Distilling network effects from Steam

José Tudón*

May 4, 2022

Abstract

This paper develops a method to estimate the demand for network goods, using minimal network data, but leveraging within-consumer variation. I estimate demand for video games as a function of individuals' social networks, prices, and qualities, using data from Steam, the largest video game digital distributor in the world. I separately identify price elasticities on individuals with and without friends with the same game, conditional on individual fixed effects and games' characteristics. I then use the discrepancies between estimated price elasticities to identify the impact of social networks. I compare my method to "traditional-IV" strategies in the literature, which require detailed network data, and find similar results. A 1% increase in friends' demands, increases demand by .13%. In counterfactual simulations, I find demand increases by about 5% from a promotional giveaway to "influencers," those users in the top 1% of popularity in the network.

Keywords: Network goods, partially observed networks, direct network externalities, network effects.

1 Introduction

Even with the recent success of the empirical literature on estimating network effects,[1] two main problems need to be tackled in the field: one practical, and one

*ITAM, Río Hondo 1, Ciudad de México 01080, México (e-mail: jtudon@itam.mx). Thanks to Michael Dinerstein, Ali Hortaçsu, Mauricio Romero, Adrián Rubli, and ITAM's applied micro workshop participants. Adrián Martínez provided outstanding research assistance.

[1]See, e.g., Björkegren (2019), De Giorgi, Frederiksen and Pistaferri (2020), Ryan and Tucker (2012).

methodological. First, consider the practical problem of collecting network data. In many settings, access to network data is prohibitively expensive, because of privacy concerns, ownership, or scale (Breza et al., 2020). Second, without a proper research strategy, the reflection problem prevents the identification of direct network effects (Manski, 1993; Rysman, 2019).[2]

I study demand network effects and propose a simple identification strategy without access to detailed network data. Instead, I leverage within-consumer variation in a three-step identification process. In the first step, the "pure" price elasticity is identified from isolated individuals' behavior in the network. In the second step, the "compounded" price elasticity—which is compounded with the network effect—is identified from the rest of the individuals. And in the third step, the network effect is identified from the discrepancy between the previous two steps' price elasticities *within* consumers. Therefore, this strategy has simpler data requirements—repeated observations for an individual—than most of the literature, which requires a complete mapping of the network. I use this strategy to study the online video game industry.

The proposed method is simple and intuitive. To fix ideas, consider Becker (1991)'s framework for social interactions, where individual demand, $d_i(p, D)$, is a function of prices, and aggregate demand, $D = \sum_i d_i(p, D) = F(p, D)$. That is, aggregate demand is a function of prices and aggregate demand, where we expect $F_p < 0$, but $F_D > 0$ for positive network effects. Therefore,

$$\frac{\partial D}{\partial p} = F_p + F_D \frac{\partial D}{\partial p} \qquad \Rightarrow \frac{\partial D}{\partial p} = \frac{F_p}{1 - F_D}.$$

This paper's strategy consists of leveraging a subset of data where $F_D = 0$, so we can estimate F_p from the observed $\frac{\partial D}{\partial p}$. Given F_p, the network effect, F_D, can then be estimated from the total effect, $F_p/(1 - F_D)$, which is identified from the complementary subset of data. In the context of online video games, individual demand increases with the number of friends who own the game. In this case, F_p can be identified from individuals who do not have friends who own the same game, and $F_p/(1 - F_D)$ can be identified from the individuals with at least one

[2]When an individuals' demand for a product increases with the number of others' demand, we have a *direct* network effect (e.g., communication networks). In contrast, *indirect* network effects arise on multi-sided markets where the individuals' demand for a product in one side of the market increases with the supply of such a product from another side of the platform, which in turn increases with individuals' demand (e.g. online platforms). For instance, if a positive network effect is ignored, the price elasticity of demand will be overestimated, because the total effect of a price change is composed of a price effect *and* a network effect.

friend with the game. Thus, the estimate has a local average treatment effect interpretation which I describe below. The main point of the paper, however, is we do not need detailed network data; we only need to know "how many friends own the game." In other words, we only need "Aggregated Relational Data (ARD)" (Breza et al., 2020) to identify network effects and price elasticities.[3]

Additionally, individual demand might be a function of other variables; for instance, $d_i(p, r, D)$, where r is quality. Then, $D = F(p, r, D)$,

$$\frac{\partial D}{\partial p} = \frac{F_p}{1 - F_D}, \qquad \text{and} \qquad \frac{\partial D}{\partial r} = \frac{F_r}{1 - F_D}.$$

Thus, as long as we can identify either F_p or F_r, we can identify F_D. If we can identify both, then they must agree with their prediction of F_D. As a result, the estimation of the network effect can be improved by using the corresponding set of moments. This paper shows this is the case when estimating F_p and F_r separately for the online video game industry.

An important concern is that consumers with and without friends might be different. I address this issue by considering consumers' purchasing behavior at the individual level, and observing their decisions for many games, some of which have also been purchased by friends, some of which have not. Moreover, individual fixed effects are also considered to absorb remaining contextual and correlated effects (Hartmann et al., 2008).

The results suggest the presence of strong network externalities: if the proportion of friends with some game j increases by 1%, then the demand for game j increases by .13%. If we ignore network effects, price elasticity is significantly overestimated. Moreover, with separate price elasticities and network externalities, managers and policymakers can simulate counterfactual pricing and marketing policies. In counterfactuals, I find strong incentives to promote the game through "influencers."

I also compare this paper's strategy to "traditional" identification strategies from the literature. Some traditional strategies use detailed network data, such as second-order friends' characteristics, as instruments for first-order friends' characteristics. I arrive at qualitatively similar conclusions with these methodologies.

This paper is related to a large empirical literature on the effects of networks on demand. To name a few, Björkegren (2019), Ryan and Tucker (2012), and

[3] ARD refers to answers to questions of the form "how many of your links have trait X?"

Tucker (2008) study telecommunications demand as a function of direct network effects, and De Giorgi, Frederiksen and Pistaferri (2020) analyze their effect on consumption. Relatedly, Bertrand, Luttmer and Mullainathan (2000) and Sacerdote (2001) offer early, credible examples of peer effects on welfare participation and student outcomes.[4] In particular, the idea of achieving identification from isolated individuals goes back at least to Fortin and Boucher (2016) for the linear-in-means model. However, this paper is the first to fully develop this method, describe the outstanding econometric problems, and offer a solution to them, especially in a structural, discrete-choice model.

This paper is also related to the empirical literature that uses minimal network data. Breza et al. (2020) use Aggregated Relational Data (ARD) to recover the parameters of a network formation model. They focus on studying the network structure. I also use ARD (e.g., "how many of your friends have game j?"), but focus on quantifying direct network effects. Indeed, ARD studies have a large practical upside, because mapping whole networks is expensive and time-consuming, as described by Breza and Chandrasekhar (2019), Banerjee et al. (2013) or Chandrasekhar, Kinnan and Larreguy (2018).

This paper also contributes to the literature about "influencers." Previous work has mostly focused on how demand reacts to the interaction between influencers' characteristics and their message. For instance, Belanche et al. (2021) study the role of congruence between what influencers promote and what they consume, and Martínez-López et al. (2020) study the brand control over the message being promoted and how organic the message might be perceived. However, the influencer marketing literature remains scarce, even as global expenditure on influencers could reach 16 trillion USD in 2022.[5] This paper adds to this literature by studying the network mechanism behind the influencers' effect on demand.

Finally, this paper is also related to work emphasizing network effects on video game demand. Dubé, Hitsch and Chintagunta (2010) find strong indirect network effects in the two-sided market of video game consoles, and Lee (2013) studies vertical integration in the presence of indirect network effects in the video game industry. By contrast, this paper studies *direct* network effects.

[4] Angrist (2014) and Rysman (2019) offer recent reviews (and critiques) on peer and network effects estimation.

[5] See `economist.com/leaders/2022/04/02`.

2 Data and empirical context

Steam is the largest video game digital distribution service in the world, with over 90 million monthly active users shopping from over 30,000 PC video games on offer.[6] In 2017, Steam generated 4.3 billion USD in sales, not counting in-app microtransactions or downloadable content, and in 2020 it recorded 7.25 million users concurrently playing games.[7]

On top of buying video games, Steam users form an undirected social network where they can "friend" each other in a similar manner to Facebook. On the Steam platform, users can connect with friends and communicate with each other, send and receive games as gifts, or coordinate to play together. By default, user profiles are public, which means users can see their friends' video game libraries, game achievements, and time spent playing video games, as well as other personal information such as name, location, and friends. To facilitate interactions, users may also form groups and designate one of them as their primary group.[8]

The data, published by O'Neill et al. (2016), contain the cross-section of the universe of 108.7 million user accounts, who collectively own 384.3 million games, and formed 3 million groups, as of March 2013. For each user, I observe their game library, friends, date of account creation, date of last activity on the platform, time spent on each video game, and group membership. And for each video game, I observe its price, rating, type of game (full game, DLC, and others), genre, developers, publishers, and date of release, among other information at the time of collection. That is, I do not observe when users bought a game or became friends with another user, but I observe a complete list of owned games and friends at the time of collection.[9]

Table 1 shows user summary statistics. In this paper, I focus on active users in 2013 for the analysis. I define "active" as having at least one friend or one game. We have over 53 million active users, though we only require a small random sample for consistency. Roughly 45% of active users have no friends but have games, and 13% of active users have no games but have friends. On average,

[6]See store.steampowered.com/about, store.steampowered.com/stats, and pcgamesn.com/steam-player-count.

[7]See pcgamesn.com, videogameschronicle.com.

[8]Users can change their privacy settings to minimize the information shared with friends and the general public. However, users are "nudged" into public profiles.

[9]Data and replication files can be found in the author's website or by request. The raw data can be found in steam.internet.byu.edu.

TABLE 1: SUMMARY STATISTICS

	MEAN	SD	P5	P25	P50	P75	P95
ALL USERS							
FRIENDS	3	14	0	0	0	1	18
OWNED GAMES	7	29	0	0	0	3	28
EXPENDITURE	83	359	0	0	0	35	390
PRICE	13	9	0	7	11	18	30
RATING	83	7	72	78	83	88	92
TIME PLAYED	360	813	0	2	38	331	1835
GROUPS	1	6	0	0	0	0	3
HAS PRIMARY GROUP	0.94	0.24	0	1	1	1	1
UNIQUE USERS (MILLIONS)	108.7						
UNIQUE GROUPS (MILLIONS)	3.0						
ACTIVE USERS							
FRIENDS	7	20	0	0	1	5	35
OWNED GAMES	13	40	0	1	3	9	60
EXPENDITURE	169	498	0	0	35	105	819
PRICE	13	9	0	7	11	18	30
RATING	83	7	72	78	83	88	92
TIME PLAYED	360	813	0	2	38	331	1835
GROUPS	2	8	0	0	0	0	7
HAS PRIMARY GROUP	0.94	0.23	0	1	1	1	1
UNIQUE USERS (MILLIONS)	53.2						
USERS WITH FRIENDS							
FRIENDS	13	25	1	2	4	13	54
OWNED GAMES	20	52	0	1	5	16	93
EXPENDITURE	255	634	0	0	45	205	1247
PRICE	12	7	0	8	12	16	25
RATING	82	5	73	79	82	85	92
TIME PLAYED	611	1008	0	30	214	782	2523
GROUPS	3	11	0	0	0	2	12
HAS PRIMARY GROUP	0.90	0.30	0	1	1	1	1
UNIQUE USERS (MILLIONS)	29.5						
USERS WITHOUT FRIENDS							
FRIENDS	0	0	0	0	0	0	0
OWNED GAMES	5	15	1	1	2	5	16
EXPENDITURE	63	189	0	5	25	53	240
PRICE	13	10	0	6	10	20	30
RATING	83	8	69	78	84	90	92
TIME PLAYED	121	457	0	0	6	53	593
GROUPS	0	0	0	0	0	0	0
HAS PRIMARY GROUP	0.00	0.00	0	0	0	0	0
UNIQUE USERS (MILLIONS)	23.6						

Notes: "All users" correspond to the full sample of 108.7 million user accounts; "Active users" are the 53 million users with at least one friend or one game; "Users with friends" are active users with at least one friend; and, "Users without friends" are active users with no friends. Expenditure and prices in US dollars; prices are as observed at data collection, and expenditure is estimated from observed prices. Rating is average percentage of "likes" across owned games. Time played is in hours, cumulative. Groups is the number of self-reported group memberships, and primary group is self-reported. See also O'Neill et al. (2016).

active users have 7 friends, 13 games, spend 169 dollars in total, pay for each game 13 dollars, play 360 hours total, joined 2 groups, and declared a primary group 94% of the time. However, statistics are skewed, showing a long tail of users with much larger consumptions, expenditures, and social networks. The table also shows users with friends have more games and play more than users without friends on average, which is consistent with positive network effects.

Table 2 shows game summary statistics. The data features 9,085 different games. The average price on the market is 11, games have an average rating of 72 "likes" out of 100, 24% of games have a multiplayer mode, and 45% of games are developed by a different entity than their publisher. By comparing these statistics with owned-games statistics from Table 1, we see users lean towards pricier and higher-rated games.

TABLE 2: GAMES SUMMARY STATISTICS

	MEAN	SD	P5	P25	P50	P75	P95
ALL GAMES							
PRICE	11	14	0	5	8	14	30
RATING	72	11	52	66	74	80	88
MULTIPLAYER	0.24	0.43	0	0	0	0	1
OUTSOURCED DEVELOPMENT	0.45	0.50	0	0	0	1	1
UNIQUE GAMES	9,085						
TOP 100 GAMES							
PRICE	19	12	5	10	20	20	40
RATING	83	8	66	79	84	89	95
MULTIPLAYER	0.68	0.47	0	0	1	1	1
OUTSOURCED DEVELOPMENT	0.49	0.50	0	0	0	1	1

Notes: Prices are in US dollars, as observed at data collection. Rating is percentage of "likes". Multiplayer is a dummy for multiplayer mode option. Outsourced development is a dummy for publisher and developer not being integrated.

In this paper, I focus on the top 100 games by popularity, which account for almost 50% of total copies sold, mainly due to computational concerns described below. In this way, I cut out the long tail of games with very small market penetration, with the caveat that the external validity of this paper is limited to popular games. Also, for the top 100, I only consider games with positive prices, because games that are "free-to-play" have different business models, which mostly rely on in-game purchases. Games in the top 100 are pricier, better rated, and offer multiplayer mode more frequently, but are outsourced at a similar rate than the rest. In the data, consumers purchase any game within the top 100 with a probability of .06.

Because of the nature of video game consumption, we expect positive social network effects. Video games can be single-player or multiplayer (or a mix). Both types of games enjoy positive network effects, but the argument can be made for a stronger effect on multiplayer games. Notably, we can construct the complete friends network with the data, but, as this paper shows, we do not need the whole network mapping to estimate network effects.

3 Linear-in-means model

I follow the peer effects literature by assuming that the demand for a network good depends on the probability of peer consumption. To be clear, this model is not a contribution of this paper; this model is commonly used in the literature (Angrist, 2014; Fortin and Boucher, 2016; Rysman, 2019). Specifically, I consider the following linear-in-means model:

$$q_{ij} = \mu_i - \alpha \log p_j + \beta n_{ij} + \gamma \log r_j + \boldsymbol{\delta}_0 \boldsymbol{x}_j + \varepsilon_{ij}, \tag{1}$$

where the quantity q_{ij} indicates if i purchased game j, μ_i is an individual fixed effect, p_j is the price of game j, n_{ij} is the proportion of friends of i that own game j, r_j is the quality of the game, measured by a rating index, ε_{ij} is a shock unobservable to the econometrician, and $\boldsymbol{x}_j$ are characteristics of game j that include genre dummies (action, adventure, RPG, etc.), days since game j was released, and indicators for multiplayer and age restrictions. The set of covariates, $\boldsymbol{x}_j$, allows me to control for other demand shifters that could correlate with price or quality. Consumers are indexed by $i = 1, \ldots, N$, and games by $j = 1, \ldots, J$.

Equation (1) has the advantage that it can be derived from a utility-maximization problem (Fortin and Boucher, 2016), and can be interpreted as the best-response function for an individual who responds to complementary social interactions. We expect $\beta > 0$ for a positive network effect.[10] Section 4 presents a binary choice model instead of equation (1)'s linear probability model.

3.1 Identification

Equation (1) is not identified, because $\mathbb{E}\left[q_{ij}|\boldsymbol{z}_{ij}\right] = \mathbb{E}\left[n_{ij}|\boldsymbol{z}_{ij}\right]$, for an appropriate vector of instruments, $\boldsymbol{z}_{ij}$. That is, we anticipate the expected outcome of i's friends to be equal to the expected outcome of i. If $\mathbb{E}\left[q_{ij}|\boldsymbol{z}_{ij}\right] \neq \mathbb{E}\left[n_{ij}|\boldsymbol{z}_{ij}\right]$, we would not have a reflection problem.[11] Indeed, taking conditional expectations and simplifying,

$$\mathbb{E}\left[q_{ij}|\boldsymbol{z}_{ij}\right] = \frac{\mu_i}{1-\beta} - \frac{\alpha}{1-\beta} \log p_j + \frac{\gamma}{1-\beta} \log r_j + \frac{\boldsymbol{\delta}_0}{1-\beta} \boldsymbol{x}_j, \tag{2}$$

[10] In the literature, β might also be known as a peer effect, a network externality, or a consumption spillover between agents. Where appropriate, I also use the term network elasticity of demand to refer to $\partial \log q / \partial \log n = \beta n / q$.

[11] In the peer-effects literature, the common approach is a group-level analog of (1), instead of an individual-level model.

even if $\mathbb{E}\left[\varepsilon_{ij}|\boldsymbol{z}_{ij}\right] = 0$.

However, identification can be achieved if we assume $\mathbb{E}\left[\varepsilon_{ij}|\boldsymbol{z}_{ij}, n_{ij} = 0\right] = 0$, so we get

$$\mathbb{E}\left[q_{ij}|\boldsymbol{z}_{ij}, n_{ij} = 0\right] = \mu_i - \alpha \log p_j + \gamma \log r_j + \boldsymbol{\delta}_0 \boldsymbol{x}_j. \tag{3}$$

The identification argument has three steps.

The first step is to identify price and rating coefficients from equation (3). We observe game ownerships and prices at the time of data collection, which means some degree of measurement error. On game ownership, this error translates into lower precision in the estimation. However, we have classical measurement error on prices, which translates into attenuation bias and is yet another reason to address price endogeneity. Then, to identify the price coefficient, α, I instrument prices with the status of integration between developers and publishers of game j as cost shifters. Software development represents the main production cost of a game, but consumers are rarely aware of the identity of the developers, as opposed to, say, publishers. If the game is developed "in-house," I expect costs to be lower compared to outsourced development. Therefore, conditional on the observable characteristics of game j, the developers of j introduce exogenous variation to the price of j. On the other hand, I assume ratings are exogenous, conditional on individual fixed effects and observable game characteristics.

Another threat to identification comes from the fact that people with no friends are different from people with some friends. However, I have three answers to this problem. First, I include individual fixed effects to control for individual heterogeneity, contextual effects, correlated effects, and, to some extent, network formation (Hartmann et al., 2008). As long as people with no friends are not systematically different in their pricing sensitivities than people with some friends, we have identification. Second, $n_{ij} = 0$ does not imply that i has no friends, but only that i has no friends with game j. That is, people who have $n_{ij} = 0$ are not rare. Indeed, no game has perfect penetration, and no individual has every game. We only require that for some game j, i's friends do not have that particular game. In the data, consumer i has game j with .06 probability, and has 7 friends, on average. Therefore, having games with no friends is not uncommon. Third, we can also identify the network effect from

[12] In other words, we require repeated observations for individuals, like in this case with multi-product adoptions, in order to address selection concerns. Alternatively, as a referee pointed out, with single-product adoptions, but multiple periods of time per individual, we would require individuals consumption to be always observable to their network.

other coefficients, such as the rating coefficients. In this case, even if people with no friends have different rating sensitivities, we would not expect such bias to be exactly the same as the bias in price sensitivities.[12]

The second step is to identify the compounded price and rating coefficients, $\alpha/(1-\beta)$ and $\gamma/(1-\beta)$, from equation (2), but conditional on non-isolated individuals:

$$\mathbb{E}\left[q_{ij}|\boldsymbol{z}_{ij}, n_{ij} > 0\right] = \frac{\mu_i}{1-\beta} - \frac{\alpha}{1-\beta}\log p_j + \frac{\gamma}{1-\beta}\log r_j + \frac{\boldsymbol{\delta}_0}{1-\beta}\boldsymbol{x}_j, \tag{4}$$

where I assume $\mathbb{E}\left[\varepsilon_{ij}|\boldsymbol{z}_{ij}, n_{ij} > 0\right] = 0$. Again, I use developer integration to identify the compounded coefficient $\alpha/(1-\beta)$.

The third step is to identify the network effect, β. With estimates of α and $\alpha/(1-\beta)$, we can back out an estimate of β. Note that we do not need γ and $\gamma/(1-\beta)$, but if we use these rating coefficients instead, the resulting β should agree with the one derived with the price coefficients.

I also compare my result with the "traditional" peer-effect instrumental-variables strategy. That is, I estimate equation (1) directly,

$$q_{ij} = \mu_i - \alpha\log p_j + \beta n_{ij} + \boldsymbol{\delta}_0\boldsymbol{x}_j + \varepsilon_{ij}, \tag{1}$$

where n_{ij} is instrumented with $n_{ij}^{(2)}$, the probability that second-degree friends have game j. In other words, let $N_i^{(2)}$ be the total number of i's second-degree friends (friends of my friends, who are not my friends), and let $N_{ij}^{(2)}$ be the number of i's second-degree friends who have purchased game j. Then, $n_{ij}^{(2)} \equiv N_{ij}^{(2)}/N_i^{(2)}$ is an instrument for n_{ij} (e.g., De Giorgi, Frederiksen and Pistaferri (2020)). The disadvantage of this method is that it requires detailed network data to construct $n_{ij}^{(2)}$, as opposed to my method, which only requires n_{ij}.

To provide yet another comparison, I also estimate equation (1) directly using group fixed effects, instead of individual fixed effects, based on the strategy by Bertrand, Luttmer and Mullainathan (2000). I use self-reported data on primary group membership by Steam users.

Finally, while individual fixed effects can control for a user adopting fewer or more games, they might not control for the likelihood of adopting certain types of games by these users. Therefore, given the richness of the data, I also estimate the linear model at the individual level. That is, for each i, the equation

$$q_{ij} = \mu_i - \alpha_i\log p_j + \beta_i n_{ij} + \gamma_i\log r_j + \boldsymbol{\delta}_{i0}\boldsymbol{x}_j + \varepsilon_{ij}, \tag{5}$$

is estimated with the "traditional" peer-effect instrumental-variables strategy and the method from this paper. Thus, I allow for parameter heterogeneity at the individual level without imposing a specific distribution for the parameters. We might have idiosyncratic price elasticities, idiosyncratic genre preferences, and so on, where different games can affect individuals in idiosyncratic ways. The results, shown in section 3.3, are similar to those from the pooled models above. While completely separate regressions for each individual is an extreme way to address concerns on heterogeneity, homophily, and correlated unobservables, we might consider intermediate steps such as random coefficients, or individual heterogeneity in a subset of parameters. However, the emphasis of this paper is not on the richness of the data, but on leveraging aggregate relational data to estimate network effects.

As a final note, equations (2) and (3) should not be estimated in a single equation model using a within transformation (even with appropriate interaction terms), because even the fixed effects differ by a factor of $(1-\beta)^{-1}$ between these equations. A full set of individual dummies, with all covariates interacted with $\mathbb{1}\{n_{ij}=0\}$ is theoretically appropriate, but not practical.

3.2 Discussion: A LATE interpretation

The estimator of this paper can be interpreted as a local average treatment effect.[13] In this case, users without any friends are "never-takers," because no matter what the rest of the network consumes, they do not change their behavior. The other extreme is the "always-takers": users for whom for any game, they always have a friend with that game. These users are always affected by their network. The rest of the users are the "compliers." Compliers change their behavior, because some of their friends bought the game.

Identification comes from compliers. We require a user to be observed in both cases where friends have and do not have a game. From the changes in demand within this user, we infer the causal effect of the network. For never-takers and always-takers, demand within users does not change, and does not provide information to identify the network effect. Then, the estimate is "local" to compliers, so its interpretation should be local, too.

Therefore, this method relies on the nature of compliers in the data. If compliers are scarce, the method has a disadvantage vis-à-vis traditional-IV strategies, for instance, because this method would have less statistical power,

[13]See Imbens and Angrist (1994), for example.

and compliers would not be representative of the general population. Another issue would be if compliers are different in important dimensions. For instance, compliers might have different tastes or might have a higher propensity to consume (beyond the effect of their networks), which again implies the estimated effect should not be generalized. However, in this specific empirical context, I find no evidence of compliers being special. Compliers represent about 53% of the population, and the type of games they consume are similar to the rest of the population.[14]

3.3 Results

To be clear, the pseudo-panel follows a random sample of .2% of active users who consider buying games $j = 1, \ldots, 100$. Table 3 shows the main results. All models include individual fixed effects and the controls described in section 3, except for column 2, which only includes group fixed effects. In all models, prices have been instrumented with an indicator for outsourced game development.

The benchmark model in column 1 presents the largest price elasticity in the estimations. As expected, the coefficients are compounded with the network effect. Column 2 shows more plausible price and network coefficients, using the approach based on Bertrand, Luttmer and Mullainathan (2000). Column 3 shows the "traditional-IV" approach, where n_{ij} has been instrumented with the proportion of second-order friends with game j. The network effect corresponds to the coefficient of n_{ij}, which is 0.83, and implies a .1 increase in the proportion of friends with game j increases the probability of buying j in by 0.08.

Columns 4 and 5 show the results of my method. Column 4 restricts the sample to $n_{ij} = 0$, and column 5 to $n_{ij} > 0$. As expected, the benchmark model yields estimates in between those of columns 4 and 5. Column 4's coefficients correspond to the price and rating effects *without* network effects, and align with those of columns 2 and 3. Column 5 shows the coefficients compounded with the network effects. The implied network effects are backed out using columns 4 and 5: let α_4 and α_5 be the coefficients from columns 4 and 5, respectively, then, the network effect is $1 - \alpha_4/\alpha_5$.

The estimated network effect from my method is 0.88 using the price coefficients, and 0.87 using the rating coefficients. Not only they agree with each other,

[14] Always-takers represent less than 2% of the population. Table 1, "Users with friends" shows essentially summary statistics for compliers, and never-takers' statistics are under "Users without friends".

Table 3: Estimation of demand and network effect

Dep var: q_{ij}	Benchmark (1)	Group FE (2)	Traditional IVs (3)	No friends (4)	Some friends (5)
log price$_j$	-0.19	-0.14	-0.06	-0.09	-0.73
	(0.003)	(0.017)	(0.003)	(0.002)	(0.025)
log rating$_j$	0.18	0.12	0.04	0.08	0.59
	(0.002)	(0.020)	(0.003)	(0.002)	(0.015)
n_{ij}		0.34	0.83		
		(0.064)	(0.004)		
Implied Net Eff		0.34	0.83		
Using prices				0.88	
Using ratings				0.87	
Implied elasticities					
price$_j$	-3.27	-2.38	-1.10	-1.54	
n_{ij}		0.35	0.85	0.91	
Controls	Yes	Yes	Yes	Yes	Yes
Individual fixed effects	Yes	No	Yes	Yes	Yes
Group fixed effects	.	Yes	.	.	.
2SLS	Yes	Yes	Yes	Yes	Yes
First-stage IV's F-stat					
log price$_j$	34,721	34,166	31,516	33,023	855
n_{ij}	.	.	2,434,143	.	.
Obs	8,170,198	8,170,198	8,170,198	6,488,419	1,681,779

Notes: Standard errors clustered at individual level. In all columns, prices are instrumented with an indicator for outsourced game development. All collumns include as controls: individual fixed effects, a set of genre dummies, days since release, and indicators for multiplayer and age restrictions. Column 1 shows the benchmark model where network effects are ignored. Column 2 estimates the network effects by including social group fixed effects. Column 3 includes n_{ij} and instruments it with the proportion of second degree friends with game j. Column 4 is estimated in the subsample where $n_{ij} = 0$, and Column 5 in the subsample where $n_{ij} > 0$. The implied network effect in columns 2 and 4 is just the coefficient of n_{ij}, while the implied network effect of my method combines the coefficients of columns 4 and 5 as $1 - \alpha_4/\alpha_5$, where α_4 and α_5 are the coefficients from columns 4 and 5, respectively.

but they are also very close to column 3's 0.83. In terms of elasticities, the implied network elastiticies of demand are around 0.9, which suggest a significant direct network externality in the platform.[15]

Moreover, the estimated price elasticity from my method is -1.54, but, if we ignore network effects, the implied price elasticity from the benchmark model is -3.27. As expected, focusing only on isolates yields a more inelastic demand than the compounded elasticity from the whole sample. Therefore, we could significantly overestimate price elasticities, and the same can be said about rating elasticities, for example.[16]

Table 4 shows the results from the individual, linear-in-means model in equation (5). The group fixed effects approach is not included, because individual

[15] To boot, most of the unreported coefficients of the controls also yield estimates of the network effect around 0.9. I do not report these results, because I do not have a strong identification argument for these coefficients.

[16] Nair (2007) reports short-run price elasticities similar to those found in this paper.

fixed effects trivially contain them. The table shows the average coefficients in the population and their standard deviations, weighted by their standard errors, because each individual regression uses at most 100 observations. To increase statistical power, I do not instrument for prices in these regressions, because, at the individual level, price endogeneity concerns are alleviated. Indeed, consumers are too small to individually affect prices, which means that price endogeneity can only come from aggregate demand shocks, conditional on observables. Results are very similar to those from the pooled models, but with considerable heterogeneity in the population.

For managers and policy makers, this means that efforts to attract consumers pay off. It also means that willingness-to-pay is significantly higher at any point of the demand curve. Thus, a two-part tariff could be more efficient.

TABLE 4: DEMAND ESTIMATES AT INDIVIDUAL LEVEL

DEP VAR: q_{ij}	BENCHMARK (1)	TRADITIONAL IVS (2)	NEW METHOD (3)
IMPLIED NET EFF		1.38	
		(0.701)	
USING PRICES			0.92
			(2.483)
USING RATINGS			0.92
			(1.684)
IMPLIED PRICE ELASTICITY	-2.52	-1.00	-2.06
	(1.085)	(2.121)	(1.993)
CONTROLS	YES	YES	YES
INDIVIDUAL COEFFICIENTS	YES	YES	YES
2SLS	.	YES	.
FIRST-STAGE IV'S F-STAT	.	107	.
OBS	8,170,198	8,170,198	8,170,198

Notes: Average coefficients across population; standard deviations in parentheses. Separate regressions for each individual. All collumns include as controls: a set of genre dummies, days since release, and indicators for multiplayer and age restrictions. Column 1 shows the benchmark model where network effects are ignored. Column 2 shows the traditional model where network effects are directly estimated by including n_{ij} in the estimating equation and instrumenting it with the proportion of second degree friends with game j. Column 3 shows the result from this paper's method where coefficients are estimated in the subsample where $n_{ij} = 0$, and the network effect is estimated by comparing those estimates with those from the subsample where $n_{ij} > 0$.

4 Discrete choice model

While the linear probability model has advantages, it can overestimate marginal effects, specially when outcomes have probabilities close to the extremes, 0 or 1. In the data, consumers purchase any random game j with a probability of .06, which warrants the exploration of a nonlinear model.

With individual data we can model the binary choice of purchasing a game or not. Consider the following analogue of equation (1):

$$u_{ij} = \mu_i - \alpha \log p_j + \beta n_{ij} + \gamma \log r_j + \boldsymbol{\delta}_0 \boldsymbol{x}_j + \varepsilon_{ij}, \tag{6}$$

where u_{ij} is the utility of consumer i from purchasing game j, and the remaining variables are as in section 3: fixed effects, prices, proportion of friends with game, ratings, and game characteristics.

Consumer i purchases j iff $u_{ij} > 0$. Let $y_{ij} \equiv \mathbb{1}\{u_{ij} > 0\}$. Assuming iid $\varepsilon \sim F$, the probability of buying j is

$$P\left[y_{ij} = 1\right] = F\left(\mu_i - \alpha \log p_j + \beta n_{ij} + \gamma \log r_j + \boldsymbol{\delta}_0 \boldsymbol{x}_j\right), \tag{7}$$

where F will later be specialized to a logit model. Thus, i's demand for j, is defined as $s_{ij} \equiv P\left[y_{ij} = 1\right]$.

Then, the log-likelihood of the data becomes

$$L = \sum_{i=1}^{N} \sum_{j=1}^{J} y_{ij} \log s_{ij} + (1 - y_{ij}) \log(1 - s_{ij}). \tag{8}$$

If we can assume a random utility model, as in (6), then the parameters can be interpreted as utility parameters. In this sense, this model is assumed to be structural, because utility parameters should be invariant to changes in policy.

4.1 Identification

As with the linear-in-means model, we have a reflection problem. In particular, a likelihood maximization program will not identify the local maximum by using first-order conditions from L, because the derivatives will not take into account the feedback from a network effect.

Consider the more compact notation

$$\boldsymbol{X}_{ij}' \boldsymbol{\theta} \equiv \mu_i - \alpha \log p_j + \gamma \log r_j + \boldsymbol{\delta}_0 \boldsymbol{x}_j,$$

where $\boldsymbol{\theta}$ is the vector of parameters, excluding β, and $\boldsymbol{X}_{ij}$ is the vector of covariates, excluding n_{ij}. We can rewrite the utility function as

$$u_{ij} = \beta n_{ij} + \boldsymbol{X}_{ij}' \boldsymbol{\theta} + \varepsilon_{ij}. \tag{6'}$$

Then, the likelihood's first-order conditions with respect to any element θ_k of $\boldsymbol{\theta}$ become

$$0 = \frac{\partial L}{\partial \theta_k} = \sum_{i=1}^{N} \sum_{j=1}^{J} \frac{y_{ij} - s_{ij}}{s_{ij}(1 - s_{ij})} \frac{\partial s_{ij}}{\partial \theta_k}.$$

However,

$$\frac{\partial s_{ij}}{\partial \theta_k} = F' \left(\beta n_{ij} + \boldsymbol{X}_{ij}' \boldsymbol{\theta} \right) \left[X_{ij}^k + \beta \frac{\partial n_{ij}}{\partial \theta_k} \right], \tag{9}$$

where X_{ij}^k is the corresponding covariate to θ_k. In this context, the reflection problem is that $\frac{\partial n_{ij}}{\partial \theta_k}$ would be ignored by a maximization program. But, in expectation, $n_{ij} = s_{ij}$, which implies that $\frac{\partial n_{ij}}{\partial \theta_k} = \frac{\partial s_{ij}}{\partial \theta_k}$ in expectation. Then,

$$\begin{aligned} \frac{\partial s_{ij}}{\partial \theta_k} &= F' \left(\beta n_{ij} + \boldsymbol{X}_{ij}' \boldsymbol{\theta} \right) \left[X_{ij}^k + \beta \frac{\partial s_{ij}}{\partial \theta_k} \right] \\ \Rightarrow \frac{\partial s_{ij}}{\partial \theta_k} &= \frac{F' \left(\beta n_{ij} + \boldsymbol{X}_{ij}' \boldsymbol{\theta} \right)}{1 - \beta F' \left(\beta n_{ij} + \boldsymbol{X}_{ij}' \boldsymbol{\theta} \right)} X_{ij}^k. \end{aligned} \tag{10}$$

A simple solution, as in the linear-in-means model, is to consider first the cases with $n_{ij} = 0$, so that

$$\begin{aligned} u_{ij} &= \boldsymbol{X}_{ij}' \boldsymbol{\theta} + \varepsilon_{ij} \\ \Rightarrow s_{ij} &= F \left(\boldsymbol{X}_{ij}' \boldsymbol{\theta} \right), \\ \Rightarrow \frac{\partial s_{ij}}{\partial \theta_k} &= F' \left(\boldsymbol{X}_{ij}' \boldsymbol{\theta} \right) X_{ij}^k \end{aligned} \tag{11}$$

for the consumer i without friends with game j.

Intuitively, we can identify $\boldsymbol{\theta}$ using equation (11), and then use that information to identify β using (10).[17]

To identify the price and rating coefficients, I use the same set of instruments from the linear-in-means model. However, because of the nature of the binary choice model, I opt for a control function approach (Petrin and Train, 2010). To create the control function, in a first stage, I regress endogenous variables on instruments and exogenous variables, and I estimate residuals. In a second stage, I include those estimated residuals as extra covariates. Standard errors can be calculated by bootstrapping.

[17] Note that from the FOC on β, $\frac{\partial s_{ij}}{\partial \beta} = F' \left(\mu_i - \alpha \log p_j + \beta n_{ij} + \gamma \log r_j + \boldsymbol{\delta}_0 \boldsymbol{x}_j \right) \beta \frac{\partial s_{ij}}{\partial \beta}$, which does not identify β.

For the estimations, I use Chamberlain (1980)'s conditional logit model to avoid the incidental parameters problem caused by the individual fixed effects (Lancaster, 2000). The rest of the threats to identification that arise here are very similar to those from the linear-in-means model which I have already discussed.

As with the linear-in-means model, I compare this paper's method with a benchmark, a group fixed effects approach, and a "traditional-IV" approach.

Finally, I also estimate these models at the individual level. For each i,

$$u_{ij} = \beta_i n_{ij} + \boldsymbol{X}_{ij}'\boldsymbol{\theta}_i + \varepsilon_{ij}, \tag{12}$$

yields a binary-choice model with idiosyncratic parameters.

4.2 Estimation

If $\boldsymbol{\theta}$ is of length K, then the first-order conditions define the following set of $2K$ moments for $K+1$ parameters $\{\boldsymbol{\theta}, \beta\}$. For $k = 1, \ldots, K$,

$$0 = \frac{\partial L}{\partial \theta_k} = \sum_{i=1}^{N}\sum_{j=1}^{J} \frac{y_{ij} - s_{ij}}{s_{ij}(1-s_{ij})} F'\left(\boldsymbol{X}_{ij}'\boldsymbol{\theta}\right) X_{ij}^k \qquad \text{if } n_{ij} = 0, \tag{13}$$

$$0 = \frac{\partial L}{\partial \theta_k} = \sum_{i=1}^{N}\sum_{j=1}^{J} \frac{y_{ij} - s_{ij}}{s_{ij}(1-s_{ij})} \frac{F'\left(\beta n_{ij} + \boldsymbol{X}_{ij}'\boldsymbol{\theta}\right)}{1 - \beta F'\left(\beta n_{ij} + \boldsymbol{X}_{ij}'\boldsymbol{\theta}\right)} X_{ij}^k \qquad \text{if } n_{ij} > 0. \tag{14}$$

Thus, estimation is done by GMM.

In practice, we can separately estimate $\boldsymbol{\theta}$ from moments (13) at the cost of efficiency. Using $\widehat{\boldsymbol{\theta}}$, we can then use moments (14) to estimate β. For consistency with section 3, I only use the price and rating moments to identify the network effect.

Finally, I use the same sample as in the linear model estimation. The size of the sample was selected because of computational concerns. The computational time for conditional logit models is roughly proportional to $JK\sum_{i=1}^{N} \min\{t_{i0}, t_{i1}\}$, where t_{i0} and t_{i1} are the number of zeros and ones. Therefore, .2% of active users and a panel of length of 100 is a reasonable compromise between speed and precision.

4.3 Results

Table 5 shows the results for a binary-choice logit model. Every column addresses price endogeneity as described. As a benchmark, column 1 shows a logit model which ignores the network effect. Column 2 estimates the network effect using group fixed effects, based on Bertrand, Luttmer and Mullainathan (2000). Column 3 identifies the network effect using the traditional instrument from the literature, the proportion of second-degree friends with game *j*. Finally, column 4 presents this paper's method, where the network effect is identified from the moments defined by the first-order conditions of a logit model with network effects.

Price elasticities roughly align with those from the linear-in-means model. I find an elastic demand curve, but the elasticity could be overestimated if we ignore network effects, as expected. Demand is also inelastic with respect to the quality of the product, measured by its rating. However, columns 2 and 3 report unrealistically low estimates of price elasticities, which favor this paper's method in column 4.

The network elasticity is estimated to be 0.13, which is lower than the one estimated with a linear probability model. When the proportion of friends with the game increases by 1%, the network externality translates to an increase in demand of .13%. For perspective, the average demand for a game in the sample is 6%. The effect is strong, as the experiments in the following section show.

Finally, Table 6 reports estimates at the individual level. Similar to the linear-in-means model, the table shows the average coefficients in the population and their standard deviations, weighted by their standard errors. In this table, prices are not instrumented to gain statistical power. At the individual level, price endogeneity concerns are alleviated as discussed above. Results are very similar to those of Table 5 for price elasticities but considerably larger for network effects. As one might expect, heterogeneity is considerable in the population, with the caveat that some of that apparent heterogeneity is a result of fewer degrees of freedom in the estimations.

Section 5 offers counterfactuals calculated with both the pooled-model estimates, which is the preferred specification, and with the individual-level estimates.

[18] See economist.com/leaders/2022/04/02/.

Table 5: Discrete-choice logit model estimation

Dep var: y_{ij}	Benchmark Logit (1)	With Group FE (2)	Traditional IV (3)	Logit + GMM (4)
log price$_j$	-1.38	-0.65	-0.51	-1.24
	(0.011)	(0.076)	(0.023)	(0.021)
log rating$_j$	1.37	0.65	0.28	0.57
	(0.009)	(0.055)	(0.017)	(0.015)
n_{ij}		2.76	4.48	2.25
		(0.043)	(0.020)	(0.023)
Implied elasticities				
price$_j$	-1.34	-0.63	-0.49	-1.20
rating$_j$	1.33	0.63	0.27	0.56
n_{ij}		0.16	0.27	0.13
Controls	Yes	Yes	Yes	Yes
Individual fixed effects	Yes	No	Yes	Yes
Group fixed effects	.	Yes	.	.
Control functions	Yes	Yes	Yes	Yes
First-stage IV's F-stat				
price$_j$	58,919	58,123	53,878	60,522
n_{ij}	.	.	2,437,944	.
Obs	8,170,198	329,564	8,170,198	8,170,198

Notes: Bootstrapped standard errors clustered at individual level with 100 repetitions. In all columns, prices are instrumented with an indicator for outsourced game development. All columns include as controls: individual fixed effects, a set of genre dummies, days since release, and indicators for multiplayer and age restrictions. Instrumented variables are regressed on instruments in a first stage, where residuals are obtained and used as control functions (Petrin and Train, 2010). Column 1 shows a logit model where the network effect is ignored. Column 2 estimates the network effects by including social group fixed effects; for computational concerns, this model is estimated in a random subsample of 2,000 groups. Column 3 also includes n_{ij} and instruments it with the proportion of i's second-degree friends who own game j. Column 4 shows this paper's logit model where the network effect is estimated using GMM and the moments defined by the FOC of a logit with network effects.

Table 6: Discrete-choice logit demands at individual level

Dep var: q_{ij}	Benchmark (1)	Traditional IVs (2)	New method (3)
Implied network elasticity		3.82	1.53
		(5.245)	(5.491)
Implied price elasticity	-1.05	-0.08	-1.30
	(1.078)	(0.140)	(2.818)
Controls	Yes	Yes	Yes
Individual coefficients	Yes	Yes	Yes
2SLS	.	Yes	.
First-stage IV's F-stat	.	32	.
Obs	8,170,198	8,170,198	8,170,198

Notes: Average coefficients across population; standard deviations in parentheses. Separate regressions for each individual. All collumns include as controls: a set of genre dummies, days since release, and indicators for multiplayer and age restrictions. Column 1 shows the benchmark model where network effects are ignored. Column 2 shows the traditional model where network effects are directly estimated by including n_{ij} in the estimating equation and instrumenting it with the proportion of second degree friends with game j. Column 3 shows this paper's logit model where the network effect is estimated using GMM and the moments defined by the FOC of a logit with network effects.

5 Counterfactual: Promotional giveaway

As *The Economist* reports, influencer marketing is on the rise, with total spending on influencers by brands probably reaching 16 trillion USD in 2022.[18] However, despite its continuing expansion, this market remains understudied. This counterfactual adds to the understanding of the influencer market by simulating demand responses from a promotional giveaway. Indeed, network effects are the mechanism behind demand growth in this market.

Consider giving a game for free to the top 1% of most popular consumers as a way to promote the game. Because of network effects, promoting consumption among some "influencers" might have net gains. The top 1% accounts for about 20% of friendship connections; that is, the probability for a random member of the network to be linked with the top 1% is 20%. Because popular consumers are also more likely to buy any game, this counterfactual consists of ensuring that the top 1% has the game.

Specifically, the probability that a random member of the top 1% buys any one game is around 14%. Therefore, if a firm j gives away its game to all members of the top 1%, then the average consumer sees its number of friends with game j increasing by 1 with a probability of $.2(1-.14)$. However, because a consumer i observes n_{ij} increasing, her consumption of j increases as well, and a feedback loop ensues. When i increases her consumption, i's friends increase theirs, which increases the consumption of i and of her second-degree friends, and so on. The total effect can be computed from the estimates in Table 5, which assumes the same effect for everyone in the population, or from Table 6, which allows for heterogeneous network effects.

The total effect is a function of the relative increase in the network size. For small games, a giveaway to the top 1% represents a big effort relative to their market share. However, for the largest game in the data, with 20% market penetration, the counterfactual giveaway represents 5% of its sales. The following figures show the net gains from the giveaway as a function of market penetration (gains net of top 1% demands).

Figure 1 uses the estimates from the pooled model, which is the preferred specification, because it is the most conservative one, and because it has more precise estimates. The figure shows that even the larger games gain about a 5% increase in sales from the giveaway. For perspective, a game would have to drop its price by 4% to achieve a 5% increase in demand.

Figure 2 uses the individual-level estimates, which display large heterogene-

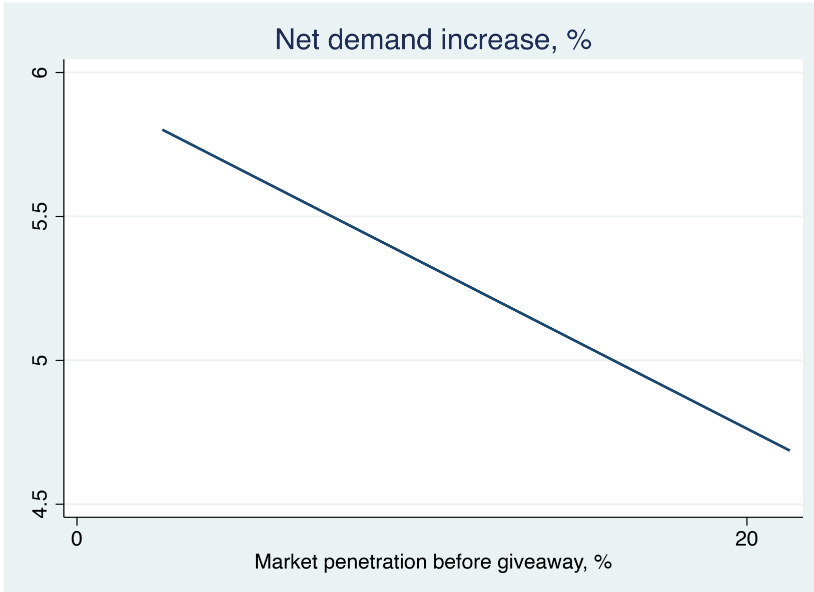

FIGURE 1: Counterfactual giveaway

ity in network effects across the population. In this case, the larger games see their demands increase by about 18%. The estimated gains in demand from the giveaway are substantially higher than in the previous case, precisely because of heterogeneity. Intuitively, because the probability of buying any game is very low, demand can only go up when it experiences a shock. And in this case, the shock is a substantial increase in the network.

6 Concluding remarks

The main contribution of this paper is to show that, with minimal network data, network effects can be credibly estimated. My method is robust in two important dimensions. First, it agrees internally by achieving similar estimates with different variables. Second, it aligns with other data-intensive strategies from the literature. Moreover, using a discrete choice model and several moments from the model, we can estimate more precisely the network effect.

In the case of Steam, price elasticity is about -1.20, while the network effect, in terms of an elasticity, is about 0.13. Disentangling the price elasticity from the network externality allows managers and policymakers to separately evaluate counterfactual pricing and marketing policies. In particular, firms have strong incentives to promote the game through influencers. As a conservative estimate,

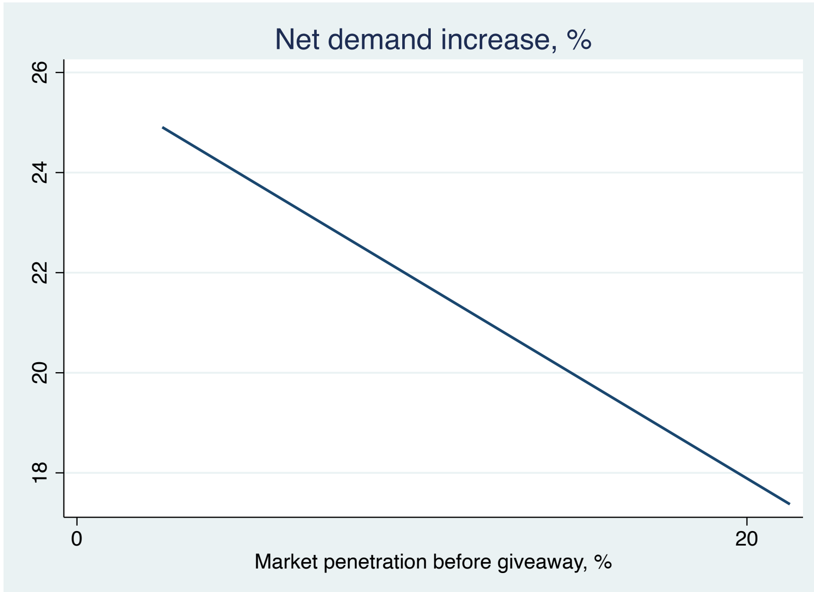

FIGURE 2: Counterfactual giveaway with individual estimates

demand increases by about 5% from a promotional giveaway to the top 1% of most popular users in the network.

References

Angrist, Joshua D. 2014. "The perils of peer effects." *Labour Economics*, 30: 98–108.

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. 2013. "The diffusion of microfinance." *Science*, 341(6144).

Becker, Gary S. 1991. "A Note on Restaurant Pricing and Other Examples of Social Influences on Price." *Journal of Political Economy*, 99(5): 1109–1116.

Belanche, Daniel, Luis V. Casaló, Marta Flavián, and Sergio Ibáñez-Sánchez. 2021. "Understanding influencer marketing: The role of congruence between influencers, products and consumers." *Journal of Business Research*, 132: 186–195.

Bertrand, Marianne, Erzo F. P. Luttmer, and Sendhil Mullainathan. 2000. "Network Effects and Welfare Cultures." *Quarterly Journal of Economics*, 115(3): 1019–1055.

Björkegren, Daniel. 2019. "The Adoption of Network Goods: Evidence from the Spread of Mobile Phones in Rwanda." *The Review of Economic Studies*, 86(3): 1033–1060.

Breza, Emily, and Arun G. Chandrasekhar. 2019. "Social Networks, Reputation, and Commitment: Evidence From a Savings Monitors Experiment." *Econometrica*, 87(1): 175–216.

Breza, Emily, Arun G. Chandrasekhar, Tyler H. McCormick, and Mengjie Pan. 2020. "Using Aggregated Relational Data to feasibly identify network structure without network data." *American Economic Review*, 1–28.

Chamberlain, Gary. 1980. "Analysis of Covariance with Qualitative Data." *The Review of Economic Studies*, 47(146): 225.

Chandrasekhar, Arun G., Cynthia Kinnan, and Horacio Larreguy. 2018. "Social networks as contract enforcement: Evidence from a lab experiment in the field." *American Economic Journal: Applied Economics*, 10(4): 43–78.

De Giorgi, Giacomo, Anders Frederiksen, and Luigi Pistaferri. 2020. "Consumption Network Effects." *The Review of Economic Studies*, 87(1): 130–163.

Dubé, Jean-Pierre H., G. J. Hitsch, and Pradeep K. Chintagunta. 2010. "Tipping and Concentration in Markets with Indirect Network Effects." *Marketing Science*, 29(2): 216–249.

Fortin, Bernard, and Vincent Boucher. 2016. "Some Challenges in the Empirics of the Effects of Networks." In *The Oxford Handbook of the Economics of Networks*. , ed. Yann Bramoullé, Andrea Galeotti and Brian Rogers, 276–302. Oxford University Press.

Hartmann, Wesley R., Puneet Manchanda, Harikesh Nair, Matthew Bothner, Peter Dodds, David Godes, Kartik Hosanagar, and Catherine Tucker. 2008. "Modeling social interactions: Identification, empirical methods and policy implications." *Marketing Letters*, 19(3-4): 287–304.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(2): 467.

Lancaster, Tony. 2000. "The incidental parameter problem since 1948." *Journal of Econometrics*, 95(2): 391–413.

Lee, Robin S. 2013. "Vertical Integration and Exclusivity in Two-Sided Markets." *The American Economic Review*, 103(7): 2960–3000.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies*, 60(3): 531.

Martínez-López, Francisco J., Rafael Anaya-Sánchez, Marisel Fernández Giordano, and David Lopez-Lopez. 2020. "Behind influencer marketing: key marketing decisions and their effects on followers' responses." *Journal of Marketing Management*, 36(7-8): 579–607.

Nair, Harikesh. 2007. "Intertemporal price discrimination with forward-looking consumers: Application to the US market for console video-games." *Quantitative Marketing and Economics*, 5(3): 239–292.

O'Neill, Mark, Elham Vaziripour, Justin Wu, and Daniel Zappala. 2016. "Condensing Steam: Distilling the Diversity of Gamer Behavior." 81–95. New York, New York, USA:ACM Press.

Petrin, Amil, and Kenneth Train. 2010. "A Control Function Approach to Endogeneity in Consumer Choice Models." *Journal of Marketing Research*, XLVI(February): 3–13.

Ryan, Stephen P., and Catherine Tucker. 2012. *Heterogeneity and the dynamics of technology adoption.* Vol. 10.

Rysman, Marc. 2019. "The reflection problem in network effect estimation." *Journal of Economics and Management Strategy*, 28(1): 153–158.

Sacerdote, Bruce. 2001. "Peer effects with random assignment: Results for Dartmouth roommates." *Quarterly Journal of Economics*, 116(2): 681–704.

Tucker, Catherine. 2008. "Identifying Formal and Informal Influence in Technology Adoption with Network Externalities." *Management Science*, 54(12): 2024–2038.