

Network effects in two-sided markets: why a 50/50 user split is not necessarily revenue optimal

Sebastian Voigt¹ · Oliver Hinz¹

Received: 22 April 2014 / Accepted: 17 June 2015 / Published online: 21 July 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Our study applies empirical scrutiny to the network effects of a leading European online dating platform. While one might expect equal gender representation on such a platform to yield the best user experience and the highest revenue per user, our analysis shows that the platform requires only 36.2 % of its user base to be female to maximize revenue, primarily because women exert stronger positive cross-side network effects on men than vice versa; this optimum results in 17.2 % higher sales than a 50/50 split. Intermediaries of two-sided markets can use our model to improve user acquisition strategies.

Keywords Network effects · Two-sided markets · Online dating · Willingness to pay · Revenue optimization

1 Introduction

In two-sided markets, an intermediary provides a platform enabling two different user groups to interact, for instance to make a transaction to satisfy their interdependent demands (Bakos and Katsamakas 2008; Ellison and Ellison 2005; Rochlet and Tirole 2003, 2006). Some two-sided online markets have expanded at a furious pace in recent years (Tucker and Zhang 2010). eBay, for example, brings together sellers and prospective buyers of different kinds of goods, Google advertisers and web users, and prosper lenders and borrowers of private loans

✉ Sebastian Voigt
voigt@emarkets.tu-darmstadt.de

Oliver Hinz
hinz@emarkets.tu-darmstadt.de

¹ Technische Universität Darmstadt, Chair of Information Systems, Electronic Markets, Hochschulstraße 1, 64289 Darmstadt, Germany

(Berger and Gleisner 2009). Eisenmann et al. (2006) provide a comprehensive list of examples for online and offline two-sided markets. Often, a neutral third party manages the platform (Yoo et al. 2002, 2007) with the commercial objective to maximize its own profits by optimally monetizing one or both user groups.

Previous research on two-sided markets indicates that the two user groups exhibit different kinds of network effects (Katz and Shapiro 1985; Liebowitz and Margolis 1994). Users may derive positive cross-side network effects (CNEs) from the participation of members on the other side of the market, which means the larger the installed user base on one side of the platform, the more attractive the service for the opposite side's users (Armstrong 2006; Li et al. 2010; Tucker and Zhang 2010). Network effects can also emerge within one user group, known as same-side network effects (SNEs). For example, a new eBay seller can have a negative effect on other sellers because he or she increases competition between sellers and may snatch away potential buyers (Kraemer et al. 2012; Li et al. 2010).

Utilizing positive network effects and mitigating negative ones is an important challenge for providers of two-sided markets. In recent years, the number of scientific studies which empirically assess such effects has been rapidly increasing (Chu and Manchanda 2013). Yoo et al. (2002) highlight the importance of identifying the magnitude of the network effects for both user groups, and state that it is difficult to estimate these effects. Knowledge of the direction and the magnitude of network effects can be used to support customer acquisition, pricing, monetization, and IT investment strategies for two-sided markets (Bakos and Katsamakas 2008; Kraemer et al. 2012; Sridhar et al. 2011).

Our empirical study examines a leading European online dating platform. Although online dating is one of the example industries in the literature on two-sided markets and seems theoretically very promising for identifying network effects (Armstrong 2006; Caillaud and Jullien 2003; Ellison and Ellison 2005; Rochet and Tirole 2003, 2006), this paper is the first to examine this industry empirically. In our case, the two user groups are heterosexual men and women. The platform enables them to search for each other, to communicate and to initiate real-life dates.

For an intermediary of a two-sided market, it is of interest to know how much future revenue and/or profit can be expected from a given user group; this datum informs effective and efficient use of limited budgets (Malthouse and Blattberg 2005; Borle et al. 2008). Prior to our research, between 35 and 41 % of the users on the platform in question were women, and the intermediary aimed to reach a 50/50 split in the near future. Naturally, one might think that equal numbers of men and women on such a platform yield the best user experience (then, every woman matches with a man) and thus the highest revenue per user for the platform intermediary. However, this does not take into consideration the differences in the user groups' willingness to pay and how CNEs and SNEs impact user behavior.

Our research aims to determine the direction and the magnitude of the different kinds of network effects on the platform and their impact on revenue, both in aggregate and of each user group individually. In addition to this empirical validation of existing theory, we propose an approach to determine the revenue-optimal ratio of men to women on the platform in light of the various existing

network effects. We show that the online dating platform in question can significantly increase its revenue with the proper balance of male and female users.

2 Network effects in two-sided markets

2.1 Previous research

Katz and Shapiro (1985) state that for many technologies, users may benefit from a growing user base. Services such as the telephone, e-mail and social networks exhibit positive network effects. These occur if two or more individuals are able to interact within this network, changing their utility of the network.

Two-sided markets have two different user groups. The intermediary provides a platform for the interaction between these groups (Berger and Gleisner 2009; Kraemer et al. 2012; Yoo et al. 2002, 2007). Usually, a user interacts only with participants from the other user group. For example, a retailer aims to sell his or her products on eBay to a certain consumer (not to another retailer), and a heterosexual man looks only for a potential female partner on Match.com.

Such two-sided markets possess network effects across user groups (CNEs) and within a single user group (SNEs). CNEs exist if the number of users on one market side influences the utility of the opposite group's users. On eBay, for example, an increased number of sellers improves the product selection and makes the platform more attractive to buyers. Similarly, having more buyers increases sellers' chance of successfully selling their items, thereby making the platform more attractive to them. SNEs exist if a user's utility is affected by the installed user base of his or her own user group (Armstrong 2006; Eisenmann et al. 2006). For example, more eBay sellers competing for a given number of potential buyers reduce each other's chances of transacting with a buyer. Depending on the investigated market, SNEs can possess either a negative effect (Dai and Kauffman 2006; Villanueva et al. 2008; Yoo et al. 2002, 2007) or a positive one (Bakos and Katsamakos 2008; Eisenmann et al. 2006) on users' utility.

To date, the literature on two-sided online markets has concentrated mainly on two research paths. The first path focuses on pricing considerations that are specific to two-sided markets experiencing network effects and examines which price structure to apply at which price level, and which user group to charge for using the services provided (Armstrong 2006; Chao and Dardenger 2013; Eisenmann et al. 2006; Jullien 2005; Parker and Van Alstyne 2005; Rochet and Tirole 2003, 2006; Rysman 2009). In the case of eBay, the platform may charge sellers, buyers, or both user groups for using the platform.

The second research path investigates the effectiveness of the intermediary's investment decisions and is more closely related to our work. Bakos and Katsamakos (2008) analyze design choices and investments such as the quality of technology, the services offered to each side, and the rules of interaction between the two user groups that create network effects in two-sided markets. Yoo et al. (2002) offer different strategies to optimize the intermediary's revenue, depending on the ownership model of the platform. Kraemer et al. (2012) find asymmetric

Table 1 Empirical studies assessing the economic results of network effects across user groups

Authors	Main research topics	Data gen.	Analysis methods	Industry/data sets	Economic dependent variables	Considers	
						CNE	SNE
Ackerberg and Gowrisankaran (2006)	NEs for banks and customers	T	Max. likelihood	ACH banking	Number of transactions	Yes	No
Brynjolfsson and Kemerer (1996)	User base on price	M	Semi-log, OLS	Spreadsheet software	Prices	Yes	No
Chacko and Mitchell (1998)	User base on company growth	M	OLS	3 technology sectors	Corporate growth rate	Yes	No
Chao and Dertinger (2013)	NEs on revenue-opt. price structure	M	Regression with IV	Portable game consoles	Associated prices	Yes	No
Chen and Xie (2007)	Implications of customer loyalty	M	Regression	Newspaper	Advertising rate, market share diff.	Yes	No
Chu and Manchanda (2013)	User base on other group's growth	T	Max. likelihood	C2C retail platform	Growth of other user group	Yes	No
Clements and Ohashi (2005)	Indirect NEs, hardw. diffusion	M	Two-stage least squares	Video game systems	Hardware and software adoption	Yes	No
Gandal et al. (2000)	Hardw. prices and softw. on diffusion	M	OLS	CD players and titles	Change in variety and sales	Yes	No
Mantrala et al. (2007)	Marketing invest on profits	M	Two-segment SURE	Newspapers	Subscriptions, ad revenue, sales	Yes	No
Nair et al. (2004)	Indirect NEs in competition	M	Monte Carlo, OLS	PDAs and software	Hardw. demand, softw. provision	Yes	No
Rysman (2004)	Importance of CNEs	M	Nested logit	Yellow pages	Consumer and advertiser demand	Yes	No
Rysman (2007)	Card usage and acceptance	M, T	Logit	Payment card transactions	Choice of favorite network	Yes	No
Shankar and Bayus (2003)	Network strength in competition	M	SEM	Video game consoles	Network strength	Yes	No

Table 1 continued

Authors	Main research topics	Data gen.	Analysis methods	Industry/data sets	Economic dependent variables	Considers	
						CNE	SNE
Wilbur (2008)	Ads on audience size and vice versa	M	Logit	TV ads	Viewer and advertiser demand	Yes	No
Asvanund et al. (2004)	Incremental value of new users	T	Logit, OLS	Peer-to-peer networks	Network value	Yes	Yes
Sridhar et al. (2011)	Optimal marketing invests with CNEs	M	DMR	Local newspaper	Demand from both sides	Yes	(Yes)
Tucker and Zhang (2010)	Installed base on listing behavior	F	Probit	Classifieds platform	Number of listings	Yes	Yes
This paper	Network effects on revenue; revenue-optimal user split	T	SURE, logit, OLS	Online dating platform	Revenue, user net gain, number of subscribers	Yes	Yes

Data generation: *F* field experiment, *M* market level data, *T* transactional company data

network effects on an eBay-like platform and assess the effectiveness of various IT and design investment features in increasing the platform value. Tucker and Zhang (2010) examine the impact of advertising the size of the user base on further participation of buyers and sellers in two-sided markets.

While these studies aim to find the intermediary's optimal strategy (e.g., in terms of revenue or platform value) to invest in IT improvements, quality, or marketing strategies, our paper searches for the revenue-optimal split between the two user groups, after empirically proving all corresponding directions of network effects, considering fixed user fees. In spite of substantial theoretical and methodological work on network effects, Wilbur (2008) as well as Kraemer et al. (2012) states that empirical analyses are still scarce due to a lack of real-life data to properly identify the effects within and across the user groups. Table 1 summarizes the results of these empirical studies and highlights the research gap and the contribution of our paper.

Our paper is related to the studies shown in Table 1, but with some notable differences. Chu and Manchanda (2013) state that previous work often focused on the benefits (or costs) a user obtains from additional users from either the same or the opposite user group, but not simultaneously from both sides. As a consequence, many studies thoroughly quantify CNEs, yet do not consider SNEs (e.g., Brynjolfsson and Kemerer 1996) or use lagged sales as a proxy for SNEs (Sridhar et al. 2011). The few existing studies that investigate both direct CNEs and SNEs use their results to model individual behavior (Tucker and Zhang 2010) or network value (Asvanund et al. 2004), while our study examines the direct impact of network effects on the intermediary's revenue, number of users and subscribers. We also notice that most studies employ data on a market level, while our study—as few others—uses a unique transactional data provided by a company. In addition, our work is the first empirical paper that studies network effects in the online dating industry.

2.2 Expected network effects on an online dating platform

2.2.1 Online dating

Three parties are involved in such a market, namely the intermediary that provides the platform and the two user groups, women and men, looking for potential partners. For reasons of simplicity (see likewise Armstrong 2006), we focus our analysis on participants looking for users of the opposite gender. Men searching for men and women searching for women are both homogeneous user groups without interaction with other user groups, and thus form a one-sided market, which is not part of our study.

Users of a dating platform clearly belong to one market side. When registering, a new user provides information on his/her gender and whether he/she is interested in meeting men or women. After this, the user typically does not change his/her role. In contrast, an eBay user can both sell and buy items at the same time, which makes it more difficult to identify the occurring network effects.

Most online dating platforms possess a ‘freemium’ pricing model. On platforms with this model, new users can create a profile for free, browse through the profiles of other users, see who visited their own profile, and send preset short messages known as ‘winks’ (such as “your picture looks nice”) to other users. However, only paying users, purchasing a subscription, can start full-text conversations with others and reply to winks. This means that at least one person (man or woman) needs to be a paying user to initiate the contact and possibly a ‘first date’ later on. A look at (each) the 100 top-grossing dating and social networking apps for iPhone (AppAnnie 2015) in the US, Japan and Germany shows that 30 out of 34 of such apps (i.e., 88 %) follow such a freemium strategy.

Kinsey et al. (1948) describe that the traditional gender role expects men to initiate contacts and women to respond. In real-life dating, women usually receive more offers from men than vice versa (Guterk et al. 1990). In addition, (Fisman et al. 2006) report from a speed dating experiment that men respond more strongly to their counterparts’ physical attractiveness. If this holds true for online dating, one can expect men’s willingness to subscribe to the paid service to be stronger than those of women.

2.2.2 CNEs

The main purpose of using an online dating platform is to look for, find, and contact potential partners of the opposite gender. Hence, users of one user group (e.g., men) care especially about the number of users on the other side (in this case: women) (Armstrong 2006; McIntyre and Subramaniam 2009; Tucker and Zhang 2010). Two-sided markets yield effects in which users in one group choose a good that affects another group’s choice of a different good (Parker and Van Alstyne 2005). For example, a woman joining the dating platform may motivate men to contact her. Thus, the utility of the platform to a paying male (female) user increases when he (she) can communicate with more women (men) (Yoo et al. 2002), which means that paying users on both market sides enjoy positive network effects from the installed user base on the opposite market side. This network effect can reflect the increased probability of finding a satisfactory match among the other side’s users (Bakos and Katsamakas 2008). Keeping the number of men constant, more women offer men a wider variety of matches (Ellison and Ellison 2005; Gehrig 1998), a greater chance of finding a unique fitting match (Caillaud and Jullien 2003), and reduce the competition between men for a specific woman (Dai and Kauffman 2006; Wang and Seidmann 1995; Yoo et al. 2002).

Previous research has shown that positive network effects leading to increased user enjoyment of the underlying service also have a positive impact on customers’ willingness to pay (e.g., Borgatti et al. 2009; Brynjolfsson and Kemerer 1996; Farrell and Saloner 1985; Katz and Shapiro 1985). Eisenmann et al. (2006) as well as Ellison and Ellison (2005) show that both user groups in two-sided markets are willing to pay more for access to a bigger network.

CNEs can also positively influence user acquisition (Villanueva et al. 2008) and retention (Chen and Xie 2007; Nitzan and Libai 2011). Single men or women are more likely to join a platform that possesses a large number of relevant users than

one that does not (Li et al. 2010). The impact of CNEs on retention or churn, however, are not trivial in the given freemium context: On the one hand, a smaller number of users of the opposite gender makes a dating service less attractive because the chances of finding a partner are lower. Thus, many users would be frustrated and may sign off earlier. On the other hand, a smaller number of users of the opposite gender may hinder people from signing up to the service (if they know in advance) or aggravate existing users' search for a fitting match, which may then lead to a longer usage lifetime for both sides.

2.2.3 SNEs

While CNEs are usually positive (but not always, as Sridhar et al. 2011 show), SNEs can be commonly found both ways in two-sided markets. For example, positive effects on each user's network utility can be found if game console owners appreciate co-playing and trading games with friends who possess the same console (Eisenmann et al. 2006), or if the platform users create a community that can provide support, collaborate, and share information with other users (Bakos and Katsamakos 2008). However, in most cases, SNEs have a negative effect on users' utility, especially in markets where users prefer fewer rivals (e.g., sellers on eBay competing for the same buyers) (Dai and Kauffman 2006; Li et al. 2010; Tucker and Zhang 2010; Wang and Seidmann 1995). Following the aforementioned idea that the utility of the online dating platform to a specific user increases when he/she can contact more users of the opposite gender, the utility of the service should decrease when there are more users of the same gender (i.e., rivals) competing for the users of the other group.

At any point in time, men and women on the platform can use the search function to check the number of users of each gender. Here, the number of users of the opposite gender is more relevant as users are looking for a partner, not a rival. Still, men/women have to option to search the platform for users of their own gender. In practice, however, they often estimate the number of rivals and the chance to find a match based on 'weak signals' such as the number of profile visits they receive from interested users of the opposite gender (Bapna et al. 2012) or the share of received messages or winks. Having too many rivals may eventually lead to fewer registrations from that user group, faster churn, and/or fewer subscriptions to the charged service. It will be interesting to see if we find substantive negative SNEs at all in our empirical study and how they differ between men and women.

3 Theoretical validation: identifying direction and magnitude of network effects

3.1 Platform and data description

In this section, we aim to empirically examine the existence and measure the magnitude of various network effects. To do so, we use customer and payment data from a leading European online dating platform that has been operational for

approximately 10 years. Users can register and create a profile for free. Every user must provide a nickname, his/her gender, place of residence and whether he or she would like to meet men or women. In addition, users can share profile pictures, age, hobbies and other personal details. All users can actively search and browse through the profiles of other (male or female) users in their vicinity. As mentioned above, we focus on the cases of men searching for women and vice versa.

The online dating platform applies the industry-typical freemium model described in Sect. 2.2: signing up and searching for other users is free of charge; however, users have to subscribe to one of the two available premium packages to be able to initiate conversations with other users. The premium packages, which we refer to as ‘Silver’ and ‘Gold’, can be purchased through a monthly subscription and can be renewed at any time. The monthly prices lie between €20 and €60 per month, depending on the length of the subscription (1–12 months; longer term packages incur a lower monthly price) and the chosen package (Gold is more expensive than Silver). The subscription prices were not changed during the entire investigated timeframe and are the same for both women and men. The Silver package allows subscribers to send messages, initiate chats and see which members are interested in their profile. The Gold package additionally highlights its subscribers in the search results and recommends users of the opposite gender in the same city with similar interests and hobbies.

Our data set covers two and a half years, from July 1st, 2010 to December 31st, 2012 (i.e., 915 consecutive days). We examine the data from one sample city of approximately 100,000 inhabitants. A total of 8923 users registered within our sample period, of which 40.8 % were women.

The analyzed payment data covers all transactions (i.e., subscriptions to a premium package) including the start and end date of the subscription, the product type (Silver or Gold), and the price. All incomplete transactions, such as fraud, chargebacks, and free upgrades (“try our Gold membership for free for 1 month!”) are excluded from the sample. The dating platform generated a total revenue of approximately €90,000 from paying users. The majority (89.3 %) of the revenue is produced by male users. Not only do men spend more money on the platform, they are also more loyal to it. The median lifetime (i.e., interval between registration and sign-off) is 102 days for a male user and 75 days for a female user. Table 2 provides a data summary and Table 3 shows the key figures on a daily basis. In addition, we show in Table 4 that Gold customers have significantly higher daily and total revenue compared to Silver users.

Table 2 Descriptive data

Number of registrations			Median user lifetime in days			Revenue share	
Women	Men	Total	Women	Men	Total	Women	Men
3640	5283	8923	75	102	86	10.7 %	89.3 %

Table 3 Descriptive data on a daily basis, $N = 915$ days

	Min.	Max.	Mean	Median	SD
Installed user base	1365	1924	1536.7	1505	133.0
Number of new registrations	1	22	7.38	7	3.61
Share of paying users in %	3.74	7.71	6.21	6.37	0.92
Share of women in %	35.5	41.1	38.4	38.3	1.1
User age (upon registration) in years	18	99	37.1	35	11.3
Lifetime (of the platform) in days	3835	4749	4292	4292	264.3

Table 4 Silver versus Gold subscriptions: comparison of duration and revenue

	Silver customers	Gold customers	P value (t test)
No. of subscriptions	595	41	–
<i>Revenue per day</i>			
Average	€1.07	€2.06	0.003
Median	€0.98	€1.31	
SD	€0.98	€4.01	
<i>Subscription length in days</i>			
Average	165	160	0.886
Median	93	152	
SD	214.9	132.8	
<i>Total revenue per subscription</i>			
Average	€133.95	€204.26	0.000
Median	€89.70	€199.26	
SD	€147.34	€156.57	

3.2 Model and variables

Kraemer et al. (2012) summarize that network effects in two-sided markets can be measured in several ways. Among them are choice models (e.g., Pavlou 2002; Rysman 2009; Stock and Yogo 2005), diffusion models (e.g., Gandal et al. 2000; Gupta et al. 2009; Chu and Manchanda 2013), vector autoregressions (e.g., Chen et al. 2001) and linear regressions (e.g., Hendel et al. 2007; Seamans and Zhu 2013). To account for the specifications of both user groups, we estimate simultaneous equation models (SURE; seemingly unrelated regression equations), as used by Mantrala et al. (2007) and Sridhar et al. (2011). To address the challenges that outliers pose for some statistical models, we use the Huber–White sandwich estimators (Huber 1967; White 1980) in all our models, thereby obviating minor concerns about the potential failure to meet assumptions, such as normality, heteroskedasticity, or observations that exhibit large residuals, leverage, or influence.

3.2.1 Dependent variables

For the purposes of our study, we consider revenue maximization on a per-user level and in total to be the primary economic variable, and aim to assess to what extent network effects (both CNEs and SNEs) describe the investigated platform's total revenue within a given timeframe. We break down the activity and revenue data on a daily basis. When a user subscribes to a premium package, we split the relevant revenue evenly over the entire subscription period. An example: a free user subscribes to a premium package from January 1st, 2012 to March 31st, 2012 (i.e., 90 days) for a total of €180, and returns to using the service for free afterwards. In our data set, he/she shows daily revenue of €2 in these 3 months, and daily revenue of zero before and after. Using this approach, revenue can be stated (for any day) as the product of the average revenue per user and the installed base (per-user group). In our regression models, we will consecutively check for network effects, first describing the daily revenue per user (*DailyRevenuePerWoman/Man*), second the net user gains (*NetGainWomen/Men*, i.e., variation of the installed base compared to the previous day), and finally the total revenue (*DailyRevenueAllWomen/AllMen/AllUsers*).

3.2.2 Independent variables

Most research models (e.g., Armstrong 2006; Bakos and Katsamakas 2008; Fudenberg and Tirole 2000; Katz and Shapiro 1985; Pang and Etzion 2012; Yoo et al. 2002) consider network effects to be linear in the size of the relevant user base. For our models in Sect. 3.3, we also employ a linear specification of network effects, counting the number of active *Men* and *Women* as the relevant user bases. Later, in Sect. 4, we use a modified model to ascertain the optimal user split between men and women.

According to the intermediary, most dating customers register in the evening and need some time setting up their profile and uploading appropriate profile pictures. We therefore assume that they start affecting other users with a time lag of 1 day (see likewise Chu and Manchanda 2013); *Men* and *Women* are therefore the number of users at the end of the previous day. In addition, we consider several control variables such as the platform lifetime in days as well as dummy variables for extraordinary TV events, seasonality, and major updates to the game. These dummy variables are set at 1 if applicable to a certain case, and 0 if not. For example, *Update2* went live on day 4088; all cases prior to the update have been labeled with 0 and with 1 as of that day. During the sample period of two and a half years, the platform underwent ten permanent game updates such as design changes and the introduction of new features. Table 5 describes the independent variables used in our model.

3.3 Identification of CNEs and SNEs

Our first analysis investigates how network effects describe the average daily revenue per user. The employed SURE model treats the average

Table 5 Description of the independent variables

Covariate	Description	Min	Max	Median	SD
<i>Women</i>	Installed base of female users on the previous day	506	782	572	63.3
<i>Men</i>	Installed base of male users on the previous day	832	1144	931	73.1
<i>PlatformLifetime</i>	Lifetime of the platform since launch (in days)	3835	4749	4292	264.3
<i>Update1</i>	Bug fixes, selected inactive users deleted	0	1	1	0.34
<i>Update2</i>	Payment website update	0	1	1	0.45
<i>Update3</i>	New payment website	0	1	1	0.49
<i>Update4</i>	Introduction of new flirt game (1)	0	1	1	0.50
<i>Update5</i>	New registration process	0	1	0	0.48
<i>Update6</i>	Introduction of monthly billing (step 1)	0	1	0	0.46
<i>Update7</i>	Introduction of monthly billing (step 2)	0	1	0	0.43
<i>Update8</i>	New Internet law implemented for payment website	0	1	0	0.37
<i>Update9</i>	Monthly billing complete	0	1	0	0.37
<i>Update10</i>	Introduction of new flirt game (2)	0	1	0	0.11
<i>TVevent1</i>	UEFA EURO 2010	0	1	0	0.11
<i>TVevent2</i>	FIFA World Cup 2012	0	1	0	0.16
<i>Winter</i>	Season	0	1	0	0.41
<i>Spring</i>	Season	0	1	0	0.40
<i>Summer</i>	Season	0	1	0	0.46
<i>Fall</i>	Season	(Omitted because of collinearity)			

DailyRevenuePerWoman and *DailyRevenuePerMan* (both in eurocent) as dependent variables. We estimate several models to ensure the robustness of our results. We begin considering only the number of users of the same gender as independent variables (SNEs; model 1) and successively include additional parameters: the number of users of the opposite gender (CNEs; model 2), platform parameters (model 3), and eventually seasonal parameters (complete model 4). Table 6 summarizes the results. We detect no change of algebraic signs for the significant variables from one model to another and thus conclude that our findings are robust. We emphasize that these results are only of descriptive nature and we can only assume causality due to the strong theoretical background available in the domain of network effects.

As we expected, we see that male users generate a much higher basic daily revenue (constant is 17.03, $p < 0.01$) compared to female users (8.35, $p < 0.01$). Without considering any network effects, adding more men to the platform would therefore be much more remunerative than adding additional women. However, our model also finds support for negative SNEs on both sides: we can see that the installed base of female users *Women* is negatively correlated ($p < 0.01$) with *DailyRevenuePerWoman*, as *Men* is with *DailyRevenuePerMan*. We can also find positive correlations between *Men* and *DailyRevenuePerWoman* as well as between *Women* and *DailyRevenuePerMan*. Both are highly significant ($p < 0.01$) and

Table 6 Results from SURE model (dependent variables: *DailyRevenuePerWoman* and *DailyRevenuePerMan* in eurocent)

Independent variables	Dependent variable: <i>DailyRevenuePerWoman</i>				Dependent variable: <i>DailyRevenuePerMan</i>			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Installed base								
<i>Women</i>	-0.00233***	-0.00327***	-0.00324***	-0.00495***		0.01414***	0.00798***	0.00664***
<i>Men</i>		0.00099***	0.00203***	0.00274***	-0.00819***	-0.01917***	-0.00772***	-0.00831***
Platform parameters								
<i>PlatformLifetime</i>			-0.00161***	-0.00172***			-0.00158***	-0.00177***
<i>Update1</i>			0.24203***	0.13408			1.92923***	1.177824***
<i>Update2</i>			0.62657***	0.6469***			0.03142	-0.09009
<i>Update3</i>			0.2999***	0.24713***			0.41002***	0.12,682
<i>Update4</i>			-0.20911***	-0.16,894***			0.21692**	0.56356***
<i>Update5</i>			0.18542***	0.18566***			-0.87221***	-0.85609***
<i>Update6</i>			-0.55129***	-0.58451***			-0.32383***	-0.3563***
<i>Update7</i>			0.57989***	0.55077***			0.95582***	0.83301***
<i>Update8</i>			0.08138***	0.10276***			0.36529***	0.24856***
<i>Update9</i>			0.60853***	0.62466***			-0.83949***	-0.54311***
<i>Update10</i>			0.27583***	0.32232***			0.5254***	0.58091***
Seasonal parameters								
<i>TVevent1</i>				-0.5141***				-1.17277***
<i>TVevent2</i>				0.11135**				0.02654
<i>Winter</i>				-0.08727**				0.05107
<i>Spring</i>				-0.07847*				0.07906
<i>Summer</i>				-0.02351*				0.41086***
<i>Constant</i>	2.77099***	2.23918***	7.3951***	8.34538***	14.97672***	16.99564***	14.79339***	17.02908***
<i>Number of observations (days)</i>	914	914	914	914	914	914	914	914

Table 6 continued

Independent variables	Dependent variable: <i>DailyRevenuePerWoman</i>				Dependent variable: <i>DailyRevenuePerMan</i>			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
R^2	0.1199	0.1259	0.5316	0.5485	0.3664	0.5229	0.7018	0.7444

* Significant at the 10 % level

** Significant at the 5 % level

*** Significant at the 1 % level

support positive CNEs. Looking at the magnitude of the network effects, we see that the positive CNEs that women exert on men are stronger than vice versa (0.00664 vs. 0.00274). Moreover, the negative SNEs effected by women are weaker than those by men (-0.00495 vs. -0.00831). While such positive CNEs could be expected, it is interesting to see that we find significant negative SNEs in both cases. Users—and primarily men—are indeed affected by stronger competition, which leads to reduced user expenditures on the focal service.

Next, we estimate a SURE model with *NetGainWomen* and *NetGainMen* as dependent variables. For each day, *NetGainWomen/Men* describes the change of the installed user base (per-user group) compared to the previous day (i.e., new registrations minus churners). Table 7 shows the results.

While we cannot find any significant CNEs (in the final model), we observe significant negative SNEs on both user sides. The more women (men) on the platform, the higher the number of churning women (men). We interpret this result as a competition effect that strengthens the negative SNEs we have seen regarding *DailyRevenuePerUser*: in case of strong competition, users do not only tend to stay free users, but they are also more likely to leave the platform. We do not observe positive reputation or popularity effects (i.e., the site growing faster as prospective customers learn that more people are using it; see Table 8 for a respective analysis).

These results are especially interesting as they indicate that each user group has a reasonable maximum size. With additional users, it becomes increasingly hard (and probably expensive) for the intermediary to acquire and keep users of a certain gender. At such a point, it may become more effective to acquire new users of the opposite user group (which brings us to the determination of the optimal split between men and women in Sect. 4.2).

We will now examine the impact on total daily revenue that additional *Women* and *Men* have. Table 9 displays the results of this analysis. Consistent to our previous analyses, we employ a SURE model to estimate *DailyRevenueAllWomen* and *DailyRevenueAllMen*, while we apply a separate OLS regression model to estimate *DailyRevenueAllUsers*.

We can see that additional women always generate additional revenue. Despite the negative SNEs leading to lower daily revenue per woman (Table 6), higher churn of female users (Table 7) and lower total revenue from women (Table 9), more women still have a positive revenue effect because of the positive CNEs they exert. While one additional female user reduces the daily *DailyRevenueAllWomen* by €0.21, it has a positive effect on the daily *DailyRevenueAllMen* of €0.69, which eventually increases *DailyRevenueAllUsers* by €0.48 per day. We see that a user's basic willingness to pay including positive CNEs overcompensates here the negative SNEs.

On the other side, despite being the main payers (independent of network effects), increasing the number of men does not always mean additional revenue. While we can find significant positive CNEs on *DailyRevenueAllWomen*, the effect of purely adding male users to *DailyRevenueAllMen* and *DailyRevenueAllUsers* is insignificant, mostly because of the aforementioned negative SNEs. Knowing of the existence of these effects, we aim to find to revenue-optimal split of men and women in the next section.

Table 7 Results from SURE model (dependent variables: *NetGainWomen* and *NetGainMen* in number of users)

Independent variables	Dependent variable: <i>NetGainWomen</i>				Dependent variable: <i>NetGainMen</i>			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Installed base								
<i>Women</i>	-0.0052943*	-0.0187988**	-0.0647049**	-0.063719**	-0.0106566	-0.0410637	-0.0410637	-0.022364
<i>Men</i>		0.0130922**	-0.0237064	-0.0366864	0.0030014	-0.0538267**	-0.0538267**	-0.0760516**
Platform parameters								
<i>PlatformLifetime</i>								
<i>Update1</i>			-0.0016744	-0.0137804**			0.0055423	-0.0073152
<i>Update2</i>			-14.541164*	-16.75659*			-16.51846***	-17.78324*
<i>Update3</i>			0.2594894	3.189207			-1.454399	2.120249
<i>Update4</i>			-2.506983	-1.784541			-3.559547*	-2.754921*
<i>Update5</i>			-0.4020995	-0.5808471			-1.340044	-1.788673
<i>Update6</i>			0.2353828	1.926006			1.081561	3.278738**
<i>Update7</i>			-1.83712**	0.2069084			-0.9842817	2.111515*
<i>Update8</i>			4.191521***	4.927273***			3.126861*	3.290054
<i>Update9</i>			1.401801	2.04894			-1.848275	-1.146902
<i>Update10</i>			-1.101929	-6813836			0.7950864	1.252844
<i>Update10</i>			-1.498372	-2.560496			-0.5833106	-2.264057
Seasonal parameters								
<i>TVevent1</i>				-5.24849				-0.0268828
<i>TVevent2</i>				0.5644956				1.164045
<i>Winter</i>				1.771118*				2.985986***
<i>Spring</i>				-0.6904099				-0.5966981
<i>Summer</i>				-0.3859537				-0.001084
Constant	3.017029*	-1.380181	81.15897**	142.6788*	6.891396*	68.70649*	68.70649*	129.7221*
Number of observations	914	914	914	914	914	914	914	914

Table 7 continued

Independent variables	Dependent variable: <i>NetGainWomen</i>				Dependent variable: <i>NetGainMen</i>			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
R^2	0.0003	0.0070	0.0932	0.1046	0.0001	0.0050	0.0898	0.1037

* Significant at the 10 % level

** Significant at the 5 % level

*** Significant at the 1 % level

Table 8 Results from SURE model (dependent variables: *NetGainWomen* and *NetGainMen*)

Independent variables	Dependent variable: <i>NetGainWomen</i>			Dependent variable: <i>NetGainMen</i>		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Installed base (total)						
<i>Users</i>	-0.0016396	-0.0443007***	-0.0497915*	-0.0033078	-0.0474156***	-0.0500245*
Platform parameters						
<i>PlatformLifetime</i>						
<i>Update1</i>		-0.0020917	-0.0148538**		0.0056722	-0.0051835
<i>Update2</i>		-15.03112***	-16.96543*		-16.36608***	-17.36849*
<i>Update3</i>		0.2033772	3.326949*		-1.436931***	1.846689
<i>Update4</i>		-2.532174**	-1.657824		-3.551705***	-3.006583*
<i>Update5</i>		-1.004848	-0.894958		-1.152406	-1.164841
<i>Update6</i>		1.528642	2.69635*		0.6789633	1.748814
<i>Update7</i>		-1.40788	0.6638192		-1.117906	1.204078
<i>Update8</i>		4.228326***	5.070342***		3.115403**	3.005915
<i>Update9</i>		1.447247	2.048838		-1.862423	-1.146699
<i>Update10</i>		-1.505206	-0.7951453		0.9206287	1.478778
<i>Update10</i>		-1.061267	-2.555395		-0.7193835	-2.274187
Seasonal parameters						
<i>TVevent1</i>			-4.914585			-0.6900263
<i>TVevent2</i>			0.3246244			1.640435
<i>Winter</i>			2.176263**			2.181358**
<i>Spring</i>			-0.5037326			-0.9674439
<i>Summer</i>			-0.2441007			-0.2828075
Constant	2.406945	90.62937***	151.0396**	4.975004	65.75831**	113.1174
Number of observations	914	914	914	914	914	914

Table 8 continued

Independent variables	Dependent variable: <i>NetGainWomen</i>			Dependent variable: <i>NetGainMen</i>		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
R^2	0.0012	0.0899	0.1035	0.0041	0.0896	0.1002

Users is the total installed base (i.e., *Men* + *Women*) at the end of the previous day

* Significant at the 10 % level

** Significant at the 5 % level

*** Significant at the 1 % level

Table 9 Results from SURE model (dependent variables: *DailyRevenueAllWomen*, *DailyRevenueAllMen* in €) and OLS regression (dependent variable: *DailyRevenueAllUsers* in €)

Independent variables	Dependent variable: <i>DailyRevenueAllWomen</i>			Dependent variable: <i>DailyRevenueAllMen</i>			Dependent variable: <i>DailyRevenueAllUsers</i>		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Installed base									
<i>Women</i>	-0.0094565***	-0.0121272***	-0.0207466***	0.1214191***	0.0768262***	0.0687561***	0.1119683***	0.0647126***	0.0480195***
<i>Men</i>	0.006163***	0.0132575***	0.0160938***	-0.1057493***	-0.002672	-0.0128498	-0.099591***	0.0105712	0.003222
Platform parameters									
<i>PlatformLifetime</i>									
<i>Update1</i>	-0.0118641***	-0.0123831***	-0.0123831***		-0.011132**	-0.0136595**		-0.0229988**	-0.0260558***
<i>Update2</i>	1.561665**	0.96887	0.96887		19.52734***	18.02646***		21.08901***	18.99346***
<i>Update3</i>	3.971721***	3.924199***	3.924199***		-2991901	-1.652735		3.672342***	2.274371*
<i>Update4</i>	1.774566***	1.370269***	1.370269***		3.258874***	0.6961245		5.033364***	2.066448***
<i>Update5</i>	-0.8161869***	-0.3984026	-0.3984026		1.27061	4.795525***		0.4548655	4.39736***
<i>Update6</i>	1.086508***	1.056412***	1.056412***		-7.615938***	-7.457788***		-6.527697***	-6.397927***
<i>Update7</i>	-2.937902***	-3.147122***	-3.147122***		-3.04403***	-3.270945***		-5.981674***	-6.415,782***
<i>Update8</i>	3.465234***	3.204152***	3.204152***		8.563072***	7.528386***		12.02826***	10.73,254**
<i>Update9</i>	0.6102715**	0.7128671***	0.7128671***		3.106759***	2.044481***		3.715196***	2.756013***
<i>Update10</i>	3.714678***	3.994181***	3.994181***		-8.110425***	-4.956301***		-4.394706***	-9.608749
<i>Update10</i>	1.422225***	1.652886***	1.652886***		4.824756***	5.087945***		6.24615***	6.739328***
Seasonal parameters									
<i>TVevent1</i>									
<i>TVevent2</i>									
<i>Winter</i>									
<i>Spring</i>									
<i>Summer</i>									
Constant	7.894175***	47.7862***	53.14111***	96.13324***	55.66326**	81.10901**	104.0282***	103.4657***	134.3192***
<i>F</i>	n/a	n/a	n/a	n/a	n/a	n/a	94.58	134.42	148.57
Number of observations	914	914	914	914	914	914	914	914	914

Table 9 continued

Independent variables	Dependent variable: <i>DailyRevenueAllWomen</i>			Dependent variable: <i>DailyRevenueAllMen</i>			Dependent variable: <i>DailyRevenueAllUsers</i>		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
R^2	0.0135	0.4211	0.4429	0.2051	0.4786	0.5530	0.1506	0.4363	0.5215

* Significant at the 10 % level

** Significant at the 5 % level

*** significant at the 1 % level

4 Practical application: determining the revenue-optimal share of men and women

4.1 Motivation and numeric example

Most two-sided markets are managed with the objective of maximizing profit generation via the paying installed user base (Yoo et al. 2002, 2007). Due to budgetary constraints, intermediaries are only able to acquire and serve a finite number of users. Such intermediaries may fail to maximize their revenue if they do not consider the network effects present on their platform. To demonstrate how such circumstances can lead to mismanagement of the platform, we will now examine a numerical example using the results from our previous analyses.

In Sect. 3.3, we found that male users spend more money on average than female users, but female users carry an additional indirect revenue potential because the positive CNEs they exert on revenue generated per male user are stronger than vice versa. In addition, the existing negative SNEs are stronger for men than for women. We apply these results to the simplified numerical example in Table 10.

For our example, we assume the network provider possesses a budget to acquire 100 users of any gender and is looking for the split between men m and women w yielding the highest overall revenue. Total revenue equals the sum of the revenue generated by both men and women:

$$Rev_{m,w} = m \text{ prob}_m \text{ fee}_m + w \text{ prob}_w \text{ fee}_w \quad \text{with } m + w = 100. \quad (1)$$

In this stylized two-sided market, men are more likely to become paying users than women ($\text{prob}_m = 6$ vs. $\text{prob}_w = 2$ %). In both groups, paying users pay the same average fee ($\text{fee}_m = \text{fee}_w = \text{€}100$). This means an average man generates revenue of $6\% \cdot \text{€}100 = \text{€}6$, while a woman generates on average only $2\% \cdot \text{€}100 = \text{€}2$. An intermediary who does not consider network effects would thus conclude that they should only acquire men as users and not a single woman.

This strategy, however, seems clearly questionable as a dating platform without women offers men no reason to become paying users on. We will now consider the impact of CNEs and SNEs upon the basic purchase likelihood. As shown in the following quadratic equation, each user's expected revenue is influenced by positive CNEs from all users of the opposite gender and negative SNEs from all other users of the same gender.

Table 10 User characteristics in a fictitious two-sided dating market

Variable name	Description	Group 1: men (m)	Group 2: women (w)
$\text{Prob}_m/\text{prob}_w$	Basic probability to become a paying user	6 %	2 %
$\text{Fee}_m/\text{fee}_w$	Avg. fee per paying user	€100	€100
$\text{CNE}_m/\text{CNE}_w$	Positive CNEs on other user group's revenue per user	€0.02	€0.06
$\text{SNE}_m/\text{SNE}_w$	Negative SNEs on same user group's revenue per user	−€0.015	−€0.01

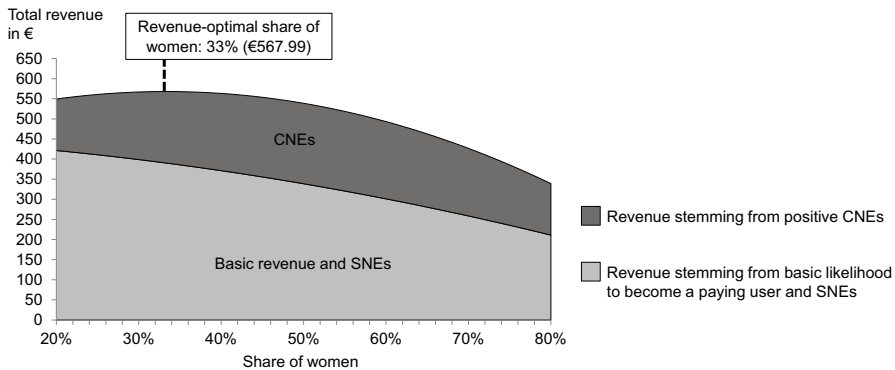


Fig. 1 Revenue-optimal share of women in the numerical example

$$\begin{aligned}
 Rev_{m,w} &= m \cdot (\text{prob}_m \text{fee}_m + \text{CNE}_w w + \text{SNE}_m \cdot (m - 1)) + w \cdot (\text{prob}_w \text{fee}_w \\
 &\quad + \text{CNE}_m m + \text{SNE}_w \cdot (w - 1)) \quad \text{with } m + w \\
 &= 100.
 \end{aligned}
 \tag{2}$$

Differentiating the revenue formula (2) with respect to m and setting the derivate to zero allows to determine the revenue-optimal split of male and female users (we expand this process in Sect. 4.2). Figure 1 shows the total revenue in our example of all male and female users combined when considering user split-dependent network effects. Changing the share of women (i.e., the horizontal axis in Fig. 1) shows two effects: First, as men generally have a higher probability for becoming paying users, the revenue stemming from this basic likelihood is highest with more men, even in light of the negative revenue impact of SNEs. Second, an elevated proportion of female users exerts CNEs, which have the highest positive revenue impact at 50 % of the user base. The CNE-induced revenue curve follows an inverted U-shaped form, and the revenue surplus is incrementally reduced with a lower/higher share of women. Taken together, these effects lead to a revenue optimum at circa 67 % men and 33 % women. Given the same total number of users, the intermediary’s revenue are 5.4 % higher than in a 50/50 user split (€567.99 vs. €538.75).

This simple example demonstrates that intermediaries with knowledge of network effects can make better business decisions, for example by identifying and profitably acquiring those customers who promise the highest revenue contribution to the network; such an identification enables the intermediary to optimize the user split on their platform.

4.2 User split optimization for the investigated platform

We will now use authentic data to empirically determine the optimal ratio of male to female users with regard to the highest possible revenue generation. Our approach is usable for platform intermediaries in two-sided markets that aim at an effective use of their limited user acquisition budgets.

In this section, we use the same data set as in Sect. 3 with slightly adjusted variables in the OLS model. First, we now use *DailyRevenuePerUser* as the

dependent variable (i.e., $DailyRevenueAllUsers/Users$) to ascertain the proportion of women that yields the best results. We also replace the previously used absolute user numbers *Men* and *Women* with a dependent variable which represents the proportion of female users, both in linear and quadratic form (*ShareOfWomen* and *ShareOfWomenSquared*, each as a percentage of total users). All other variables remain the same as those in Table 6. Table 11 shows the results of the employed regression model.

Table 11 shows positive linear and negative quadratic influence of the proportion of female users on total revenue per user, which yields a single point of female-

Table 11 Results from OLS regression

Independent variables	Dependent variable: <i>DailyRevenuePerUser</i>		
	Model 1	Model 2	Final model 3
User split			
<i>ShareOfWomen</i>	334.2746***	312.4332***	346.8351***
<i>ShareOfWomenSquared</i>	-468.437***	-428.2298***	-479.6163***
Platform parameters			
<i>PlatformLifetime</i>		0.00136***	0.00159***
<i>Update1</i>		0.06716	0.29307***
<i>Update2</i>		0.13269*	-0.27801***
<i>Update3</i>		-0.25703***	-0.17709**
<i>Update4</i>		-0.36184***	-0.14882
<i>Update5</i>		-0.5221***	-0.70081***
<i>Update6</i>		0.29947***	-0.05956
<i>Update7</i>		-0.68634***	-0.53678***
<i>Update8</i>		0.33535**	0.40668**
<i>Update9</i>		0.53921***	0.69158***
<i>Update10</i>		0.21367	0.36379*
Seasonal parameters			
<i>TVevent1</i>			0.38896*
<i>TVevent2</i>			-0.07567
<i>Winter</i>			-0.08448
<i>Spring</i>			0.49053***
<i>Summer</i>			0.24884***
Constant	-56.39442***	-58.46802***	-65.29321***
<i>F</i>	162.70	347.90	258.91
Number of observations	1005,275	1005,275	1005,275
R^2	0.0002	0.0034	0.0035
Optimum (highest revenue dep. on share of women)	35.7 %	36.5 %	36.2 %

Dependent variable: *DailyRevenuePerUser* (in eurocent)

* Significant at the 10 % level

** Significant at the 5 % level

*** Significant at the 1 % level

dependent maximized revenue. The regression formula (3) slightly differs from the equation used in the numeric example (2) as we are not restricted to a total of 100 users and have different variables compared to the previously used, simplified model. Differentiating the abridged regression formula (3) shown below with respect to *ShareOfWomen* and setting the derivate to zero (4) yields a critical point at a proportion of female users of 36.2 % (5). As we can easily see from Eq. (4), the function's second derivative is negative and the identified point is therefore a local maximum in terms of revenue: the desired revenue-optimal proportion of female users.

$$\text{Daily Revenue Per User} = \beta_0 + \beta_1 \text{ Share Of Women} + \beta_2 \text{ Share Of Women Squared} + \dots + \varepsilon \quad (3)$$

$$\frac{d \text{ Daily Revenue Per User}}{d \text{ Share Of Women}} = 346.8351 + 2 \cdot (-479.6163) \cdot \text{Share Of Women} = 0 \quad (4)$$

$$(\text{Optimal}) \text{ Share Of Women} = \frac{-346.8351}{-2 \cdot 479.6163} = 36.2\% \quad (5)$$

Similar to the previous approach, the logit regression model in Table 12 aims to assess the proportion of women yielding the highest share of paying users. In both analyses, we find similar results: a share of women of 34.6 % leads to the highest share of premium subscribers, while 36.2 % maximizes the intermediary's revenue. Below this optimum, additional female users contribute higher utility to the overall network (through positive network effects) than men, leading to either more subscribers or additional revenue. When the optimum is surpassed, adding more men to the platform will be more valuable than adding more women.

Figure 2 illustrates the relationship between the share of female users and its correlation to the expected total revenue. Like in our numerical example, the curve has an inverted U-shape. As shown previously in Table 3, the proportion of women fluctuated between 35.5 and 41.1 % over the course of our 915-day observation period. However, our results show that, given a constant number of users, the intermediary's total daily revenue will be approximately 2 % higher with a women's share of 36.2 % compared to the historical maximum of 41.1 %, and a full 17.2 % higher compared to a 50/50 split. Our results display that the intermediary in our study should abandon its previous goal of reaching a 50/50 user split.

5 Discussion

5.1 Research contributions

Our study investigates users' spending behavior on an online dating platform. Despite progress in gender equality, findings from more than 60 years ago (Kinsey et al. 1948) still seem to apply. Asymmetric societal norms still exist in people's mate searching behavior that prevent women from making the first move (Bapna

Table 12 Results from logit regression analysis (dependent variable: *IsPayer*)

Independent variables	Dependent variable: <i>IsPayer</i>		
	Model 1	Model 2	Model 3
User split			
<i>ShareOfWomen</i>	168.1591***	198.7777***	219.8497***
<i>ShareOfWomenSquared</i>	−232.8991***	−285.3661***	−317.5482***
Time parameters			
<i>PlatformLifetime</i>		0.0007249***	0.0008232***
<i>UserLifetime</i>		−0.000601***	−0.0006021***
<i>Update1</i>		0.0038217	0.0220544
<i>Update2</i>		0.0106529	−0.0783018**
<i>Update3</i>		−0.0309305	0.0075203
<i>Update4</i>		−0.1802073***	−0.1577811***
<i>Update5</i>		−0.0754961**	−0.1610922***
<i>Update6</i>		0.1274491***	0.0378605
<i>Update7</i>		−0.1883186***	−0.1205443***
<i>Update8</i>		0.1423438**	0.1926842***
<i>Update9</i>		0.1261103**	0.1554679***
<i>Update10</i>		0.0661558	0.0442279
Seasonal parameters			
<i>TVevent1</i>			0.2236641***
<i>TVevent2</i>			−6.59e-06
<i>Winter</i>			0.0374548*
<i>Spring</i>			0.1818007***
<i>Summer</i>			0.0576748***
Constant	−33.53169***	−40.6881***	−44.57342***
Number of observations	1005,275	1005,275	1005,275
Pseudo R^2	0.0021	0.0172	0.0175
Optimum (highest share of payers dep. on share of women)	36.1 %	34.8 %	34.6 %

* Significant at the 10 % level

** Significant at the 5 % level

*** Significant at the 1 % level

et al. 2012; Fisman et al. 2006; Piskorski 2012). This aspect certainly accounts for the results of our study, where we could see men are more likely willing to pay for online dating services assuming a sufficient installed user base of women than vice versa.

Our study identifies the existence and the magnitude of the various network effects in this market. Estimating a SURE model reveals positive CNEs in both directions: having more female users increases the average revenue per male user and the total revenue generated by men, and vice versa. A larger choice set of potential partners increases the chance of a free user finding someone he/she is

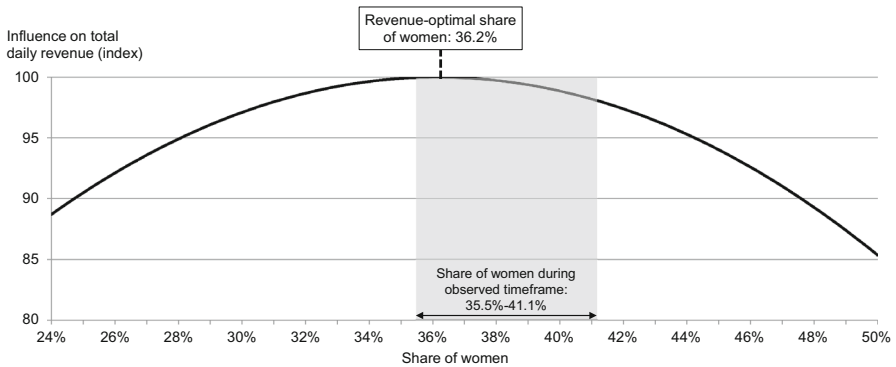


Fig. 2 Revenue-optimal share of women

interested in and eventually subscribing to a premium package that allows him/her to send messages to other platform users.

Furthermore, we find negative SNEs for both men and women. Increasing the number of women reduces the revenue per woman, leads to a higher churn of female users, and reduces the total revenue generated from the installed base of women. Increasing the number of men reduces the revenue per man and leads to a higher churn of male users while there is no significant negative impact on total revenue. In the given freemium model, free users might be deterred from purchasing the premium package if there are too many people competing for a given number of users of the opposite gender. Existing premium users who send messages to potential partners may receive fewer answers if there is too much competition, and become frustrated. As a consequence, the share of customers who renew their subscription may decrease and the number of users leaving the platform may increase.

In addition, we observe that the positive CNEs that women exert are stronger than the SNEs on the women's side. As long as the number of male users clearly exceeds female ones, more women always mean extra revenue. For men, we could not find statistical support in this case. As a combination of positive CNEs and substantial negative SNEs, the total revenue impact of solely increasing the number of male users is not significant.

5.2 Practical contributions

We have shown that operators of two-sided markets who aim to optimize their revenue can use information on network effects to acquire, manage and monetize their user base more effectively. We present an approach that determines the optimal split between the two user groups in terms of revenue and the number of premium subscribers. We find a positive linear and a negative quadratic influence of the proportion of women on revenue and number of subscribers. Thus, the utility of incremental women (who are the user group that exert the strongest CNEs) for the entire network follows an inverted U-shape. This is in line with the work by Bapna and Umyarov (2012) who discovered in an online social music network that the

strength of influence decreases with a user's number of friends. Our model can be easily extended, for example to other regional markets, or to any other two-sided platform with network effects.

We find that a female proportion of the user base of 34.6 % leads to the highest share of premium subscribers, while the revenue-optimal proportion of women was 36.2 %. As the platform's share of female users was at circa 40 % and therefore above the revenue optimum at the end of our observation timeframe, acquiring more male users promised higher future revenue at that time. Our findings are in conflict with the intermediary's initial strategy to achieve a 50/50 user split between women and men. Our results indicate that the optimal user split generates 17.2 % more revenue with the same number of users than the targeted, intuitive 50/50 split.

The company that provided us with the data used the results from our analysis to develop a decision support system that assesses the expected customer lifetime value of an additional male or female user. The system is continuously collecting information to provide an updated assessment of the revenue-optimal share of women at any time. This allows the company to identify those users who promise the highest incremental value for the platform and to adapt its customer acquisition strategy accordingly. Based on our static results from the end of our observation timeframe, the network intermediary relocated its marketing budgets to acquire more male users; it adapted the costs per install (CPI) for new users according to the CLV projection for new users and launched a marketing campaign that primarily targeted male singles.

5.3 Limitations

The limitations of our study offer several avenues for interesting future research. First, there are other possible ways to measure the influence of network effects on revenue. Besides the employed SURE, OLS and logit regression models, a random-effects or fixed-effects panel data model would also be appropriate. Alternatively, a hazard model could be used to better understand the dynamics of the development. In our work, we employed a linear—and for the optimization problem additionally a quadratic—specification of network effects. Apart from these forms, logarithmic and polynomial relationships (or combined functions; Asvanund et al. 2004) between the installed user base and dependent economic variables are also possible.

Second, unobserved causes may exist that could bias the estimates of network effects (Liu et al. 2007). It is hard to imagine that omitted variables could easily reverse the assessed direction of the network effects. However, the estimated coefficients may still be biased in terms of their magnitude (Kraemer et al. 2012). We tried to employ instrumental variables but failed to identify valid orthogonal variables. This would have certainly helped us to build additional confidence in our results.

Third, the observed network effects are likely to strongly depend on the underlying price model. In our study, we observe that additional female users increase male users' willingness to pay more strongly than vice versa. Depending on the magnitude of the (asymmetric) CNEs, several researchers suggest increasing the price difference between the two user groups on a two-sided market (e.g., Strauss 1999) or to charge only one user group and give away the service to the other under

certain conditions (e.g., Armstrong 2006; Bakos and Katsamakas 2008; Caillaud and Jullien 2003)—a possible (but in practice hardly used) strategy for dating platforms that might yield a different revenue-optimal user split. Jullien (2005) provides a list of possible price models for intermediaries in two-sided markets. Empirical testing of the impacts of a price model change (e.g., moving from a subscription-based to a transaction-based pricing model) upon the revenue-optimal user split would be a worthwhile supplement to the examination carried out in our study.

Lastly, not only price model changes but also price level changes may alter the network effects and thereby the revenue-optimal proportion of women in the user base of such a platform. In our case, the intermediary kept prices fixed during the entire observation period; however, a price change may result in a new optimal user split, depending on each user group's respective price sensitivity.

6 Summary and conclusion

This study's objectives were to empirically assess the influence of CNEs and SNEs on revenue in a two-sided online network and to derive the revenue-optimal split between the two user groups, men and women. Therefore, we investigated a leading online dating platform's user activity and payment data over a period of two and a half years. Our sample covered 8923 users in one city who spent approximately €90,000 by subscribing to one of the premium packages offered by the platform provider.

In general, men are more willing to pay for dating services than women (if the installed base of women is sufficiently large). In addition, we observed that both user groups (i.e., male and female users) exert positive CNEs with regard to revenue and user enrollment of the other group; however, the positive CNEs women exert on revenue generation per man are stronger than vice versa. Moreover, we identified negative SNEs which lead to lower revenue per user and an increased churn rate on a market side, when that side exclusively grows.

Operators of two-sided markets can use information regarding asymmetric network effects such as these to acquire, manage, and monetize their user base more effectively. For the online dating platform in our study, we calculated the revenue-optimal user split and found that a female proportion of the user base of 36.2 % yielded 17.2 % more revenue than a 50/50 split for the same total number of users. Our model is transferrable not only to other online dating platforms, but to all kinds of two-sided markets with network effects. Platform intermediaries can use the results from this optimization problem to develop more efficient user acquisition and monetization strategies.

7 Executive summary

In two-sided markets such as the online dating industry in question, two different user groups interact and generate various network effects, which can be either positive or negative. Users may derive positive cross-side network effects from the participation

of the other user group; for example, the more women are on a dating platform, the more attractive the service is for men. In addition, same-side network effects can (usually negatively) impact the utility of the platform for the group's users if the size of that group becomes too large. Capitalizing on positive network effects and mitigating negative ones is an important challenge for providers of two-sided platforms. In our work, we analyze activity and payment information for over 8900 online dating users over a two-and-a-half-year period. We show that positive cross-side and negative same-side network effects have a significant impact on the revenue generated per user. We use these results to determine the revenue-optimal ratio of women to men on the platform. There is a natural inclination to think that an equal number of men and women (i.e., a 50/50 split) yields the best user experience and thus the highest revenue per user for the platform intermediary. However, our analysis shows that the revenue-optimal proportion of female users on the platform is a mere 36.2%, mainly because (a) men have a higher basic willingness to pay for the service than women, and (b) women exert stronger positive cross-side network effects on the on-platform spending habits of men than vice versa. The identified optimum yields 17.2% higher revenue than the 50/50 split the platform provider initially aimed for. Academics and practitioners can use our framework to quantify network effects, determine the revenue-optimal ratio of users in any two-sided market, and develop more effective customer acquisition and monetization strategies.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Akerberg, Daniel A., and Gautam Gowrisankaran. 2006. Quantifying equilibrium network externalities in the ACH banking industry. *RAND Journal of Economics* 37(3): 738–761.
- AppAnnie. 2015. iOS Top App Charts 4 January 2015, <http://www.appannie.com/apps/ios/top/?device=i-phone>. Accessed 05 Jan 2015.
- Armstrong, Mark. 2006. Competition in two-sided markets. *RAND Journal of Economics* 37(3): 668–691.
- Asvanund, Atip, Karen Clay, Ramayya Krishnan, and Michael D. Smith. 2004. An empirical analysis of network externalities in peer-to-peer music-sharing networks. *Information Systems Research* 15(2): 155–174.
- Bakos, J.Yannis, and Evangelis Katsamakas. 2008. Design and ownership of two-sided networks: implications for Internet platforms. *Journal of Management Information Systems* 25(2): 171–202.
- Bapna, Ravi, Jui Ramaprasad, Galit Shmueli, and Akhmed Umyarov (2012): One-way mirrors and weak-signaling in online dating: a randomized field experiment. https://www.business.utah.edu/sites/default/files/media/ravi_bapna_0.pdf. Accessed 09 Jun 2015.
- Bapna, Ravi, and Akhmed Umyarov. 2012. Do your online friends make you pay? A randomized field experiment in an online music social network, NBER Working Paper Series 2012. <http://www.heinz.cmu.edu/download.aspx?id=3660>. Accessed 09 Jun 2015.
- Berger, Sven C., and Fabian Gleisner. 2009. Emergence of financial intermediaries in electronic markets: the Case of Online P2P Lending. *BuR Business Research* 2(1): 39–65.
- Borgatti, Stephen B., Ajay Mehra, Daniel J. Brass, and Giuseppe Labianca. 2009. Network analysis in the social sciences. *Science* 323: 892–895.
- Borle, Sharad, Siddharth S. Singh, and Dipak C. Jain. 2008. Customer lifetime value measurement. *Management Science* 54(1): 100–112.

- Brynjolfsson, Erik, and Chris F. Kemerer. 1996. Network externalities in microcomputer software: an econometric analysis of the spreadsheet market. *Management Science* 42(12): 1627–1647.
- Caillaud, Bernard, and Bruno Jullien. 2003. Chicken and egg: competition among intermediation service providers. *RAND Journal of Economics* 34(2): 309–328.
- Chacko, Mani, and Will Mitchell. 1998. Growth incentives to invest in a network externality environment. *Industrial and Corporate Change* 7(4): 731–744.
- Chao, Yong, and Timothy Derdenger. 2013. Mixed bundling in two-sided markets in the presence of installed base effects. *Management Science* 59(8): 1904–1926.
- Chen, Yuxin, Chakravarthi Narasimhan, and Z. John Zhang. 2001. Individual marketing with imperfect targetability. *Marketing Science* 20(1): 23–41.
- Chen, Yuxin, and Jinhong Xie. 2007. Cross-market network effect with asymmetric customer loyalty: implications for competitive advantage. *Marketing Science* 26(1): 52–66.
- Chu, Junhong, and Puneet Manchanda. 2013. Quantifying cross-network effects in online consumer-to-consumer platforms, working paper, University of Michigan.
- Clements, Matthew T., and Hiroshi Ohashi. 2005. Indirect network effects and the product cycle: video games in the US, 1994–2002. *The Journal of Industrial Economics* 53(4): 515–542.
- Dai, Qizhi, and Robert J. Kauffman. 2006. To be or not to B2B: evaluating managerial choices for e-procurement channel adoption. *Information Technology Management* 7(2): 109–130.
- Eisenmann, Thomas R., Geoffrey Parker, and Marshall W. Van Alstyne. 2006. Strategies for two-sided networks. *Harvard Business Review* 84(10): 92–101.
- Ellison, Glenn, and Sara Fisher Ellison. 2005. Lessons about Markets from the Internet. *Journal of Economic Perspectives* 19(2): 139–158.
- Farrell, Joseph, and Garth Saloner. 1985. Standardization, compatibility, and innovation. *RAND Journal of Economics* 16(1): 70–83.
- Fisman, Raymond, Sheena S. Iyengar, Emir Kamenica, and Itamar Simonson. 2006. Gender differences in mate selection: evidence from a speed dating experiment. *Quarterly Journal of Economics* 121(2): 673–697.
- Fudenberg, Drew, and Jean Tirole. 2000. Pricing a network good to deter entry. *Journal of Industrial Economics* 48(4): 373–390.
- Gandal, Neil, Michael Kende, and Rafael Rob. 2000. The dynamics of technological adoption in hardware/software systems: the case of compact disc players. *RAND Journal of Economics* 31(1): 43–61.
- Gehrig, Thomas. 1998. Competing markets. *European Economic Review* 42(2): 277–310.
- Gupta, Sunil, Carl F. Mela, and Jose M. Vidal-Sanz. 2009. The value of a “free” customer. https://faculty.fuqua.duke.edu/~mela/bio/papers/Gupta_Mela_Vidal-Sanz_2009.pdf. Accessed 09 Jun 2015.
- Gutek, Barbara A., Aaron Groff Cohen, and Alison M. Konrad. 1990. Predicting social-sexual behavior at work: a contact hypothesis. *Academy of Management Journal* 33(3): 560–577.
- Hendel, Igal, Aviv Nevo, and François Ortalo-Magne. 2007. The relative performance of real estate marketing platforms: MLS versus FSBOMadison.com. *American Economic Review* 99(5): 1878–1898.
- Huber, Peter J. 1967. The behavior of maximum likelihood estimates under nonstandard conditions. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Berkeley, 221–233.
- Jullien, Bruno. 2005. Two-sided markets and electronic intermediaries. *CESifo Economic Studies* 51(2–3): 233–260.
- Katz, Michael L., and Carl Shapiro. 1985. Network externalities, competition, and compatibility. *American Economic Review* 75(3): 424–440.
- Kinsey, Alfred C., Wardell B. Pomeroy, and Clyde E. Martin. 1948. *Sexual behavior in the human male*. Philadelphia: Saunders.
- Kraemer, Tim, Oliver Hinz, and Bernd Skiera. 2012. Measuring the economic success of marketing investments in two-sided markets. In: Proceedings of the 34th INFORMS Marketing Science Conference, Boston.
- Liebowitz, S.J., and Stephen E. Margolis. 1994. Network externality: an uncommon tragedy. *Journal of Economic Perspectives* 8(2): 133–150.
- Li, Shengli, Yipeng Liu, and Subhajyoti Bandyopadhyay. 2010. Network effects in online two-sided market platforms: a research note. *Decision Support Systems* 49(2): 245–249.
- Liu, Qing, Thomas Otter, and Greg M. Allenby. 2007. Investigating endogeneity bias in marketing. *Marketing Science* 26(5): 642–650.

- Malthouse, Edward C., and Robert C. Blattberg. 2005. Can we predict customer lifetime value? *Journal of Interactive Marketing* 19(1): 2–16.
- Mantrala, Murali K., Prasad A. Naik, Shrihari Sridhar, and Esther Thorson. 2007. Uphill or downhill? *Locating the firm on a profit function*, *Journal of Marketing* 71(2): 26–44.
- McIntyre, David P., and Mohan Subramaniam. 2009. Strategy in network industries: a review and research agenda. *Journal of Management* 35(6): 1494–1517.
- Nair, Harikesh, Pradeep Chintagunta, and Jean-Pierre Dubé. 2004. Empirical analysis of indirect network effects in the market for personal digital assistants. *Quantitative Marketing and Economics* 2(1): 23–58.
- Nitzan, Irit, and Barak Libai. 2011. Social effects on customer retention. *Journal of Marketing* 75(6): 24–38.
- Pang, Min-Seok, and Hila Etzion. 2012. Analyzing pricing strategies for online services with network effects. *Information Systems Research* 23(4): 1364–1377.
- Parker, Geoffrey G., and Marshall W. Van Alstyne. 2005. Two-sided network effects: a theory of information product design. *Management Science* 51(10): 1494–1504.
- Pavlou, Paul A. 2002. Institution-based trust in interorganizational exchange relationships: the role of online B2B marketplaces on trust formation. *Journal of Strategic Information Systems* 11(3–4): 215–243.
- Piskorski, Mikdaj Jan. 2012. Social failures and social solutions: evidence from OkCupid. <http://sie-pr.stanford.edu/system/files/shared/Piskorski-SIEPR.pdf>. Accessed 09 Jun 2015.
- Rochet, Jean-Charles, and Jean Tirole. 2003. Platform competition in two-sided markets. *Journal of the European Economic Association* 1(4): 990–1029.
- Rochet, Jean-Charles, and Jean Tirole. 2006. Two-sided markets: a progress report. *RAND Journal of Economics* 37(3): 645–667.
- Rysman, Marc. 2004. Competition between networks: a study of the market for yellow pages. *Review of Economic Studies* 71(2): 483–512.
- Rysman, Marc. 2007. An empirical analysis of payment card usage. *Journal of Industrial Economics* 55(1): 1–36.
- Rysman, Marc. 2009. The economics of two-sided markets. *Journal of Economic Perspectives* 23(3): 125–143.
- Seamans, Robert, and Feng Zhu. 2013. Responses to entry in multi-sided markets: the impact of Craigslist on local newspapers. *Management Science* 60(2): 476–493.
- Shankar, Venkatesh, and Barry L. Bayus. 2003. Network effects and competition: an empirical analysis of the home video game industry. *Strategic Management Journal* 24(4): 375–384.
- Sridhar, Shrihari, Murali K. Mantrala, Prasad A. Naik, and Esther Thorson. 2011. Dynamic marketing budgeting for platform firms: theory, evidence, and application. *Journal of Marketing Research* 48(6): 929–943.
- Stock, James H., and Motohiro Yogo. 2005. Testing for weak Instruments in linear IV regression. In David W.K. Andrews and James H. Stock (eds.): identification and inference for econometric models: essays in honor of Thomas Rothenberg, 80–108. Cambridge: Cambridge University Press.
- Strauss, Steven. 1999. Marketing strategies for products with cross-market network externalities, ISBM Report 14-2000, Yale School of Management.
- Tucker, Catherine, and Juanjuan Zhang. 2010. Growing two-sided networks by advertising the user base: a field experiment. *Marketing Science* 29(5): 805–814.
- Villanueva, Julian, Shijin Yoo, and Dominique M. Hanssens. 2008. The impact of marketing-induced versus word-of-mouth customer acquisition on customer equity growth. *Journal of Marketing Research* 45(1): 48–59.
- Wang, Eric T.G., and Abraham Seidmann. 1995. Electronic data interchange: competitive externalities and strategic implementation policies. *Management Science* 41(3): 401–418.
- White, Halbert. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica. Journal of the Econometric Society* 48(4): 817–838.
- Wilbur, Kenneth C. 2008. A two-sided, empirical model of television advertising and viewing markets. *Marketing Science* 27(3): 356–378.
- Yoo, Byungjoon, Vidyanand Choudhary, and Tridas Mukhopadhyay. 2002. A model of neutral B2B intermediaries. *Journal of Management Information Systems* 19(3): 43–68.
- Yoo, Byungjoon, Vidyanand Choudhary, and Tridas Mukhopadhyay. 2007. Electronic B2B marketplaces with different ownership structures. *Management Science* 53(6): 952–961.