Journal of Economic Behavior and Organization 150 (2018) 43–61

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

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

Demand effects of consumers' stated and revealed preferences

Per Engström[a,*], Eskil Forsell[b]

[a] Department of Economics, Uppsala University, Box 513, Uppsala SE-751 20, Sweden
[b] Department of Economics, Stockholm School of Economics, Box 6501, Stockholm SE-113 83, Sweden

ARTICLE INFO

Article history:
Received 21 March 2017
Revised 8 April 2018
Accepted 13 April 2018

JEL classification:
C21
C26
D83
M31

Keywords:
Peer effects
Observational learning
Network effects
Stated preferences
Revealed preferences
eWOM
Google play
Android apps
Regression discontinuity design
Instrumental variable analysis

ABSTRACT

Knowledge of how consumers react to different signals is fundamental to understanding how markets work. The modern electronic marketplace has revolutionized the possibilities for consumers to gather detailed information about products and services before purchase. Specifically, a consumer can easily – through a host of online forums and evaluation sites – estimate a product's quality based on either (i) what other users say about the product (*stated preferences*) or (ii) how many other users that have bought the product (*revealed preferences*). In this paper, we compare the causal effects on demand from these two signals based on data from the biggest marketplace for Android apps, Google play. This data consists of daily information, for 42 consecutive days, of more than 500,000 apps from the US version of Google play. Our main result is that consumers are much more responsive to other consumers' revealed preferences, compared to others' stated preferences. A 10 percentile increase in displayed average rating only increases downloads by about 3%, while a 10 percentile increase in displayed number of downloads increases downloads by about 25%.

1. Introduction

The existing economic literature on consumer peer effects and electronic word of mouth (eWOM) has largely focused either on what other consumers (or experts) say about products, or to what extent other consumers buy the products. Economists tend to think that "talk is cheap" and that true preferences are reflected by actions rather than words. From this perspective, it seems natural that consumers' actual choices should impact others more than their stated tastes. In this study, we compare the effects of these two quality related signals: what people say about a product and whether a product is in high demand. In particular, we will ask which type of signal that has the most influence on potential consumers: other consumers' stated (eWOM) or revealed preferences (observational learning).[1]

* Corresponding author.
E-mail addresses: per.engstrom@nek.uu.se (P. Engström), eskil.forsell@phdstudent.hhs.se (E. Forsell).

[1] Chen et al. (2011) and Li and Wu (2013) study similar questions. However, their identification strategies are very different from ours.

https://doi.org/10.1016/j.jebo.2018.04.009
0167-2681/

The online revolution has facilitated a dramatic increase in the commercial use of user feedback systems. For almost any type of product or service – e.g. a vacation, a smart phone or a movie – it is now possible to easily access a host of relevant measures of its popularity, such as online user reviews, average ratings and number of users.

Researchers are becoming increasingly interested in the workings of eWOM and consumer peer effects.[2] A central question is to what extent these quality related signals actually affect prospective consumers. Our findings indicate that consumers react to both stated and revealed preferences but that their reaction to revealed preferences is much stronger.

Identification of causal effects like these is tricky. Ratings and other user generated content, from a specific site, is not the only source of information that may influence consumers. Information about a product's quality and popularity may also come from other sources such as friends, experts or bloggers. High quality of a product may hence induce both high demand, and high user rating, which confounds a causal interpretation of the association between user statistics and demand indicators. A large number of studies have therefore used experimental and quasi-experimental setups to measure the causal effects of eWOM and other related peer effects.[3,4]

The first part of our analysis concerns the effect of others' stated preferences. In order to identify the effect of average user ratings, we utilize the specific way that *Google play* presented the ratings of its apps at the time our data was collected. In the list view the consumer only observed the average user rating rounded off to the closest half star. A single rating ranged from 1 to 5 stars, in whole star increments, which meant that the presented average rating only took on values in the range: $(1.0; 1.5; 2.0; \ldots; 4.5; 5.0)$. An app that had the exact rating 3.249 would thus be rounded down to 3.0, while an app with exact rating 3.251 would be rounded up to 3.5. Standard results from the regression discontinuity design (RD) literature states that if the app distribution around the thresholds $(1.25; 1.75; \ldots; 4.75)$ is continuous, the causal effect of presented average ratings will be identified. Intuitively we may think of each threshold as a local randomized experiment since apps slightly below and slightly above are, under some conditions, almost identical in expectation. This specific application of RD was pioneered in two recent studies that, independently, applied the same technique on the restaurant review site, Yelp.com. This site presents average user ratings in a way that is similar to that of Google play. The two previous studies find that higher user ratings leads to higher reservation rates (Anderson and Magruder, 2012) and higher revenue (Luca, 2011).

The RD strategy has a particular advantage in this type of setting, where the discontinuity thresholds appear repeatedly throughout the whole range of the forcing variable (average rating in our case). Formally an RD estimate has a local average treatment effect (LATE) interpretation around the identifying threshold. If the threshold appears in a region where mostly "unrepresentative" elements locate, a LATE estimate may not be a good proxy for the average treatment effect (ATE), which is often more relevant. The fact that we use 8 different thresholds, spread out evenly throughout the whole range of average rating, makes the estimated effect more generalizable.

This first part of our study, where we study stated preference, makes several contributions to the literature on eWOM. First of all, our analysis is based on a much larger dataset than the two previous studies that use a similar identification strategy. Luca (2011) uses 854 restaurants (quarterly data) in his main RD specification while Anderson and Magruder (2012) have access to data from 328 restaurants (daily data) for their main outcome variable (reservations). We use a dataset from *Droidmeter* that includes publicly available daily information for all apps on Google play. Our main RD specification is based on data for 42 consecutive days covering about 50,000 apps.[5] Our large dataset allows for a rich set of RD specifications and robustness checks.

An additional contribution is that we study a novel market that is interesting in itself. The app-market is global and is growing very rapidly: from Google play's introduction in 2008 more than 25 billion apps had been downloaded by September 26, 2012. The one billion threshold was reached as late as in 2010 which indicates a very rapid increase since then.[6] Furthermore, given Google play's position as the official Android marketplace, the ratings on Google play are potentially a very important source of information for consumers.

We find that the causal effect of an additional half star corresponds to a 0.15–0.20 per mille point increase in the daily rating rate (daily per mille change in number of ratings). We use the daily rating rate as a proxy for the daily download rate as we do not have exact data for the latter measure. The baseline rating rate is 4.25 per mille per day. Given this, the per mille point increase translates into a relative effect of a 4–5% increase in the daily rating rate as a result of an additional half star.

As a robustness check we also perform the analysis with actual downloads, measured in intervals (more on this below), as the outcome variable. The estimated effect is similar (but less precisely estimated) which indicates that the daily rating rate is a good proxy for the daily download rate.

[2] For a seminal theoretical (economics) contribution see Avery et al. (1999) and for a recent survey see Chan and Ngai (2011).

[3] For the effects of expert reviews see: Reinstein and Snyder (2005) (movies), Friberg and Grövist (2012) (wine) and Hilger et al. (2011) (wine). For general peer effects see: Duflo and Saez (2002) and Sorensen (2006) (retirement decisions/health plan); Hesselius et al. (2009) (sickness absence); Moretti (2011) (movies). And for average user rating's effect on demand see: Melnik and Alm (2002), Jin and Kato (2006), Resnick et al. (2006), Lucking-Reiley et al. (2007), Cabral and Hortacsu (2010) (ebay auctions) and Chevalier and Mayzlin (2006) (books).

[4] For the effects of observational learning, see Bandura (1977) for a seminal psychological contribution, and Banerjee (1992) and Bikhchandani et al. (1992) for seminal economics contributions. Prominent empirical studies that isolate the causal effects of other's choices include, Cai et al. (2009), Salganik et al. (2006) and Tucker and Zhang (2011).

[5] The total number of apps available at Google play is more than 10 times larger than this. However, most of the apps have very few downloads, and even fewer ratings, which makes them unsuitable to include in the RD analysis.

[6] Source: the official android blog (http://officialandroid.blogspot.com/).