this correlation is natural as questions are asked by people interested in purchasing a product, it will be useful to check if this result holds for other platforms. In addition, there may be specific types of questions that are more predictive of sales than others. Finally, sentiments of questions and answers can serve as additional predictors of sales. The value of the number of offers as a sales proxy can potentially be increased by weighting sellers differently (e.g. based on the size of their assortment), as well as by accounting for the possibility that sales are likely to be stronger correlated with the number of sellers for products with a lower price dispersion (through a more uniform distribution of sales across sellers).

Aknowledgements
The article was prepared within the framework of the Academic Fund Program at the National

Research University Higher School of Economics (HSE University) in 2019 — 2020 (grant №19-01-070) and within the framework of the Russian Academic Excellence Project «5-100».

References

Ackerberg, D., Lanier Benkard, C., Berry, S., Pakes, A., 2007. Econometric tools for analyzing market outcomes. Handb. Econom. 6, 4171–4276.

Antipov, E.A., Pokryshevskaya, E.B., 2016. Rank-sales relationship in electronic commerce: Evidence from publicly available data on 11 product categories. Electron. Commer. Res. Appl. 16. https://doi.org/10.1016/j.elerap.2015.11.005

Bae, Y.H., Gruca, T.S., Lim, H., Russell, G.J., 2020. The size-rank relationship for market shares of consumer packaged goods. Appl. Econ. 1–9.

Bhuyan, S., 2020. Is there market power in the US brewing industry? Eur. J. Appl. Econ. 17, 67–79.

Brynjolfsson, E., Hu, Y., Simester, D., 2011. Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. Manage. Sci. 57, 1373–1386.

Brynjolfsson, E., Hu, Y.J., Smith, M.D., 2006. From niches to riches: Anatomy of the long tail. Sloan Manage. Rev. 47, 67–71.

Bürkner, P.C., 2018. Advanced Bayesian multilevel modeling with the R package brms. The R Journal, 10, 395-411.

Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A., 2017. Stan. A Probabilistic Program. Lang. J. Stat. Softw. 76.

Chevalier, J., Mayzlin, D., 2006. The effect of word of mouth on sales: Online book reviews. J. Mark. Res. 43, 345–354.

Gelman, A., Goodrich, B., Gabry, J., Vehtari, A., 2019. R-squared for Bayesian regression models. Am. Stat. 73, 307–309.

Morais, J., Thomas-Agnan, C., Simioni, M., 2018. Using compositional and Dirichlet models for market share regression. J. Appl. Stat. 45, 1670–1689.

Nevo, A., 2001. Measuring Market Power in the Ready-to-Eat Cereal Industry. Econometrica 69, 307–42.

Öugüt, H., Onur Tacs, B.K., 2012. The influence of internet customer reviews on the online sales and prices in hotel industry. Serv. Ind. J. 32, 197–214.

Sun, M., 2012. How does the variance of product ratings matter? Manage. Sci. 58, 696–707.

Vehtari, A., Gelman, A., Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat. Comput. 27, 1413–1432.

Yang, Y., Park, S., Hu, X., 2018. Electronic word of mouth and hotel performance: A meta-analysis. Tour. Manag. 67, 248–260.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Evgeny A. Antipov

Elena B. Pokryshevskaya