Social Influence in Product Choice and Market Competition: Evidence from a Mobile Communication Network

Qi Wu *

December 22, 2020

Please click here for the most recent version.

Abstract

Social influence is an important driver of consumption behavior, but its effect on firm competition and pricing is understudied. While social influence may create an incentive for firms to reduce initial prices to attract a larger customer base, it can also result in firms charging higher prices in the future because of a social differentiation effect. This paper investigates whether and how social influence affects product choices and firm competition, drawing on a novel dataset that consists of large scale de-identified mobile call records from a city in China. I first identify social influence using a new identification strategy that exploits the partially overlapping network of friends and residential neighbors and the intertemporal variation in friend circles. I find that the purchasing probability for a phone model doubles with 10 percent more friends using the same model. Consumers are more likely to conform to wealthier friends and choose visually distinct features, suggesting that status-seeking motivation may be an important driver of social influence. I then evaluate how social influence affects firm competition by building and estimating a structural model that incorporates social influence in consumer demand. I find that social influence favors high-quality products while reducing low-quality products' market share. In addition, a small price drop of a product would lead to larger gains through quantity expansion by peers. Social influence, on average, reduces initial prices by 0.7 percent and increases subsequent prices by 0.1 percent. It also increases the total profits of new products by 3.4 percent and increases consumer surplus by about 1.7 percent.

Keywords Consumer Demand, Social Influence, Pricing Strategy, Smartphone, High-tech Product

JEL Codes D12, D22, D91, L11, L63

*Department of Economics, Cornell University. Email: qw98@cornell.edu. I am grateful to Panle Jia Barwick, Eleonora Patacchini, Benjamin Leyden, and Giulia Brancaccio for invaluable advice, guidance and support. I would like to thank Nahim Bin Zahur, Jean-Francois Houde, Abhishek Ananth, Qinshu Xue, Yimeng Tang, Zihan Hu, Evan Riehl, Matteo Benetton, Luming Chen, Seth Sanders, Penny Sanders, Jorgen Harris and participants at Cornell IO Workshop for helpful comments. Financial support from the C.V. Starr Fellowship and Sage Fellowship is gratefully acknowledged. I declare I have no relevant or material financial interests that relate to the research described in this paper. All errors are my own.

1 Introduction

Social influence is an important driver of decision making and seamlessly shapes our preferences (Arnold, 2017; Ovide, 2020). The rapid growth of internet technologies and social media platforms have revolutionized our daily interactions and made social influence ubiquitous in areas of human life, including buying consumer goods and services, buying houses, purchasing financial assets, etc.(Bailey et al., 2018a; Lancieri and Sakowski, 2020). Peers' choices can not only be actively shared on platforms such as Pinterest and Instagram,[1] but also be passively disclosed through their digital footprints recorded by platforms such as Facebook and Twitter.[2] Therefore, recent innovations in mobile communication and social media have enhanced the potential role of social influence in consumption decision more than ever.

Social influence not only affects consumer behavior, it could also change firm competition in product markets. The impact of firms' responses to social influence on competition is not clear a priori. If firms respond to social influence by lowering prices to invest in their consumer base, this could enhance competition and benefit consumers. On the other hand, more friends choosing a certain product could create social conformity and add one additional horizontally differentiated feature to the product, thus softening the competition. As the potential power of social influence grows, it is important to understand the impact of social influence on the nature of competition and consumer welfare.

There is a rich literature on the importance of peer effects in consumption (Aral et al., 2009; Bandiera and Rasul, 2006; Conley and Udry, 2010; Giorgi, 2018). However, it is still a long-standing challenge to provide a causal analysis of the social influence and separate it from other confounding factors, particularly sorting on correlated tastes in the empirical literature. In addition, on the supply side, there is a growing theoretical literature studying how firms may react to take advantage of social influence from uniform pricing competition (Cabral, 2011; Economides et al., 2004) to personal pricing, based on node centrality measures (Fainmesser and Galeotti, 2015; Leduc et al., 2017). However, there is little empirical evidence on the impact of peers' choices on market competition and firm pricing. Specifically, does social influence differentially affect demand for high

[1]According to an Instagram consumer study in 2017, 72% of consumers report buying fashion and beauty products based on Instagram posts. More details can be found here: https://www.retaildive.com/news/study-instagram-influences-almost-75-of-user-purchase-decisions/503336/

[2]For example, the U.S. social media company Twitter recently added a feature that displays the source where each tweet is sent from, where a user tweets from the web or a mobile phone. If a user sends a tweeter on a phone, whether he uses Twitter's iOS or Android apps, or a third-party service. The Chinese version of twitter - Weibo – adopted a similar feature where the tweeting handset is displayed to the followers.

and low-quality products? Does it intensify or moderate market competition? In the era of big data, new data sources available from the information and communications technology (ICT) industry make it possible to better understand these questions.

In this paper, I first quantify social influence using a novel dataset that consists of a large-scale mobile call data from a provincial city in China from November 2016 to October 2017 to construct individuals' network of friends and their phone choices. I develop new identification strategies that exploit the partially overlapping network of friends and residential neighbors and the intertemporal variation in friend circles. Next, to assess how firms' pricing behavior responds to social influence, I develop a new structural model that embeds peer spillovers on demand and sheds light on how the demand side spillovers affect supply side incentives. These types of spillovers have not previously been considered in the empirical industrial organization literature. I estimate the model combining the non-conventional micro-level call data with traditional market-level sales data in the Chinese smartphone market. In counterfactual simulations, I explore how social influence affects consumer tastes for quality as well as the pricing behavior of firms.

The call data provide three important pieces of information. It tracks subscribers' handset weekly, and I use this information to infer new phone purchases from changes in the phones used. Among 2.3 million users, I identify around 20.3% individuals who change from non-smartphones or older smartphones to newer smartphones, and these individuals constitute the sample of the study.[3] The data also provides an accurate set of products that consumers are considering at the time of purchase. In addition, The data give me all the mobile call detail records between the users and the call contacts, which allows me to construct individuals' set of real-world social contacts. I examine social influence by looking at the impact of peers' phone ownership on new buyers' choice probability. I measure peers' influence on a new buyer as the fraction of his or her social contacts using a particular phone three months prior to the phone change. Lastly, besides the social space, the call data also allows me to track people spatially over time. This provides individuals' workplace and residential locations.

I begin with a reduced-form analysis that relates each individual's phone choice to his or her friends' past phone choices. I find strong evidence of social influence in the smartphone market using micro-level call data. A 10 percent increase in the share of friends using a given product doubles the average choice probability (1.6 percent) conditional on purchasing, after controlling for sorting on correlated observables and unobserved

[3] The change rate is consistent with a national marketing survey conducted by Penguin Intelligence in September 2017 as described in Section 2

phone tastes. I exploit the partially overlapping structure of contacts and residential neighbors to construct two instrumental variables for the share of friends – the choices and average phone attributes of the residential neighbors of the peers – to partial out the spurious correlation from correlated tastes. A rich set of controls helps to partial out unobserved preferences towards different phones including individual characteristics, the interaction of individual and phone characteristics (for example, older people might prefer phones with larger screen sizes). I also add residential neighborhood by brand fixed effects to capture heterogeneous demand due to income effects and product by month fixed effects to capture seasonality and product-specific demand shocks. My 2SLS estimates are almost identical to the OLS results with extensive controls, which confirms the strength of the controls and provides evidence that the result is not purely driven by unobserved correlated tastes in demand for products.

The intertemporal variation in friend circles also allows me to conduct a falsification test. I construct a similar measure of the lagged shares of peers' choices based on new buyers' future friend network and compare the impact of current friends and future friends. Under correlated tastes, both types of friends should matter since they should share similar preferences with a given individual. I find that the coefficient on future friends is insignificant and order-of-magnitude smaller than current friends, which confirm that the effect is purely driven by unobserved correlated tastes.

To better understand the underlying mechanism, I document considerable heterogeneous effects of social influence across peer groups and by product type. I find suggestive evidence that social influence is motivated by status-seeking. Specifically, consumers are more likely to be influenced by affluent friends in both relative and absolute level. In terms of the attractiveness of product features, I find that people tend to conform more to visible features (e.g. bigger screen and more color options) than hidden functions (higher CPU speed and better screen resolution) conditional on prices and all other features. For the intersection of friends and coworkers, new coworkers who are a possible source of new information are not as influential as pre-existing coworkers. Moreover, new coworkers' impact is insignificantly different from zero, further providing evidence that is inconsistent with the information sharing channel.

My reduced-form analysis points to the importance of social influence in smartphone choices. To understand the impact of social influence on market competition and pricing strategies, I set up and estimate a structural model of demand and supply. In the demand model, I extend the specification in Berry et al. (2004) to include preferences for peers' choices from earlier period as a separate attribute in the utility function. The model allows me to recover a measure of the preference for peers as the utility gain due to complementary value between the individual and the peers, including conformity, based on the suggestive evidence on status-seeking, and benefits of common application usage on the same phone.

Social influence generates two effects in demand and would modify firm incentives. On the one hand, a dynamic nature in demand occurs as a consequence of the social influence – peers' decisions connect demand today with demand tomorrow. I call it the "social multiplier effect". On the other hand, it adds to another dimension of product differentiation, making people less price sensitive. I call it the "social differentiation effect". The results suggest that social influence plays a sizable role in demand. The willingness to pay for a one percent increase in share of friends is equivalent to 9 dollars (3.6 percent of the average price of 250 dollars). The other estimation results are intuitive: on average consumers prefer smartphones with a larger screen, better camera resolution, higher CPU speed, and lighter weight, ceteris paribus. The average price elasticity among all products is about -2.9.

I assume a static demand system for the following reasons. First, after 2015 the smartphone market has become stablized with a slight decline in new sales. Second, 89 percent people are mobile users and the penetration rate of smartphones among consumers remain quite stable at around 50 percent since 2015.[4] Third, low replacement cost makes Chinese smartphone users replace their phones more frequently than global users. People replace phones every 2 to 3 years (Lu, 2017). Mobile phones with high configuration at low prices are springing up, providing Chinese mobile phone users with more options, driving the user demand and shortening replacement cycle.[5] So, with relatively low switching cost, a static demand model captures well a mature market where people frequently replace smartphones to serve their needs. I include month dummies to capture seasonality and demand shocks.

On the supply side, I use a two-period pricing model to evaluate the peer impact on firm dynamic pricing. I allow the marginal costs to change over time to capture changes of the technology frontier. Then the counterfactual analysis isolates the role of social influence on prices holding all other factors constant. In the model, firms choose the optimal prices for each phone in each period to maximize the expected discount profits. Pricing in the first period will take into account the potential social multiplier effect and

[4] Mobile phone internet user penetration in China 2015-2025, Published by Statista Digital Market Outlook, July 17, 2020 https://www.statista.com/statistics/309015/china-mobile-phone-internet-user-penetration/

[5] According to the China Mobile Consumer Survey 2018 released by global accounting and consulting firm Deloitte, nearly 80 percent of Chinese users bought their current phones in 2017 compared to just 58 percent of global users.

social differentiation effect through peers. Correspondingly, these two effects alter firm incentives. Firms would have the investment incentive to reduce the initial prices and then have the harvest incentive to increase prices later.

Based on the model estimates, I conduct counterfactual simulations to address the research questions of whether social influence is the same for high-quality vs. low-quality products and how it would change the prices. In the counterfactual scenario, I set the social influence to be zero. To see the impact on demand for different qualities, I re-estimate the demand (market share) for all the products, holding other factors such as prices as fixed. The results show that without social influence, high-quality products experience the biggest drop in market share. It suggests that social influence favors high-quality products and pushes low-quality products to smaller shares. This is because social influence magnifies the perceived quality difference. In the next counterfactual, I re-optimize the prices in the first and second periods by simulating both the demand and supply sides. On average, I find that social influence reduces the introductory prices by 0.7 percent higher and increases the second-period prices by 0.05 percent. Overall, it increases firm profits by 3.4 percent and increases consumer surplus by 1.7 percent. These findings suggest that with a higher degree of spillover among consumers, firms have a strong incentive to grab higher demand at the beginning and engage in fiercer price competition.

The paper contributes to three strands of literature. The first strand focuses on the literature on peer effects in consumption. From conspicuous consumption (Giorgi, 2018; Veblen, 1899) to product adoption (Aral et al., 2009; Bandiera and Rasul, 2006; Conley and Udry, 2010), social influence is one of the important themes in consumer choices. While these papers make important connections between consumer demand and social influence, few take the additional step to explore the role of social influence on the nature of competition and social welfare. A closely related paper is Bailey et al. (2019), which studies the social influence in phone adoption using Facebook data in the U.S. cellphone market. They find that consumers who are younger and less-educated are more influential to Facebook friends' product choices in the U.S. market and thus qualitatively suggest that network effects would affect the nature of competition and enhance consumer welfare without using phone price and attribute information. I complement their study by looking at social influence in China, a fast-growing economy. In this setting, the characteristics of influential consumers are quite different from those in the U.S. market – middle-aged and affluent individuals are more influential. Moreover, this paper is one of the first structural analyses that quantifies to what extent social influence affects demand, market competition and firm pricing.

Second, this paper relates to literature on quality preference for products. Smallwood and Conlisk (1979) shows that theoretically low-quality products could dominate the market when consumers put too much weight on others' consumption. Amaldoss and Jain (2005a) shows that in conspicious goods market, if firms are asymmetric in terms of quality, in the presence of "social effects" such as status-seeking, markets tend to prefer high-quality products and vanish the market share of low-quality products. However, theory predictions rely on model specification and parameter values. Under different assumptions, different market outcomes would arise. This paper provides the first empirical analysis that examines how demand for quality is affected by social influence.

Third, this paper explores the aggregate effects of peer spillovers on market competition and firm pricing, which is in the spirit of network goods and network effects literature in industrial organization. Seminal work by Katz and Shapiro (1985) and Farrell and Saloner (1986) suggest that global network externalities (e.g., from platforms) would soften competition and grant market power to firms with large installation bases when firms compete on quantities. Under oligopoly, local network externality (e.g., social influence) could change the degree of price competition (Cabral, 2011; Economides et al., 2004) and lead to market segmentation (Banerji and Dutta, 2009). Recent advancements in network literature have been limited to the theory side as well. A small but growing theory literature shows that firms can price discriminate based on node centrality (Chen et al., 2018; Leduc et al., 2017) or degree of susceptibility (Fainmesser and Galeotti, 2015). However, model predictions depend on restictive assumptions of the parameters. With detailed network data, this paper provides the first empirical analysis of the impact of social influence on firm dynamic pricing.

Finally, the paper relates to a growing literature that uses mobile communication networks to study decision-making in economics. With geocoded social interaction data from mobile phone trackers, scholars have explored topics including restaurant choices (Athey et al., 2018), migration and human mobility (Barwick et al., 2019; Blumenstock, 2018; Blumenstock et al., 2015), and the housing market (Bailey et al., 2018b; Buchel et al., 2019). Closely related papers study communication technology adoption and acquisition, including studies on phone adoption in the last decade in developing countries (Bjorkegren, 2018), carrier switching behavior (Hu et al., 2019), and contagion product purchase in carriers (Ma et al., 2015). The current study complements findings for high-tech products and shows the importance of utilizing new data sources from digitization along with traditional data in understanding market outcomes.

This paper proceeds in eight sections. In Section 2, I give background on the industry and describe the data and sample. In Section 3, I provide the reduced-form analysis to show the existence of social influence in consumer choices and explore heterogeneous analysis for the mechanism of social influence. Section 4 outlines the demand model, the two-period pricing model and Section 5 describes the estimation method and results. In Section 6, I compute demand, prices, firm profits and consumer surplus in the counterfactual scenario. Section 7 provides a few robustness checks. Section 8 concludes.

2 Industry background and Data

2.1 Overview

This paper studies the Chinese smartphone industry, an ideal setting to study phone purchases for a few reasons. First of all, this industry has experienced rapid growth by 30 times in sales in the past decade as in Figure 3.[6] After 2015, the market becomes saturated with a slight decline in demand in new sales. Domestic brands and international brands engage in the fierce competition in pricing and advertising. Second, China's mobile phone market has become a red ocean with nearly-saturated segments. Mobile phones with different combinations of features and low prices offer consumers more options to purchase and low replacement cost, shortening the replacement cycle (Deloitte, 2018). Third, unlike in the U.S., sales of smartphones are much less carrier-dependent. Most phones are sold contract-free: 25 percent of sales are through a carrier in the sample period, including stand-alone and bundle sales. However, the subscription rate of phone bundles in the observed carrier is about 5 to 10 percent in the sample period. The prepaid bills of Chinese users account for over 50 percent, and the rate of contract phones have no advantage over prepaid bills in China (Deloitte, 2018). Such a low fraction of contract phones simplify the firm's pricing decision without considering carriers as intermediaries. Lastly, smartphones have a relatively high penetration rate in China. According to marketing research,[7] in 2016, 45.4 percent population has ever used smartphone once a month. On average, people use smartphones 78 minutes per day in 2016 and 98 minutes per day in 2017. Smartphones become an important daily communication necessity and influence social life at a substantial level.

[6] The shipment volume of smartphones is 16 million in 2009 and 473 million in 2017.

[7] *EMarketer foxmedia.co.uk*, retrieved from Statista.com

2.2 Data

The data come from three main sources. The first two data sets come from one major mobile communication service provider in a provincial city in China. It takes about 30 to 65 percent market share.[8] The third piece is market-level data from a marketing research data vendor. The rest are hand-collected data to supplement the main datasets.

Mobile Communication Data The first set of data comes from a major carrier in one provincial city in China. It provides us micro-data about transitions between cellphone devices, a dynamic call network, and phone usage.

- Mobile Device Weekly Tracker As a part of the technical process, the carrier generates phone device logs when a user accesses its service. I observe a weekly tracker of mobile devices for 2.3 million users from November 2016 to October 2017. In each week, it keeps track of a user's most-frequently-used device. It provides a brand and model name associated with each device, such as "Samsung A8" and "Huawei Mate 9". Besides, it also tracks each user's monthly plan subscription. Demographic information including age, gender, and birth county is supplemented from the phone sim card registration records.

- Call Detail Records For billing purposes, the mobile carrier records data for each transaction, called Call Detailed Records (CDRs). It includes the universe of calls from and to the carrier's users from November 2016 to October 2017. For each call, it reports an anonymous identifier of the sender and receiver, a timestamp and the call duration. The call frequency and duration are aggregated to the pairwise weekly level. It provides a unique social network based on calls. Moreover, instead of a snapshot of the network, I observe the dynamics of the network, which is one of the key variations that I exploit to achieve identification. Based on active phone use during daytime (9 am-6 pm) and nighttime (10 pm-7 am), primary work and residential locations are identified.[9]

Quarterly Mobile Phone Tracker The second set of data is market-level data from IDC Research which covers all smartphone sales in China between Q1 2009 and Q2 2019.[10] I observe sales, the average national price (ASP)at the handset model by year-quarter

[8] A market share range is provided to keep the city and carrier anonymous.

[9] Services include voice calls, SMS, and data browsing. Working location is the most frequently used location from 9 am to 6 pm in a given week; the residential location is the most frequently used location from 10 pm to 7 am. Typical traffic and commuting hours are excluded to avoid misclassifying.

[10] https://www.idc.com

level.[11]

Hand-collected Attributes I supplement the IDC data with hand-collected data from two online electronics listing and rating websites: ZOL and GSMArena. For each model, I obtain a comprehensive set of phone attributes ranges from display to performance, including CPU clock speed, screen size, battery capacity, main camera resolution, 4G connection, and weight, etc.

Hand-collected House Price To measure the socioeconomic status of consumers, I supplement the micro-data with hand-collected house prices as a proxy for income levels from one major real estate listing platform AnJuKe.com. I observe the monthly average per square meter house price for all residential communities specified at the main street addresses. By March 2018, it covers 64% and 21% of the blocks in urban and surrounding rural areas respectively.[12] I geocode the communities and merge the prices to the residential locations identified from the carrier with a radius of 1 kilometer, the average distance between two streets bordering a block. The average house price is about 13931.97 RMB (2184.05 USD) per square meter.

2.3 Sample

Sample Construction and Peer Group There are 3 million individuals who use valid mobile devices (brand and model) to begin with. To avoid classifying multiple device holders as new buyers, I drop individuals who hold multiple devices, for example, 'A-A-B-A-B-A'. This excludes about 11 percent users. Moreover, to exclude carrier-related sales, I drop individuals who are on phone bundle plans. This brings the sample size from 2.7 million to 2.68 million. Lastly, to make sure the phones are not used for temporary, I focus on individuals who have weekly records for at least 2 months. This leaves 2.3 million phone users. Sample selection details can be found in Table A8.

Relying on the weekly tracker of devices, I identify the newly-made choices during the sample period through the change of devices. A *phone change* is identified if the following criteria hold. First, an individual uses at least two devices in the sample periods; Second, there is no re-occurrence of a previously held device; Third, the old

[11] ASP is the average end-user (street) price paid for a typically configured mobile phone. ASP includes all freight, insurance, and other shipping and handling fees such as taxes (import/export) and tariffs that are included in vendor or channel pricing. Point-of-sale taxes (e.g., VAT or sales tax) are generally excluded. Additional subsidies offered by mobile operators are not factored into this price.

[12] I obtain 4302 residential communities from AnJuke.com. It matches 708 blocks out of 1406 in the city with 592 out of 790 in the urban part.

and the new device are held for at least one month, respectively. I identify 550,120 new buyers among 2.3 million users during the sample period. New buyers constitute the sample of the study as I know their exact purchase decisions and an accurate set of products they consider at the time of purchase. Figure A2 illustrates the top 100 frequent replacement sequences of devices.

Call networks reflect the real social connections (Bjorkegren, 2018; Blumenstock, 2018). To make call contacts a more reliable proxy for social contacts, I only include contacts who have at least 6 calls per year as in Onnela et al. (2007) to filter accidental calls. To further remove accidental calls, I remove calls less than 16 seconds (the 10th percentile of the call distribution). Table A9 reports the process of the call contact selection. I end up with 172 million pairs of unique call parties. The peer group of interest for new buyer i at time t consists of all social contacts she makes calls to or receives calls from at in the prior three months, i.e., from $t-3$ to $t-1$.[13] I only focus on contacts within a fixed window – three months – before the purchase to make peer groups comparable regardless of the purchase timing. Without a fixed window, the number of friends would grow with the purchase time mechanically, which makes the peer groups incomparable.

Summary Statistics Table 1a reports the summary statistics for the sample and compares the sample demographics and subscription fee with China Family Panel Studies (CFPS) Dataset in 2014, a national representative survey that offers indicator for people who ever used a cellphone or not. In the sample, the average age is 39.21 years old, which is similar to the national representative. There are 35 percent female users in the sample, which is lower than the national average 46 percent. In the sample, 61 percent individuals living in urban area, which are quite comparable to the national representative ratio 64 percent. In the sample, the average monthly fee is about 67.79 RMB (10.13 USD), and a bit higher than 61.39 RMB (9.18 USD) in the CFPS. However, for users who spend at least 30 RMB (4.54 USD) per month,[14] the sample average fee is 75.65 RMB (11.45 USD), similar to 72.84 RMB (11.02 USD) in CFPS. The sample age distribution is a bit different from the national representative ratio because the sample focuses on people with stable subscription and exclude students who are likely to be economic dependent.

Table 1b shows the summary statistics for new buyers and the rest in the sample. In terms of gender ratio and age, there is no systematic difference between the two groups. Among new buyers, 34% of them are female, and the average age is 39. 59 percent of the individuals are in an urban area, which is slightly smaller than 61 percent among

[13] I use one-way contact as the baseline definition of a friend. An alternative definition of reciprocal communications delivers robust results in Section 7.

[14] 30 RMB is the lowest fee for plans with data volumes.

the rest of the sample. The average monthly fee is 69.25 RMB (10.48 USD) for the new buyers, similar to 67.36 RMB (10.19 USD) for the rest. On average, one consumer has 64 friends in the peer group regardless of the mobile carrier. The last row compares the fraction of same-carrier contacts between buyers and non-buyers. 44 percent of them use the same mobile carrier as the buyers, similar to 43 percent, the fraction for non-buyers. The similar same-carrier fraction suggests no systematic selection bias in terms of peer coverage between buyers and non-buyers. Table A4 compares consumers with higher fraction within-carrier friends and those with a lower fraction. There is no big systematic difference between the two. Consumers with more same-carrier friends are slightly more likely to be female and about 1 years old younger than those with fewer same-carrier friends. There is no difference in terms of the spatial distribution between urban and rural areas.

In addition to consumer demographics, I also examine the phone ownership by brand and consumer phone changing behavior by operating systems to see the sample representativeness. Table 2 reports the market shares by brand and the rate of phone change in the sample to national representative surveys. The upper panel compares the market share among new phone buyers in the sample to new sales in the IDC data in Q2 2017. Huawei and OPPO possess 21.73 percent and 19.75 percent, similar to their national shares of 21.54 percent and 18.42 percent. Vivo and Apple have 17.98 percent and 10.98 percent, which are slightly higher than the national shares of 14.74 percent and 7.33 percent. For Xiaomi, the share in the sample of 10.82 percent is slightly smaller than its national share of 13.03 percent. Although the shares are slightly different, the top-five brands and their ranking order in the sample are the same as those in IDC data. Moreover, the phone change patterns are quite comparable with a large marketing survey on smartphone usage and replacement behavior in China in 2017 conducted by Penguin Intelligence Research. The overall phone change rate in 12 months is 20.3 percent in the sample, with 19 percent for Android users and 21.26 percent for IOS users. It is similar to 16 percent for Android users and 23.5 percent for IOS users in the marketing survey.

Product Given various variants for each model and similar models released in different years, I group phone models based on the closeness of major characteristics, as described in Appendix B.3. Extremely expensive and cheap models are excluded before grouping. For example, I drop ultra-luxury phones targeted as high-end gifts such as Huawei Mate 9 Porsche Design, whose release price at 1317 USD (9000 RMB) (compared to the initial release prices of iPhones at around 990 USD). I also drop phones cheaper than 67 USD (450 RMB), such as phones from domestic brand Sugar, LaJiao, etc. I end up with 62 models. Table 3 shows primary phone attributes for products available for markets, including the price, phone age, camera resolution, screen size, screen resolution, CPU clock speed, weight, battery capacity, and fingerprint. The phone age is the number of quarters since released in Q3 2017. The phones range from newly released models with age zero to old products with age 16 quarters. On average, the products are 5 quarters away from release. The average price is about 250.89 USD. The main camera resolution captures the functionality of phone cameras, and on average, it is 13.3 Megapixel. The average phone screen size is 5.34 inches, and there is relatively small variation among smartphones. The screen resolution has relatively more variation than its size, and the average resolution is about 1.79 pixels. The larger the pixel it covers, the better resolution it becomes. The CPU clock speed reflects the phone's computing and operating speed and the quality of the chipset. The average CPU speed is 1.8 GHz. On average, the phone's weight is 146.79 grams. Phones' weight depends on the material of the body and the screen. It is costly to make the screen thinner and reduce weight. The battery capacity is one important functional measure of phones, and a larger capacity indicates longer standing time. The average battery capacity is 3.2 Ah. Fingerprint function is one of the innovations on the screen bio-touch technology. On average, 69 percent of the models allow for fingerprint recognition.

Table 4 shows that the phone change in the sample reflects phone upgrading instead of switch back to an older spare phone. It compares the features of the old handset and the new handset for new buyers. About 20 percent of users upgrade from 2G or 3G network compatible handsets to 4G compatible handsets. For key phone features, on average, the new phones are all improved than the old phones.

3 Existence of social influence in Product Choice

I start with the reduced-form analysis to show the existence of social influence. Let us denote individuals by i, peers of individual i by $m(i)$, products by j and time by t.[15] I explore the existence of social influence on the product choices starting from the following linear probability model:

$$y_{ijt} = \beta s_{m(i),j,t-3} + \boldsymbol{Z_i X_j} \gamma_1 + \boldsymbol{Z_i} \gamma_2 + \boldsymbol{Z_{m(i),j}} \gamma_3 + \lambda_{R(i)f(j)} + \eta_{jt} + \varepsilon_{ijt} \tag{1}$$

where product j is a smartphone model, month t = 1,...10. The dependent variable y_{ijt} takes value one if individual i chooses product j at month t, zero otherwise. As described

[15] The data is organized at individual by alternative level as in Table A1.

in Section 2.3, $m(i)$ is the peer group of individual i. The main variable of interest, $s_{m(i),j,t-3}$, measures social influence. It is the share of social contacts that choose (use or change to) alternative j prior to i's choice among the total number of social contacts. Because it is possible to have reverse causality if contemporaneous peer choices are used, I focus on the impact of peers three months before the purchase. Consider, for example, individual i purchases product j in month t and I use the share of his or her peers who use product j at $t-3$.[16] The lagged structure also reflects that it takes time for social influence to come into effect and for individuals to make purchase decision.

$\boldsymbol{X_j}$ is a vector of major product attributes including screen size, weight, battery capacity, CPU clock speed and camera resolution. $\boldsymbol{Z_i}$ is a vector of individual characteristics including gender, age and dummy variable for residing in urban area. The interaction terms of individual characteristics and primary phone attributes capture differential preference towards smartphone features. For example, female is interacted with camera resolution as female would prefer phones with better selfie quality. Age is interacted with screen size to account for older people may prefer larger screen. $\boldsymbol{Z_{m(i),j}}$ is a vector of the average demographics of friends using each alternative, capturing the contextual exogenous effects from social contacts. These variables are ij specific. For person i, product j, it includes the average female ratio, average age and urban ratio among i's social contacts using product j. To capture the income effect, I include the residential neighborhood-by-brand fixed effects, $\lambda_{R(i)f(j)}$, where $R(i)$ is the residential neighborhood of individual i and $f(j)$ represents the smartphone firm (i.e., brand) of j. In addition, to capture seasonality and product-specific shocks in demand, I include product by month fixed effects, η_{jt}. ε_{ijt} is an i.i.d error term. β is the parameter of interest and captures social influence in consumer product choices. However, there are challenges that could be contaminated its causal interpretation. I discuss the challenges in detail in the following subsection.

3.1 Addressing Sorting on Correlated Tastes

A long-standing identification challenge with observational data is to differentiate social influence from correlated tastes, which render individuals and their contacts' to form a friendship as well as to conduct similar behavior. For example, one chooses a phone because his friends are using that particular phone. However, such correlation could result from high-tech loving preference instead of social influence stemming from the behavior. The key issue is to show that the correlation among consumers and their

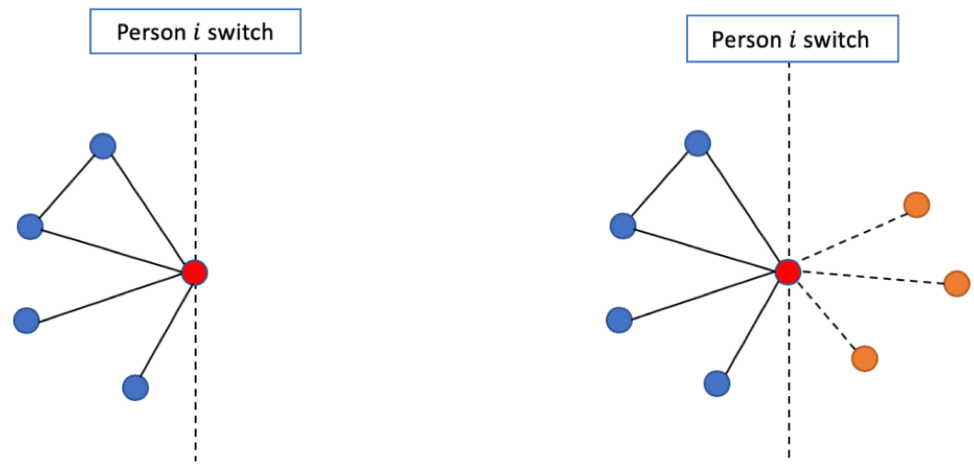
[16] Robustness checks are available when using $t-1$, $t-2$ etc. as the end period in Section 7.

peers' decisions is not driven by sorting on both observed and unobserved correlated preferences.

To deal with the challenge, I develop several strategies to address sorting on observed and unobserved correlated tastes. On the one hand, I include two sets of controls to deal with sorting on observed tastes. First, leverage the network structure, I am able to separate social influence from contextual exogenous effects by directly including controls of the average demographics of friends ($\boldsymbol{Z}_{m(i),j}$). Second, to account for differential preference towards smartphone features, I include a full set of interactions of individual characteristics and primary phone attributes. For example, age is interacted with a full set of attributes to account for differential needs and enthusiasm towards technological features among the young and the old. On the other hand, to address the unobserved correlated tastes, several strategies are taken to mitigate the concern as below by exploiting the intertemporal variation in contacts as well as the partially overlapping network of contacts and residential neighbors.

First, I use a novel falsification test to show the existence of social influence by comparing the effect of two groups of contacts relative to one's purchase timing: current friends vs. future friends. The underlying assumption is that sorting or homophily is about innate characteristics of consumers that are static at least during one year, while the behavioral impacts of social influence is sequential. If the effect is driven by social influence, I would expect to see the following sequence. In essence, one person makes the purchase, then followed by communication with the other person, the other person makes a similar purchase. However, if the effect is purely driven by homophily or unobserved correlated tastes, then two persons could choose products independently, regardless of the time sequence of choices or when they become friends. Then let us consider the current friends and future friends for each consumer as illustrated in Figure 1. By the time of the phone change, the blue dots on the left-hand side are friends one already knows before his phone acquisition i.e., 'current friends,' while the orange dots on the right-hand side are friends he makes afterward, i.e., 'future friends.' The current friends' choices correlate with the buyers' choices could be due to social influence or sorting. However, the future friends' choice would be correlated with the buyer's choice only because they share similar tastes i.e., sorting.

To the extent that the unobserved correlated tastes are static in the sample period, I expect to see current contacts have a similar impact as future friends if the effect in model 1 is driven by sorting. That is, the difference between the impacts of the current contacts and future contacts suggests the existence of social influence. To put into the



(a) Current Friends - Peer Group **(b)** Future Friends

Figure 1: Falsification Test Illustration

Notes: Figure 1 shows conceptual idea for the falsification test to separate social influence from sorting on unobserved tastes. Blue dots on the left-hand side are old friends known prior to the phone change, i.e. current friends; Orange dots on the right-hand side are new friends one makes after the phone change, i.e. future friends.

formal presentation, in model 2, I test the difference between β_1 and β_2.[17]

$$y_{ijt} = \beta_1 s_{m(i)j,t-3} + \beta_2 s_{m'(i)j,t-3} + \boldsymbol{Z_i X_j}\gamma_1 + \boldsymbol{Z_i}\gamma_2 + \boldsymbol{Z_{m(i),j}}\gamma_3 + \lambda_{R(i)f(j)} + \eta_{jt} + \varepsilon_{ijt} \quad (2)$$

where $m(i)$ denotes the current friends and $m'(i)$ denotes the future friends.

To check the assumption that unobserved correlated tastes are about innate characteristics and time-invariant, I show there is no systematic change in the composition of contacts over time. The idea is that if there is a sudden change in the unobserved tastes, I expect to see changes in the social network, and the composition of the contacts. Table A5a and A5b show that there is no systematic differences in observed characteristics of contacts made before and after the change. Thus, no difference in the observed pre-determined characteristics of current and future friends implies no changes in the unobserved tastes.

Second, I construct individual taste controls from choices of same-old-brand users and future friends. A natural way to partial out unobserved time-invariant tastes for smartphones is to include individual fixed effects (Iyengar et al., 2011; Nair et al., 2010).

[17]Falsification tests are also conducted when control for current and future friend characteristics. The results barely change.

Although the data does not allow to include such individual fixed effects, I construct individual taste controls to account for innate preference for smartphones based on overall consumers' phone change patterns and the revealed preferences. Building on the falsification test, the first control variable is the share of future friends using each alternative prior purchase. It indeed provides a unique control for pair-wise correlated tastes. As discussed earlier, the future friends' choices, along with extensive control of its demographic shares, capture sorting on both observable and unobservables through revealed preference. If the main estimate remains stable after adding such controls, it provides evidence that the effects are unlikely to be driven by sorting.

The second control variable is the share of same-old-brand non-contacts, the share of non-contact consumers who replace from the same old brand as the new buyers into each alternative in the earlier month. For example, individual i used to choose Samsung A1 and now purchases OPPO R9 Plus. I look at the non-contacts of i who used to choose Samsung and calculate the share of these past Samsung users choosing OPPO R9 Plus eventually. The share of same-old-brand non-contacts helps to capture the common phone tastes through the revealed preference from actual subsequent choices. The subsequent choices carrying the same taste for the previous brand serve as sufficient statistics for preferences towards specific models. I exclude social contacts from consumers sharing the same old brand to make sure variations in subsequent choices are not affected by one's social network.

Third, utilizing the exogeneity of signal coverage quality across buildings and the partially overlapping structure of call contacts and residential neighbors, I use the choices and the average phone attributes of friends' residential neighbors as instrumental variables for share of friends $s_{m(i)j,t-3}$. A residential neighbor is a person living in the same building (location), a smaller geographical unit than the residential block (neighborhood).

Figure 2 illustrates the idea of the instruments. Phone purchaser i is a friend of person A and D, who reside in one residential building. E is a friend of D and lives in the same building as D. The instrumental variables for A and D's choices exploit the information of the phones of their residential neighbors B and C, who are not direct friends of i or friends of friends of i. To formalize the presentation, let us denote the individual i's choice as y_{ij} and y_{ij} takes value one if individual i chooses product j and zero otherwise. Denote individual i's phone attributes as X_{ik}. Denote individual i's social contacts in peer group as $m(i)$ and i's residential neighbors as $NB(i)$. Then my instrumental variables for $s_{m(i),j,t-3}$ in Model 1 can be defined as $s_{NB(m(i)),j}$, the share of i's contacts' residential neighbors using j:

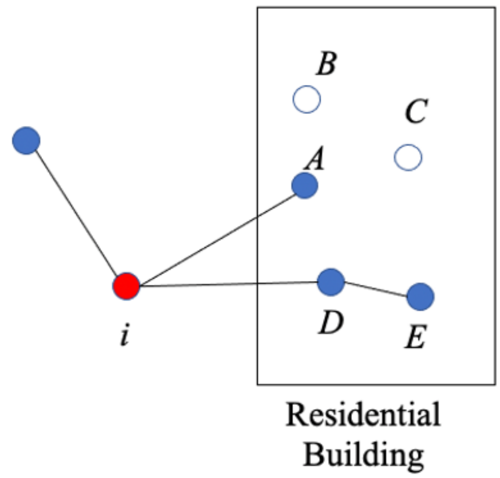


Figure 2: Instrumental Variables Illustration

Notes: Figure 2 shows conceptual idea for the instrumental variables. Phone purchaser i has friends A and D. E is a friend of D. A, B, C, D and E live in the same residential building. The instrumental variables for A and D's choices is constructed from the phone choices of their residential neighbors B and C, who are not direct friends of i or friends of friends of i.

$$s_{NB(m(i)),j} = \frac{1}{|m(i)|} \sum_{m \in m(i)} \frac{\sum_{l \in NB(m)} y_{lj}}{|NB(m)|},$$

and $\boldsymbol{x}_{NB(m(i)),j}$, the average phone attributes of the residential neighbors of i's contacts who use j:

$$\boldsymbol{x}_{NB(m(i))j} = \frac{1}{|m(i)|} \sum_{m \in m(i)} \frac{\sum_{l \in NB(m)} y_{mj} X_{lk}}{|NB(m)|}$$

where $m(i)$ is the set of i's peer group, $m \in m(i)$ is individual i's peer, $l \in NB(m)$ is a neighbor of peer m.

The identification assumption for the friends' residential neighbor instruments is that they must satisfy the relevance and exclusion restriction conditions. The relevance condition is satisfied by two possible factors. First, the correlation between residential neighbors arises due to supply side effects such as common exposure to advertising in nearby stores and elevators. Second, the correlation could also occur due to common signal exposure in the residential building. Local signal quality varies across locations in the same neighborhood due to different distances to nearby cell towers and middle obstructions such as trees and buildings.[18] Research shows that phone's antenna performance is

[18] Morin (2013) suggests that the further away from a cell tower, the weaker your cell phone signal is going to be. Obstructions between phones and the cell tower can cause cell signal issues, including

vital for the phone's ability to ensure radio coverage, especially in low signal situations. Technical reports indicate that mobile coverage and antenna reception affects both voice and data transmission. A phone's internal components (e.g. processor, memory) generate electrical noise that affects reception, and the antenna performance of the different models vary considerably even across popular smartphones (Commission for Communications Regulation, 2018; Pedersen, 2016). Thus, people living in the same residential building with weak signal condition would choose phones or certain phone features that help overcome the problem and provide stronger reception. As the coverage exposure is determined by the base station structures designed by the mobile operator, the neighbor effects are local and exogenously affected by the geographical variation of coverage quality.

The exclusion restriction requires that consumers are not directly affected by their friends' residential neighbors. To make sure I break the direct interactions between the consumer and the friends' neighbors, I drop those friends and friends' friends living in the same residential building as the phone purchaser's friends. In addition, residential neighborhood-by-brand fixed effects in Model 1 also controls for the time-invariant neighborhood-specific common preferences.

3.2 Results

Now I present the results for the baseline model with gradual controls. Table 5 reports the results for the linear probability models for smartphone model choice (see equation 1). In column (1), I only include the residential neighborhood fixed effects to control for spatial and income related factors. In col (2), I further control for the contextual effects by including friend demographic shares. I find that the exogenous contextual effects matter and bring down the main estimates by about one third. In column (3) to (6), I control for product-by-month fixed effects to capture any supply side effects such as marketing. The R-squared increases from 0.013 in column (2) to 0.022 in column (3), while the main estimate does not drop much. This is partly because that the social influence is measured by the lagged outcome of friends within a fixed time window, and the social influence does not vary much seasonally. In column (4), I control additionally for sorting by including the corresponding share of future friends. It raises the R-squared by three times, however, barely changes the main estimate. In column (5), I control for

mountains, hills, large buildings, and even trees. In addition, the building materials at home may be causing varying amounts of cell phone signal interference. For example, metal siding, concrete, and wire mesh can cause significant signal loss. At the same time, wood and drywall generally allow the signal to pass through more easily.

the individual phone taste by adding the share of same-brand non-contacts in the earlier market. The coefficient on this taste control is 0.73, suggesting that a 10 percent increase in the share of same-brand non-contacts using a given alternative is associated with a 7 percent increase in the choice probability. It also increases the explanatory power of the model as the R-squared goes from 0.065 to 0.098. It suggests that common brand preferences explains a large proportion of product choice. Despite the large effect from brand preferences, the main effect remains fairly stable at around 0.10 to 0.11. In the last column, the main effect remains stable even after adding the taste controls together into one regression. Column (6) is the result from the preferred specification. The point estimate is 0.10, suggests that a 10 percent increase in the share of friends using a given product would increase the choice probability by 1 percentage point, which almost doubles the average choice probability (1.6 percent).

Table 6 reports the falsification test for correlated taste. There might be attenuation issues in the key regressors – the share of (current) friends and share of future friends – if the purchase timing is too early or too late in the data. To alleviate the concern, I report the results using the sample of all new buyers in odd columns, and in even columns, I restrict to the subsample of new buyers who change their phone in the middle of the sample period, from the fourth to the eighth month. In all columns, individual residential-by-brand fixed effects, product by month fixed effects, and friend demographic controls are included.[19] Although the two regressors have similar means and standard deviations as reported in Appendix Table A2, column (1) and (2) suggest that the impact of current friends is 10 times bigger than the impact from future friends. This finding provides evidence that the effect is not purely driven by sorting. In column (3) and (4), I further add in individual taste control of the share of same old-brand non-contacts in the earlier market, and the future friend's impact diminishes and becomes less precise. However, in contrast, the impact of current friends remains stable and robust. In column (1), the impact from future friends is 0.008 and significant at 5 percent level. However, in column (3), future friends' impact goes down to 0.005 and becomes insignificant. A similar change also features in column (4) compared to column (2): the main estimate remains stable at 0.10, but future friends become non-influential after controlling for individual phone tastes. Such findings support the conjecture that social influence exists, and it is hard to be reconciled by sorting on correlated tastes. It also further comforts that a rich set of controls effectively control for unobserved phone tastes. In addition, treating future contacts as a control for the unobserved tastes, the positive social influence still goes through.

[19] Results barely change when demographic controls for both current and future friends are all included.

In Table 7, I report the 2SLS results in comparison to the OLS results. Column (1) and (2) show the comparison with only residential fixed effects, while column (3), (4) and (5) include all controls described in the preferred specification. To make it comparable to the 2SLS counterparts, In columns (2) and (4), I use the friends' residential neighbors' choices and their phone attributes (average CPU clock speed and 4G compatibility) as the instrumental variables. The F-tests for the significance of the instruments reported at the bottom of Table 7 suggest that the instrumental variables are strong and statistically significant. Column (1) reports the estimates using specification of Table 5 column (1). Column (2) reports the IV counterparts to column (1), and it delivers a slightly smaller estimate than the OLS counterpart in column (1). The reduction in the main estimate suggests that the instruments help remove the upward bias. The F-statistics is 580.5, suggesting the instruments are strong. Column (3) carries the OLS result from the specification in Table 5 column (3). Column (4) reports the first stage estimates. Share of friends' neighbors and the average CPU speed and 4G connection of friends' neighbors significantly correlate with the share of friends, which suggests a valid first stage relevance. Column (5) reports the IV counterpart to column (4) when adding a full set of controls, including interactions of individual-product characteristics, individual residential-by-brand fixed effects, product-by-month fixed effects, and friend demographic controls. The main estimate is 0.106, not statistically different from the OLS estimate 0.10 in column (3). The similar magnitude of OLS and 2SLS estimates suggests that the rich set of controls in the main specification does help control unobserved individual tastes. The main estimate is quite robust at 0.10 across several different specifications.

As the average conditional choice probability for a particular product is about 1.6 percent, demand for a given model doubles with a 10 percent increase in friends' share using that particular product conditional on purchasing. Conversations with a marketing expert in the industry at IDC suggest that a successful marketing campaign leads to a 4 percent increase in the smartphone market. Therefore, a 1 percent increase from a 10 percent increase in friend shares, i.e., about 2 to 3 same-carrier friends or 6 friends in general, is quite a sizable impact.

3.3 The Influencer, Affluent Friends and Status-Seeking

It has been a challenge to understand the underlying mechanism behind the social influence with observational data in the literature due to a lack of information on peers. With rich information on both the friends' choices and friends' demographics, I explore the possible underlying mechanism behind the social influence by examining the heterogeneous effects across peer groups and product types.

Based on the literature and the content, three possible channels are considered: information sharing, status-seeking, and attraction by the same operating system in the context of smartphones. The first two channels are usually discussed in the peer effect literature. One possible channel is conformity, as high-tech products like smartphones are considered status symbols in developing countries (Dey et al., 2016; Jain, 2017; Katz and Sugiyama, 2005). As a signaling device, people would be attracted by the style and visual features and be better off by conforming to a particular group of friends and choosing the same product as friends. Moreover, information sharing would allow people to know the features and functions of phones and update beliefs about product quality. Such a process would trigger the consumption of certain products. The information sharing channel is consistent with the "word of mouth" notion in marketing. Lastly, for smartphone specific features, people may prefer to use the same phone as their friends to utilize the same features shared by the same operating system. Although these channels are far from complete, I try to use the social network, and detailed information on socio-economic status and product attributes to enrich the understanding of the behavioral motivations. To do so, I stratify peers into different socio-demographic groups and examining heterogeneous influence by peer groups and product attributes.

Status-Seeking and the Reference Group First, I stratify peers into different groups by their socio-demographic conditions and examine the heterogeneous effects from different groups. I find stronger heterogeneous effect by income levels. Table 8 reports the results when use per square meter house price as income proxy. An alternative income measure –average monthly plan fee is used and the results are reported in Robustness Table 26. Column 1 compares the influence of friends of different *absolute* income levels. Three categories – high, middle and low – are considered if the friend's income measure is above, within and below one standard deviation of the distribution. The coefficient on high income and low income are statistically different. A 10 percent increase in the high income friends is 1.5 times the impact than a 10 percent increase in the low income friends.

Next, I stratify peers into two groups *relative* to the consumer (ego): more affluent than the ego and less affluent friends. A peer is considered as more affluent than the ego if the income measure is larger than the ego by at least one standard deviation of the distribution, and otherwise as similar or less affluent. Table 8 Column 2 reports the result using house price as income proxy. It suggests that friends with relative higher house price is 2.5 times influential than friends of similar of lower house price. Taking the results from both absolute income and relative income, it suggests that people tend to conform to their wealthy friends, which is consistent with the status good hypothesis.

Status-Seeking and Product Attributes Attraction Along with studies of consumption of status symbol such as luxury goods, people favor visible features (Heffetz, 2012; Veblen, 1899) due to psychological demostration effects. To further investigate the mechanism, I classify product attributes into two groups: visual and hidden features. Visual features highlight the horizontal differentiation that are less quality-representative, including the average screen size, the number of colors available, and the number of cameras for each brand among products released from 2012 to 2017. In contrast, the hidden features refer to vertical attributes representing the phone quality that affect phone performance, but not easily seen without experiencing. The average vertical features of all models released by each firm in the past five years represent the overall quality of the brand. So I focus on CPU clock speed, and screen resolution.

Table 9 reports the heterogeneous effects of peers by phone attributes. Interestingly, social influence faciliates the demand from better visual feature, instead of functional features. In Table 9, in order to disentangle effects by product attributes, I replace the product-by-month fixed effects with brand-by-vintage-by-month three-way fixed effects, while controlling for all key product features in all columns. Taking other features as constant, a 10 percent increase in the friend share would lead an additional increase of 0.56 percentage point in the choice probability for models with a bigger screen compared to models with smaller screen. Similarly, models of more color options, more cameras attract higher demand through peers. However, this is not true for hidden functionality such as higher CPU speed and better screen resolution. Hence, taking together with the findings in affluent peers, it suggests that people tend to conform to peers due to status-seeking.

Information Sharing It is possible that one learn about the products from peers and then make the purchase. This is usually hard to distinguish without experiments. I provide suggestive evidence that is not consistent with the information sharing channel by examining heterogeneous effects by peers who are possible source of new information.

I observe coworkers of these new phone buyers and the job movements.[20] I look at the intersection of friends and coworkers. Newly joined coworkers are possible sources

[20]There are about 8% job changers during the sample period as documented in Barwick et al. (2019), a separate work using same data.

of new information and provide new information on phones. Assume that conditional on the workplace neighborhood one works in, new coworkers joining the workplace is exogenous to one's phone choices. Thus, as a first check, I focus on consumers who have at least one recently joined coworker prior to phone change and exploit the exogenous shifts in the coworker composition to see the information vs. conformity channel. If the effect is driven by information, I expect to see that newly joined coworkers have a bigger influence than the pre-existing ones on one's phone choice as they are the new shock to the coworker circle and convey new information about the products. However, if it is driven by conformity, I would like to expect a higher influence from the pre-existing coworkers than coworkers who recently joined. Another possible source of new information is new friends. I compare the influence by friendship length. On average, friends in the peer group are known for thirty weeks. Then I define longer (shorter) friendship as friends who know more (less) than thirty weeks. If the effect from friends with shorter relationship is stronger than that from those with longer relationship, I cannot reject the hypothesis that the effect is driven by information.

Results in Table 11 Column 1 suggest that the pre-existing coworkers have stronger influence, while the newly joined coworkers' influence is not precisely estimated and not statistically different from zero. However, the sample size drops dramatically due to the fact that only 8 percent of people are changing jobs. This also makes the mean of "share new coworker" quite small than that of "share pre-existing coworker". On caveat of interpreting the comparision is that there is not enough variation among new coworkers. However, Column 2 provides another piece of evidence without the problem of sample attrition. Column 2 suggests friends who are known for a relatively longer time have higher influence than those known for shorter time. Although the two variables have similar mean and standard deviation, they show different influence over the model choice. Taking the two pieces of evidence together, people are more likely to choose the product used by friends and coworkers that they know relative longer, thus it is suggestive that the social influence is not consistent with information sharing channel.

Operating System Compatibility It is possible that people would like to choose the same product as their friends because they can share the same mobile operating system to facilitate communication. So far, there are three major mobile operating systems - IOS, Android and others.[21] In particular, Apple's IOS shows such operating system effect because it enables users to connect through its unique features such as FaceTime and iMessage. To investigate the effect through operating system, I consider the share of

[21] Others includes Unix, BlackberryOS etc.

current friends using the same operating system as each given alternative. If operating system entails users to adopt similar products, then having a larger share of friends using the same operating system would increase not only the chance of choosing one particular model, but also models of the same OS. However, the remaining variation across products within the same OS would not be explained by OS effect alone.

Specifically, I include "Share Same OS" and "Share Friend" into the same regression. If the social influence is driven by OS effect, I would expect the OS effect be statistically significant and the main coefficient to decrease. Table 10 reports the estimate for OS effects. In column (1), the same OS effect is about 0.004, a much smaller impact than the social influence. It suggests that among users choosing the same product in the same month, there is a slightly small increase in demand induced by a larger number of peers using the same operating system. Moreover, the main effect is rather stable at around 0.10 as in Table 5 column (6). Table 10 column (2) reports the effect due to same brand effect. The same brand effect is about 0.02, and significant at 1 percent level. This suggests that a 1 percent increase in the share of friends using the same brand would additionally increase the conditional choice probability by 0.002 percentage points. Such increase could be driven by the preference of using the same brand product or same smartphone application on the same brand phone with friends. However, the main effect remains stable at around 0.09, suggesting the social influence is not fully absorbed by the same operating system and brand effect.

To summarize, I find sizable social influence in consumer demand that a 10 percent increase in the share of friends using a given product would increase the average choice probability by 0.01, which almost doubles the average choice probability conditional on purchasing. This result remains robust after controlling for sorting on correlated observables, unobserved neighborhood characteristics, and unobserved phone tastes. Then, I explore the underlying mechanisms of the social influence in smartphone choices by examining the heterogeneous effects by peers wealth and product characteristics. I find that people tend to conform to affluent peers both in relative and absolute levels. Visual attributes of phones capture higher influence. Information from new colleagues and new friends are not as important as suggested by the information sharing channel. Although I cannot exclude the possibility that consumers choose the same product as their friends due to the same operating system, the effect remains at a small magnitude. Therefore, the results show suggestive evidence for status-seeking motivation and using same operating system behind the social influence.

4 Structural Model for Smartphones with Social Influence

To move from individual spillover to aggregate effects on demand and firm competition, I need a framework to evaluate preferences and understand how these individual-level effects translate into firm incentives. To do so, I develop and estimate a model for demand and a two-period dynamic pricing model, incorporating the social influence. The model will allow me to perform counterfactual simulations to examine how social influence affects the demand for products of different qualities and how pricing strategy changes when firms compete under social influence.

4.1 Demand

As suggested in Section 3.3, social influence is more likely to work through conformity due to status-seeking. I incorporate the social influence into the random-coefficient discrete choice model to describe smartphone demand and to quantify the complementary value of peers consumption due to preferences for conformity and common operating system.

A conditional choice problem is considered as I focus on the set of new buyers who provide me exact purchase events and an accurate set of products at the time of purchase. The model can be extended to incorporate the extensive margin by adding an outside option of not purchasing the handset each month and expanding the sample size to all users. However, this extension requires more restrictive assumptions on how the extensive margin decision (purchase or not) is affected by peers and the relationship between social influence at the extensive and the intensive margin. Moreover, for non-new buyers, without exact purchase timing and no accurate information on the duration of phone possession, it requires a lot of data imputation. Since the study's focus is on the social influence in product choice, the conditional choice problem suits the need without the cost of imposing complicated assumptions on the extensive margin and data imputation.

A market is defined as a urban/suburban/rural geographical area[22] by month. In each market, conditional on purchasing, each consumer choose from J_t models to maximize utility. Indirect utility of individual i buying product j in market t is a function of

[22]There are five urban districts in the city proper and eighteen surrounding rural counties in total. I consider the five urban districts as one urban market, five suburban counties and satellite cities as one suburban market, and the rest as a rural area.

product attributes, share of peers beforehand and individual demographics:

$$u_{ijt} = \bar{u}(p_{jt}, X_{jt}, \xi_{jt}, s_{m(i),j,t-3}, D_i) + \varepsilon_{ijt} \tag{3}$$

Then, I specify $\bar{u}(p_{jt}, X_{jt}, \xi_{jt}, s_{m(i),j,t-3}, D_i)$ as

$$\bar{U}_{ijt} = \alpha_i p_{jt} + \sum_{k=1}^{K} X_{jk}\beta_{ik} + \theta s_{m(i),j,t-3} + \zeta_{f(j)} + \eta_t + \xi_{jt} + \varepsilon_{ijt} \tag{4}$$

where $s_{m(i),j,t-3}$ is the share of friends of i using product j 3 months prior to the phone change. It reflects two new features of the model that social influence captures. On the one hand, it allows for the intertemporal social multiplier effects between consumers. Peers in $m(i)$'s consumption at $t-3$ will affect i's decision at t. In this way, even though consumers are myopic, social influence generates a dynamic nature in demand. On the other hand, social influence enters as an additional product feature that captures the complementary value between consumer and the peers. It increases the horizontal product differentiation, which would soften the competition.

Consumer i is described by $w_i = (y_i, \boldsymbol{D}_i, \boldsymbol{\nu}_i)$, where y_i is income proxied by house prices, $\boldsymbol{D}_i$ includes age and total call minutes, and $\boldsymbol{\nu}_i$ is unobserved independent standard normal taste shocks. Total call minutes reflect the usage intensity of the users. Assume that ν_i is independent of the unobserved quality shock ξ_{jt}.

To reflect the motivation of conforming to wealthier friends and the value of using applications on the same phone, I enrich the model with the heterogeneous value of the share of friends by individuals' income and usage intensity. So the indirect utility becomes

$$\begin{aligned}\bar{U}_{ijt} = \alpha_i p_{jt} + \sum_{k=1}^{K} X_{jk}\beta_{ik} + \bar{\theta} s_{m(i),j,t-3} + \theta_{inc} s_{m(i),j,t-3}\mathbf{1}\{y_i > p75\} + \theta_{use} s_{m(i),j,t-3} Ncalls_i \\ + \zeta_{f(j)} + \eta_t + \xi_{jt} + \varepsilon_{ijt}\end{aligned} \tag{5}$$

where $\mathbf{1}\{y_i > p75\}$ takes value one if the phone buyer' income (house price) belongs to the top 25th percentile of the distribution. $\bar{\theta}$ is the base social influence. θ_{inc} captures the additional utility gain of high income individuals to conform to friends. θ_{use} reports the additional utility gain for intensive users when choosing the same product as friends.

I define individual i's marginal utility for one hundred dollar α_i is defined as

$$\alpha_i = \bar{\alpha} + \alpha_1 \mathbf{1}\{y_i > p75\} + \sigma_p \nu_{ip} \tag{6}$$

The first term in random coefficient α_i is the base price sensitivity $\bar{\alpha}$. The second component $\alpha_1 \mathbf{1}\{y_i > p75\}$ captures the change of the disutility from price if income belongs to the top 25% of the income distribution. One would expect α_1 to be negative since wealthy consumers are less price sensitive. The third term is a random shock which captures idiosyncratic factors that influence price elasticity, such as assets accumulated in the past. ν_{ip} is assumed to follow the standard normal distribution, and σ_p is the dispersion parameter.

X_{jt} is a vector of observed product attributes, including a constant term, screen size, weight, main camera resolution, CPU clock speed (X_{jk}). I define individual i's taste for attribute k as:

$$\beta_{ik} = \bar{\beta} + \boldsymbol{D}_i \boldsymbol{\beta}_{Dk} + \sigma_k \nu_{ik} \tag{7}$$

which follows a random normal distribution with mean $\bar{\beta}_k$ and standard deviation σ_k. Different consumers may have different tastes due to unobserved demographics or idiosyncratic preference. To capture rich preference heterogeneity, I interact phone attributes with individual age.[23] Similar to the discussion in Section 3, for example, it accounts for preferences that older people may prefer to buy phones with larger screens. I also allow random tastes for the screen size, camera resolution and CPU clock speed in addition to price, and assume dispersions for other attributes to be 0.

ξ_{jt} is the unobserved product attributes that are observable to both firms and consumers but unobserved to the econometrician, such as product quality perceived by consumers. $\zeta_{f(j)}$ are brand dummies, captures brand-specific permanent shock for j, $f(j)$ is the brand for product j. η_t are area-by-month fixed effects. Finally the idiosyncratic preference shock ε_{ijt} is assumed to be i.i.d across (i, j, t) and follow type I extreme value distribution.

To facilitate the discussion on identification and estimation below, I rewrite the utility function as:

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt} \tag{8}$$

where

$$\delta_{jt} = \bar{\alpha} p_{jt} + \sum_{k=1}^{K} X_{jk} \bar{\beta}_k + \zeta_{f(j)} + \eta_t + \xi_{jt} \tag{9}$$

[23] Interactions with call minutes are rarely significant.

$$\mu_{ijt} = (\alpha_1 \mathbf{1}\{y_i > p75\} + \sigma_p \nu_{ip}) p_{jt} + \sum_{k=1}^{K} X_{jk} (\boldsymbol{D}_i \boldsymbol{\beta}_{Dk} + \sigma_k \nu_{ik})$$
$$+ \bar{\theta} s_{m(i),j,t-3} + \theta_{inc} s_{m(i),j,t-3} \mathbf{1}\{y_i > p75\} + \theta_{use} s_{m(i),j,t-3} Ncalls_i \quad (10)$$

μ_{ijt}, the individual-specific utility, depends on individual characteristics and the peers past choices. δ_{jt}, the mean utility captures only product by market specific components.

I use $\boldsymbol{\theta}_1$ to denote parameters in δ_{jt}, which I call linear parameters, and $\boldsymbol{\theta}_2$ to denote parameters in μ_{ijt}, which I call nonlinear parameters, following Berry et al. (1995). The nonlinear parameters are individual specific and include: $\boldsymbol{\theta}_2 = \{\bar{\theta}, \theta_{inc}, \theta_{use}, \alpha_1, \beta_{age,2}, \beta_{age,3}, \beta_{age,4}, \sigma_p, \sigma_2, \sigma_3, \sigma_4, \sigma_5\}$, where $\bar{\theta}, \theta_{inc}, \theta_{use}$ measure social influence, α_1 are the how the marginal utility for price change for high income, $\beta_{age,2}, \beta_{age,3}, \beta_{age,4}$ are the parameters capturing how the marginal utility for phone screen size, camera resolution and CPU clock speed change with age, and $\sigma_p, \sigma_2, \sigma_3, \sigma_4, \sigma_5$ are the parameters that measure dispersions in random tastes for price, screen size, camera resolution, CPU clock speed and weight.

Thus, the conditional choice probability that i chooses product j becomes:

$$P_{ijt}(Y_i = j | \boldsymbol{X}, \boldsymbol{p}, s_{m(i),j,t-3}, w_i, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{exp(\delta_{jt} + \mu_{ijt})}{\sum_{j'=1}^{J} exp(\delta_{j't} + \mu_{ij't})} \quad (11)$$

I use the individual conditional choice probabilities for form maximum likelihood and estimate the nonlinear parameters.

Let A_{jt} be the set of consumer characteristics such that j has the highest utility for consumers in this set. That is, $A_{jt} = \{\nu_i | u_{ijt}(s_{m(i)}, y_i, \nu_i, \boldsymbol{D}_i) \geq u_{ij't}(s_{m(i)}, y_i, \nu_i, \boldsymbol{D}_i), \forall j'\}$ Then aggregate individual choice probabilities to obtain the market share of product j at the market t:

$$s_{jt}(\boldsymbol{X}, \boldsymbol{p}, \boldsymbol{s_m}, \theta_1, \theta_2) = \int_{i \in t, A_{jt}} P_{ijt}(\boldsymbol{X}, \boldsymbol{p}, \nu_i, s_{m(i)}, \theta_1, \theta_2) dF(\nu_i, s_{m(i)}) \quad (12)$$

where $\boldsymbol{s_m}$ is a vector of share of friends for individuals. I use the market shares for mean utility inversion in the estimation following Berry et al. (1995) and Goolsbee and Petrin (2004).

I choose a static demand system for the following reasons. First, as discussed in Section 2, the Chinese smartphone market has been saturated after 2015, and the demand becomes stabilized with a slight decline in new sales. Second, the smartphone market is saturated with domestic products in all product segments that provide various functional features at relatively low prices. This market feature remarkably reduces the replacement cost, making Chinese smartphone users replace their phones more frequently than global users.[24] More phone options at low cost essentially shorten the replacement cycle. Third, smartphones have a stable penetration rate of around 50 percent since 2015.[25] This suggests that with relatively low switching cost, consumers are more likely to replace their phones at their need without much intention to delay. Therefore, a static demand is a feasible option to estimate in twelve-month data and captures the market features well.

4.2 Supply: Two-Period Pricing Model

Firms compete on prices. Pricing decisions are crucial for smartphone firms, especially the release prices determines the pricing trajectory and the profits over the life cycle based on the following two facts that I document from the market-level data. First of all, market-level data suggests that the average life cycle of a model is 4 to 5 quarters since 2015. Notably, more than 50 percent of the revenue comes from the first 3 quarters, i.e., the half life cycle. So the release prices are the most relevant prices at the demand peaks. Second, although the phones' prices are going down over time (Figure 4), the release prices for top-five brands remain stable and slightly increase over time, as suggested in Figure 5. Figure 4 plots the prices for all models released after 2015 by quarters since release. Each light blue line in the background indicates the pricing trajectory for a model. The dark blue line is the median price across all these models, suggesting that prices decline over time. When zoom into the pricing pattern for top brands and their top models in each year in Figure 5, it is interesting that the release prices are not necessarily going down. Instead, the release prices are relatively stable for Apple and OPPO phones and increasing over time for Huawei and vivo phones. Therefore, release prices are crucial decisions for firms as it determines both the pricing path and the profit path over the life cycle. To keep the model tractable, I focus on the pricing stage while abstracting away from early-stage decisions such as product entry decision and phone attribute choice.

I adopt a dynamic pricing model with two periods. Two-period is chosen to allow me to capture the decision of release prices and keep the model flexible to capture price drop over time while remaining computationally tractable. Firms choose optimal prices for smartphone models in each period to maximize the expected discount profits. In the

[24] According to the China Mobile Consumer Survey 2018 released by global accounting and consulting firm Deloitte, nearly 80 percent of Chinese users, bought their current phones in 2017 compared to just 58 percent of global users.

[25] Mobile phone internet user penetration in China 2015-2025, Published by Statista Digital Market Outlook, July 17, 2020, https://www.statista.com/statistics/309015/china-mobile-phone-internet-user-penetration/

first period, pricing takes into account the social multiplier effect and the possible social differentiation effect in the second period. It is equivalent to assuming that Period 2 in my data is the end of product life, or firms only care about the first two periods of their life cycle and do not play the game after Period 2. I divide the sample period into two: Q4 2016, Q1 2017 as Period 1 and Q2 2017, Q3 2017 as Period 2. Among 62 products, 35 of them are new released after Q2 2016.[26] So a cost estimation using all models would fit the two periods of the actual life cycle. To isolate the impact of social influence on the pricing strategy while controlling for these other factors, I allow marginal costs to change over time and enable price elasticity to respond to social influence.

Let $\boldsymbol{p}_1, \boldsymbol{p}_2$ denote prices in Period 1 and Period 2, $\boldsymbol{mc}_1, \boldsymbol{mc}_2$ denote marginal costs in each period. A firm f maximizes the expected discount profit:

$$W_f = \sum_{j \in J_f} (p_{j1} - mc_{j1}) Q_{j1} + \delta (p_{j2} - mc_{j2}) Q_{j2} \tag{13}$$

where

$$Q_{j1} = \sum_{t \in \{T=1\}} s_{jt} M_t, \qquad Q_{j2} = \sum_{t \in \{T=2\}} s_{jt} M_t \tag{14}$$

t denotes an market (area by month), and T denotes the period. s_{jt} is the aggregate market share of product j in market t in Equation 12. M_t is the market size of t, proxied by the total number of mobile users including new buyers and non-buyers in each month. δ is the discount factor. J_f represents the products offered by firm f. $p_{j2} = p_{j2}(\boldsymbol{Q}_1(\boldsymbol{p}_1))$ is a function of the first-period prices. The SPNE prices are solved using backward induction starting from Period 2. The first-order condition for Period 2 becomes

$$p^*_{j2} = mc_{j2} - \underbrace{\frac{\partial Q_{j2}(\boldsymbol{p}, \boldsymbol{X}, \boldsymbol{s_{m1}})}{\partial p_{j2}} \times Q_{j2}}_{\text{Price Markup due to Social Differentiation Effect}} = p^*_{j2} + [\Delta^{-1}_{f2} \times \boldsymbol{Q}_2]_j \tag{15}$$

where Δ_{f2} is a J-by-J matrix, whose (j, r) element is $\frac{\partial Q_{r2}}{\partial p_{j2}}$ if j and r are produced by the same firm and zero otherwise. The second term in Equation 15 is the price markup and represents how much the optimal price chosen by a firm deviates from the competitive price (equal to the marginal cost). The markup includes a semi-demand elasticity to price, that takes into account the social influence on the firm's pricing strategies. The

[26] There are 16, 8, and 6 new products in Q3 2016, Q4 2016, and Q1 2017, respectively. Q3 is usually a season of model release since Apple releases new products in September and competing firms usually follow Apple's timeline to introduce new models. 5 new products in the second period (Q2 and Q3 2017).

semi-elasticity term differs from the counterpart without social influence as it considers the peer's choices $\boldsymbol{s}_{m1}$. Specifically, if $\theta > 0$, more friends using a given product would create "social differentiation effect", i.e., making people more likely to choose the product due to the social complementary value and become less sensitive to prices. The social differentiation effect intensifies the horizontal product differentiation among products and provides an additional markup than the case without social influence. Such additional differentiation would lead firms to increase prices to "harvest".

The first-order condition for Period 1 becomes

$$p_{j1}^{*} = mc_{j1} - [\Delta_{f1}^{-1} \times \tilde{\boldsymbol{Q}}_1]_j \tag{16}$$

$$\tilde{\boldsymbol{Q}}_1 = \boldsymbol{Q}_1 - \delta \underbrace{\frac{\partial \boldsymbol{Q}_2(\boldsymbol{p}, \boldsymbol{X}, \boldsymbol{s_{m0}})}{\partial \boldsymbol{p}_1}}_{\substack{\text{Inter-temporal} \\ \text{Social Multiplier Effect}}} \times \underbrace{\left\{\Delta_{f2}^{-1} \times \boldsymbol{Q}_2 + Diag(\Delta_{f2}^{-1} \times \boldsymbol{Q}_2)\right\}}_{\substack{\text{Price Markup due to Social Differentiation Effect} \\ \text{in Period 2}}} \tag{17}$$

where Δ_{f1} is a J-by-J matrix, whose (j, r) element is $\frac{\partial Q_{r1}}{\partial p_{j1}}$ if j and r are produced by the same firm and zero otherwise. Note that for new products released in Period 1, the lagged share of friends terms are all zero. So own price semi-elasticities, i.e., diagonal terms of Δ_{f1}, for these new products are not affected by social influence. Therefore, in Equation 16, the major influence comes into effect through the adjusted quantity sold $\tilde{\boldsymbol{Q}}_1$.

The inter-temporal partial derivatives $\frac{\partial \mathbf{Q}_2}{\partial \mathbf{p}_1}$ is a function of the social influence θ, and it can be obtained through analytical derivation as the following

$$\frac{\partial Q_{j2}}{\partial p_{r1}} = \int_i \frac{ds_{ij2}}{dp_{r1}} dF(i) = \int_i \theta s_{ij2}(1 - s_{ij2}) \left(\sum_{l \in m(i)} \frac{\partial s_{lj1}}{p_{r1}} \right) dF(i) \tag{18}$$

$$\frac{\partial s_{lj1}}{p_{r1}} = \begin{cases} \alpha_l s_{lj1}(1 - s_{lj1}) & \text{, if } j = r \\ \alpha_l s_{lj1} s_{lr1} & \text{, if } j \neq r \end{cases}$$

where $F(i)$ is the distribution of individuals, $l \in m(i)$ is a friend of individual i, s_{lj1} is friend l's individual choice probability for product j in Period 1. In Equation 18, θ enters the inter-temporal semi-elasticity, indicating that the dynamic nature of demand arise due to the social influence. I call it "intertemporal social multiplier effect". Such impact enters Equation 17 thus Equation 16 affecting the first period pricing decision.

When social influence is present, the multiplier effect provides firm incentive to invest in consumer base in the initial period, which can then be leveraged to enact price increase in later periods. Therefore, the model predicts that social influence would leads to lower introductory prices to "invest", which can be tested in the counterfactual simulation. Details on model prediction illustration can be found in Appendix D.

Assume that marginal cost depends on product characteristics, brand fixed effects, month fixed effects and a product-time specific shock.

$$ln(mc_{jT}) = W_{jT}\phi + \omega_{jT} \tag{19}$$

where W_{jT} includes log of phone attributes, firm dummies and a second-period dummy, $T = 1, 2$. The second-period dummy captures the fact that the technology frontier is moving and the marginal cost of existing products goes down. ω_{jT} stands for unobserved cost shock to model j in period T. Combining Equations 15, 16 and 19 yields

$$ln\begin{bmatrix} p_{j1} + [\Delta_{f1}^{-1} \times \tilde{\boldsymbol{Q}}_1]_j \\ p_{j2} + [\Delta_{f2}^{-1} \times \boldsymbol{Q}_2]_j \end{bmatrix} = \begin{bmatrix} W_{j1} \\ W_{j2} \end{bmatrix} \phi + \begin{bmatrix} \omega_{j1} \\ \omega_{j2} \end{bmatrix} \tag{20}$$

which I bring to data for estimation.

5 Estimation

5.1 Estimation Procedure

Parameter of Interest and Identification Similar to the reduced-form analysis, θ is the parameter of interest and captures the local consumption externality among consumers. It is modelled as the same in general for all products and all consumers. As discussed in section 3, a rich set of controls help to account for the unobserved correlated tastes. The controls include interaction terms of individual characteristics and phone attributes, average peer characteristics, residential neighborhood by brand fixed effects, and product by month fixed effects. In the utility specification, the random coefficients and the interaction terms with user demographics serve the same function to capture heterogeneous preferences for phones. The contextual exogenous effects from peers also collapse into this part in the utility specification because it captures correlation in terms of demographics. Since aggregating market shares to neighborhood level would be too demanding,[27] the market (area by month) dummies in the mean utility serves to capture

[27] At neighborhood level, market shares are tiny.

the differential income effects at a cruder level than the neighborhood. The mean utility part in Equation 9 captures the product by market fixed effects as a whole. Given the rich model specification, I take the share of friends as exogenous and social influence is identified from the variation in friends' phone choices among consumers conditional on product tastes.

Estimation In the demand model, there are two sets of parameters to be estimated. $\boldsymbol{\theta}_1$ collects parameters in δ_{jt} (Equation 9), also called "linear parameters"; $\boldsymbol{\theta}_2$ collects parameters in μ_{ijt} (Equation 10), also called "nonlinear parameters". $\boldsymbol{\theta}_1 = \{\bar{\alpha}, \bar{\beta}_1, \bar{\beta}_2, \bar{\beta}_3, \bar{\beta}_4, \bar{\beta}_5, \gamma\}$, $\boldsymbol{\theta}_2 = \{\bar{\theta}, \theta_{inc}, \theta_{use}, \alpha_1, \beta_{age,2}, \beta_{age,3}, \beta_{age,4}, \sigma_p, \sigma_2, \sigma_3, \sigma_4, \sigma_5\}$, where 2 to 5 represents phone screen size, camera resolution, CPU clock speed, weight and battery capacity. γ represents a vector of 7 brand fixed effects and 30 market fixed effects.[28] There are 43 linear parameters and 12 nonlinear parameters to estimate.

Following Goolsbee and Petrin (2004), the estimation is conducted in two steps. In the first step, I maximize the simulated likelihood subject to a constraint to find the nonlinear terms and product by market-specific constants. In the second step, I recover the linear parameters. In the first step, I do not maximize it over the entire space of $(\boldsymbol{\theta}_2, \delta)$ directly. Instead, in the spirit of Berry et al. (2004), I conditional on $\boldsymbol{\theta}_2$ and solve for the vector $\delta_{jt}(\boldsymbol{\theta}_2)$ market by market that matches observed market shares to those predicted by the model. It is equivalent to maximize the simulated likelihood subject to a constraint.

Specifically, let s^N denote the market share observed in the data. At each $\boldsymbol{\theta}_2$ and for each market t, I use a contraction mapping routine to solve for

$$\delta_{jt}^{h+1}(\boldsymbol{\theta}_2) = \delta_{jt}^{h}(\boldsymbol{\theta}_2) + lns_{jt}^N - lns_{jt}(\boldsymbol{\theta}_2, \delta_{jt}) \tag{21}$$

where $s_{jt}(\boldsymbol{\theta}_2, \delta_{jt})$ is j's model predicted share in market t at δ_{jt} and $\boldsymbol{\theta}_2$, s_{jt}^N is the observed market share from the data. Because the fixed effects exist and are unique, the δ_{jt} that sets this objective function to zero exists and is known to be the unique minimum.

In the second step, I deal with the endogeneity in price and market share using instrumental variable approach. Two sets of instruments are constructed. The first set is the BLP instruments. It includes the number of products on the market in the same year by the same firm, and number of products in the same year by the rival firm. They capture the competition intensity that affects firms' pricing decisions. The second set of instruments is the differentiation IVs following Gandhi and Houde (2019). They

[28] 7 brands include Apple, Huawei, Xiaomi, OPPO, vivo, Samsung, and others. 30 market fixed effects include the interaction of 3 areas (urban/subruban/rural) by 10 months.

capture the substitution and competition along the product characteristics space. Non-price attributes are assumed to be orthogonal to ξ_{jt}. Details for estimation routine can be found in Appendix C.

On the supply side, as reported in Table A6, the prices go down during the sample period, as the average (release) price is 266.56 dollars in the Period 1 and 244.27 dollars in Period 2. The average depreciation rate is about 0.89. With the observed prices and demand estimates, the marginal costs are estimated using Equations 15 and 16. The discount factor δ is set to be 0.95. Static and inter-temporal demand semi-elasticities are computed using observed data and the demand estimates. The cost parameters ϕ are obtained using Equation 20 when assuming a normal distributed cost shock.

5.2 Estimation Results

To facilitate computation, the estimation is done in a random sample of 5,000 new buyers, which gives me 187,316 observations at individual-model level and 1,142 observations at product-market level. Table 12 reports the estimation results from my demand model. I present coefficients on Share Friend, interaction terms and key phone attributes as well as the parameters that measure the dispersion in random coefficients. Table 13 reports alternative model specifications and the main estimates are quite stable. The log-likelihood is highest in the specification in Table 12.

For an average consumer, having a one percent increase in the share friend would increase the utility by 0.038 evaluated at the mean of high income fraction 0.21 and the average call duration 3727 minutes (0.01*(2.815+(-0.247)*0.21+0.302*3.272). The initial estimated mean price coefficient -0.911, coefficient on price interaction with high income 0.20 and the price dispersion coefficient 0.0001 give the aggregate price elasticity as -1.04, which is below the industry estimate.[29] Following Berry et al. (2004) and Gentzkow (2007), I calibrate the price dispersion parameter σ_p and re-estimate the mean price coefficient $\bar{\alpha}$ such that the model predicted aggregate price elasticity matches the industry estimate. Then, the mean price coefficient becomes -1.032, and the price dispersion is calibrated to be 0.6.

The estimated coefficient on the lagged share friend is 2.815, statistically significant at 1 percent level. Thus, the willingness to pay for a one percent in share friend is 0.036 (0.038/1.032). That is, a one percent increase in share friend is equivalent to a 3.6 percent reduction in price. The average price for a smartphone is 250 dollars (1759 RMB). Thus

[29] A marketing survey of P.I. Research suggests that the aggregate price elasticity for smartphones is -1.74 in 2017.

all else equal, a one percent increase in share friend is equivalent to a price drop by 9 dollars (63.3 RMB).

Coefficients on key attributes are also intuitively signed and significant. All else equal, consumers on average favor products with larger screen, higher camera resolution, a faster CPU speed and a lighter weight. For example, I find that the willingness to pay for a one-mega pixel increase in camera resolution is about 45 dollars (325.3 RMB) for an average consumer. Similarly, an increase in the screen size by 0.1 inches is equivalent to a price decrease of 16.7 dollars (120.5 RMB), while an increase in the CPU speed by 0.1 GHz is equivalent to a price drop of 13.0 dollars (93.6 RMB). In the estimation, I include 7 brand dummies including Apple, Huawei, OPPO, Samsung, Xiaomi, Vivo and a group of all other brands. Apple posssess a larger brand value followed by Huawei, Vivo, Xiaomi and OPPO, while Samsung is relative less attractive.

Table 14 reports the predicted market share among compared to the actual market share. The upper panel shows the market shares for models by the release year, and the lower panel aggregates models by brand. The predicted shares mimic well the actual shares, suggesting a good fit of the model.

The model captures rich preference heterogeneity and delivers reasonable substitution patterns across products. Table 15 reports the median own and cross-price elasticities for top 10 popular products in Q4 2016. The median own-price elasticites ranges from -0.06 to -7.49, with a mean of -2.9. The table suggests reasonable substitution patterns. For example, a 1 percent increase in price for iPhone 6 leads to 0.23 percent increase in iPhone 5s and 0.14 percent increase in OPPO R9s Plus, which are considered as "high-end" products in the same category. In contrast, it leads to less increase in low-end products such as Vivo 37 and Redmi 3S. 1 percent increase in price of Xiaomi I 4 leads to a larger demand increase in similar products such as Vivo Y37 and OPPO R7, while smaller increase in iPhone 6s and OPPO R9s Plus.

Table 16 reports the demand semi-elasticities of social influence for the top five products in Q3 2017: OPPO R7, Huawei P8, Vivo V3 Max, iPhone 6s and Xiaomi MI 4. Element in row i column j shows the average percentage change in the market share of product j with a 10 percent increase in the share friend of product i. It suggests that all else equal, an 10 percent increase in the share friend leads to about 0.7-0.8 percent increase in its own demand, while it also leads to about 0.01-0.02 percent decrease in competitors' demand. This illustrates an important competition source due to social influence. Increasing one's own peer ownership not only enhances its own demand, but also intensifies competition and decreases rival's demand.

Table 17 reports the results from the regression of the (log of) estimated marginal costs on the smartphone characteristics. Many variables enter with significantly coefficients and with the anticipated sign. I find it costs more to build larger screen, better camera resolution, lighter weight, and higher CPU speed into a new smartphone. This finding is consistent with the industry teardown reports. IHS Teardown research and industry reports from other sources (Nellis, 2017; Segan, 2017; Su-Hyun, 2020) suggest that the bill of material break down for a typical smartphone suggest that display, body, camera and processor are the most expensive and account for more than half the cost of components.[30] The coefficients suggest that having one percent increase in the CPU clock speed, camera resolution and the screen size will increase marginal cost by 0.876 percent, 0.578 percent, and 5.013 percent. Reducing the weight by 1 percent will increase the marginal cost by 2.77 percent. These cost estimates are also close to studies in the smartphone industry. Wang (2018) finds that in 2014, one percent increase in CPU clock speed, camera resolution, and display size will increase marginal cost by 0.793 percent, 0.485 percent, and 0.503 percent respectively. My estimates of CPU speed and camera resolution are quite close to Wang (2018), except for a larger estimate for screen size. With recent development in technology, each inch of display embeds multiple sensors such as touching sensor and face recognition which are costly and consistent with the industry cost breakdown. Thus, the larger estimate of 5.013 captures the increasing costs per inch of the display.

The coefficient on the second-period dummy is -0.23, significant at 1 percent level. This captures the drop in marginal cost of an existing product due to the moving of technology frontier. Coefficients on brand dummies reflect the relative cost compared to the "Others" group. Apple has higher marginal cost than most of the brands, followed by Samsung and OPPO. Huawei and Vivo have marginal costs in the middle level, and Xiaomi has lower marginal costs.

6 Counterfactual Simulations

I conduct the counterfactual simulations to address the research questions of interest: How does social influence affect demand for quality? Is it the same for all products? What is the effect of social influence on firm pricing strategies? With demand and cost estimates, I simulate the demand and prices in the absence of social influence to shed light to these questions empirically.

[30] For example, according to the estimates of iPhone X, display takes 4.5 percent, camera takes about 9 percent, chipset and memory takes about 16 percentof the total cost. For Samsung Galaxy S20, a 6.87-inch AMOLED display even takes 75 dollars per unit, which is about 15 percent of the total costs.

6.1 Is social influence different for high-quality vs. low-quality products?

Theories suggest that if firms are asymmetric in terms of quality, in the presence of "social effect", markets tend to disproportional favor high or low quality products (Amaldoss and Jain, 2005b; Smallwood and Conlisk, 1979). As high-tech products like smartphones are a combination of several key features which essentially determine the product quality, it is important to empirically evaluate the impact of social influence for the demand of quality. To address this question, I conduct a counterfactual simulation on the demand side where I set the social influence to be zero, holding the pricing decisions constant. Specifically, consider the utility function in Equation 4, I set $\theta = 0$ and recalculate the individual choice probabilities and aggregate to market shares of distinct products, holding all other factors constant.

In terms of product quality, here I consider two measures: the mean utlity δ_j and the unobserved ξ_j. Mean utility is a linear combination of price, non-price attributes, and brand fixed effects, representing the overall attractiveness of a product to consumers. The second measure is the unobserved quality estimated as a residual from the second stage of the demand system. It captures the unobserved demand shifters such as brand image.

Figure 6 reports the percent change in demand by quality when social influence is present compared to the counterfactual case when social influence is absent. Figure 6a on the left-hand side plots the percent change in market shares against the mean utilities of each product. It suggests that social influence increase the market share of high-quality products by 5 to 26 percent, while reduces demand of low-quality products by 2 to 10 percent. Figure 6b on the right-hand side plots the average percent change in market shares against the unobserved quality ξ. Similarly, I find that social influence favors high-quality products and reduces demand for low-quality products.

Given that social influence is the same for all products in the model, the heterogeneous impact on products of different quality can be explained by the difference in consumer base due to different levels of product attractiveness. The differential attraction gets amplified through peers again. That is, popular high-quality products would engage in more customers to purchase through social influence than unpopular low-quality products. Table 18 reports the conditional market share among new buyers for products above median quality and below median quality. The market share for above-median products is 55.2 percent in the counterfactual case, while increases to 56.9 percent with social influence. In contrast, the market share for below-median products, social influence reduces their market shares from 44.8 percent to 43.1 percent. The gap in demand between the two groups of products enlarges from 10.3 percent to 13.8 percent with social influence. Figure 7 confirms that high quality products are slightly positively associated with bigger demand. Thus, the counterfactual simulation suggests that social influence magnifies the perceived quality difference and disproportionally favors high-quality products. Consumers benefit as social influence increases the average perceived quality level.

6.2 What is the impact of social influence on firm pricing?

To study the role of social influence on firm behavior, specifically, the dynamic pricing strategy, I set the social influence to be zero and re-optimize the equilibrium prices in the first and second periods by simulating both the demand and supply side. Intuitions from the demand and supply model suggest: On the one hand, the social multiplier effect generates more demand in the second period. Such an effect provides firms an incentive to use low release prices as a tool to invest in the consumer base in the first period. On the other hand, more friends using a particular product would create social differentiation effects, making consumers less sensitive to prices, thus providing firms incentive to increase the prices in the second period. These predictions can not be checked directly using data but can be tested through counterfactual simulations.

Here I focus on prices of 30 new products introduced from Q3 2016 to Q1 2017, holding the other products' prices as fixed and compare the release prices and second-period prices to the counterfactual optimal prices.[31] In the counterfactual simulation, $\theta = 0$ for all products and zero lagged share of friends for products in Period 1. I solve for the new equilibrium prices backward until reach the convergence of prices in Period 1 and Period 2. Starting with a guess of release prices p_1^0 and a guess of second-period prices p_2^0, in each iteration, I solve for the sales, static semi-elasticities in Period 1 and 2 and the inter-temporal demand semi-elasticities $\frac{dQ_2}{dp_1}$, then derive the equilibrium prices p_1^1 and p_2^1 according to Equation 15 and Equation 16. Next, update p_1^1 and p_2^1 as the starting prices and solve for new equilibrium prices. Repeat these two steps until the convergence is achieved between the starting prices and the solved prices. Detailed simulation procedures are described in Appendix C.2.

Table 19 first row shows the average release prices, second-period prices, total profits, and consumer surplus in the counterfactual scenario without social influence. The second row shows the counterparts when social influence is present. The third row

[31] There are 5 new products released in Period 2 and I assume firms take their entry as given.

shows the percent change using the counterfactual case as the baseline. Column 1 suggests firms' "investment" incentive. The average release price with the social influence is 266.56 dollars, 0.7 percent lower than the average counterfactual of 268.43 dollars, which is consistent with the theory prediction. Column 2 suggests the "harvest" incentive that the second-period average price with social influence is 250.54 dollars, 0.05 percent higher than the counterfactual average of 250.41 dollars. In addition, the counterfactual total profits are 127.801 million dollars, lower than the profits of 132.172 million dollars with social effects. The consumer surplus without social influence is 75.938 million dollars, about 1.7 percent lower than 77.25 million dollars with social influence. These findings suggest that social influence provides firms investment incentives to compete by reducing release prices at the beginning, which can then be leveraged to enact price increases in subsequent periods. Overall, social influence raises both consumer surplus and firm profits, thus enhances the social welfare.

Heterogeneous Effects Across Products To understand the incentive for firms to adopt the invest–harvest pricing strategy in the presence of social influence, I first examine the heterogeneity in price adjustments and profit gains. Table 20 reports the heterogeneous price adjustment by the unobserved product quality. I group the products into three groups of low, middle and high quality, using the 30th and 60th percentile of ξ distribution. The upper panel shows the adjustments of release prices. It suggests that when social influence is present, the average release price for low-quality products is reduced by 7.5 dollars (6 percent), changing from 132.5 dollars down to 125.0 dollars. The prices for middle-quality products are on average 3.7 dollars (1.78 percent) lower, from 211.56 dollars without social influence to 207.87 dollars with social influence. The last column suggests that the average release price for high-quality products is reduced by about 1.04 dollars (0.29 percent), from 367.56 to 366.51 dollars. Thus, social influence leads to a more massive price drop for low-quality products. The lower panel reports the magnitude of price adjustments when social influence is present, compared to the counterfactual case. Although the overall second-period price adjustment size is 10 times smaller than in Period 1, there is still variation by quality levels. High-quality products experience the largest increase in second-period price (0.193 dollars), followed by middle-quality products (0.105 dollars) and low-quality products (0.068 dollars).

The heterogeneous price adjustments by unobserved quality can be explained by their difference in price elasticity. Figure 9 reports that low-quality products in the data are associated with more elastic demand. To maintain competitiveness, low-quality products have the incentive to drop prices to a bigger magnitude in the first-period to engage consumers. This finding is consistent with the general theory prediction on penetration pricing that high price elasticity of demand in the short run is the desirable condition of an early low-price policy, i.e., a high degree of sales responsiveness to reductions in price (Dean, 1950).

Next I explore heterogeneous effects in profits among products to understand firm's incentive to alter the pricing strategy with social influence. Table 21 reports the average profit of products by different qualities. First of all, the average profit of a product increases by 3.42 percent, gaining about 0.125 million dollars in the city. Interestingly, an average product of high-quality and middle-quality benefits more than an average low-quality product with social influence. Due to the relatively less elastic demand, high-quality products have the incentive not to drop introductory prices by a large magnitude and slightly increase second-period prices due to the social differentiation effects. In this way, by adopting the penetration pricing strategy with social influence, high-quality products benefit the most. Moreover, an average product of all quality levels are benefiting with social influence, which provides all products the incentive to adopt the invest–harvest pricing strategies.

Decompose Consumer Surplus Lastly, I try to understand the change in consumer surplus due to changes in the pricing strategy. Since social influence enters the utility function additively, a positive influence parameter would mechanically increase consumer welfare. Therefore, I try to decompose the change in consumer surplus into two parts. One is the change due to the a nonzero influence parameter – "addition effect"; the other is the change due to price adjust, holding fixed the influence parameter –"price effect". Table 22 reports the decomposition of change in consumer surplus. In Period 1, the increase in consumer surplus is related to the nonzero social influence since new products have zero lagged friend share. In Period 2, turning on the social influence increases the consumer surplus by 0.29 percent, while the increased prices reduce consumer surplus by 0.01 percent. Overall, the benefit in consumer surplus in the presence of social influence remains positive. It is consistent with the finding that the increase in the second-period prices is 10 times lower than the price drop in the first period. Two possible reasons could limit the size of the price increase. First, the overall demand is relative elastic with the average price elasticity of -2.9 so that the benefits of expanding quantities dominate the benefits from increasing prices. Second, one caveat is that the distribution of the lagged share of friends is skewed distributed, with a large fraction of zeros. Such data feature could mitigate the social differentiation effects through peers in the second period.

7 Robustness Checks

I perform three robustness checks for the baseline estimate from the following perspectives. First, I check if the estimate of the social influence is robust to an alternative definition of friend. Second, given the skewed distribution of "Share Friend", I use an alternative dummy variable as the key regressor and see if the causal interpretation go through. Third, I provide other robustness checks including alternative time lags and heterogeneous effects by peers' monthly subscription fee.

7.1 Alternative Friend Definition: Reciprocal Contacts

As discussed in Section 2, I show the baseline result using an alternative definition of friends, reciprocal contacts. A reciprocal contact is call contact that both calls and being called by the individual. Thus, it captures a possibly stronger relationship than the one-way contact. Appendix table A10 shows the communication pattern for the two friend definitions among the selected contacts. Consistent with the communication literature (Onnela et al., 2007), call frequency and duration are right-skewed. Table A11 shows the network size under different social contact definition. Among reciprocal contacts, the same-carrier fraction of friends is on average 64 percent, which is higher than 44 percent if use the baseline one-way contact definition. This is consistent with findings in the telecom research that closer friends tend to use same carrier.

Table 23 reports the baseline result using reciprocal contacts. The OLS results are similar to estimates in Table 5. The 2SLS results are slightly smaller than the OLS counterparts. The causal impact from peers still go through and the estimate is about 0.08 to 0.10 percentage points.

7.2 Alternative Regressor: Friend Dummy

Given the skewed distribution of Share Friend variable, I also checked the alternative regressor – a dummy variable that takes value one if there's at least one friend using the alternative 3 months before the phone change, and zero otherwise. Table 24 reports the baseline result. After adding various controls, the main estimate become stable across different specifications. Having at least one friend using a given product would increase the average choice probability by 1 percentage point conditional on purchasing. Table 25 reports the result of using three strategies discussed in Section 3 addressing the correlated tastes concern. I find consistent evidence that after controlling for the common brand tastes, the 2SLS estimate is similar to the OLS counterpart.

7.3 Other Robustness

Alternative Time Lags Figure 10 reports the baseline coefficient in the preferred specification with alternative time lags before phone change. "-3" corresponds to $t-3$ in the main text. Each point reports the point estimate and the error bar shows the confidence interval from a separate regression. It suggests that the main coefficient remains stable since 2 months before phone change.

Alternative Income Proxy Table 26 shows the heterogeneous effects of peers using average monthly plan fee as income proxy. It shows similar result as in Table 8.

8 Concluding Remarks

This paper examines how social influence affects demand, market competition, and firm pricing strategies. To this end, I first show the existence of social influence in product choices in a large scale mobile communication network together with data from the Chinese smartphone market. I develop three strategies to address the correlated tastes, including comparing different friend groups, constructing controls for phone tastes, and using instrumental variables to partial out the correlated tastes. I find that conditional on purchasing, a 10 percent increase in a given alternative doubles the average purchasing probability, which is as sizable as 25% of the effect of a successful marketing campaign. I also find suggestive evidence that social influence is motivated by status-seeking incentives. Social influence works stronger through wealthier peers and products with visually distinct attributes (such as bigger screen size and more color options) than hidden functions (higher CPU speed and better screen resolution).

Next, going from individual spillover to aggregate effects, I develop and estimate a structural model for the demand and a two-period dynamic pricing model, incorporating the social influence. I conduct counterfactual simulations where I reduce the impact of social influence to be zero to study how it affects the demand for high-quality and low-quality products differently in the market and how it affects the pricing strategy. I find that an increase in one product's peer ownership would strengthen its own demand while reduces the rival's demand at the same time. Moreover, social influence increases the demand for high-quality products and reduces demand for low-quality products. These results suggest the pro-competition effect of social influence. On the supply side, counterfactual prices suggest that social influence reduces the introductory prices by 0.7 percent and increases the second-period price by 0.05 percent. Overall, it increases firm profits by 3.42 percent and increases consumer surplus by 1.7 percent. The price change is pronounced for products of elastic demand. In general, this finding suggests that with a higher degree of spillover among consumers, firms have a strong incentive to grab higher demand at the beginning and engage in fiercer price competition.

With the rapid growth of digitization and social media, new data sources are becoming available now. This paper showcases a future research direction of combining conventional market-level data with unconventional but new microdata such as social network data to study the competition and welfare consequence of the growing communication and influence from peers and opinion leaders. Other aspects utilizing social network information, such as social targeting, are also important topics for future research.

References

Amaldoss, W. and S. Jain (2005a). Pricing of conspicuous goods: A competitive analysis of social effects. *Journal of Marketing Research 42*(1), 30–42.

Amaldoss, W. and S. Jain (2005b). Pricing of conspicuous goods: A competitive analysis of social effects. *Journal of Marketing Research 42*(1), 30–42.

Aral, S., L. Muchnik, and A. Sundararajan (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences 106*(51), 21544–21549.

Arnold, A. (2017). *4 Ways Social Media Influences Millennials' Purchasing Decisions*. Retrieved April 28, 2020, from Forbes.com https://www.forbes.com/sites/andrewarnold/2017/12/22/4-ways-social-media-influences-millennials-purchasing-decisions/.

Athey, S., D. Blei, R. Donnelly, F. Ruiz, and T. Schmidt (2018, May). Estimating heterogeneous consumer preferences for restaurants and travel time using mobile location data. *AEA Papers and Proceedings 108*, 64–67.

Bailey, M., R. Cao, T. Kuchler, and J. Stroebel (2018a). The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy 126*(6), 2224–2276.

Bailey, M., R. Cao, T. Kuchler, and J. Stroebel (2018b). The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy 126*(6), 2224–2276.

Bailey, M., D. M. Johnston, T. Kuchler, J. Stroebel, and A. Wong (2019, May). Peer effects in product adoption. Working Paper 25843, National Bureau of Economic Research.

Bandiera, O. and I. Rasul (2006). Social networks and technology adoption in northern mozambique. *Economic Journal 116*(514), 869–902.

Banerji, A. and B. Dutta (2009). Local network externalities and market segmentation. *International Journal of Industrial Organization 27*(5), 605–614.

Barwick, P. J., Y. Liu, E. Patacchini, and Q. Wu (2019, May). Information, mobile communication, and referral effects. Working Paper 25873, National Bureau of Economic Research.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–890.

Berry, S., J. Levinsohn, and A. Pakes (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy 112*(1), 68–105.

Bjorkegren, D. (2018, 07). The Adoption of Network Goods: Evidence from the Spread of Mobile Phones in Rwanda. *The Review of Economic Studies 86*(3), 1033–1060.

Blumenstock, J. E. (2018). Estimating economic characteristics with phone data. *AEA Papers and Proceedings 108*, 72–76.

Blumenstock, J. E., G. Cadamuro, and R. On (2015). Predicting poverty and wealth from mobile phone metadata. *Science 350*, 1073–1076.

Buchel, K., M. V. Ehrlich, D. Puga, and E. Viladecans-Marsal (2019). Calling from the outside: The role of networks in residential mobility. Working Paper wp2019-1909, CEMFI.

Cabral, L. (2011, 01). Dynamic Price Competition with Network Effects. *The Review of Economic Studies 78*(1), 83–111.

Chen, Y.-J., Y. Zenou, and J. Zhou (2018). Competitive pricing strategies in social networks. *The RAND Journal of Economics 49*(3), 672–705.

Commission for Communications Regulation (2018, June). *Mobile Handset Performance (Voice)*. Retrieved from www.comreg.ie.

Conley, T. G. and C. R. Udry (2010, March). Learning about a new technology: Pineapple in ghana. *American Economic Review 100*(1), 35–69.

Dean, J. (1950). *Pricing Policies for New Products*. Harvard University Press.

Deloitte (2018). *Chinese Consumers At the Forefront of Digital Technologies: China Mobile Consumer Survey 2018*.

Dey, B. L., K. Sorour, and R. Filieri (2016). Icts in developing countries: Research, practices and policy implications.

Economides, N., M. Mitchell, and A. Skrzypacz (2004). Dynamic oligopoly with network effects.

Fainmesser, I. P. and A. Galeotti (2015, 07). Pricing Network Effects. *The Review of Economic Studies 83*(1), 165–198.

Farrell, J. and G. Saloner (1986). Installed base and compatibility: Innovation, product preannouncements, and predation. *American Economic Review 76*(5), 940–55.

Gandhi, A. and J.-F. Houde (2019, October). Measuring Substitution Patterns in Differentiated Products Industries. NBER Working Papers 26375, National Bureau of Economic Research, Inc.

Gentzkow, M. (2007, June). Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review 97*(3), 713–744.

Giorgi, G. D. (2018). Consumption Network Effects. 2018 Meeting Papers 692, Society for Economic Dynamics.

Goolsbee, A. and A. Petrin (2004). The consumer gains from direct broadcast satellites and the competition with cable tv. *Econometrica 72*(2), 351–381.

Heffetz, O. (2012). Who sees what? demographics and the visibility of consumer expenditures. *Journal of Economic Psychology 33*(4), 801 – 818.

Hu, M. M., S. Yang, and D. Y. Xu (2019). Understanding the social learning effect in contagious switching behavior. *Management Science 65*(10), 4771–4794.

Iyengar, R., C. Van den Bulte, and T. W. Valente (2011). Opinion leadership and social contagion in new product diffusion. *Marketing Science 30*(2), 195–212.

Jain, H. (2017). *Decoding purchase journey of Indian smartphone buyer*. Retrieved June 20, 2017, from https://www.linkedin.com/pulse.

Katz, J. E. and S. Sugiyama (2005). *Mobile Phones as Fashion Statements: The Co-creation of Mobile Communication's Public Meaning*, pp. 63–81. London: Springer London.

Katz, M. L. and C. Shapiro (1985). Network externalities, competition, and compatibility. *The American Economic Review 75*(3), 424–440.

Lancieri, F. and P. M. Sakowski (2020, Forthcoming). Competition in digital markets: A review of expert reports. *Stanford Journal of Law, Business and Finance*.

Leduc, M. V., M. O. Jackson, and R. Johari (2017). Pricing and referrals in diffusion on networks. *Games and Economic Behavior 104*(C), 568–594.

Lu, T. (2017). *Smartphone Users Replace Their Device Every Twenty-One Months*. Retrieved October 13, 2017, from Counterpoint Research https://www.counterpointresearch.com.

Ma, L., R. Krishnan, and A. L. Montgomery (2015). Latent homophily or social influence? an empirical analysis of purchase within a social network. *Management Science 61*(2), 454–473.

Marlow, C. (2009). *Maintained Relationships on Facebook*. Retrieved March 9, 2009, from https://overstated.net/2009/03/09/maintained-relationships-on-facebook/.

Morin, J. (2013). *How to Improve Cell Phone Reception in Your Home*. Retrieved April 4, 2013, from https://www.homes.com.

Nair, H. S., P. Manchanda, and T. Bhatia (2010). Asymmetric social interactions in physician prescription behavior: The role of opinion leaders. *Journal of Marketing Research 47*(5), 883–895.

Nellis, S. (2017). *Apple's iPhone X has higher margin than iPhone 8: analysis*. Retrieved November 06, 2017, from https://www.reuters.com.

Onnela, J.-P., J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences 104*(18), 7332–7336.

Ovide, S. (2020). *You Are Being Influenced*. Retrieved April 28, 2020, from NY-Times https://www.nytimes.com/2020/04/28/technology/digital-influencers-coronavirus.html.

Pedersen, G. (2016, 9). *Mobile Phone Antenna Performance 2016*. Denmark: Nordic Council of Ministers.

Segan, S. (2017). *Teardown Reveals iPhone X Parts Cost $370*. Retrieved November 8, 2017, from https://www.pcmag.com.

Smallwood, D. E. and J. Conlisk (1979). Product quality in markets where consumers are imperfectly informed. *The Quarterly Journal of Economics 93*(1), 1–23.

Su-Hyun, S. (2020). *What makes up smartphone prices?* Retrieved June 9, 2020, from https://www.thejakartapost.com.

Veblen, T. (1899). *The Theory of the Leisure Class*. McMaster University Archive for the History of Economic Thought.

Wang, P. (2018). Innovation Is the New Competition: Product Portfolio Choices with Product Life Cycles. Working papers.

Figures and Tables

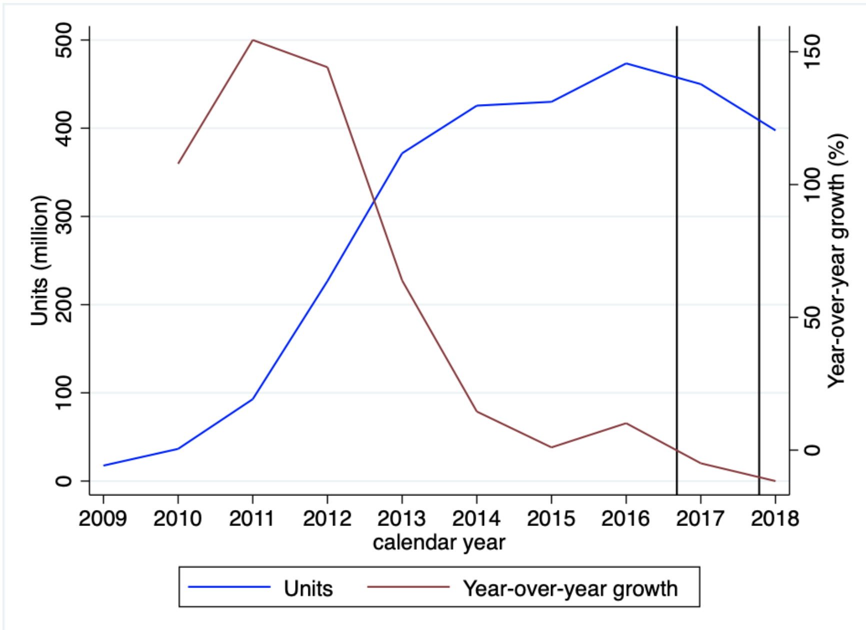


Figure 3: Sales in Chinese Smartphone Market
Notes: The figure plots the annual sales and year-over-year growth rate of smartphones in China. The blue line is the trend for sales; the red line is the growth trend in sales. Data Source: IDC Quarterly Mobile Phone Tracker.

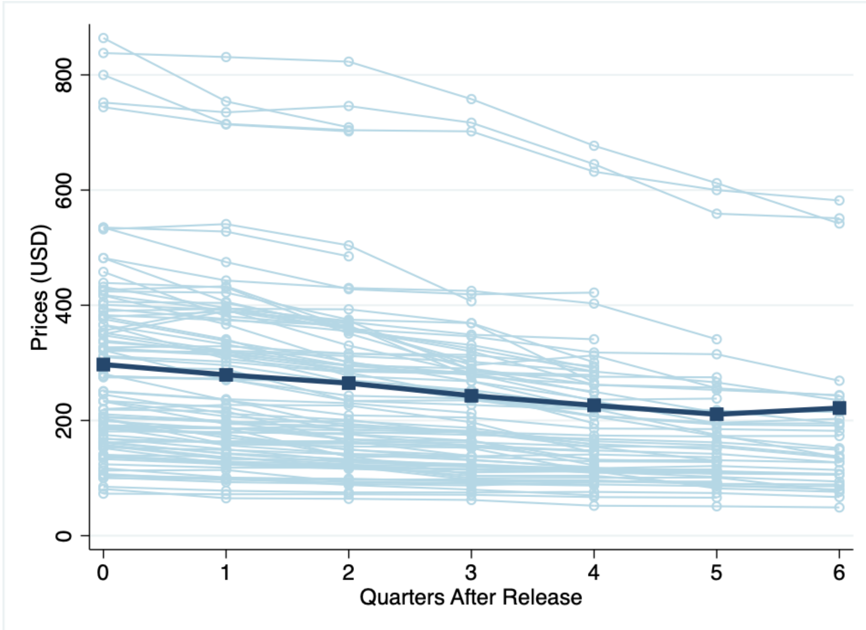


Figure 4: Median Prices Since Release

Notes: The figure plots the prices by quarters from release for models released after 2015. The dark blue line represents the median prices among all these models. Each light blue indicates a model. Data Source: IDC Quarterly Mobile Phone Tracker.

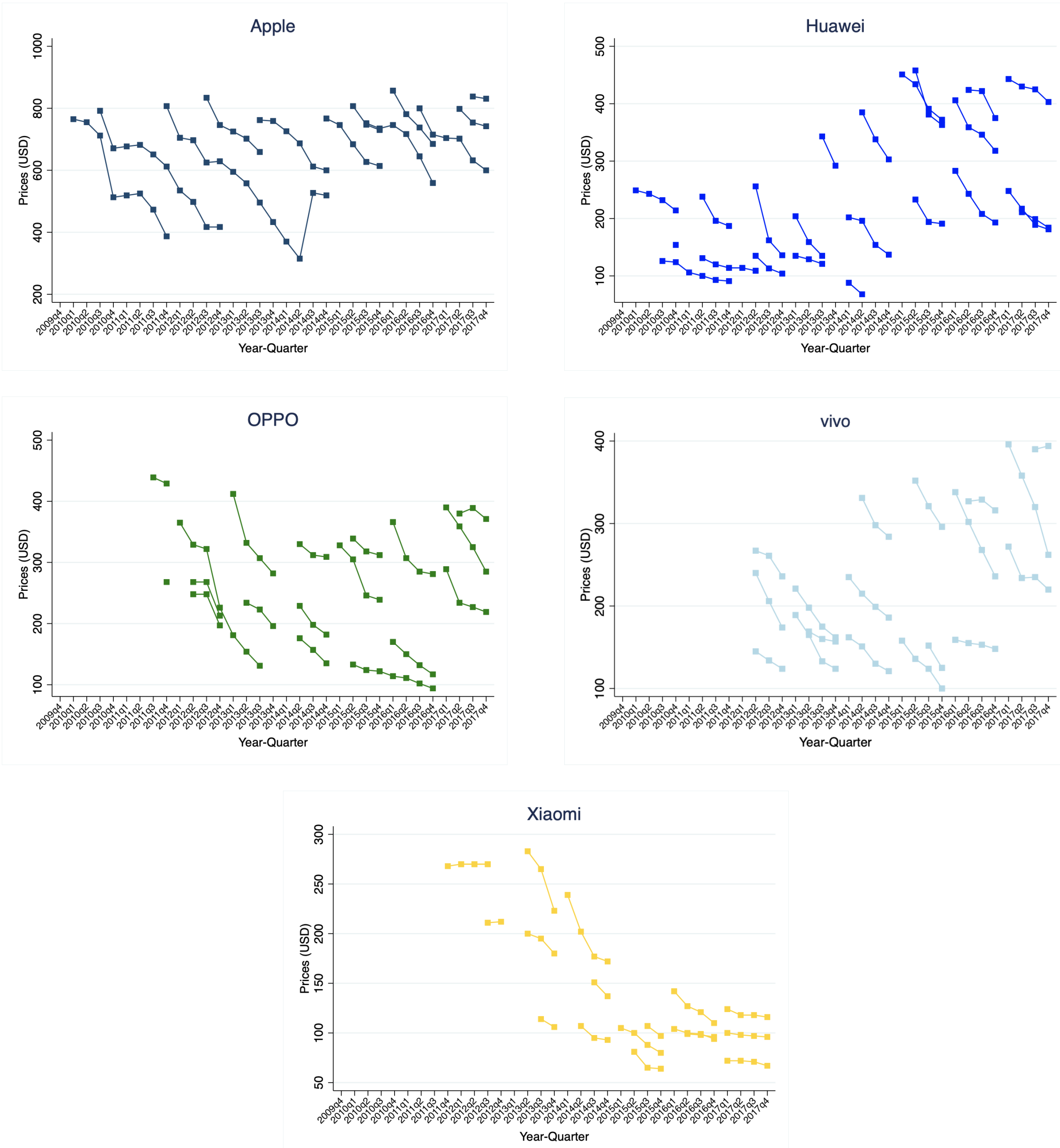


Figure 5: Price Trend By Brand.
Notes: The figure plots the prices of 3 most popular products in each year of Apple, Huawei, OPPO, vivo and Xiaomi. Data Source: IDC Quarterly Mobile Phone Tracker.

Table 1: Summary Statistics: Users

(a) Consumer Representativeness

	Users		National CFPS 2014	
	Mean	Std. Dev.	Mean	Std. Dev.
Demographics				
Female	0.35	0.48	0.46	0.50
Age (midpoint)	39.31	12.46	39.58	14.07
Age 25-34	0.29	0.45	0.23	0.42
Age 35-44	0.26	0.44	0.24	0.43
Age 45-59	0.26	0.44	0.27	0.45
Age above 60	0.08	0.28	0.09	0.29
Urban	0.61	0.49	0.64	0.48
Monthly Subscription Fee				
All range	67.79	64.67	61.39	62.13
Exceeds 30 RMB	75.65	64.93	72.84	62.71

(b) Non-Buyers vs. New Buyers

	Non-Buyers			New Buyers				
	Mean	SD	N	Mean	SD	N	Diff.	t-stat
Female	0.35	0.47	1,542,702	0.34	0.47	481,464	0.01^{***}	7.16
Age (midpoint)	38.25	13.22	1,542,787	39.32	12.59	481,623	-1.07^{***}	-49.51
Age 25-34	0.29	0.45	1,556,118	0.30	0.46	486,296	-0.00^{***}	-6.25
Age 35-44	0.23	0.42	1,556,118	0.24	0.43	486,296	-0.02^{***}	-22.77
Age 45-59	0.23	0.42	1,556,118	0.26	0.44	486,296	-0.03^{***}	-41.80
Age above 60	0.08	0.26	1,556,118	0.07	0.26	486,296	0.00^{***}	7.91
Urban	0.61	0.49	1,274,249	0.59	0.49	426,437	0.02^{***}	25.92
Avg. monthly fee	67.36	64.81	1,582,046	69.25	64.19	491,624	-1.89^{***}	-15.49
Frac. same-carrier contacts	0.43	0.49	1,656,518	0.44	0.50	497,607	-0.09^{***}	-31.72

Notes: The users restricts to individuals with a valid handset brand and model during the sample period. N. users = 2,380,331. 'Age' uses the midpoint of each age range. 'Urban' is a dummy for individuals who live in an urban area. The last two columns in panel (a) present the national average and standard deviation reported in 2014 CFPS among individuals with phone-related expenses that exceed 30 RMB per month, weighted by representative national weights. *** p<0.01, ** p<0.05, * p<0.1.

Table 2: Consumer Representativeness: Phone Ownership and Changes

	Sample	National
Market share of new sales		IDC 2017Q2
Huawei	21.73%	21.54%
OPPO	19.75%	18.42%
Vivo	17.89%	14.74%
Apple	10.98%	7.33%
Xiaomi	10.82%	13.03%
Samsung	4.71%	3.81%
Phone change rate		P.I. Research 2017
Android users	19%	16%
IOS users	21.26%	23.50%
Overall	20.3%	-

Notes: The table compares sample moments to moments in national sales data and a national marketing survey. The sample includes individuals with valid handset brand and model during the sample period. N. users = 2,380,331. "Phone change" is identified based on the criteria described in the text. The upper panel compares the market shares by brand among phone changers to the market share of new sales by brand in the IDC data in 2017Q2. The lower panel compares the phone change rate to a large marketing survey on smartphone usage and replacement behavior in China in 2017 conducted by Penguin Intelligence Research.

Table 3: Summary Statistics: Product Attributes

Variable	Mean	SD	Min	Max	SD/Mean
Price (USD)	250.89	154.097	67	708	0.602
Phone Age by Q3 2017(quarters)	5.17	2.96	0	16	0.573
Camera - main (mega pixel)	13.30	2.72	8	29	0.204
Display - Screen size (inch)	5.34	0.33	4	6.01	0.062
Display - Screen Resolution (total pixels)	1.79	0.43	0.41	2.33	0.242
Performance - CPU clock speed (GHz)	1.80	0.25	1.2	2.5	0.136
Design - Weight (g)	146.79	18.67	95.38	180	0.127
Battery capacity (Ah)	3.20	0.54	1.56	4.1	0.169
BioTouch: Fingerprint	0.69	0.32	0	1	0.459

Notes: The table reports phone attributes for models available for markets. N. products = 62 after grouping phone models based on the closeness of major characteristics as described in Appendix B.3. Composite model "other" in each market is also included.

Table 4: Change to Upgrade: Phone features Old vs New

	Old Phone		New Phone			
	Mean	SD	Mean	SD	Diff	t-stats
Network 4G	0.73	0.44	0.93	0.25	0.20	295.8
Camera - main (mega pixel)	10.8	3.95	12.96	3.8	2.35	342.66
Screen size (inch)	5	0.75	5.27	0.56	0.28	254.73
Screen resolution (total pixels)	1.37	0.83	1.65	0.77	0.3	219.93
CPU speed (GHz)	1.62	0.41	1.8	0.4	0.21	277.59
Weight (g)	149.78	23.99	156.53	20.25	6.86	163.47
Battery capacity (Ah)	2.61	0.77	3.02	0.72	0.42	312.54
Fingerprint	0.36	0.48	0.72	0.45	0.38	452.78

Notes: The table compares the features of the old handset and the new handset for all new buyers.

Table 5: Baseline Result: Social Influence in Product Choice

Dep. var. Prob i chooses phone j at time t	(1)	(2)	(3)	(4)	(5)	(6)
Share Friend	0.18***	0.12***	0.11***	0.11***	0.10***	0.10***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Share Future Friend				0.01**		0.005
				(0.004)		(0.003)
Share Same-old-brand					0.73***	0.73***
					(0.04)	(0.04)
Observations	4,218,170	4,218,170	4,218,170	4,218,170	4,218,170	4,218,170
R-squared	0.010	0.013	0.022	0.065	0.098	0.098
Resid. Neighbohood x brand FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes	Yes
Product x month	No	No	Yes	Yes	Yes	Yes

Notes: One unit of observation is an individual-model pair. "Share Friend" is the share of friends using phone j three months prior to time t. "Share Future Friend" is defined analogously, except using people who befriend individual i after time t. In other words, this is the fraction among the set of future friends who are using phone j at time $t-3$. "Share of Same-old-brand" is defined using non-friend new-phone buyers who shared the same phone brand as individual i's old phone model. This variable is the fraction of these users who use phone model j at time $t-3$. "Controls" include individual characteristics, the interaction of individual by phone attributes, and the average characteristics of peers as described in Section 3 Model 1. Product-by-month fixed effects are included in Columns 3-6. Column 6 is the preferred specification. Standard errors are in parentheses and clustered at the neighborhood by model level. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Falsification Test: Social Influence vs. Correlated Tastes

Dep. var. Prob i chooses phone j at time t	(1)	(2)	(3)	(4)
Share Friend	0.11***	0.11***	0.10***	0.10***
	(0.01)	(0.01)	(0.01)	(0.01)
Share Future Friend	0.01**	0.01**	0.005	0.008
	(0.004)	(0.006)	(0.003)	(0.007)
Share Same-old-brand			0.73***	0.73***
			(0.04)	(0.04)
Observations	4,218,170	2,082,518	4,218,170	2,082,518
R-squared	0.065	0.072	0.098	0.105
Resid. Neighbohood x brand FE	Yes	Yes	Yes	Yes
Friend control	Yes	Yes	Yes	Yes
Product x month FE	Yes	Yes	Yes	Yes
Middle Months	No	Yes	No	Yes

Note: One unit of observation is an individual-model pair. Columns 2 and 4 restrict to the subsample of individuals who change phones in the middle of the sample period (the fourth to the eighth month) to allow for enough observations on future friends . "Share Friend" is the share of friends using phone j three months prior to time t. "Share Future Friend" is defined analogously, except using people who befriend individual i after time t. In other words, this is the fraction among the set of future friends who are using phone j at time $t-3$. "Share of Same-old-brand" is defined using non-friend new-phone buyers who shared the same phone brand as individual i's old phone model. This variable is the fraction of these users who use phone model j at time $t-3$. "Controls" include individual characteristics, the interaction of individual by phone attributes, and the average characteristics of peers as described in Section 3 Model 1. Product-by-month fixed effects are included in all columns. Standard errors are in parentheses and clustered at the neighborhood by model level. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: IV results: Social Influence in Product Choice

Dep. var. Prob i chooses phone j at time t	(1) OLS	(2) IV	(3) OLS	(4) First stage	(5) IV
Share Friend	0.18***	0.13***	0.11***		0.11***
	(0.01)	(0.02)	(0.01)		(0.01)
Share Friends' Neighbors				0.11***	
				(0.01)	
Friends' neighbors avg. CPU speed				-0.02**	
				(0.01)	
Friends' neighbors avg. 4G				-0.04***	
				(0.01)	
Observations	4,218,170	4,218,170	4,218,170	4,218,170	4,218,170
R-squared	0.010	–	0.022	–	–
Resid. Neighbohood x brand FE	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes
Product x month FE	No	No	Yes	Yes	Yes
Weak IV test (F-stat)	–	580.5	–	–	639.83

Notes: One unit of observation is an individual-model pair. Columns 1 and 3 report the OLS estimates specified as in Table 5 columns 1 and 3. Columns 2 and 5 report the 2SLS counterparts using the choices and average phone attributes of the residential neighbors of friends as IV for 'Share Friend'. Column 4 reports the first-stage for column 5. "Share Friend" is the share of friends using phone j three months prior to time t. "Controls" include individual characteristics, the interaction of individual by phone attributes, and the average characteristics of peers as described in Section 3 Model 1. Standard errors are in parentheses and clustered at the neighborhood by model level. *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Mechanism: Status-Seeking by Peers' Income Levels

Dep. var. Prob i chooses phone j at time t	(1)	(2)
Share Friend of		
HP $>$ 19000	0.051***	
	(0.005)	
HP 5300-19000	0.045***	
	(0.006)	
HP $\leq$ 5300	0.033***	
	(0.010)	
Higher House Price		0.058***
		(0.010)
Similar or Lower House Price		0.023**
		(0.010)
Observations	4,002,782	4,002,782
R-squared	0.096	0.098
Residence Neighborhood FE	Yes	Yes
Controls	Yes	Yes
Product by month	Yes	Yes

Notes: The table compares the social influence by friends of different income levels. One unit of observation is an individual-model pair. Key independent variable "Share Friend" is re-constructed from friends in different reference group by house price per square meter. 5300 and 19000 RMB per square meter are the 25th and 75th percentile of the distribution. "Higher" refers to friends whose house prices per square meters are at least one standard deviation (2000 RMB) higher than the new buyer's house price, otherwise belongs to "Similar or Lower". Own house price are included in column 2. Standard errors are in parentheses and clustered at the neighborhood by model level. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Mechanism: Status Seeking by Visual Attributes

Dep. var.	Visual Attributes			Hidden Attributes	
Prob i chooses phone j at time t	(1)	(2)	(3)	(4)	(5)
Share Friend	0.056***	0.094***	0.066***	0.094***	0.120***
	(0.008)	(0.013)	(0.008)	(0.012)	(0.012)
Bigger Screen	-0.003				
	(0.002)				
Share Friend x Bigger Screen	0.056***				
	(0.012)				
More color option		-0.001			
		(0.002)			
Share Friend x More color option		0.037*			
		(0.019)			
Three cameras			-0.001		
			(0.002)		
Share Friend x Three cameras			0.052***		
			(0.015)		
High CPU Speed				0.004	
				(0.003)	
Share Friend x High CPU Speed				0.021	
				(0.025)	
Better Screen Resolution					-0.030***
					(0.003)
Share Friend x Better Screen Resolution					-0.016**
					(0.007)
Constant	0.003	-0.009	-0.009	-0.011	0.012
	(0.012)	(0.013)	(0.017)	(0.016)	(0.011)
Observations	4,218,170	4,082,100	4,218,170	4,218,170	4,218,170
R-squared	0.096	0.096	0.097	0.096	0.096
Controls	Yes	Yes	Yes	Yes	Yes
Residence neighborhood FE	Yes	Yes	Yes	Yes	Yes
Brand x month	Yes	Yes	Yes	Yes	Yes

Notes: The table reports the heterogeneous effects of social influence by product attributes. One unit of observation is an individual-model pair. "Share Friend" is the share of friends using phone j three months prior to time t. In all columns, I control for key product attributes (screensize, camera resolution, CPU speed, weight and price), interactions of individual-product characteristics, friend demographic shares, and share of same-old-brand non-contacts. Brand-by-month fixed effects are controlled. Standard errors are in parentheses and clustered at neighborhood by brand level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 10: Mechanism: Same Operating System

Dep. var. Prob i chooses phone j at time t	(1)	(2)	(3)
Share Friend	0.101***	0.085***	0.085***
	(0.009)	(0.007)	(0.007)
Share Same OS as j	0.004**		0.001
	(0.002)		(0.001)
Share Same Brand as j		0.020***	0.019***
		(0.005)	(0.004)
Share Same-old-brand	0.729***	0.725***	0.725***
	(0.039)	(0.039)	(0.039)
Observations	4,218,170	4,218,170	4,218,170
R-squared	0.098	0.098	0.098
Resid. Neighbohood x brand FE	Yes	Yes	Yes
Friend control	Yes	Yes	Yes
Product x month FE	Yes	Yes	Yes

Notes: The table reports the additional effects of social influence from same operating system and same brand. One unit of observation is an individual-model pair. "Share Same OS(Brand) as j" is the share of friends use or change to the same operating system (brand) as the given product three months prior to the phone change. Standard errors are in parentheses and clustered at the neighborhood by model level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 11: Mechanism: Inconsistent with Information Sharing

Dep. var. Prob i chooses phone j at time t	(1)	(2)
Share Friend		
and Pre-exist coworker	0.039***	
	(0.010)	
and Newly joined coworker	0.009	
	(0.027)	
Longer Relationship		0.084***
		(0.03)
Shorter Relationship		0.050
		(0.03)
Observations	273,358	4,218,170
R-squared	0.099	0.096
Resid. Neighbohood x brand FE	Yes	Yes
Controls	Yes	Yes
Product by month	Yes	Yes

Notes: The table reports the heterogeneous effects of social influence by information sources. One unit of observation is an individual-model pair. "Share Friend" is the share of friends using phone j three months prior to time t. Longer friendship considers friends who start the first call in week 30 or earlier, otherwise shorter. Standard errors are in parentheses and clustered at the neighborhood by model level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 12: Demand Estimates

	Est.	S.E.
First Stage Parameters		
Share Friend	2.815	0.153
Interactions		
Share Friend x (Income >75th percentile)	-0.247	0.111
Share Friend x Call minutes (per thsd)	0.302	0.051
Price x (Income >75th percentile)	0.204	0.043
Screen size x Age	0.029	0.001
Camera x Age	-0.003	0.000
CPU Speed x Age	0.029	0.002
Deviations		
σ_p Price	0.600	n.a.
σ_2 Screen size	0.822	0.055
σ_3 Camera	0.000	0.009
σ_4 CPU Speed	0.002	0.062
σ_5 Weight	0.000	0.005
Log likelihood	-9954.1122	
Observations	187,316	
Second Stage Linear Parameters		
Price	-1.032	0.110
Screen size	0.693	0.164
Camera resolution	0.187	0.018
Weight	-0.011	0.003
CPU Speed	0.538	0.164
Apple	(omitted)	(omitted)
Huawei	-2.209	0.311
OPPO	-2.509	0.237
Samsung	-2.723	0.265
Xiaomi	-2.458	0.311
Vivo	-2.433	0.259
Observations	1,142	

Notes: First stage parameters are obtained using 187,316 individual-model observations from a 1% random sample of 5,000 new buyers. σ_p is calibrated to be 0.60 so that the aggregate price elasticity equals to the industry estimate of -1.74. 1,142 product-market fixed effects are estimated out from the first stage constrained simulated likelihood maximization. The second stage is estimated including 7 brand fixed effects (Apple, Huawei, Xiaomi, OPPO, vivo, Samsung and others), 30 market fixed effects and phone ages on the estimated product-market fixed effects obtained in the first stage. Linear parameters are obtained through 2SLS IV regression. Cragg-Donald Wald F statistics is 46.64.

Table 13: Robustness: Other Demand Specifications

	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.
First Stage Parameters										
Share Friend	3.409	0.134	3.502	0.156	2.821	0.176	2.822	0.176	2.779	0.180
Interactions										
Price x (Income >75th percentile)	0.131	0.040	0.131	0.040	0.142	0.034	0.240	0.026	0.148	0.030
Share Friend x (Income >75th percentile)			-0.317	0.284						
Share Friend x Call minutes (per thsd)					0.327	0.057	0.323	0.057	0.277	0.058
Screensize x Age									0.029	0.002
Camera x Age									-0.004	0.000
CPU speed x Age									0.033	0.002
Deviations										
σ_p Price	0.0002	0.0301	0.0002	0.0301	0.0002	0.0293	0.0002	0.027	0.000	0.026
σ_2 Screensize									0.866	0.063
σ_3 Camera resolution									0.000	0.011
σ_4 CPU speed									0.005	0.078
σ_5 Weight									0.001	0.006
Log likelihood	-10610.279		-10609.643		-10591.232		-10562.735		-10491.5947	
Second Stage Linear Parameters										
Price	-1.174	0.126	-1.175	0.126	-1.206	0.132	-1.185	0.128	-1.108	0.128
Screen size	0.936	0.189	0.935	0.189	0.948	0.197	0.913	0.192	0.946	0.191
Camera resolution	0.187	0.020	0.187	0.020	0.184	0.021	0.171	0.021	0.190	0.021
Weight	-0.016	0.003	-0.016	0.003	-0.017	0.003	-0.016	0.003	-0.015	0.003
CPU speed	0.666	0.189	0.668	0.189	0.736	0.197	0.751	0.192	0.589	0.191

Notes: This table reports the result using other demand specifications. First stage parameters are obtained using 187,316 individual-model observations from a 1% random sample of 5,000 new buyers. 1,142 product-market fixed effects are estimated out from the first stage constrained simulated likelihood maximization. The second stage is estimated including 7 brand fixed effects (Apple, Huawei, Xiaomi, OPPO, vivo, others and fringe), 30 market fixed effects and phone ages on the estimated product-market fixed effects obtained in the first stage. Linear parameters are obtained through 2SLS IV regression.

Table 14: Model Fit: Share Among New Buyers

	N. models	Data all	Predicted
By Vintage			
Models 2017	11	9.65%	10.26%
Models 2016	34	50.41%	49.95%
Models 2015	10	13.19%	12.96%
Models before 2015	6	11.94%	12.34%
Fringe	1	6.50%	5.61%
By Brand			
Top-Five brands	36	71.04%	71.34%
Apple	6	6.74%	6.22%
Huawei	11	18.01%	18.64%
Xiaomi	6	9.41%	9.23%
OPPO	5	18.90%	18.93%
vivo	8	17.98%	18.33%
Other	25	12.90%	12.35%
Samsung	2	3.11%	3.00%
Lenovo	4	0.47%	0.47%
CoolPad	4	0.65%	0.62%
Meizu	3	3.70%	3.49%
LeTV	2	1.42%	1.36%
Nokia	5	0.41%	0.39%
ZTE	1	0.09%	0.09%
Nubia	1	0.12%	0.11%
Gionee	1	2.72%	2.63%
360	1	0.22%	0.21%
Fringe	1	6.50%	5.61%

Notes: This table reports the actual and predicted share for models of different release years and by brand. The actual share is the share among all new buyers. The predicted share is obtained using a 1% random sample of 5,000 new buyers.

Table 15: Median Own and Cross-Price Elasticities

Model	Apple iPhone 6s	OPPO R7	Apple iPhone 5s	Vivo V3 Max	Huawei P8	OPPO R9s Plus	Huawei Mate 8	Vivo X6s Plus	Xiaomi MI 4	Vivo Y37	Xiaomi Redmi 3S
Apple iPhone 6s	-4.168	0.094	0.228	0.060	0.107	0.137	0.017	0.015	0.025	0.010	0.000
OPPO R7	0.074	-2.216	0.115	0.059	0.024	0.032	0.015	0.015	0.007	0.008	0.000
Apple iPhone 5s	0.098	0.065	-2.373	0.048	0.032	0.035	0.015	0.016	0.008	0.007	0.000
Vivo V3 Max	0.071	0.082	0.116	-2.020	0.022	0.031	0.016	0.014	0.007	0.008	0.000
Huawei P8	0.298	0.116	0.277	0.078	-6.287	0.105	0.031	0.017	0.027	0.010	0.000
OPPO R9s Plus	0.782	0.168	0.323	0.118	0.113	-7.492	0.030	0.030	0.022	0.013	0.000
Huawei Mate 8	0.060	0.094	0.173	0.076	0.041	0.037	-3.281	0.059	0.015	0.010	0.000
Vivo X6s Plus	0.073	0.127	0.236	0.088	0.030	0.049	0.080	-4.342	0.015	0.012	0.001
Xiaomi MI 4	0.003	0.026	0.049	0.019	0.020	0.015	0.008	0.006	-1.197	0.050	0.000
Vivo Y37	0.009	0.081	0.116	0.056	0.020	0.025	0.016	0.015	0.278	-2.208	0.000
Xiaomi Redmi 3S	0.009	0.034	0.025	0.028	0.000	0.004	0.007	0.009	0.002	0.003	-1.645

Notes: The table reports the median own and cross-price elasticities across markets for top 10 popular products in Q4 2016. The rows and columns are ranked by the descending order of the market shares. Cell entries i, j where i indexes row and j column, gives the percent change in market share of model i with one percent change in price of j. Each entry represents the median of the elasticities from the 30 markets (urban/subruban/rural by 10 months).

Table 16: Marginal Effects of Lagged Friend Share on Purchase Probabilities (Estimated Percentage Changes)

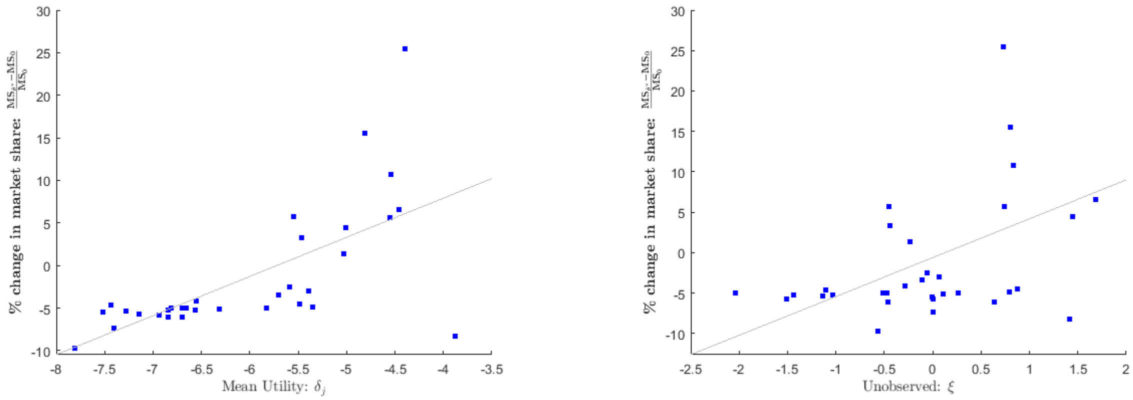
	OPPO R7	Huawei P8	Vivo V3 Max	iPhone 6s	Xiaomi MI4
OPPO R7	0.670	-0.016	-0.018	-0.013	-0.013
Huawei P8	-0.016	0.764	-0.015	-0.014	-0.015
Vivo V3 Max	-0.018	-0.015	0.630	-0.012	-0.013
iPhone 6s	-0.013	-0.014	-0.012	0.641	-0.013
Xiaomi MI4	-0.013	-0.015	-0.013	-0.013	0.681

Notes: The table reports the average percentage change in the purchase probabilities arising from increasing the lagged share of friends by 10 percent for the top products in brand Apple, Huawei, OPPO, vivo and Xiaomi in Q4 2016. Because they are percentage changes, they do not sum up to one. Cell entries i, j where i indexes row and j column, gives the percentage change in market share of product j with a 10 percent increase in share of friends using product i.

Table 17: Marginal Costs

Y = Ln(mc)	Full Dynamics (T =2)	
Ln(X)	Est.	S.E.
Screen size	5.013	1.704
CPU Speed	0.876	0.243
Battery capacity	0.327	0.359
Camera Resolution	0.578	0.134
Weight	-2.772	0.952
T=2	-0.230	0.073
Baseline = Others		
Apple	1.752	0.174
Huawei	0.0810	0.139
OPPO	0.465	0.148
Samsung	0.525	0.210
Xiaomi	-0.656	0.231
Vivo	0.244	0.147
Observations	115	

Notes: The table reports the cost coefficients from a log-log specification. "T=2" is a dummy for the second period. The number of observation is 115, including 57 models available in Period 1 (Q4 2016 and Q1 2017) and 58 new models available in Period 2 (Q2 2017 and Q3 2017).



(a) **(b)**

Figure 6: Social Influence on Demand By Quality

Notes: The figure plots percent change in market share by unobserved quality with social influence compared to the case without social influence. X-axis in (a) is the mean utility from demand; X-axis in (b) is unobserved quality estimated from the demand system. Y-axis is the average percent change in market share when social influence is present compared to when social influence is absent for each product across all markets.

Table 18: Social Influence Enlarges Demand Gap between High vs. Low-Quality Products

	ξ		
Market share	Below Median	Above Median	Gap
$\theta = 0$	0.448	0.552	0.103
$\theta = \theta^*$	0.431	0.569	0.138

Notes: The table reports the market shares for products below and above the median quality (ξ). "Gap" is the difference between market share of above-median products and below-median products. $\theta = 0$ represents the counterfactual scenario without social influence. $\theta > 0$ represents the case with social influence.

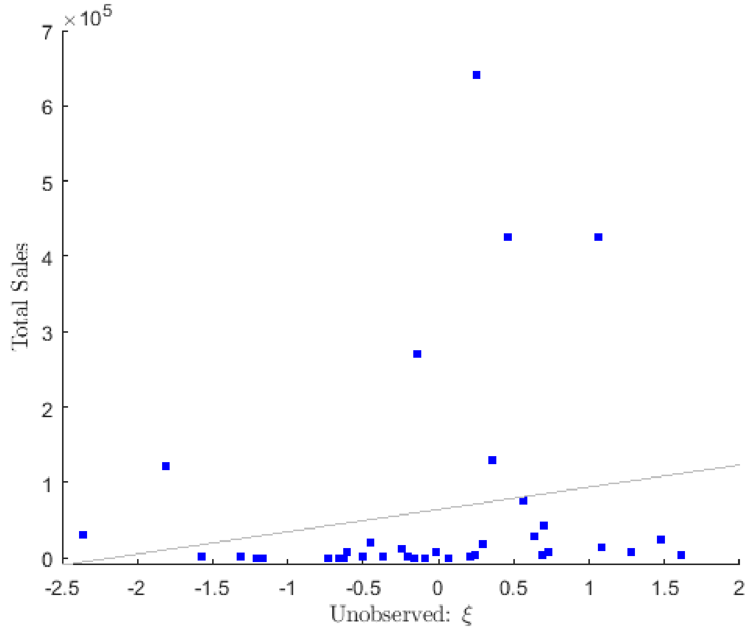


Figure 7: Total Sales By Quality

Notes: X-axis is ξ, the unobserved quality estimated from the demand system. Y-axis is the total sales for each product during the sample period.

Table 19: Counterfactual Prices, Profits and CS Without Social Influence

	Average P1 (USD)	Average P2 (USD)	Total Profits (Million)	CS (Million)
$\theta = 0$	268.432	250.407	127.801	75.938
$\theta = \theta^*$	266.555	250.544	132.172	77.250
$Y_{\theta^*} - Y_0$	-1.876	0.137	4.371	1.312
$(Y_{\theta^*} - Y_0)/Y_0$	-0.70%	0.05%	3.42%	1.70%

Notes: The table reports the counterfactual prices, profits and consumer surplus when social influence is set to zero. The last row reports the percent change of each variable taking the counterfactual scenario as the baseline.

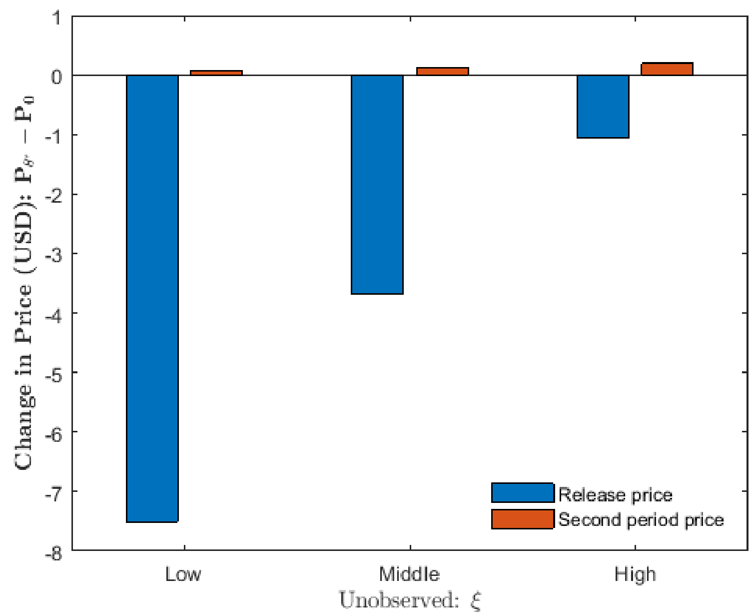


Figure 8: Heterogeneous Price Changes Due to Social Influence

Notes: The figure reports the average price changes by product quality due to social influence,taking the counterfactual scenario as the baseline. ξ is the unobserved quality estimated from the demand system. Three quality levels are grouped based on the 30th and 60th percentile of ξ distribution. The blue bar on the left-hand side for each quality level is the change in release prices; the orange bar on the right-hand for each quality level is the change in second-period price.

Table 20: Heterogeneous Price Changes Due to Social Influence

	Average	ξ Low	ξ Middle	ξ High
Panel A: Release Price (USD)				
$\theta = 0$	268.432	132.505	211.562	367.555
$\theta = \theta^*$	266.555	125.000	207.866	366.507
$p_{\theta^*} - p_0$	-1.876	-7.505	-3.696	-1.048
$(p_{\theta^*} - p_0)/p_0$	-0.70%	-6.00%	-1.78%	-0.29%
Panel B: Second-period Price (USD)				
$\theta = 0$	250.407	122.098	211.721	340.353
$\theta = \theta^*$	250.544	122.167	211.826	340.546
$p_{\theta^*} - p_0$	0.137	0.068	0.105	0.193
$(p_{\theta^*} - p_0)/p_0$	0.05%	0.06%	0.05%	0.06%

Notes: The table reports the average prices with and without social influence by product quality. ξ is the unobserved quality estimated from the demand system. Three quality levels are grouped based on the 30th and 60th percentile of ξ distribution. Panel A reports the release prices, i.e., P1. Panel B reports the second-period prices, i.e., P2. The last row in each panel reports the percent change of each variable, taking the counterfactual scenario as the baseline.

Table 21: Heterogeneous Average Profit Changes Due to Social Influence

Average Profits (Million)	ALL	ξ Low	ξ Middle	ξ High
$\theta = 0$	3.776	0.806	4.940	5.297
$\theta = \theta^*$	3.651	0.926	4.715	5.034
$\pi_{\theta^*} - \pi_0$	0.125	-0.120	0.225	0.263
$\pi_{\theta^*} - \pi_0/\pi_0$	3.42%	-1.40%	4.77%	5.22%

Notes: The table reports the average profit with and without social influence by product quality. ξ is the unobserved quality estimated from the demand system. Three quality levels are grouped based on the 30th and 60th percentile of ξ distribution. The last row reports the percent change of each variable, taking the counterfactual scenario as the baseline.

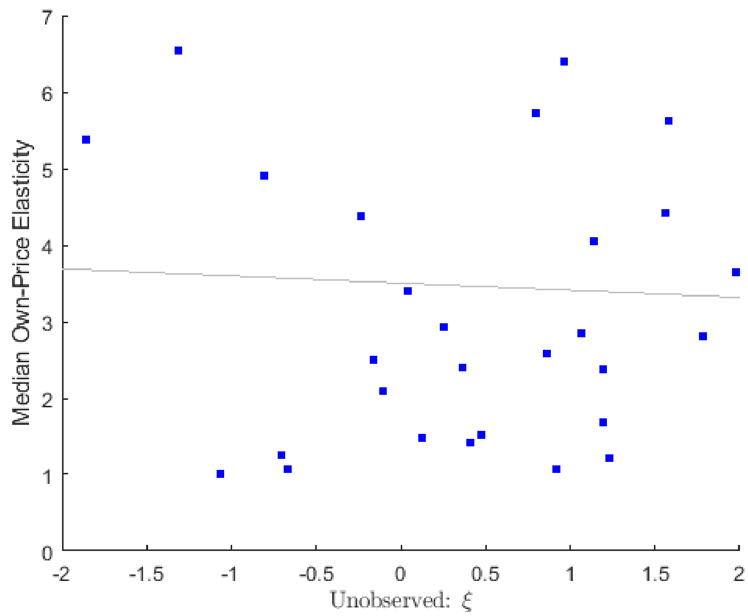


Figure 9: Own-Price Elasticity and Unobserved quality ξ

Notes: The x-axis is ξ, the unobserved quality estimated from the demand system. Y-axis is the median own-price elasticity for each product across markets calculated using the demand estimates.

Table 22: Decompose ΔCS Due to social influence

	CS Period1 (Million)	CS Period2 (Million)
Addition Effect		
$\theta = 0, P = P(0)$	39.484	36.454
$\theta = \theta^*, P = P(0)$	39.484	36.561
ΔCS_θ	0	0.29%
Price Effect		
$\theta = \theta*, P = P(\theta^*)$	40.691	36.559
ΔCS_P	2.97%	-0.01%
$\Delta CS_\theta + \Delta CS_P$	2.97%	0.28%

Notes: The table decomposes the change in consumer surplus due to social influence. The first panel shows the addition effect due to the inclusion of a positive term of share of friends in the utility specification. The second panel shows the change in consumer surplus due to adjustment in pricing strategies, holding the social influence constant.

Table 23: Baseline Robustness: Reciprocal Contacts

Dep. var. Prob i chooses phone j at time t	(1) OLS	(2) OLS	(3) IV
Share Friend	0.10***	0.09***	0.08***
	(0.01)	(0.01)	(0.01)
Share Future Friend		0.003	
		(0.003)	
Share Same-old-brand	0.73***	0.73***	0.74***
	(0.04)	(0.04)	(0.04)
Observations	4,218,976	4,218,976	4,171,236
R-squared	0.096	0.096	–
Resid. Neighbohood x brand FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Product x month FE	Yes	Yes	Yes
J test (J-stat)	–	–	466.6
Weak IV test (F-stat)	–	–	1380

Notes: The table reports the robustness check using reciprocal contacts as the friend definition. One unit of observation is an individual-model pair. Regressors are defined in the same way as in Model 1 using reciprocal contact definition. Columns 1 and 2 report the OLS estimates specified as Table 5 columns 5 and 6. Column 3 reports the 2SLS counterparts using the choices and average phone attributes of the residential neighbors of reciprocal friends as IV for 'Share Friend'. Standard errors are in parentheses and clustered at the neighborhood by model level. *** p<0.01, ** p<0.05, * p<0.1.

Table 24: Baseline Robustness: Alternative Regressor Friend Dummy

Dep. var. Prob i chooses phone j at time t	(1)	(2)	(3)	(4)	(5)	(6)
Friend	0.03***	0.01***	0.01***	0.01***	0.01***	0.01***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Future Friend				0.001***		-0.000
				(0.000)		(0.000)
Share Same Brand					0.73***	0.73***
					(0.01)	(0.01)
Observations	4,218,976	4,218,976	4,218,976	4,218,976	4,218,976	4,218,976
R-squared	0.009	0.015	0.062	0.062	0.095	0.095
Residential Cell FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes	Yes
Product x month FE	No	No	Yes	Yes	Yes	Yes

Notes: The table reports the robustness check using a dummy variable as the key regressor. One unit of observation is an individual-model pair. 'Friend' is a dummy takes value one if there is at least a friend in the peer group that uses or changes to j three months prior to time t, zero otherwise. 'Future Friend' takes value one if there is a friend known after the phone purchase that uses or changes to j three months prior to time t, zero otherwise. "Share of Same-old-brand" is defined using non-friend new-phone buyers who shared the same phone brand as individual i's old phone model. This variable is the fraction of these users who use phone model j at time $t-3$. "Controls" include individual characteristics, the interaction of individual by phone attributes, and the average characteristics of peers as described in Section 3 Model 1. Standard errors are in parentheses and clustered at the neighborhood by model level. *** p<0.01, ** p<0.05, * p<0.1.

Table 25: Addressing Correlated Tastes: Alternative Regressor Friend Dummy

Dep. var. Prob i chooses phone j at time t	(1) OLS	(2) OLS	(3) IV
Friend	0.01***	0.01***	0.01***
	(0.001)	(0.001)	(0.002)
Future Friend	0.001**	-0.000	
	(0.000)	(0.000)	
Share Same-old-brand		0.73***	0.73***
		(0.01)	(0.04)
Observations	4,218,976	4,218,976	4,218,976
R-squared	0.062	0.095	–
Resid. Neighbohood x brand FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Product x month FE	Yes	Yes	Yes
Weak IV test (F-stat)	–	–	532.1

Notes: One unit of observation is an individual-model pair. Variables are the same as in Table 24. Columns 1 and 2 report the OLS estimates specified as columns 5 and 6 Table 24. Column 3 reports the 2SLS counterpart using the choices and average phone attributes of the residential neighbors of friends as IV for "Friend". Standard errors are in parentheses and clustered at the neighborhood by model level. *** p<0.01, ** p<0.05, * p<0.1.

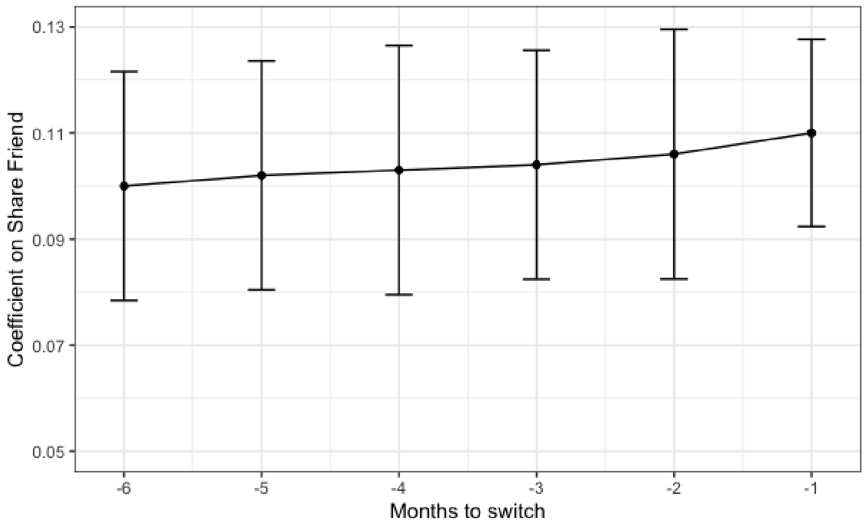


Figure 10: Baseline Robustness: Alternative time lags

Notes: The figure plots the coefficient on "Share Friend" using alternative time lags $t-6$, $t-5$, $t-4$, $t-2$ and $t-1$. That is, the share of friends using phone j 6/5/4/2/1 months prior to time t. Each point is the point estimate and the error bar represents the confidence interval in a separate regression using the new regressor in the preferred specification as column 6 in Table 5.

Table 26: Robustness: Status-seeking

Dep. var. Prob i chooses phone j at time t	(1)	(2)
Share Friend of		
Fee >136	0.074***	
	(0.012)	
Fee 18-136	0.061***	
	(0.011)	
Fee $\leq$ 18	0.051***	
	(0.008)	
Higher Fee		0.083***
		(0.010)
Similar or Lower Fee		0.050***
		(0.010)
Observations	4,128,580	4,128,580
R-squared	0.096	0.098
Residence Neighborhood FE	Yes	Yes
Controls	Yes	Yes
Product by month	Yes	Yes

Notes: The table compares the social influence by friends of different income proxied by monthly fee. One unit of observation is an individual-model pair. Key independent variable "Share Friend" is reconstructed from friends in different reference group by monthly plan fee. 18 and 136 RMB are the 25th and 75th percentile of the distribution. "Higher" refers to friends whose monthly plan fees are at least one standard deviation (40 RMB/mon) higher than the phone buyer's fee, otherwise belongs to "Similar or Lower". Own monthly plan fee is included in column 2. Standard errors are in parentheses and clustered at the neighborhood by model level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix

A Summary Statistics

Table A1: Data Structure Example

pid	month	product	choice	ShareFriend	Female	ShareFemale
103001	4	1	0	0.13	1	0.25
103001	4	2	1	0.4	1	0.23
103001	4	3	0	0.05	1	0.4
103001	4	4	0	0.1	1	0.6
103001	4	5	0	0.1	1	0.1
103001	4	6	0	0.07	1	0.2
103001	4	7	0	0.1	1	0.3
103001	4	8	0	0.05	1	0.1

Table A2: Summary Statistics: Current vs. Future Friends

Variable	Obs	Mean	Std. Dev.	Min	Max
Share Friend	4,218,170	0.016	0.076	0	1
Share Future Friend	4,218,170	0.014	0.072	0	1

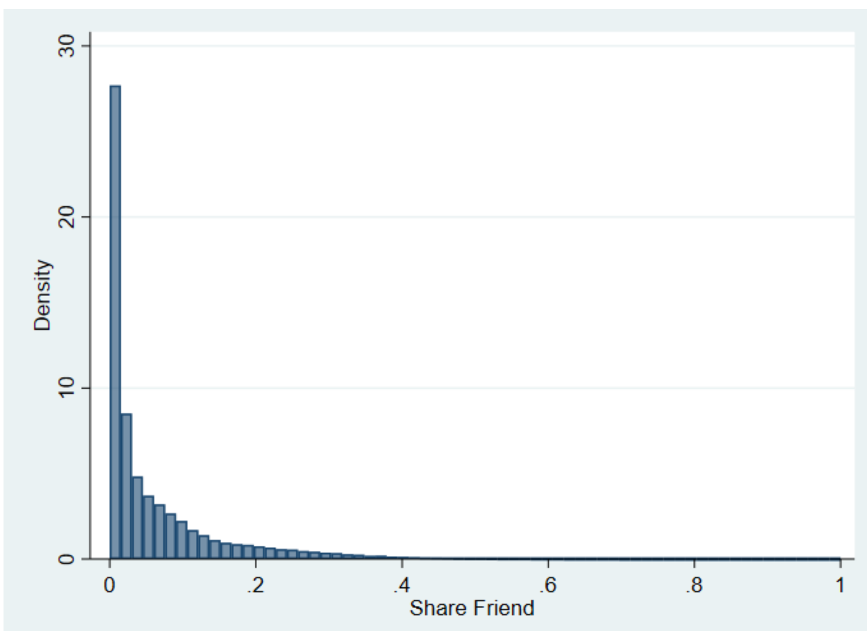


Figure A1: Distribution of Share Friend

Table A3: Summary Statistics: Share of Friends by Different Groups

	Mean	SD	N
Share Longer Friendship	0.009	0.040	4,861,999
Share Shorter Friendship	0.007	0.034	4,861,999
Share Higher HP	0.002	0.021	4,846,788
Share Lower or Similar HP	0.007	0.038	4,846,788
Share Pre-existing Coworkers	0.0159	0.089	278,628
Share newly-joined Coworkers	0.0002	0.010	278,628

Table A4: Balance Test by same-carrier fraction

	Below Median			Above Median				
	Mean	SD	N	Mean	SD	N	Diff.	t-stat
Female	0.33	0.47	1,025,318	0.35	0.48	952,021	-0.02***	-36.43
Age (midpoint)	36.72	12.91	1,025,432	38.44	13.28	952,164	-1.72***	-92.18
Reside in urban	0.50	0.50	837,973	0.49	0.50	827,597	0.01***	8.85
Work in urban	0.50	0.50	701,327	0.50	0.50	708,290	0.00***	4.16
Born outside the Province	0.59	0.49	1,040,873	0.61	0.49	985,527	-0.02***	-29.88

Notes: The table shows comparison of covariates by the fraction of same-carrier baseline one-way contacts. The cutoff is the median of the distribution, 34 percent.

Table A5: Friend and Pairwise Characteristics: Current vs. Future Friends

(a) Friend Characteristics

	Current friends			Future friends				
	Mean	SD	N	Mean	SD	N	Diff.	t-stat
Female	0.31	0.46	1,096,494	0.31	0.46	573,116	0.01***	8.30
Age (midpoint)	39.58	11.24	1,096,808	38.36	11.34	573,268	1.22***	66.31
Reside in Urban	0.55	0.50	1,047,993	0.56	0.50	555,081	-0.01***	-16.81

(b) Pairwise Characteristics

	Current friends			Future friends				
	Mean	SD	N	Mean	SD	N	Diff.	t-stat
Same gender	0.62	0.49	1,075,047	0.60	0.49	561,032	0.02***	19.94
|Age A - Age B|	9.66	9.05	1,076,029	10.29	9.05	561,421	-0.63***	-42.46
Both urban	0.45	0.50	981,469	0.46	0.50	520,835	-0.01***	-6.96
Urban-rural	0.10	0.30	981,469	0.11	0.31	520,835	-0.01***	-16.01
Rural-urban	0.10	0.30	981,469	0.11	0.31	520,835	-0.01***	-12.63
Both rural	0.35	0.48	981,469	0.33	0.47	520,835	0.02***	25.64
N. calls per month	1.70	0.91	1,157,182	1.78	0.97	611,774	0.32***	84.52

Notes: One observation is a call link A-B, where A is the phone changer. Characteristics of B is reported in panel (a). Difference in observables between A and B is reported in panel (b) .

Table A6: Summary Statistics: Prices of New Products

	Mean	SD	N
P1 (Release price)	266.555	174.845	34
P2	244.267	162.579	34

Notes: The table shows the release price and the latest prices in Period 2 (Q2 2017-Q3 2017) for products released from Q2 2016 to Q1 2017.

Table A7: New Buyer Demographics By Month of Purchase

Month of purchase	Dec 2016	Jan 2017	Feb 2017	March 2017	April 2017	May 2017	June 2017	July 2017	Aug 2017	Sep 2017
Female	0.36	0.37	0.38	0.39	0.38	0.38	0.37	0.36	0.34	0.35
	(0.48)	(0.48)	(0.49)	(0.49)	(0.49)	(0.48)	(0.48)	(0.48)	(0.47)	(0.48)
Age (midpoint)	37.34	40.02	40.17	39.79	39.51	39.09	38.78	38.67	38.69	37.99
	(13.18)	(11.85)	(12.38)	(12.68)	(12.82)	(12.58)	(12.57)	(12.82)	(12.70)	(12.85)
Urban	0.57	0.59	0.61	0.60	0.58	0.58	0.59	0.57	0.57	0.56
	(0.50)	(0.49)	(0.49)	(0.49)	(0.49)	(0.49)	(0.49)	(0.49)	(0.49)	(0.50)
Avg month plan fee	7.33	9.52	8.01	7.44	7.23	7.85	8.35	7.95	7.94	7.65
	(9.43)	(10.69)	(9.45)	(9.01)	(8.89)	(9.28)	(9.89)	(9.45)	(9.67)	(9.34)
House price per square meter	1908.46	1958.74	1975.49	1949.96	1918.95	1936.06	1945.66	1930.38	1924.68	1911.32
	(715.25)	(716.68)	(716.30)	(718.35)	(718.38)	(714.92)	(712.90)	(717.16)	(716.86)	(714.76)
Total duration (minutes) of calls	3065.70	4201.96	3811.81	3412.43	3244.01	3517.17	3765.36	3446.18	3368.51	3158.82
	(3820.19)	(4082.76)	(4305.47)	(3794.86)	(3724.91)	(3932.16)	(4200.31)	(4347.01)	(3975.64)	(3829.16)
Total number of calls	1996.16	2784.30	2466.27	2202.62	2115.02	2326.03	2471.61	2263.69	2242.57	2105.50
	(2496.71)	(2671.36)	(2666.13)	(2424.63)	(2413.91)	(2661.62)	(2818.69)	(2654.53)	(2676.26)	(2624.48)

Notes: The table shows the demographic information (gender, age, urban), income proxies (monthly plan fee, house price) and phone use intensity (total duration and number of calls in one year) for new buyers by the month of purchase. There is no obvious compositional difference among new buyers in different months.

B Research File for Sample Construction

B.1 New Buyer Sample

Relying on the weekly tracker of devices, I identify the newly made choices during the sample period through the change of devices. A phone change is identified if the following criteria hold:

1. One sim card experienced more than one devices (brand + model) in the sample periods

2. There is no re-occurrence of a previously held device

3. Holding the new device for at least one month

4. holding at least one previous device for at least one month

Table A8: Sample Selection

	N. Users	% remain
	3,061,230	
Mobile devices	2,740,754	89.53%
(-) multiple-device holders*	2,740,650	89.53%
(-) users contract with phone bundle and "one sim dual terminal" plans	2,685,837	87.45%
(-) users observed less than 2 months	2,380,331	77.76%

Notes:Multiple-device holders are identified if one sim card experiences several devices and switch back and forth between them. ("A-A-B-A-B-A")

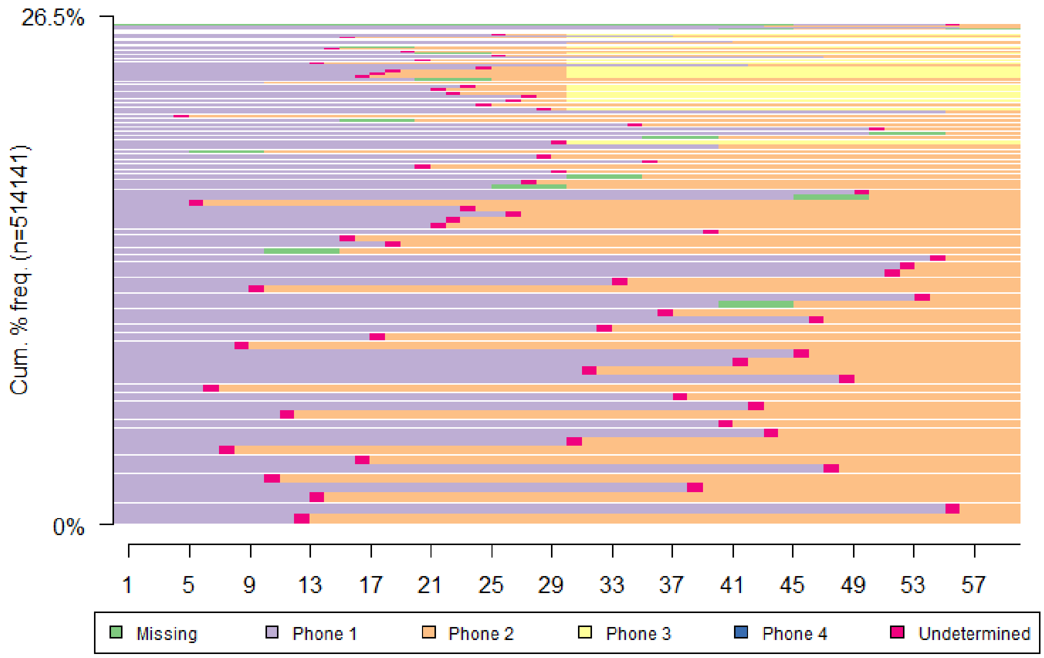


Figure A2: Phone Change: Top 100 Frequent Phone Sequences

Notes: The figure shows the top 100 most frequent sequence of phone sequences in the weekly tracker data for phone buyers. The top 100 patterns accounts for 26.5 percent among 514,141 new buyers. For example, the bottom segment is the most frequent pattern that uses "phone 1" for 12 weeks, then undetermined device for 1 week, followed by "phone 2" for 44 weeks.

B.2 Dyad selection and Contact definition

Call records capture the real-world social contacts. To rule out accidental calls from unknown parties and business entities, two levels of filtering are conducted to exclude links that are infrequently contacted. The criteria are chosen based on both total call frequency and duration in twelve months. A pair of call contacts (i, m) are excluded if either of the following two criteria hold:

1. total call duration is less than 10th percentile of the nonzero call distribution (16 seconds).

2. on average call each other less than one call per quarter.

Table A9 reports the process of the call contact selection. Limiting the minimum call duration in sample period to be 16 seconds helps to remove potential accidental calls by around 10% from the raw data. The average quarterly call frequency criteria further excludes around 45% of the pairs. In this way, accidental and infrequent call contacts are filtered out. So after two steps of selection, I end up with 172 million pairs of call contacts.

Table A9: Call Contact Selection

	N. dyads	% remain
All pairs	390,209,050	
Total duration at least 16 seconds	353,449,502	90.58%
On average at least one call per quarter	172,843,963	44.30%

After dropping infrequent links, I refer a call contact as a social contact. Analogous to Onnela et al. (2007) and Marlow (2009), based on the feature of the CDRs, I refine the following definition for friends to represent closer friendship and greater frequency of interaction.

Baseline (“Friend 1"): A link represents directional communication if the user called to the friend at the other end of the link at least once during observation period (whether or not the calls were reciprocated).

Reciprocal (“Friend 2"): A link represents reciprocal (mutual) communication, if the user both initiated a call to the friend at the other end of the link, and also received a call from them during the observation period.

Table A10: Dyad-level: Call time and Frequency

	N. of pairs	Mean	SD
Friend1: Baseline one-way			
Frequency	172,843,963	18.16	66.71
Duration (seconds)	172,843,963	1724.82	11747.24
Friend2: Reciprocal			
Frequency	100,784,483	27.68	85.40
Duration (seconds)	100,784,483	2628.69	13669.70

Notes: Table A10 shows the communication pattern for the two contact definitions. Distribution for frequency and call time are right-skewed. Frequency (Duration) is the number of calls (seconds) one users calls the other in the sample period. Bottom 10% extreme numbers are excluded.

Table A11: User-Level: Network Size

	N	Mean	SD
Friend1: Baseline one-way			
N. Friend1	2,186,716	64.54	93.64
N. same-carrier contacts	2,186,716	22.42	37.62
Same-carrier fraction	2,160,915	0.44	0.64
Friend2: Reciprocal			
N. Friend2	1,837,531	47.85	63.54
N. same-carrier contacts	1,837,531	20.21	16.36
Same-carrier fraction	1,837,531	0.64	0.30

B.3 Product Grouping and Selection

I focus on call device tracker data, 2016Q4-2017Q3. 82 percent of users' device are matched with models from the IDC tracker data in sample period. There are many variants for each model and similar models released in different years. Given the large number of models, I first group models based on the closeness of major characteristics. Then identify the unique models and its market share in the call device tracker data.

First I drop extremely expensive/cheap handset before grouping and selection. For example, I drop ultra-luxury phones targeted as high-end gifts, such as the Huawei Mate 9 Porsche Design, whose release price at 1317 USD (9000 RMB) (compared to initial release prices of iPhones at around 990 USD). I also drop phones cheaper than 67 USD (450 RMB) such as phones from domestic brand Sugar, LaJiao etc. Product lines are divided based on the release price.

Grouping Firms release model variants to increase demand and price discriminate with a low costs. For the same base model, variants usually come with slightly different features such as storage capacity RAM and ROM. For these model variants, I treat them as the same model. Another proliferation is that for non-frontier models, firms introduce models with slightly different features at low cost by combining different components together. Similar to Wang (2018), I group models in the same product line into clusters based on a distance measure and identify the earliest released model as the unique model in each cluster. Consider model A and B from the same brand and same product line, the distance between A and B is measured as a Euclidean distance along six dimensions normalized by the standard deviations :

$$D_{A,B} = \sqrt{\frac{1}{6}\sum_{k=1}^{6}\frac{x_k^A - x_k^B}{SD(x_k)}}$$

where the six major attributes are CPU clock speed, camera resolution, screen size, screen resolution, battery capacity and fingerprint function. Then using the data-driven K-Means clustering algorithm, models in each product line are classified into clusters, such that models within the same cluster are as similar as possible (i.e., high intra-class similarity), whereas models from different clusters are as dissimilar as possible (i.e., low inter-class similarity). As a result, 464 models are grouped as 167 models.

Product Selection After grouping models, I focus the major models that take 70 percent of (the new purchase) market share in each market. Then I collapse the rest into a composite fringe product so that there is one in each market. Attributes of the composite product are obtained with share-weighted average within each group.

Table A12: Product Grouping and Selection

	N. models
In CDR device tracker (include variants)	849
Combine variants, have at least 25 users	564
Merged with IDC on sale + attributes	464
After grouping	167
Top 70% share in each market	62

C Estimation and Counterfactual Simulation Procedures

C.1 Demand Estimation Routine

For each individual, R = 1000, fix a set of draws $\{v_i^r\}_{r=1}^R$ and income level $\{y_i^r\}_{r=1}^R$ from a log-normal distribution estimated using survey data. In each market (month), randomly draw 500 consumers, each with a vector of demographic and income information. Gender, age, and the urban dummy are randomly draw from the survey data, weighted by the national representative weights. After drawing the income from the log-normal distribution, I assign a high income dummy which equals to 1 if it passes the 75th percentile. Conditional on the gender, age, urban, high income dummy and month of purchase, randomly draw the share of friends vector for each alternative from the sample of new buyers. Note that in the estimation procedure, the share of friends vector is random draw from the sample, however, in the counterfactual analysis, this vector is generated in the model through the lagged structure.

The estimation proceeds in two steps. In the first step, I conduct steps 1-5 find the nonlinear terms $\boldsymbol{\theta}_2$ and product by market-specific constants δ_{jt}; in the second step, conduct step 6 to recover linear parameters $\boldsymbol{\theta}_1$.

1. Start with some initial guess for non-linear parameter $\boldsymbol{\theta}_2^0$;

2. Inverse demand: start with an initial guess δ^0.

   Given $\{y_i^r, \nu_i^r\}_{r=1}^R$, $\boldsymbol{\theta}_2^0$ and δ^0, calculate model predicted individual choice probabilities from each draw

$$P_{ijt}^r(Y_i = j | y_i^r, \nu_i^r, s_{it-3}, \boldsymbol{X}, \boldsymbol{p}, \delta^0, \boldsymbol{\theta}_2^0) = \frac{1}{\sum_r} \frac{exp(\delta_{jt}^0 + \mu_{ijt}^r)}{\sum_{j'=1}^J exp(\delta_{j't}^0 + \mu_{ij't}^r)}$$

   where $\mu_{ijt} = (\bar{\alpha} + \sigma_p \nu_{tip}) p_{jt} + \theta s_{i,j,t-3}$.

   Calculate the average as the model predicted conditional choice probability of person i choosing alternative j:

$$\bar{P}_{ijt} = \frac{1}{R} \sum_i^R P_{ijt}^r(\delta^0, \boldsymbol{\theta}_2^0)$$

   Then aggregate to predicted market shares $s_{jt}(\delta^0, \boldsymbol{\theta}_2^0)$.

$$s_{jt}(\delta^0, \boldsymbol{\theta}_2^0) = \frac{1}{N} \sum_{i \in m} \bar{P}_{ijt}(Y_i = j)$$

   Iterate over the contraction mapping until δ converges:

$$\delta_{jt}^{h+1} = \delta_{jt}^h + ln s_{jt}^N - ln(s_{jt}(\delta^h, \theta_2^0))$$

   Denote the converged mean utility as $\delta(\boldsymbol{\theta}_2^0)$.

3. Substitute that $\delta(\boldsymbol{\theta}_2^0)$ for δ^0 into the model's predictions for the individual conditional choice probability,

$$\bar{P}_{ijt}(\delta(\boldsymbol{\theta}_2^0), \boldsymbol{\theta}_2^0)) = \frac{1}{R} \sum_i^R P_{ijt}^r(\delta(\boldsymbol{\theta}_2^0), \boldsymbol{\theta}_2^0)$$

   The simulated likelihood function of the sample becomes

$$SLL(\delta(\boldsymbol{\theta}_2^0), \boldsymbol{\theta}_2^0)) = \sum_{i=1}^N \sum_{j=0}^J ln \bar{P}_{ijt}(\delta(\boldsymbol{\theta}_2^0), \boldsymbol{\theta}_2^0))$$

4. Choose $\boldsymbol{\theta}_2$ and $\delta(\boldsymbol{\theta}_2)$ that maximize the constrained simulated likelihood. For each guess of $\boldsymbol{\theta}_2$, repeat step 1-3.

$$\max_{\hat{\delta}(\theta_2),\theta_2} SLL(\hat{\delta}(\theta_2),\theta_2) = \sum_{i=1}^{N}\sum_{j=0}^{J} ln\left[\frac{1}{R}\sum_{i}^{R} P_{ijt}^{r}(\hat{\delta}(\theta_2),\theta_2)\right]$$

s.t.

$$s_{jt}^{N} - s_{jt}(\boldsymbol{\theta}_2,\delta_{jt}) = 0$$

5. Estimate linear parameters using two-stage IV regression:

$$\delta_{jt} = X_{jt}\bar{\beta} + \zeta_{f(jt)} + \eta_t + \xi_{jt}$$

C.2 Counterfactual Simulation Procedure: Supply

I solve for the new equilibrium prices backward in two steps.

1. Initial guess of products prices p_1^0.
2. In period 1, find new individual demand (500 consumers) given p_1^0
3. Obtain total sales in period 1: $Q_1(p_1^0)$ for each model
4. Calculate new semi-elasticity $\frac{dQ_1}{dp_1}$ with new demand shares according to analytical form.
5. Inner loop at p_1^0,
   (a) initial guess prices in period 2 p_2^0
   (b) Simulate friend choices in period 1, and obtain lagged friend share for period 2: $lagshare_2(p_1^0)$
   (c) Based on lagged friend share in period 2, obtain individual demand in period 2: $q_i(lagshare_2(p_1^0), p_2^0)$
   (d) Calculate total sales in period 2 $Q_2(p_1^0, p_2^0)$
   (e) Calculate $\frac{dQ_2}{dp_2}$ with new demand shares according to analytical form.
   (f) Calculate new equilibrium price in period 2 according to

$$p_2^1 = mc_2 - (\frac{\partial Q_{j2}}{\partial p_{j2}} \times Ownership)'^{-1} \times Q_{j2} \tag{22}$$

(g) Calculate $||p_2^1 - p_2^0||$ for all products

(h) Repeat until the distance fall below the tolerance level; Obtain $p_2^*(p_1^0)$

6. Take $p_2^*(p_1^0)$ as given, use $\frac{dQ_2}{dp_1}(p_1^0, p_2^*)$ (obtained outside the counterfactual loops)

7. Calculate equilibrium price in period 1 according to F.O.C.

$$p_1^1 = mc_1 - ((\frac{\partial Q_{j1}}{\partial p_{j1}}) * Ownership)'^{-1} \times \tilde{Q}_1$$

where

$$\tilde{Q}_1 = Q_{j1} - \beta Q_{j2} \times \frac{\partial Q_{j2}}{\partial p_{j2}}^{-1} \frac{\partial Q_{j2}}{\partial p_{j1}} + \beta Q_{j2} \frac{\partial p_{j2}}{\partial p_{j1}}$$

8. Calculate $||p_1^1 - p_1^0||$ for all new products

9. Repeat until the distance fall below the tolerance level; Obtain p_1^*

D Prices and Social Influence: Model Prediction Illustration

I simplify the product life cycle into two periods. A firm f maximizes the expected discount profit

$$W_f = \sum_{j \in J_f} (p_{j1} - mc_{j1}) Q_{j1} + \delta (p_{j2} - mc_{j2}) Q_{j2} \tag{23}$$

where δ is the discount factor. J_f represents the products offered by firm f, including products that are newly released in Period 1. $p_{j2} = p_{j2}(\mathbf{Q}_1(\boldsymbol{p}_1))$ is a function of the introductory prices. The optimal prices are solved using backward induction starting from Period 2. The first-order conditions are

$$mc_{j2} = p_{j2}^* + [\Delta_{f2}^{-1} \times \mathbf{Q}_2]_j \tag{24}$$

$$\begin{aligned} mc_{j1} &= p_{j1}^* + \left[\Delta_{f1}^{-1} \times \left\{ Q_{j1} + \delta \sum_{r \in J_f} (p_{r2} - mc_{r2}) \frac{\partial Q_{r2}}{\partial p_{j1}} + \delta Q_{j2} \frac{\partial p_{j2}}{\partial p_{j1}} \right\} \right] \\ &= p_{j1}^* + \left[\Delta_{f1}^{-1} \times \left\{ \mathbf{Q}_1 - \delta \frac{\partial \mathbf{Q}_2}{\partial \mathbf{p}_1} [\Delta_{f2}^{-1} \times \mathbf{Q}_2] - \delta Diag\left(\frac{\partial \mathbf{Q}_2}{\partial \mathbf{p}_1}\right) [\Delta_{f2}^{-1} \times \mathbf{Q}_2] \right\} \right]_j \end{aligned} \tag{25}$$

where Δ_{ft} is a J-by-J matrix, whose (j,r) element is $\frac{\partial Q_{rt}}{\partial p_{jt}}$, $t = 1,2$. The inter-temporal partial derivatives $\frac{\partial \mathbf{Q}_2}{\partial \mathbf{p}_1}$ is a function of social influence θ. Its diagonal terms are

$$\frac{\partial Q_{j2}}{\partial p_{j1}} = \sum_m M_m \int_{i\in m} \frac{dS_{ij2}}{dp_{j1}} dF(i) = \sum_m M_m \int_{i\in m} \theta S_{ij2}(1-S_{ij2}) \left[\sum_{l\in m(i)} \alpha_l S_{lj1}(1-S_{lj1}) \right] dF(i)$$

where S_{ij2} is the choice probability of person i choosing j in period 2. (j,r)th element:

$$\frac{\partial Q_{j2}}{\partial p_{r1}} = \sum_m M_m \int_{i\in m} \frac{dS_{ij2}}{dp_{r1}} dF(i) = \sum_m M_m \int_{i\in m} \theta S_{ij2}(1-S_{ij2}) \left[\sum_{l\in m(i)} \alpha_l S_{lj1}(S_{lr1}) \right] dF(i)$$

Social influence and second-period prices

$$p_{j2}^* = mc_{j2} - [\Delta_{f2}^{-1} \times \mathbf{Q}_2]_j$$

Ignore time subscript t=2 for now. Denote the price quantity derivative as $\Delta_{jj} = \frac{\partial Q_j}{\partial p_j}$. At individual level, denote $\Delta_{i,jj} = \frac{\partial S_{ij}}{\partial p_j}$.

$$\Delta_{i,jj} = \alpha_i (1-S_{ij}) S_{ij} < 0$$

The own price elasticity for product j, ϵ_{jj}, is decreasing in individual share S_{ij}.

$$|\epsilon_{jj}| = |\Delta_{jj}| \frac{p_j}{Q_j} = \int_i |\alpha_i (1-S_{ij}) p_j| dF(i)$$

When $\theta > 0$, S_{ij} increases and own price elasticities decrease. So when $\theta > 0$, the optimal prices in second period are higher than the counterfactual optimal prices when $\theta = 0$.

Social influence and release prices

$$p_{j1}^* = mc_{j1} - \left[\Delta_{f1}^{-1} \times \left\{ \mathbf{Q}_1 - \underbrace{\delta \frac{\partial \mathbf{Q}_2}{\partial \mathbf{p}_1} [\Delta_{f2}^{-1} \times \mathbf{Q}_2] - \delta Diag\left(\frac{\partial \mathbf{Q}_2}{\partial \mathbf{p}_1} \right) [\Delta_{f2}^{-1} \times \mathbf{Q}_2]}_{\theta > 0} \right\} \right]_j$$

The gap in optimal prices between $\theta = 0$ and $\theta = \theta^* > 0$ becomes

$$\begin{aligned} p_{j1}^{\theta=0} - p_{j1}^{\theta=\theta^*} &= \left[-\Delta_{f1}^{-1}(\boldsymbol{Q}_1^0 - \boldsymbol{Q}_1^{\theta^*}) + \Delta_{f1}^{-1} \times \delta \left\{ \frac{\partial \mathbf{Q}_2}{\partial \boldsymbol{p}_1}[\Delta_{f2}^{-1} \times \boldsymbol{Q}_2] + Diag\left(\frac{\partial \mathbf{Q}_2}{\partial \boldsymbol{p}_1}\right)[\Delta_{f2}^{-1} \times \boldsymbol{Q}_2] \middle| \theta^* \right\} \right]_j \\ &\approx \left[\Delta_{f1}^{-1} \times \delta \left\{ \frac{\partial \boldsymbol{Q}_2}{\partial \mathbf{p}_1}[\Delta_{f2}^{-1} \times \boldsymbol{Q}_2] + Diag\left(\frac{\partial \boldsymbol{Q}_2}{\partial \boldsymbol{p}_1}\right)[\Delta_{f2}^{-1} \times \boldsymbol{Q}_2] \middle| \theta^* \right\} \right]_j > 0 \end{aligned} \tag{26}$$

Note that the first term $\boldsymbol{Q}_1^0 - \boldsymbol{Q}_1^{\theta^*}$ in first line in Equation 26 is driven by the price effect $p_{j1}^{\theta=0} - p_{j1}^{\theta=\theta^*}$ through dynamic channel, not the direct effect of the change of θ, and is isomorphic to the price changes, I ignore this part when evaluating the effect of change θ on first period prices. Δ_{f1}^{-1} is not a function of θ because for all new introduced products because the lagged shares are all zero. So the sign of the price gap is determined by the term inside the curly bracket. As discussed in earlier part, $[\Delta_{f2}^{-1} \times \boldsymbol{Q}_2]$ becomes more negative when $\theta > 0$. Note that $\frac{\partial Q_2}{\partial \boldsymbol{p}_1}$ is function of θ with negative diagonal values. So the price gap as the product of two negative terms is positive. That is, when $\theta > 0$, the optimal introductory prices are lower than the counterfactual optimal prices.